

Информатика и её применения

Том 19 Выпуск 3 Год 2025

СОДЕРЖАНИЕ

Методы условно-оптимальной фильтрации по байесову критерию для наблюдаемых неявных стохастических систем	
<i>И. Н. Синицын</i>	2
Стабилизация траектории линейной системы со скачкообразным дрейфом при целочисленных ограничениях на управление	
<i>А. В. Босов</i>	10
Оценка риска метода блочной пороговой обработки при решении обратных статистических задач с данными, заданными на случайной сетке	
<i>О. В. Шестаков</i>	19
Оптимизация скоростного режима грузового поезда по критерию среднего ущерба от аварий	
<i>А. В. Борисов, А. Н. Игнатов, В. А. Борисов</i>	27
Модель развития многопользовательской системы связи при наращивании пропускной способности сети	
<i>Ю. Е. Малашенко, И. А. Назарова, М. В. Козлов</i>	36
Разработка модели для расчета производительности планирования передач в сетях 5G NR с учетом сигнального трафика	
<i>Э. С. Сопин, А. И. Назарьин, С. Я. Шоргин</i>	46
Многовариантное резервирование с учетом логико-топологических особенностей транзисторных схем	
<i>С. Ф. Тюрин, М. С. Никитин, Ю. А. Степченков, Ю. Г. Дьяченко</i>	55
Классификация малых наборов данных большой размерности	
<i>А. А. Грушо, Н. А. Грушо, М. И. Забежайло, В. В. Кульченков, Е. Е. Тимонина</i>	67
Автоматизация разметки имплицитных логико-семантических отношений: возможности и ограничения	
<i>А. А. Гончаров, П. В. Ярошенко</i>	73
Иерархия Акоффа и задачи извлечения знания из текстов	
<i>И. М. Зацман</i>	82
Об авторах	90
Правила подготовки рукописей	92
Requirements for manuscripts	95

МЕТОДЫ УСЛОВНО-ОПТИМАЛЬНОЙ ФИЛЬТРАЦИИ ПО БАЙЕСОВУ КРИТЕРИЮ ДЛЯ НАБЛЮДАЕМЫХ НЕЯВНЫХ СТОХАСТИЧЕСКИХ СИСТЕМ

И. Н. Сеницын¹

Аннотация: Разработаны методы синтеза условно-оптимальных фильтров (УОФ) по байесову критерию (БК) для обработки информации во взаимосвязанных наблюдаемых непрерывных и дискретных неявных негауссовских стохастических системах (СтС), приводимых к явным. Дан краткий обзор работ по условно-оптимальной фильтрации по среднеквадратичному, энергетическому и сложному статистическому критериям для явных и неявных, непрерывных и дискретных нелинейных СтС. Представлены методы приведения неявных СтС для гладких и разрывных неявных функций. Разработаны точные методы синтеза БК УОФ для приведенных дифференциальных и разностных уравнений. Особое внимание уделено нормальным БК УОФ для аддитивных и параметрических возмущений. Рассмотрены вопросы эквивалентности гауссовских и негауссовских возмущений при синтезе БК УОФ. Сформулированы направления дальнейших обобщений и применений.

Ключевые слова: байесов критерий (БК); неявная стохастическая система; нормальный БК-условно-оптимальный фильтр (НБК УОФ); стохастическая система (СтС); условно-оптимальный фильтр (УОФ)

DOI: 10.14357/19922264250301

EDN: VTVRXX

1 Введение

Теория условно-оптимальной фильтрации в среднеквадратичном смысле стохастических процессов (СтП), описываемых явными стохастическими дифференциальными уравнениями, нашла применение при решении задач обработки информации в реальном масштабе времени. Для неявных дифференциальных СтС, допускающих приведение к явным дифференциальным, развиты соответствующие методы условно-оптимальной фильтрации. Так, в [1] разработаны математические модели наблюдаемых неявных дифференциальных СтС, приводимых к дифференциальным, для задач фильтрации. Получены базовые уравнения нелинейных УОФ. Развитие методов УОФ при авторегрессионных помехах в наблюдениях для некоррелированных и авторегрессионных помех на случай неявных непрерывных и дискретных гауссовских СтС, приводимых к явным, дано в [2]. Представлены математические модели неявных непрерывных и дискретных гауссовских и негауссовских СтС, приводимых к явным. Для приведенных неявных СтС в условиях, когда наблюдения не оказывают влияния на объект наблюдения и описываются явными уравнениями, разрешенными относительно производных (разностей), получены уравнения базовых методов УОФ. Статья [3] посвящена методам условно-оптималь-

ного синтеза фильтров для неявных дискретных СтС, приводимых к явным, путем эквивалентной гауссовской и негауссовской линеаризации. Предполагается, что наблюдения не влияют на объект наблюдения и описываются дискретными нелинейными уравнениями с некоррелированными и авторегрессионными помехами. Дан обзор работ в области синтеза УОФ для явных и неявных дискретных наблюдаемых СтС. Приведены основные модели дискретных неявных СтС и методы их эквивалентной линеаризации. Получены уравнения УОФ. В качестве примера рассмотрены УОФ для авторегрессионных приведенных уравнений.

В [4] изложены методы синтеза нелинейных нормальных УОФ (НУОФ) по среднеквадратичному критерию для обработки информации во взаимосвязанных наблюдаемых неявных непрерывных и дискретных гауссовских и негауссовских СтС, приводимых к явным. Представлены уравнения наблюдаемых непрерывных и дискретных явных и неявных гауссовских и негауссовских СтС, методы приведения неявных гладких и разрывных СтС к явным. Для взаимосвязанных неявных объектов и явных систем наблюдения разработаны методы синтеза НУОФ, основанные на решении методом нормальной аппроксимации совместных приведенных уравнений объекта, системы наблюдения и уравнений линейных и нелинейных УОФ для гауссовских и негауссовских СтС. В каче-

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

стве дискретных неявных СтС, приводимых к дискретным, рассмотрены нелинейные регрессионные и авторегрессионные уравнения.

В [5] рассмотрены методы синтеза нелинейных УОФ по сложному статистическому критерию (ССК) для обработки информации во взаимосвязанных наблюдаемых непрерывных и дискретных неявных негауссовских СтС, приводимых к явным. Дан краткий обзор работ по УОФ по среднеквадратичному критерию для явных и неявных СтС и ССК УОФ для явных систем. Представлены методы приведения для гладких и разрывных неявных функций. Разработаны точные методы синтеза ССК УОФ для приведенных дифференциальных, регрессионных и авторегрессионных уравнений. Рассмотрены две задачи по применению основных результатов к задачам непрерывной и дискретной ССК-фильтрации в случае аддитивных шумов и неявных функций, допускающих эквивалентную статистическую линеаризацию. Получены обобщения фильтров Калмана и Калмана–Бьюси. Предложены возможные обобщения точных и приближенных методов.

Рассмотрим методы БК-синтеза УОФ для обработки информации во взаимосвязанных объектовых наблюдаемых непрерывных и дискретных негауссовских СтС, приводимых к явным. Для явных СтС, соответствующих БК, УОФ рассмотрены в [6–8].

2 Уравнения приведенных объектовых наблюдаемых неявных стохастических систем

Обобщая [1–5], рассмотрим векторную неявную дифференциальную СтС с нелинейностями, описываемыми гладкими детерминированными функциями размерности $(n^\Phi \times 1)$:

$$\Phi = \Phi(t, Z_t, \dot{Z}_t, \dots, Y_t^{(k)}, U_t) = 0,$$

$$Z(t_0) = Z_0, \dot{Z}(t_0) = \dot{Z}_0, \dots, Y^{(k)}(t_0) = Z_0^{(k)}. \quad (1)$$

Уравнение для стохастических возмущений U_t возьмем в виде, разрешенном относительно белых шумов:

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

Здесь $a^U(t, U_t)$ и $b^U(t, U_t)$ — $(n^Z \times 1)$ - и $(n^Z \times n^V)$ -мерные функции; V_t^U — белый шум в строгом смысле, допускающий представление вида

$$V_t^U = \dot{W}_t^U, \quad W_t^U = W_0^U(t) + \int_{R_0^q} c^U(u)P^0(t, du),$$

где $c^U = c^U(u)$ — известная векторная функция той же размерности, что и W_0^U ; интеграл при любом $t \geq t_0$ представляет собой стохастический интеграл по центрированной пуассоновской мере $P^0(t, \mathcal{A})$, независимой от W_0^U и имеющей независимые значения на попарно непересекающихся множествах; \mathcal{A} — борелевское множество пространства R_0^q с выколотым началом. Уравнение (2) понимается в смысле Ито и имеет единственное среднеквадратичное решение [7, 9, 10].

Для гладких функций в (1), допускающих стохастические производные Ито до h -го порядка, выполним следующие преобразования. Будем дифференцировать по t левые части уравнений (1) по обобщенной формуле Ито [7, 9, 10] до тех пор, пока не появятся производные белого шума. В результате получим следующую систему неявных нелинейных дифференциальных уравнений:

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

Далее введем составной вектор $[\bar{Z}_t^T Z_t'^T]^T$, где $\bar{Z}_t = [Z_t^T \dot{Z}_t^T \dots Z_t^{(k-1)T}]^T$, и вспомогательный вектор Z_t' , определяемый уравнениями (3) в виде функционального уравнения

$$\Psi(\bar{Z}_t, Z_t') = 0. \quad (4)$$

Как показано в [1–3], если неявные гладкие вектор-функции и конечные неявные уравнения (1) и (4) допускают гладкую замену переменных, то они приводятся к системе, состоящей из векторного стохастического дифференциального уравнения Ито и функционального неявного векторного уравнения (**теорема 2.1**):

$$d\bar{Z}_t = a^\Pi(t, \bar{Z}_t) dt + b^\Pi(t, \bar{Z}_t) dW_0 + \int_{R_0^q} c^\Pi(t, \bar{Z}_t, u) P^0(dt, du), \quad \Psi^\Pi(\bar{Z}_t, Z_t', t) = 0.$$

Приводимые неявные дискретные регрессионные и авторегрессионные СтС описываются разностными уравнениями вида (**теоремы 2.2 и 2.3**):

$$\begin{aligned} \bar{Z}_{k+1} &= \omega_k^\Pi(\bar{Z}_k, V_k); & \Psi_k^\Pi(\bar{Z}_k, Z_k') &= 0; \\ \bar{Z}_{k+1} &= a_k^\Pi(\bar{Z}_k) + b_k^\Pi(\bar{Z}_k) V_k; & \Psi_k^\Pi(\bar{Z}_k, Z_k') &= 0. \end{aligned}$$

Здесь функции $\omega_k^\Pi(\bar{z}_k, v_k)$, $a_k^\Pi(\bar{z}_k)$ и $b_k^\Pi(\bar{z}_k)$ имеют размерности $(n_{\bar{z}} \times 1)$, $(n_{\bar{z}} \times 1)$ и $(n_{\bar{z}} \times n_v)$ соответственно, $k = 1, 2, \dots$

В условиях теоремы 2.1 предположим, что объектовая неявная СтС допускает приведение к явной

СтС, измерительная система вполне наблюдаема, причем уравнение наблюдения разрешено относительно вектора наблюдения. Введем новые обозначения для составных векторов состояния объекта X_t , вспомогательного вектора Z'_t и системы наблюдения Y_t . Тогда в качестве приведенных уравнений состояния объекта и наблюдений будем рассматривать для приведенных дифференциальных СтС следующие уравнения (**теорема 2.4**):

$$\left. \begin{aligned} \dot{X}_t &= \varphi^\Pi(X_t, Y_t, t) + \psi^\Pi(X_t, Y_t, t) V, \\ \Psi^\Pi(X_t, Z'_t, t) &= 0; \\ \dot{Y}_t &= \varphi_1(X_t, Y_t, t) + \psi_1(X_t, Y_t, t) V, \quad V = \dot{W}. \end{aligned} \right\} (5)$$

Здесь

$$V = \dot{W}, \quad W(t) = W_0(t) + \int_{R_0^q} c(u) P^0(t, du),$$

где $c(u)$ — векторная функция размерности $(n^W \times 1)$ q -мерного аргумента u ; интеграл при любом $t \geq \geq t_0$ представляет собой стохастический интеграл по центрированной пуассоновской мере $P^0(t, \mathcal{B})$, независимой от СтП W_0 и имеющей независимые значения на непересекающихся множества. Интенсивность $\nu = \nu(t)$ СтП $W(t)$ в этом случае определяется формулой:

$$\begin{aligned} \nu(t) &= \nu_0(t) + \nu'(t), \\ \nu'(t) &= \int_{R_0^q} c(u) c(u)^T \nu_P(t, u) du, \end{aligned} (6)$$

где $\nu_0(t)$ и $\nu_P(t, u)$ — интенсивности СтП W_0 и P^0 .
Для дискретного случая в условиях теорем 2.2 и 2.3 приведенные уравнения состояния и наблюдения будут следующими (**теоремы 2.5 и 2.6**):

$$\left. \begin{aligned} X_{k+1} &= \omega_k^\Pi(X_k, Y_k, V_k), \quad \Psi^\Pi(X_k, Z'_k) = 0, \\ Y_k &= \omega_{1k}(X_k, Y_k, V_k); \end{aligned} \right\} (7)$$

$$\left. \begin{aligned} X_{k+1} &= \varphi_k^\Pi(X_k, Y_k) + \psi_k^\Pi(X_k, Y_k) V_k, \\ \Psi_k^\Pi(X_k, Z'_k) &= 0; \\ Y_k &= \varphi_{1k}(X_k, Y_k) + \psi_{1k}(X_k, Y_k) V_k, \end{aligned} \right\} (8)$$

где V_k — дискретные негауссовские в общем виде белые шумы с известной одномерной характеристической функцией $h_k = h_k(\varepsilon; t)$.

3 Байесов критерий оптимизации

Как известно [7, 9, 10], в основе БК оптимизации СтС лежит следующий общий принцип оценки качества СтС. Качество решения задачи оценивания

в каждом конкретном случае определяется некоторой функцией потерь $l = l(X_t, \hat{X}_t)$, значение которой зависит от конкретных реализаций процесса X_t и его оценки \hat{X}_t . Эта функция называется функцией потерь. Качество решения задачи оценивания в среднем для данной реализации процесса X_t при всех возможных реализациях оценки \hat{X}_t определяется условным математическим ожиданием функции потерь

$$\rho(A | X_t) = M [l(X_t, \hat{X}_t) | X_t]. (9)$$

Эту величину обычно называют условным риском. Условный риск зависит от оператора A , определяющего оценку \hat{X}_t , и от реализации процесса X_t . Наконец, среднее качество решения задачи оценивания при всех возможных реализациях X_t и \hat{X}_t характеризуется математическим ожиданием условного риска (9), равным безусловному ожиданию функции потерь

$$\rho(A) = M [\rho(A | X_t)] = M [l(X_t, \hat{X}_t)] (10)$$

и называемым средним риском. Установив таким образом меру качества оценки, можно за критерий оптимальности СтС принять критерий минимума среднего риска:

$$M [l(X_t, \hat{X}_t)] = \min. (11)$$

Если требуется обеспечить наилучшее решение задачи оценивания при некоторых дополнительных условиях, заданных в виде

$$C_j(A) = M [\varphi_j(X_t, \hat{X}_t)] = c_j \quad (j = \overline{1, Q}), (12)$$

то, применяя метод неопределенных множителей Лагранжа, можно свести задачу оценивания к отысканию минимума следующей величины:

$$\begin{aligned} \rho_1(A) &= \rho(A) + \sum_{j=1}^Q \lambda_j C_j(A) = \\ &= M \left[l(X_t, \hat{X}_t) + \sum_{j=1}^Q \lambda_j C_j(X_t, \hat{X}_t) \right]. \end{aligned} (13)$$

Эту величину можно рассматривать как средний риск для новой функции потерь

$$l_1(X_t, \hat{X}_t) = l(X_t, \hat{X}_t) + \sum_{j=1}^Q \lambda_j \varphi_j(X_t, \hat{X}_t), (14)$$

содержащей неопределенные параметры $\lambda_1, \dots, \lambda_Q$.

Замечание 3.1. Формулы (9)–(14) для векторов \hat{X}_t (\hat{X}_k) имеют место и для векторов \hat{Z}'_t (\hat{Z}'_k). Последние вычисляются на основе функциональных соотношений $\Psi(\hat{X}_t, \hat{Z}'_t) = 0$ и $\Psi(X_k, Z'_k) = 0$.

4 Основные результаты

4.1 Непрерывные негауссовские по байесову критерию условно-оптимальные фильтры

Пусть сначала исходная неявная дифференциальная СтС приведена к дифференциально-функциональной системе (5). Требуется по результатам наблюдений $Y_{t_0}^t = \{Y(\tau) \mid t_0 \leq \tau \leq t\}$ найти наилучшую оценку $\hat{X}_t = \hat{X}(t)$ процесса X_t в классе допустимых оценок, определяемых уравнением фильтра Пугачёва

$$d\hat{X}_t = \alpha_t \xi(\hat{X}_t, Y_t, t) dt + \beta_t \eta(\hat{X}_t, Y_t, t) dY_t + \gamma_t dt \quad (15)$$

при заданных структурных функциях $\xi = \xi(X_t, Y_t, t)$ и $\eta = \eta(X_t, Y_t, t)$, отображающих $R^{n_y + n_x + 1}$ в R^q и $R^{s n_y}$ соответственно. Неизвестные коэффициенты фильтра $\alpha_t = \alpha(t)$, $\beta_t = \beta(t)$ и $\gamma_t = \gamma(t)$ представляют собой матрицы размерности $(n_x \times q)$, $(n_x \times s)$ и $(n_x \times 1)$.

Будем считать условно-оптимальным такой фильтр из класса допустимых (15), который минимизирует дифференциал dI_t по времени t байесова критерия I_t :

$$dI_t = \min, \quad I_t = Ml(X_t, \hat{X}_t)$$

при достаточно гладких положительных функциях потерь $l(X_t, \hat{X}_t)$, $l \in C^3(R^{n_x} \times R^{n_x})$. Для нахождения неизвестных коэффициентов $\alpha_t = \alpha(t)$, $\beta_t = \beta(t)$ и $\gamma_t = \gamma(t)$ фильтра выпишем выражение для дифференциала dI_t , используя обобщенную формулу Ито [7]. Имеем

$$dI_t = M \left[l_{\hat{X}_t}^T \varphi^\Pi + l_{\hat{X}_t}^T (\alpha_t \xi + \beta_t \eta \varphi_1) + l_{X_t \hat{X}_t} : \sigma_{12} + \frac{1}{2} l_{X_t X_t} : \sigma_{11} + \frac{1}{2} l_{\hat{X}_t \hat{X}_t} : \sigma_{22} \right] dt + \int_{R_0^q} \left[l(X_t + \psi^\Pi c, \hat{X}_t + \beta_t \eta \psi_1 c) - l(X_t, \hat{X}_t) - l_{\hat{X}_t}^T \psi^\Pi c - l_{\hat{X}_t}^T (\beta_t \eta \psi_1 c) \right] \mu_P(dt, dv), \quad (16)$$

где

$$\sigma_{11} = \psi^\Pi \nu \psi^{\Pi T}; \quad \sigma_{12} = \psi^\Pi \nu (\beta_t \eta \psi_1)^T; \\ \sigma_{22} = (\beta_t \eta \psi_1) \nu (\beta_t \eta \psi_1)^T;$$

$$\mu_P((t_1, t_2], A) = M P((t_1, t_2], A) = \\ = \int_{t_1}^{t_2} \nu_P(s, A) ds, \quad A \subset R^q;$$

$$l_{X_t}^T = \left[\frac{\partial l}{\partial X_1}, \dots, \frac{\partial l}{\partial X_{n_x}} \right];$$

$$l_{\hat{X}_t}^T = \left[\frac{\partial l}{\partial \hat{X}_1}, \dots, \frac{\partial l}{\partial \hat{X}_{n_x}} \right];$$

$$l_{X_t \hat{X}_t} = \left[\frac{\partial^2 l}{\partial X_i \partial \hat{X}_j} \right]; \quad l_{X_t X_t} = \left[\frac{\partial^2 l}{\partial X_i \partial X_j} \right];$$

$$l_{\hat{X}_t \hat{X}_t} = \left[\frac{\partial^2 l}{\partial \hat{X}_i \partial \hat{X}_j} \right];$$

$$f_{X_t X_t} : D_t = [\text{tr}(f_{ij, X_t X_t} D_t)], \quad f = [f_{ij}];$$

$$f_{ij, X_t X_t} = \left[\frac{\partial^2 f_{ij}}{\partial X_k \partial X_s} \right];$$

$$D_t = [D_{ij}] \quad (i, j, k, s = \overline{1, n_x}); \quad (17)$$

$\nu_P(t, dv)$ — интенсивность пуассоновского процесса $P([0, t], A)$. В (16), (17) производные l_{X_t} , $l_{\hat{X}_t}$, $l_{X_t X_t}$ и $l_{\hat{X}_t \hat{X}_t}$ вычисляются в точке (X_t, \hat{X}_t) ; функции φ^Π , φ_1 , ψ^Π и ψ_1 — в точке (X_t, Y_t, t) ; $\psi^\Pi c$ и $\psi_1 c$ — в точке (X_t, Y_t, t, v) , а ξ и η — в точке (\hat{X}_t, Y_t, t) .

Выражение для дифференциала dI_t содержит линейные и квадратичные по β_t слагаемые в σ_{12} и σ_{22} , а также линейные по α_t слагаемые. Поэтому, предполагая, что $l_{\hat{X}_t \hat{X}_t}$ — положительно определенная матрица для всех X_t, \hat{X}_t , коэффициент β_t определим из условия минимума dI_t по β_t :

$$\frac{\partial dI_t}{\partial \beta_t} = M \left[l_{\hat{X}_t} \varphi_1^T \eta^T + \left(l_{X_t \hat{X}_t} \psi^\Pi + l_{\hat{X}_t \hat{X}_t} \beta_t \eta \psi_1 \right) \nu \psi_1^T \eta^T \right] dt + M \int_{R_0^q} \left[\frac{\partial l(X_t + \psi^\Pi c, \hat{X}_t + \beta_t \eta \psi_1 c)}{\partial \beta_t} - l_{\hat{X}_t} \psi_1 c^T \eta^T \right] \mu_P(dt, dv) = 0, \quad (18)$$

где

$$\frac{\partial l(X_t + \psi^\Pi c, \hat{X}_t + \beta_t \eta \psi_1 c)}{\partial \beta_t} = \left[\frac{\partial l(X_t + \psi^\Pi c, \hat{X}_t + \beta_t \eta \psi_1 c)}{\partial \beta_{ij}} \right]$$

есть матрица размерности $(n_x \times s)$ ($i = \overline{1, n_x}$, $j = \overline{1, n_y}$).

Определить коэффициенты α_t аналогично тому, как это сделано для β_t , нельзя, так как этот минимум не существует для $\alpha_t \in R^{n_x r}$.

Коэффициент α_t в (15) будем определять из условия

$$M(l_{\hat{X}_t} \xi^T) = 0, \quad (19)$$

которое должно выполняться в каждый момент времени t . Для этого необходимо и достаточно, чтобы условие (19) выполнялось в начальный момент времени t_0 и производная $M(l_{\hat{X}_t} \xi^T)$ по t в каждый момент времени тождественно равнялась нулю. В результате после дифференцирования (19) по обобщенной формуле Ито для коэффициента α_t получаем следующее уравнение:

$$\begin{aligned} & M(l_{\hat{X}_t} \xi^T) + M(l_{\hat{X}_t X_t} \varphi^{\Pi} \xi^T) + \\ & + M[l_{\hat{X}_t \hat{X}_t} (\alpha_t \xi + \beta_t \eta \varphi_1) \xi^T] + \\ & + M[l_{\hat{X}_t} (\alpha_t \xi + \beta_t \eta \varphi_1)^T \xi_{\hat{X}_t}^T] + M(l_{\hat{X}_t} \varphi_1^T \xi_{Y_t}^T) + \\ & + \frac{1}{2} M[(l_{\hat{X}_t} \xi^T)_{sv} : \sigma] + \\ & + \int_{R_0^q} M[l_{\hat{X}_t} (X_t + \psi^{\Pi} c, \hat{X}_t + \beta_t \eta \psi_1 c) \times \\ & \times \xi (Y_t + \psi_1 c, \hat{X}_t + \beta_t \eta \psi_1 c, t)^T - \\ & - l_{\hat{X}_t} \xi^T - l_{\hat{X}_t X_t} \psi^{\Pi} c \xi^T - l_{\hat{X}_t} \psi_1 c^T \xi_{Y_t} - \\ & - l_{\hat{X}_t \hat{X}_t} (\beta_t \eta \psi_1 c) \xi^T - l_{\hat{X}_t} (\beta_t \eta \psi_1 c)^T \xi_{\hat{X}_t}^T] \nu_P(t, dv) = \\ & = 0, \quad (20) \end{aligned}$$

где

$$\sigma = B_t \nu B_t^T, \quad B_t = [\psi^{\Pi T} \psi_1^T (\beta_t \eta \psi_1)^T]^T. \quad (21)$$

Коэффициент γ_t в (15) может быть определен из условия несмещенности оценки $M X_t = M \hat{X}_t$ в каждый момент времени, т. е.

$$\gamma_t = M \varphi^{\Pi} - \alpha_t M \xi - \beta_t M (\eta \varphi_1). \quad (22)$$

Таким образом, получено следующее утверждение.

Теорема 4.1. Пусть векторный СтП $[X_t^T Y_t^T]^T$ удовлетворяет приведенным стохастическим дифференциальным уравнениям (5) и имеет конечные одномерные вероятностные моменты. Тогда при отсутствии ограничений на коэффициент α_t значения коэффициентов α_t , β_t и γ_t БК УОФ (15) определяются согласно (18)–(22).

Пусть теперь матричный коэффициент α_t лежит в некотором компакте

$$\alpha_t \in Q \subset R^{n_x q}, \quad \alpha_t = [\alpha_{ij}] \quad (i = \overline{1, n_x}, j = \overline{1, r}). \quad (23)$$

Будем по-прежнему определять его из условия минимума дифференциала dI_t в каждый момент времени t . В этом случае коэффициент α_t определится из условия

$$\alpha_t = \arg \min_{\alpha_t \in Q} M(l_{\hat{X}_t}^T \alpha_t \xi). \quad (24)$$

Таким образом, приходим к следующему результату.

Теорема 4.2. Пусть векторный СтП $[X_t^T Y_t^T]^T$ удовлетворяет стохастическим дифференциальным уравнениям (5) и имеет конечные одномерные моменты. Тогда оценка \hat{X}_t описывается уравнением (15) с ограничением (23). При этом значения коэффициентов усиления α_t , β_t и γ_t фильтра по БК (11) определяются уравнениями (18), (22) и (24).

В частности, если коэффициенты усиления α_{ij} , $\alpha_t = [\alpha_{ij}]$, ограничены по модулю:

$$|\alpha_{ij}| \leq \alpha_{ij}^m, \quad \alpha_{ij}^m > 0 \quad (i = \overline{1, n_x}; j = \overline{1, r}),$$

т. е. компакт Q представляет собой параллелепипед, то решением задачи (24) будет

$$\alpha_t = -n \operatorname{sign} M(l_{\hat{X}_t} \xi^T), \quad \alpha_{ij}^m = [\alpha_{ij}^m].$$

Замечание 4.1. Уравнение (19) имеет простой геометрический смысл. Рассмотрим сужение функций l и ξ на пространстве R^{n_x} оценок \hat{X}_t . В любой точке \hat{X}_t этого пространства вектор ξ должен быть ортогонален (в смысле скалярного произведения, задаваемого оператором математического ожидания) вектору нормали $l_{\hat{X}_t}$ к поверхности уровня функции l , проходящей через эту точку \hat{X}_t , т. е. вектор ξ должен касаться поверхностей уровня функции l .

4.2 Эквивалентность непрерывных по байесову критерию условно-оптимальных фильтров

При отсутствии скачкообразных компонент в уравнениях (5), т. е. когда $\psi^{\Pi} c = \psi_1 c = 0$, уравнения (18) и (20) для коэффициентов α_t и β_t упрощаются:

$$M[l_{\hat{X}_t} \varphi_1^T \eta^T + (l_{X_t \hat{X}_t} \psi^{\Pi} + l_{\hat{X}_t \hat{X}_t} \beta_t \eta \psi_1) \nu \psi_1^T \eta^T] = 0;$$

$$\begin{aligned} & \mathbf{M} \left(l_{\hat{X}_t} \xi_t^T \right) + \mathbf{M} \left(l_{\hat{X}_t X_t} \varphi^{\Pi} \xi_t^T \right) + \\ & \quad + \mathbf{M} \left[l_{\hat{X}_t \hat{X}_t} (\alpha_t \xi + \beta_t \eta \varphi_1) \xi_t^T \right] + \\ & \quad + \mathbf{M} \left[l_{\hat{X}_t} (\alpha_t \xi + \beta_t \eta \varphi_1)^T \xi_{\hat{X}_t}^T \right] + \mathbf{M} \left(l_{\hat{X}_t} \varphi_1^T \xi_t^T \right) + \\ & \quad + \frac{1}{2} \mathbf{M} \left[\left(l_{\hat{X}_t} \xi_t^T \right)_{vv} : \sigma \right] = 0. \end{aligned}$$

При этом коэффициент γ_t определяется уравнением (22).

Полученные уравнения могут быть положены в основу замены негауссовского белого шума эквивалентным гауссовским белым шумом $\nu_{\vartheta} = \nu_{\vartheta}(t)$.

Замечание 4.2. Задача определения структурных коэффициентов БК УОФ (15) сводится к задаче определения одномерного распределения вектора $[X_t^T Y_t^T \hat{X}_t^T]^T$, определяемого уравнениями (5) и (15). Точные и приближенные методы для приведенных уравнений описаны в [1–5].

4.3 Непрерывные нормальные по байесову критерию условно-оптимальные фильтры

Применим приближенный метод [4] определения нормального одномерного распределения вектора $Z_t = [Y_t^T X_t^T \hat{X}_t^T]^T$, удовлетворяющего уравнениям (5) и (22), записанного в виде

$$\dot{Z}_t = a^{\Pi}(Z_t, t) + b^{\Pi}(Z_t, t) V.$$

Здесь введены следующие обозначения:

$$\begin{aligned} a^{\Pi} &= a^{\Pi}(Z_t, t) = \begin{bmatrix} \varphi_1 \\ \varphi^{\Pi} \\ \alpha_t \xi + \beta_t \eta \varphi_1 + \gamma_t \end{bmatrix}; \\ b^{\Pi} &= b^{\Pi}(Z_t, t) = \begin{bmatrix} \psi_1 \\ \varphi^{\Pi} \\ \beta_t \eta \psi_1 \end{bmatrix}. \end{aligned}$$

В результате приходим к уравнениям **теорем 4.3** и **4.4** для нормального распределения $\mathcal{N} = \mathcal{N}(m^Z, K^Z)$. Вектор математического ожидания m_t^Z и ковариационная матрица K_t^Z определяются из следующих уравнений:

$$\begin{aligned} \dot{m}_t^Z &= B_1 = \mathbf{M}_{\mathcal{N}} a^{\Pi}, & m_{t_0}^Z &= m_0^Z; \\ \dot{K}_t^Z &= B_2 = B_{21} + B_{21}^T + B_{22}, & K_{t_0}^Z &= K_0^Z, \end{aligned}$$

где

$$B_{21} = \mathbf{M}_{\mathcal{N}} a^{\Pi} (Z_t^T - m_t^{ZT}) = \left[\frac{\partial}{\partial m_t^Z} B_1^T \right]^T K_t^Z;$$

$$B_{22} = \mathbf{M}_{\mathcal{N}} b^{\Pi T} \nu) t) b^{\Pi},$$

а интенсивность белого шума определяется согласно (6).

4.4 О дискретных по байесову критерию условно-оптимальных фильтрах

Уравнения БК УОФ теорем 4.3 и 4.4 для приведенных дискретных систем (7) и (8) получаются предельным переходом из теорем 4.1 и 4.2 при $N = \min_{ij} \alpha_{ij}^m \rightarrow \infty$. Соответствующие результаты, в том числе для нормальных БК УОФ и обобщенных фильтров Калмана–Бьюси и Калмана, подробно рассмотрены в [4, 5, 7].

5 Заключение

Разработаны методы условно-оптимальной фильтрации по БК для обработки информации в реальном масштабе времени для неявных непрерывных и дискретных СтС, приводимых к явным. Предполагается, что наблюдения влияют на объект наблюдения. Особое внимание уделено приближенному синтезу БК УОФ на основе метода нормальной аппроксимации.

Полученные методы допускают обобщение на случай приводимых неявных СтС, зависящих от случайных параметров, определяемых каноническими представлениями [9, 11, 12], эрмитарных СтС [11–15], а также управляемых СтС [16–20].

Наконец, отметим, что разработанные методы БК УОФ нуждаются в дальнейшем развитии на случай, когда требование гладкости функции потерь можно ослабить.

Литература

1. Сеницын И. Н. Методы условно-оптимальной фильтрации и экстраполяции в наблюдаемых неявных стохастических системах // Информатика и её применения, 2024. Т. 18. Вып. 4. С. 2–9. doi: 10.14357/19922264240401. EDN: TFRJYK.
2. Сеницын И. Н. Условно-оптимальная фильтрация и экстраполяция в неявных дифференциальных гауссовских стохастических системах при автокоррелированной помехе в наблюдениях // Информатика и её применения, 2024. Т. 18. Вып. 4. С. 19–25. doi: 10.14357/19922264240403. EDN: CVUETK.
3. Сеницын И. Н. Дискретное условно-оптимальное оценивание в неявных наблюдаемых стохастических системах // Системы и средства информатики, 2024. Т. 34. № 4. С. 16–30. doi: 10.14357/08696527240402. EDN: TLHYUA.
4. Сеницын И. Н. Методы нормальной условно-оптимальной фильтрации для наблюдаемых неявных стохастических систем // Информатика и её применения, 2025. Т. 19. Вып. 1. С. 44–51. doi: 10.14357/19922264250106. EDN: TEPFUE.
5. Сеницын И. Н. Методы условно-оптимальной фильтрации по сложному статистическому критерию для наблюдаемых неявных стохастических систем // Информатика и её применения, 2025. Т. 19. Вып. 2. С. 27–34. doi: 10.14357/19922264250204. EDN: QMIXFW.

6. Синицын И. Н., Мошук Н. К., Шин В. И. Условно-оптимальная фильтрация в стохастических дифференциальных системах по бейсовым критериям // Докл. Акад. наук, 1993. Т. 330. № 4. С. 437–440.
7. Синицын И. Н., Шин В. И., Корепанов Э. Р. Теория условно-оптимальной фильтрации стохастических процессов по сложно-статистическим критериям // Системы и средства информатики, 1993. № 5. С. 106–120.
8. Синицын И. Н., Шин В. И., Мошук Н. К. Общая теория условно-оптимальной фильтрации процессов в стохастических дифференциальных системах // Системы и средства информатики, 1995. № 7. С. 75–85.
9. Пугачёв В. С. Теория случайных функций и её применение к задачам автоматического управления. — 3-е изд. — М.: Физматлит, 1962. 884 с.
10. Синицын И. Н. Фильтры Калмана и Пугачёва. — 2-е изд. — М.: Логос, 2007. 776 с.
11. Пугачёв В. С., Синицын И. Н. Теория стохастических систем. — 2-е изд. — М.: Логос, 2004. 1000 с.
12. Синицын И. Н. Канонические представления случайных функций. Теория и применения. — 2-е изд. — М.: ТОРУС ПРЕСС, 2023. 816 с.
13. Колмановский В. Б., Носов В. Р. Устойчивость и периодические режимы регулируемых систем с последствием. — М.: Наука, 1981. 448 с.
14. Финогенко И. А. О неявных функционально-дифференциальных уравнениях в банаховом пространстве // Динамика нелинейных систем. — Новосибирск: Наука, 1983. С. 151–164.
15. Азбелев Н. В., Максимов В. П., Рахматулина Л. Ф. Введение в теорию функционально-дифференциальных уравнений. — М.: Наука, 1991. 280 с.
16. Борисов А. В. Оптимальная фильтрация в системах с вырожденными шумами в наблюдениях // Автоматика и телемеханика, 1998. № 11. С. 32–45.
17. Босов А. В. Управление линейным выходом марковской цепи по квадратичному критерию. Случай полной информации // Информатика и её применения, 2022. Т. 16. Вып. 2. С. 19–26. doi: 10.14357/19922264220203. EDN: FEQKUN.
18. Коновалов М. Г., Разумчик Р. В. Синтез управления двумерным случайным блужданием с эталонным стационарным распределением // Информатика и её применения, 2022. Т. 16. Вып. 2. С. 109–117. doi: 10.14357/19922264220214. EDN: WVMOVH.
19. Босов А. В. Исследование робастности численных аппроксимаций фильтра Вонэма // Информатика и её применения, 2023. Т. 17. Вып. 2. С. 41–49. doi: 10.14357/19922264230206. EDN: BGILKR.
20. Босов А. В. Оптимальная фильтрация состояния нелинейной динамической системы по наблюдениям со случайными запаздываниями // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 8–17. doi: 10.14357/19922264230302. EDN: CFVYJM.

Поступила в редакцию 23.12.2024

Принята к публикации 15.08.2025

BAYES CRITERION CONDITIONALLY OPTIMAL FILTERING METHODS FOR OBSERVABLE IMPLICIT STOCHASTIC SYSTEMS

I. N. Sinitsyn

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

Abstract: The paper is devoted to methods of conditionally optimal filter (COF) synthesis by Bayes criterion (BC) in continuous and discrete implicit non-Gaussian stochastic systems (StS) reducible to explicit. A short COF survey by mean square, energetic, and complex statistical criteria for explicit and implicit continuous and discrete StS is given. Reduction methods for smooth and nonsmooth implicit functions are developed. Exact and approximate (based on normal approximation and statistical linearization) methods for BC COF in reducible implicit continuous and discrete StS are considered. Special attention is paid to normal BC COF. The problem of equivalence of non-Gaussian noises in BC COF is discussed. Future directions of research and applications are presented.

Keywords: Bayes criterion (BC); conditionally optimal filter (COF); implicit StS; normal COF; stochastic system (StS)

DOI: 10.14357/19922264250301

EDN: VTVRXX

References

1. Sinitsyn, I. N. 2024. Metody uslovno-optimal'noy fil'tratsii i ekstrapolyatsii v nablyudaemykh neyavnykh stokhasticheskikh sistemakh [Conditionally optimal filtering and extrapolation methods for observable implicit stochastic systems]. *Informatika i ee Primeneniya* — *Inform. Appl.* 18(4):2–9. doi: 10.14357/19922264240401. EDN: TFPJYK.
2. Sinitsyn, I. N. 2024. Uslovno-optimal'naya fil'tratsiya i ekstrapolyatsiya v neyavnykh differentsial'nykh gaussovskikh stokhasticheskikh sistemakh pri avtokorrelirovannoy pomekhe v nablyudeniyakh [Conditionally optimal filtering and extrapolation for differential Gaussian implicit

- stochastic systems at autocorrelated noise in observations]. *Informatika i ee Primeneniya — Inform. Appl.* 18(4):19–25. doi: 10.14357/19922264240403. EDN: CVUETK.
3. Sinitsyn, I. N. 2024. Diskretnoe uslovno-optimal'noe otsenivanie v neyavnykh nablyudaemykh stokhasticheskikh sistemakh [Discrete conditionally-optimal estimation in observable implicit stochastic systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(4):16–30. doi: 10.14357/08696527240402. EDN: TLHYA.
 4. Sinitsyn, I. N. 2025. Metody normal'noy uslovno-optimal'noy fil'tratsii dlya nablyudaemykh neyavnykh stokhasticheskikh sistem [Normal conditionally-optimal filtering methods for implicit stochastic systems]. *Informatika i ee Primeneniya — Inform. Appl.* 19(1):44–51. doi: 10.14357/19922264250106. EDN: TEPFUE.
 5. Sinitsyn, I. N. 2025. Metody uslovno-optimal'noy fil'tratsii po slozhno-statisticheskomu kriteriyu dlya nablyudaemykh neyavnykh stokhasticheskikh sistem [Complex statistical criterion conditionally optimal filtering methods for observable implicit stochastic systems]. *Informatika i ee Primeneniya — Inform. Appl.* 19(2):27–34. doi: 10.14357/19922264250204. EDN: QMIXFW.
 6. Sinitsyn, I. N., N. K. Moshchuk, and V. I. Shin. 1993. Conditionally optimal filtering of processes in stochastic differential systems according to Bayesian criteria. *Dokl. Math.* 47(3):528–533.
 7. Sinitsyn, I. N., V. I. Shin, and E. R. Korepanov. 1993. Teoriya uslovno-optimal'noy fil'tratsii stokhasticheskikh protsessov po slozhno-statisticheskim kriteriyam [Theory of conditionally optimal filtering of stochastic processes according to complex statistical criteria]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 5:106–120.
 8. Sinitsyn, I. N., V. I. Shin, and N. K. Moshchuk. 1995. Obshchaya teoriya uslovno-optimal'noy fil'tratsii protsessov v stokhasticheskikh differentsial'nykh sistemakh [General theory of conditionally optimal filtering of processes in stochastic differential systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 7:75–85.
 9. 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]. 3rd ed. Moscow: Fizmatgiz. 884 p.
 10. Sinitsyn, I. N. 2007. *Fil'try Kalmana i Pugacheva* [Kalman and Pugachev filters]. 2nd ed. Moscow: Logos. 776 p.
 11. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 908 p.
 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.
 13. Kolmanovskiy, V. B., and V. R. Nosov. 1981. *Ustoychivost' i periodicheskie rezhimy reguliruemyykh sistem s posledstviem* [Stability and periodic modes of regulated systems with consequences]. Moscow: Nauka. 448 p.
 14. Finogenko, I. A. 1983. O neyavnykh funktsional'no-differentsial'nykh uravneniyakh v banakhovom prostranstve [On implicit functional differential equations in a Banach space]. *Dinamika nelineynykh sistem* [Dynamics of nonlinear systems]. Novosibirsk: Nauka. 151–164.
 15. Azbelev, N. V., V. P. Maksimov, and L. F. Rakhmatulina. 1991. *Vvedenie v teoriyu funktsional'no-differentsial'nykh uravneniy* [Introduction to the theory of functional differential equations]. Moscow: Nauka. 280 p.
 16. Borisov, A. 1998. Optimal filtering in systems with degenerate noises in observations. *Automat. Rem. Contr.* 59(11,1):1526–1537. EDN: LFDCVH.
 17. Bosov, A. V. 2022. Upravlenie lineynym vykhodom markovskoy tsepi po kvadraticnomu kriteriyu. Sluchay polnoy informatsii [Linear output control of Markov chain by square criterion. Complete information case]. *Informatika i ee Primeneniya — Inform. Appl.* 16(2):19–26. doi: 10.14357/19922264220203. EDN: FEQKUN.
 18. Konovalov, M. G., and R. V. Razumchik. 2022. Sintez upravleniya dvmernym sluchaynym bluzhdaniem s etalonnym stacionarnym raspredeleniem [Controlling a bounded two-dimensional Markov chain with a given invariant measure]. *Informatika i ee Primeneniya — Inform. Appl.* 16(2):109–117. doi: 10.14357/19922264220214. EDN: WVMOBH.
 19. Bosov, A. V. 2023. Issledovanie robustnosti chislennykh approksimatsiy fil'tra Vonema [Robustness investigation of the numerical approximation of the Wonham filter]. *Informatika i ee Primeneniya — Inform. Appl.* 17(2):41–49. doi: 10.14357/19922264230206. EDN: BGILKR.
 20. Bosov, A. V. 2023. Optimal'naya fil'tratsiya sostoyaniya nelineynoy dinamicheskoy sistemy po nablyudeniym so sluchaynymi zapazdyvaniyami [Nonlinear dynamic system state optimal filtering by observations with random delays]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):8–17. doi: 10.14357/19922264230302. EDN: CFVYJM.

Received December 23, 2024

Accepted August 15, 2025

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; sinitsin@dol.ru

СТАБИЛИЗАЦИЯ ТРАЕКТОРИИ ЛИНЕЙНОЙ СИСТЕМЫ СО СКАЧКООБРАЗНЫМ ДРЕЙФОМ ПРИ ЦЕЛОЧИСЛЕННЫХ ОГРАНИЧЕНИЯХ НА УПРАВЛЕНИЕ*

А. В. Босов¹

Аннотация: Рассматривается целочисленный вариант задачи стабилизации линейной стохастической дифференциальной системы, дрейфующей в скачкообразно изменяющемся направлении, формируемом цепью Маркова. Цель управления формализуется квадратичным функционалом качества. В зависимости от условий возможны варианты с полной информацией (состояние цепи известно) и косвенными наблюдениями (состояние системы служит косвенным наблюдением за неизвестным состоянием цепи). Особенность постановки заключается в целочисленных ограничениях на допустимые значения управления. В отличие от решенной ранее задачи без ограничений, для «целочисленной» постановки условия существования решения не выполняются, поэтому исследуется ε -оптимальное решение. Найти ε -оптимальное управление можно в результате дискретизации оптимального решения в задаче без ограничений и применения смешанно-целочисленного нелинейного программирования. Однако стохастический характер задачи и большое число вариантов переключений не позволяют гарантировать вычислительную реализуемость решения методом динамического программирования. Для практической реализации использован метод релаксации: рассчитывается эвристическая аппроксимация как результат целочисленного преобразования ε -оптимального управления в задаче без ограничений. Предложены три варианта таких преобразований. Проведен численный эксперимент на той же прикладной модели, что использовалась в предыдущих работах по управлению без ограничений (динамика положения простого механического привода). Результаты расчетов прежде всего подтверждают применимость предлагаемых решений с точки зрения целевой задачи стабилизации, а также позволяют сравнить характер стратегий релаксации.

Ключевые слова: стабилизация линейной системы; квадратичный функционал качества; динамическое программирование; управление с обратной связью; фильтр Вонэма; смешанно-целочисленное нелинейное программирование; метод релаксации; механический привод

DOI: 10.14357/19922264250302

EDN: YZENCC

1 Введение

Задачи оптимизации и, в частности, оптимального управления состоянием динамической системы встречаются в многочисленных научных, технических, инженерных, финансовых, государственных и прочих приложениях. Небольшая, но интересная часть этих задач имеет особые свойства, описываемые дискретными (конечными, счетными) ограничениями [1]. Для решения достаточно успешно используются методы смешанно-целочисленного нелинейного программирования (mixed-integer nonlinear programming, MINP) [2]. В этих методах сочетаются интеллектуальный перебор в пространствах целочисленных ограничений и, следовательно, комбинаторная сложность и методы нелинейной оптимизации. Поскольку речь об оптимальном управлении, то к последним естественно отнести принцип максимума [3] и динами-

ческое программирование [4]. Задачи оптимального управления с ограничениями, состоящими в том, что состояние, измерения и/или управления принимают целые значения, становятся значительно сложнее. Это смешанные задачи, где есть и оптимальное управление, и целочисленная оптимизация.

С формальной точки зрения ничего принципиально нового смешанный тип задач не приносит. На первом плане остаются обычные для оптимального управления вопросы: выпуклая ли задача и можно ли применять критерии оптимальности; если задача невыпуклая, то какими методами сделать ее выпуклой и преобразовать в решаемую задачу оптимизации. Но на втором плане возникают вопросы целочисленной оптимизации, сложность которых скорее вычислительного характера, чем фундаментального. Целочисленная оптимизация — это выбор из конечного или счетного множества вари-

* 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, ABosov@frcscs.ru

антов, т.е. комбинаторика, перебор и NP-сложные задачи. Успешно применяются существующие в большом числе и очень разнообразные эвристические подходы, позволяющие перевести нереализуемые оптимальные решения полного перебора в сферу поиска приемлемых, практически реализуемых аппроксимаций. Общие сведения по этой тематике хорошо отражены в обзоре [1].

Задача MINP, которой посвящена данная статья, относится к области стохастического оптимального управления. Приложений MINP здесь обнаружить не удастся, обычно исследователи решают задачи в детерминированной постановке, когда оптимальное управление и оптимальная траектория существуют и задаются или хотя бы аппроксимируются обычными функциями, а не распределениями вероятностей. Рассматриваемая задача управления состоит в стабилизации линейной системы около траектории, формируемой марковской цепью, т.е. дрейфующей в скачкообразно изменяющихся направлениях. Кроме того, на состояние системы влияет аддитивный гауссовский шум, так что целевое положение еще требуется оценить по косвенному зашумленному измерению. Решение обеспечивает:

- (1) квадратичный функционал, задающий цель стабилизации в виде среднего квадрата отклонения состояния от целевой марковской траектории;
- (2) фильтр Вонэма [5], формирующий оптимальную оценку цепи;
- (3) принцип разделения, сводящий задачу к постановке с полной информацией и решению уравнения Беллмана.

Подробно это решение изучено в [6–8], а в данной статье на управление накладывается дополнительное целочисленное ограничение. Соответственно, в круг изучаемых вопросов входит принципиальная возможность стабилизации системы целочисленным управлением, существование решения и методы практической реализации стратегии стабилизации.

Стохастическая постановка препятствует использованию большинства методов MINP; перспективным оказывается лишь метод релаксации (relaxing integrality). Его идея состоит в том, чтобы сначала отказаться от целочисленных ограничений и решить задачу оптимизации на более широком классе, а затем полученное решение адаптировать к целочисленным ограничениям [9, 10]. Этот подход хорош еще и тем, что позволяет очень удобно оценить возможные потери при аппроксимации, так как «нижняя» граница получается автомати-

чески при решении оптимизационной задачи со снятыми (релаксированными) ограничениями.

2 Стабилизация системы со скачкообразным дрейфом при целочисленных ограничениях на управление

Для постановки задачи на каноническом вероятностном пространстве $(\Omega, \mathcal{F}, \mathcal{P}, \mathcal{F}_t)$, $t \in [0, T]$, рассмотрим линейную дифференциальную стохастическую систему с состоянием $z_t \in \mathbb{R}^{n_z}$:

$$dz_t = a_t y_t dt + b_t z_t dt + c_t u_t dt + \sigma_t dw_t, \quad z_0 = Z. \quad (1)$$

Под действием управления $u_t \in \mathbb{Z}^{n_u}$ траектории z_t формируются в направлениях (дрейфуют), которые определяет марковский скачкообразный процесс $y_t \in \mathbb{R}^{n_y}$ — цепь с конечным числом состояний и значениями в множестве единичных координатных векторов $\{e_1, \dots, e_{n_y}\}$ в \mathbb{R}^{n_y} . Здесь \mathbb{R}^n — пространство n -мерных векторов с вещественными элементами; \mathbb{Z}^n — пространство n -мерных векторов с целыми элементами. Предполагаются известными распределение π_0 начального состояния $y_0 = Y$ цепи y_t , матрица интенсивностей переходов Λ_t и функции времени $a_t \in \mathbb{R}^{n_z \times n_y}$, $b_t \in \mathbb{R}^{n_z \times n_z}$, $c_t \in \mathbb{R}^{n_z \times n_u}$ и $\sigma_t \in \mathbb{R}^{n_z \times n_w}$. Возмущение $w_t \in \mathbb{R}^{n_w}$ — стандартный векторный винеровский процесс, начальное условие $Z \in \mathbb{R}^{n_z}$ — гауссовская случайная величина с известным математическим ожиданием и ковариацией.

Цель стабилизации формализует целевой функционал

$$J(U_0^T) = \mathbb{E} \left\{ \int_0^T \|P_t y_t + Q_t z_t + R_t u_t\|_{S_t}^2 dt + \|P_T y_T + Q_T z_T\|_{S_T}^2 \right\}, \quad (2)$$

где $U_0^T = \{u_t = U_t(z_t), 0 \leq t \leq T\}$, $P_t \in \mathbb{R}^{n_J \times n_y}$, $Q_t \in \mathbb{R}^{n_J \times n_z}$, $R_t \in \mathbb{R}^{n_J \times n_u}$, $S_t \in \mathbb{R}^{n_J \times n_J}$, $S_t \geq 0$, $S_t = S'_t$, $0 \leq t \leq T$, — заданные матричные функции; весовая функция $\|x\|_S^2 = x' S x$ для симметричной неотрицательно определенной матрицы S , единичной матрице $S = \mathbf{1}$ соответствует евклидова норма $\|x\|_1^2 = |x|^2$, x' — транспонированная матрица (вектор) x . Таким образом, класс допустимых управлений определяется как класс управлений с обратной связью по состоянию z_t [11], причем вариант $u_t = U_t(z_t)$ означает, что состояние цепи y_t неизвестно, а состояние системы z_t , таким образом, служит косвенным наблюдением.

Постановка с управлением $u_t \in \mathbb{R}^{n_u}$, а также вариант с полной информацией (допустимым управлением $u_t = U_t(y_t, z_t)$ и принципом разделения) изучена в [6–8]. Условия оптимальности таковы: w_t, y_t, Y и Z независимы в совокупности; u_t — случайный процесс с конечным вторым моментом; функции $\Lambda_t, a_t, b_t, c_t, \sigma_t, P_t, Q_t, R_t$ и S_t кусочно-непрерывны; выполнены условия невырожденности $R'_t S_t R_t > 0$ и $\sigma_t \sigma'_t > 0$.

В контур оптимального управления $u_t^* = U_t(z_t^*)$ входит оценка $\hat{y}_t = \mathbb{E}\{y_t | \mathcal{F}_t\}$ состояния цепи по косвенным наблюдениям z_t , которую задает фильтр Вонэма [5]:

$$\begin{aligned} d\hat{y}_t &= \Lambda'_t \hat{y}_t dt + (\text{diag}(\hat{y}_t) - \hat{y}_t \hat{y}'_t) a'_t (\sigma_t \sigma'_t)^{-1/2} \times \\ &\times (\sigma_t \sigma'_t)^{-1/2} (dz_t - a_t \hat{y}_t dt - b_t z_t dt - c_y u_t dt), \\ \hat{y}_0 &= \mathbb{E}\{Y\}. \end{aligned} \quad (3)$$

Оптимальное управление без целочисленных ограничений имеет вид:

$$\begin{aligned} u_t^* &= -\frac{1}{2} (R'_t S_t R_t)^{-1} (c'_t (2\alpha_t z_t^* + \beta \hat{y}_t) + \\ &+ 2R'_t S_t (P_t \hat{y}_t + Q_t z_t^*)); \end{aligned} \quad (4)$$

$$\left. \begin{aligned} \frac{d\alpha_t}{dt} - (M_t^\alpha \alpha_t + \alpha'_t (M_t^\alpha)') + N_t^\alpha - \\ - \alpha'_t c_t (R'_t S_t R_t)^{-1} c'_t \alpha_t = 0, \quad \alpha_T = Q'_T S_T Q_T; \\ \frac{d\beta_t}{dt} + \beta_t \Lambda'_t + M_t^\beta - N_t^\beta \beta_t = 0, \\ \beta_T = 2Q'_T S_T P_T; \\ \frac{\partial \gamma_t}{\partial t} + y' \Lambda_t \Lambda'_t + M_t^\gamma y = 0, \quad \gamma_T y = y' P'_T S_T P_T y; \\ \frac{\partial \delta_t}{\partial t} + \text{tr} \{ \sigma'_t \alpha_t \sigma_t \} = 0, \quad \sigma_T = 0, \end{aligned} \right\} (5)$$

где

$$\begin{aligned} M_t^\alpha &= Q'_t S_t R_t (R'_t S_t R_t)^{-1} c'_t; \\ N_t^\alpha &= Q'_t (S_t - S_t R_t (R'_t S_t R_t)^{-1} R'_t S_t) Q_t; \\ M_t^\beta &= 2 \left((a'_t - P'_t S_t R_t (R'_t S_t R_t)^{-1} c'_t) \alpha_t + \right. \\ &\quad \left. + P'_t (S_t - S_t R_t (R'_t S_t R_t)^{-1} R'_t S_t) Q_t \right); \\ N_t^\beta &= Q'_t S_t R_t (R'_t S_t R_t)^{-1} c'_t + \alpha_t c_t (R'_t S_t R_t)^{-1} c'_t; \\ M_t^\gamma y &= N_t^\gamma (y) = \\ &= y' \beta'_t (a_t - c_t (R'_t S_t R_t)^{-1} R'_t S_t P_t) y + \\ &+ y' P'_t (S_t - S_t R_t (R'_t S_t R_t)^{-1} R'_t S_t) P_t y - \\ &- \frac{1}{4} y' \beta'_t c_t (R'_t S_t R_t)^{-1} c'_t \beta_t y. \end{aligned}$$

Параметры из уравнений (5) нужны для определения функции Беллмана

$$\begin{aligned} V_t = V_t(z) = \inf_{U_t^T} \mathbb{E} \left\{ \int_t^T \|P_s y_s + Q_s z_s + R_s u_s\|_{S_s}^2 ds + \right. \\ \left. + \|P_T y_T + Q_T z_T\|_{S_T}^2 \right\} \end{aligned}$$

в виде $V_t^* = z' \alpha_t z + z' \beta_t \hat{y}_t + \gamma_t \hat{y}_t + \delta_t$, так что на оптимальной траектории z_t^* цена управления $V_0^* = Z' \alpha_0 Z + Z' \beta_0 (\hat{y}_0) + \gamma_0 \hat{y}_0 + \delta_0$. Таким образом, нижняя граница функционала (2) для управлений $u_t \in \mathbb{R}^{n_u}$ имеет вид:

$$J((U^*)_0^T) = \inf J(U_0^T) = \mathbb{E}\{V_0^*\}. \quad (6)$$

Теперь, имея детальное решение задачи без ограничений $u_t \in \mathbb{R}^{n_u}$, исследуем решение задачи с дискретными ограничениями $u_t \in Z^{n_u}$.

Изучение управления традиционными детерминированными моделями MINP начинается с упрощения [12], когда u_t принимает всего два значения (бинарное управление), т. е. $u_t \in \{-1, 1\}$, и заданного конечного числа «переключений» (переключением естественно назвать момент, когда u_t изменяет значение). Можно обобщить это ограничение и считать бинарным каждый элемент $(u_t)_k \in \{-1, 1\}$, $k = \overline{1, n_u}$. Такое сужение класса допустимых управлений в совокупности с выпуклым квадратичным критерием позволяет утверждать, что оптимальное решение существует, а во многих типовых постановках и найти его. Есть множество примеров решения задач оптимального управления, когда эффективно применяется принцип максимума и из ограничения на управление $u(t) \in [-1, 1]$ получается, что оптимальное управление имеет вид $u(t) = \text{sign}(f(t))$. Таким образом, оптимальное управление превращается в последовательность переключений, целочисленные ограничения выполняются автоматически. Решение сводится к поиску оптимальной последовательности моментов переключений. Задачи такого рода могут быть весьма разнообразны, вплоть до управления гибридными системами переменной размерности [13]. Кроме того, для поиска приближенного решения можно применять дискретизацию, сводить задачу к перебору возможных последовательностей переключений и применять эвристические методы MINP.

Отказ от ограничений на число переключений, даже без стохастической составляющей, приводит в эти рассуждения сингулярность и, как следствие, приводит к несуществованию оптимального решения. Проблема в известном эффекте Фуллера — ситуации, когда оптимальное управление

переключается бесконечное число раз на конечном интервале. Более того, в [14] эта проблема связана с квадратичным критерием, аналогичным рассматриваемому (2). Эта известная и до сих пор изучаемая задача сама по себе заслуживает обзоров [15]. Изучение этого феномена продолжается и в очень современных приложениях, например в навигации подводных дронов [16], но в детерминированных постановках. Ясно, что наличие в модели (1) винеровского процесса w_t феномен Фуллера только усугубляет. Поэтому первый вывод в отношении решаемой задачи стабилизации — это отсутствие оптимального решения, т. е., хотя $\inf J(U_0^T)$ для управлений $u_t \in \mathbb{Z}^{n_u}$ существует, поскольку нижняя граница $J(U_0^T)$ определена в (6), а функционал (2) — квадратичный по u_t , но управления $(U^{**})_0^T$, для которого минимум достигается $J((U^{**})_0^T) = \inf J(U_0^T)$ на классе $u_t \in \mathbb{Z}^{n_u}$, вообще говоря, не существует, по крайней мере в классе целочисленных кусочно-непрерывных функций.

Далее можно воспользоваться традиционным приемом, задав заранее число переключений. Оптимальное решение у такой задачи будет, но исходный смысл дрейфа за марковской цепью будет искажен. Поэтому дальнейшей целью будет поиск ε -оптимального решения и борьба с лишними переключениями («болтанкой»), которая должна приводить к практически реализуемым управлениям. Получить такое решение с любой точностью можно в виде аппроксимации оптимального управления без ограничений (4) равномерной дискретизацией. Поскольку почти все траектории оптимального управления u_t^* из (4) кусочно-непрерывны, так как кусочно-непрерывны решения обыкновенных дифференциальных уравнений (5), то аппроксимация u_t^* кусочно-постоянным управлением $u_t^{*D} = u_t^*(t_i)$, $t \in [t_i, t_{i+1})$ (D — discrete), при равномерном разбиении интервала управления $0 = t_1 < \dots < t_i < t_{i+1} < \dots < t_N = T$ с шагом $t_{i+1} - t_i = \delta$, $i = 0, N-1$, может быть выполнена так, что разница между значением (6), обеспечиваемой оптимальным управлением u_t^* , и аппроксимацией u_t^{*D} на почти всех траекториях системы (1) будет меньше любого заданного ε . Это следует из записи целевого функционала (2), полученной как результат разделения задач управления и фильтрации:

$$J(U_0^T) = \mathbb{E} \left\{ \int_0^T \|P_t + \hat{y}_t + Q_t z_t + R_t u_t\|_{S_t}^2 dt + \|P_T \hat{y}_T + Q_T z_T\|_{S_T}^2 + \int_0^T \|P_t (y_t - \hat{y}_t)\|_{S_t}^2 dt + \|P_T (y_T - \hat{y}_T)\|_{S_T}^2 \right\}. \quad (7)$$

Непрерывность всех фигурирующих в (7) функций t , включая неявно входящие в решение z_t , и независимость от любого управления оценки \hat{y}_t , определяемой фильтром Вонэма (3), и ее точности $y_t - \hat{y}_t$ гарантируют, что выбором шага разбиения δ можно обеспечить выполнение неравенства

$$|J((U^*)_0^T) - J((U^{*D})_0^T)| < \varepsilon.$$

Теперь для решаемой задачи стабилизации можно утверждать существование ε -оптимального целочисленного управления u_t^{*I} (I — integer). Действительно, такое управление есть среди всех управлений вида $(u_t^I)_k = [(u_t^{*D})_k]$, $\pm 1, \pm 2, \dots, \pm M$, $k = \overline{1, n_u}$, $t \in [t_i, t_{i+1})$, где $[x]$ — целая часть числа x . Конечность M гарантирует существование «верхней» границы — значение функционала (2) на нулевом управлении u_t . Управление u_t^{*I} находится среди кусочно-постоянных функций, которые в каждой точке разбиения t_i заданы заменой управления без целочисленных ограничений на ближайшее слева или справа целое каждой координаты, отстоящее от него на $\pm 1, \pm 2$ и т. д.

Отметим, что даже при оптимизации только среди управлений, отклоняющихся от u_t^{*D} на величину, меньшую 1, хотя бы по одной координате, число вариантов будет значительным. Но все же это рассуждение обосновывает существование оптимального управления u_t^{*I} , так как число вариантов управлений u_t^I конечно. Из этого же вытекает принципиальная возможность вычислить управление u_t^{*I} простым перебором всех вариантов или эффективным приближением любым эвристическим методом MINP. Заметим только, что, в отличие от традиционных детерминированных задач, в рассматриваемой постановке управление ищется в форме обратной связи, так что комбинаторная оптимизация должна быть выполнена для каждой возможной траектории z_t (вместе с соответствующей траекторией оценки \hat{y}_t или цепи y_t), что в реальности сильно ограничивает варианты. Таким образом, остается единственный рабочий вариант — формировать приближение к u_t^{*I} путем преобразования u_t^{*D} в кусочно-целочисленное.

3 Алгоритм MINP для стабилизации системы со скачкообразным дрейфом

Сформулированная выше идея использования u_t^{*D} , ε -оптимального управления без ограничений, для эвристического выбора алгоритма u_t^I (реализуемой аппроксимации u_t^{*I} , ε -оптимального управления с целочисленными ограничениями), т. е.

использование решения задачи управления, полученного на более широком классе допустимых управлений, известна давно как метод релаксации (relaxed control problem) [9, 10]. Самый простой вариант этого метода в рассматриваемой задаче дает управление

$$(u_t^I)_k = \text{round}((u_t^{*D})_k), \quad k = \overline{1, n_u}, \quad (8)$$

где $\text{round}(x)$ — функция целочисленного округления числа x .

Такое управление следует стратегии стандартного округления (standard rounding) [1]. Его недостаток — отсутствие противодействия эффекту Фуллера: сколько раз управление без ограничений пересечет середину некоторого целочисленного интервала, столько раз и будет выполнено переключение в целочисленном управлении. Применяемый фильтр Вонэма (3), борясь с ошибками наблюдения, формируемыми винеровским процессом w_t , очевидно (что продемонстрировано в примере ниже), реагируя на смену состояния цепи y_t даже при очень точных наблюдениях, имеет некоторый период, когда оценка носит переходный характер, реагируя на скачкообразное изменение y_t . Вслед за оценкой колебаться будет и стратегия (8). Для борьбы с этим недостатком предлагаются два приема [1]: сумма округлений (sum of rounding) и сумма округлений с порогом (sum of rounding with a different threshold). Адаптированные к рассматриваемой задаче эти стратегии принимают вид:

$$(u_t^I)_k = \begin{cases} \text{round}((u_t^{*D})_k), \quad k = \overline{1, n_u}, \\ \text{если } \left| (u_{t-l}^{*D})_k - \text{round}((u_t^{*D})_k) \right| + \\ \quad + \sum_{l=0}^{L-2} \left| (u_{t-l}^{*D})_k - (u_{t-l}^I)_k \right| \geq \frac{L}{2}; \\ (u_{t-1}^I)_k \quad \text{иначе;} \end{cases} \quad (9)$$

$$(u_t^I)_k = \begin{cases} \text{round}((u_t^{*D})_k), \quad k = \overline{1, n_u}, \\ \text{если } \lambda_{L-1} \left| (u_{t-l}^{*D})_k - \text{round}((u_t^{*D})_k) \right| + \\ \quad + \sum_{l=0}^{L-2} \lambda_l \left| (u_{t-l}^{*D})_k - (u_{t-l}^I)_k \right| \geq \frac{L}{2}; \\ (u_{t-1}^I)_k \quad \text{иначе.} \end{cases} \quad (10)$$

Оба управления (9) и (10) реализуют одну и ту же стратегию, состоящую в том, что управление $(u_t^I)_k$ на шаге t для k -й координаты меняется (выполняется переключение со старого значения $(u_{t-1}^I)_k$ на новое $(u_t^I)_k$, полученное округлением оптимального без ограничений) только в том случае, когда необходимость переключения подтверждена на L

последних шагах (моментах времени). В управлении (9) требуется, чтобы разница между управлениями без ограничений и округленными L раз была больше 0,5, включая текущий. Управление (10) добавляет веса для придания большей значимости менее отдаленным моментам: предполагается, что $\sum_{l=0}^{L-1} \lambda_l = 1, \lambda_l > \lambda_{l-1} > 0$. Таким образом, реализуется менее инертное управление, которое должно быть ближе к варианту (8).

Отметим, что учет числа переключений в методе релаксации — стандартный эвристический прием [1], призванный обеспечить целочисленному управлению меньшую вычислительную сложность, когда нет ограничения на число переключений. Можно число переключений добавить аддитивным членом к целевому функционалу (2): интеграл $\int_0^T |u_{t-} - u_t|^2 dt$ для кусочно-непрерывных реализаций u_t даст ноль, а для целочисленных — число переключений с учетом их величины. Но в стохастической постановке использовать эту идею затруднительно, поэтому можно исходить из качественного определения, что стратегия (8) порождает максимальное число переключений, стратегия (10) — более инертная, учитывает «соседние» управления, но отдает предпочтение более позднему, стратегия (9) — самая инертная.

4 Численный эксперимент

Для выполнения практического эксперимента воспользуемся моделью кран-балки, описывающей перемещения по рельсу тележки крана для подъема груза [6–8]. Единственное внесенное изменение касается параметров управления, с тем чтобы придать смысл задаче целочисленного управления, уменьшив за счет штрафов и коэффициентов усиления абсолютные значения управляющего воздействия (в примере ниже большая часть управлений принимает значения в интервале $(-10; 10)$, что дает смысл целочисленному ограничению). Положение тележки по отношению к начальному (крайнему) задает скалярная переменная X_t . В это положение тележка стремится попасть со скоростью V_t . Скорость регулирует сила, которая линейно зависит от текущего положения, текущей скорости, неконтролируемого входного воздействия y_t (марковская цепь, моделирующая последовательность номеров линий, занимаемых под погрузку) и управление u_t :

$$\left. \begin{aligned} dX_t &= V_t dt, \quad t \in (0, T]; \\ dV_t &= aX_t dt + bV_t dt + cy_t dt + hu_t dt + \\ &\quad + \sqrt{g} dw_t. \end{aligned} \right\} \quad (11)$$

Качество управления $J(U_0^T)$ для различных стратегий стабилизации

$J((U_{y_t}^{*D})_0^T)$	$J((U_{y_t}^{I_1})_0^T)$	$J((U_{y_t}^{I_2})_0^T)$	$J((U_{y_t}^{I_3})_0^T)$	$J((U_{y_t}^{*D})_0^T)$	$J((U_{y_t}^{I_1})_0^T)$	$J((U_{y_t}^{I_2})_0^T)$	$J((U_{y_t}^{I_3})_0^T)$
10,88	11,15	11,30	11,33	17,96	18,21	18,19	18,25

Марковская цепь y_t имеет три состояния, матрицу интенсивностей

$$\Lambda_t = \Lambda = \begin{pmatrix} -0,5 & 0,5 & 0 \\ 0,5 & -1 & 0,5 \\ 0 & 0,5 & -0,5 \end{pmatrix}$$

и начальное распределение $\pi_0 = (1, 0, 0)'$. Модель (11) приводится к виду (1) обозначениями

$$z_t = (X_t, V_t)'; \quad a_t = c = (c_1, c_2, c_3);$$

$$b_t = \begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix}; \quad c_t = h; \quad \sigma_t = \sqrt{g}$$

(здесь учтено, что для расчета оценки фильтра Вонэма (3) нужно использовать только dV_t). Начальные условия X_0 и V_0 предполагаются независимыми гауссовскими случайными величинами с нулевым средним и дисперсиями $\sigma_X^2 = \sigma_V^2 = 1$. В расчетах параметры модели выбраны следующими: $a = -0,25$; $b = -0,5$; $g = 0,01$; $c_1 = -1,5$; $c_2 = -0,5$; $c_3 = 0,5$; $h = 100$; $T = 10$.

Цель управления состоит в отслеживании координатой X_t положения Cy_t , $C = (-6, -2, 2)$, и формализуется целевым функционалом

$$J(U_0^T) = \mathbb{E} \left\{ \int_0^T (|Cy_t - X_t|^2 + 0,1 |u_t|^2) dt \right\}. \quad (12)$$

Дискретизация выполнялась с шагом $\delta = 0,001$, для оценки значения целевого функционала моделировались 1000 траекторий (11), по которым выполнялось усреднение и получалась оценка (12) для разных вариантов управлений.

Оптимальное управление без ограничений реализовано в двух вариантах: дискретизованное (ε -оптимальное) с полной информацией о состоянии цепи $(U_{y_t}^{*D})_0^T$, которое получается из (4) заменой оценки \hat{y}_t фильтра Вонэма (3) точным значением y_t , и с косвенными наблюдениями, т. е. с оценкой \hat{y}_t (использовалась устойчивая аппроксимация [17]) $(U_{y_t}^{*D})_0^T$. Целочисленные управления — по три стратегии для каждого варианта наблюдения: стратегии простого округления $(U_{y_t}^{I_1})_0^T$ и $(U_{y_t}^{I_1})_0^T$ согласно (8), стратегии суммарного округления $(U_{y_t}^{I_2})_0^T$ и $(U_{y_t}^{I_2})_0^T$ согласно (9), стратегии суммарного округления с порогом $(U_{y_t}^{I_3})_0^T$ и $(U_{y_t}^{I_3})_0^T$ согласно (10). Интервал для стратегий округления выбран равным 3, веса для стратегии с порогом — 1,5; 1; 0,5. Результаты расчетов приведены в таблице.

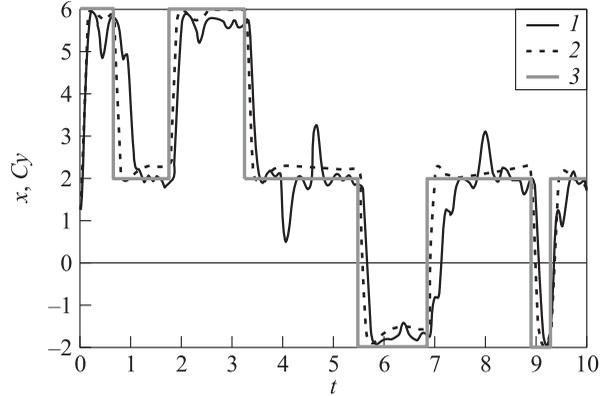


Рис. 1 Типовая траектория привода: 1 — $(U_{y_t}^{*I})_0^T$; 2 — $(U_{y_t}^{*I})_0^T$; 3 — Cy_t — направление дрейфа

Отметим очень небольшую разницу между оптимальным и целочисленным управлением в целом. Оставался актуальным вопрос обоснованности отказа от вычисления ε -оптимального целочисленного управления $(U_{y_t}^{*I})_0^T$ или $(U_{y_t}^{*I})_0^T$. Выше указывалось на принципиальную возможность такой оптимизации путем полного перебора конечного числа вариантов. Близость результатов в таблице — вполне достаточное основание, чтобы не пытаться решать такую вычислительно трудную задачу ради возможности «вклиниться» между оптимальным (по методу релаксации) и эвристическим (путем округления) решениями.

Характер, который демонстрируют стабилизируемые траектории системы (11) и разница между моделями наблюдений, иллюстрирует рис. 1. На рис. 2 приведены примеры траекторий управления, полученных по разным стратегиям округления и проиллюстрировано различие в их характере.

5 Заключение

Первое, что подтвердил выполненный расчет, — это возможность стабилизации рассматриваемой системы (линейное состояние, марковские и гауссовские возмущения) в положении, задаваемом цепью, с использованием исключительно целочисленных управлений. Разница между оптимальным непрерывным и приближенным целочисленным управлением видна, но проигрыш незначителен.

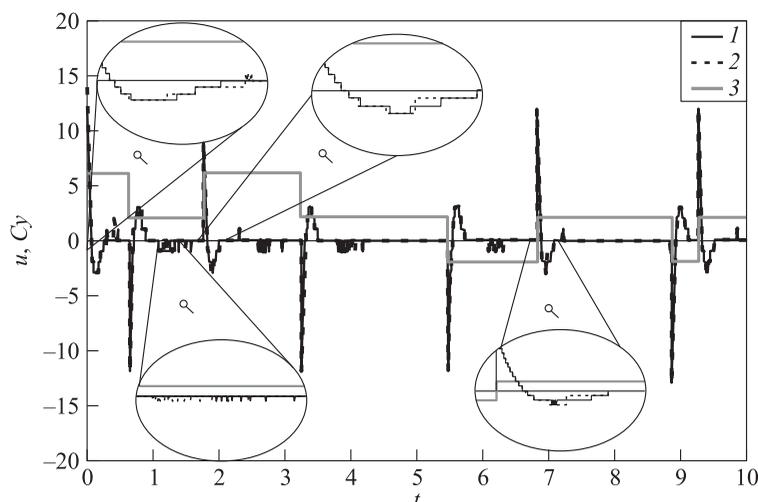


Рис. 2 Траектории управлений: 1 — $(U_{y_t}^{I_2})_0^T$; 2 — $(U_{y_t}^{I_3})_0^T$; 3 — Cy_t — направление дрейфа

Ясно, что этот проигрыш будет расти при увеличении стоимости управления и переключения. Но принципиальная возможность стабилизации сохранится в любом случае, что и является главным результатом.

Кроме того, согласно представленным расчетам, надо отметить почти двукратную разницу в качестве управления в моделях с полной информацией и неполной, которая одинаково существенна как для управлений без ограничений, так и для целочисленных стратегий. Эта разница вполне согласуется с экспериментами, выполненными ранее в [6–8].

Литература

1. Belotti P., Kirches C., Leyffer S., Linderoth J., Luedtke J., Mahajan A. Mixed integer nonlinear optimization // *Acta Numer.*, 2013. Vol. 22. P. 1–131. doi: 10.1017/S0962492913000032.
2. Sahinidis N. V. Mixed-integer nonlinear programming 2018 // *Optim. Eng.*, 2019. Vol. 20. P. 301–306. doi: 10.1007/s11081-019-09438-1.
3. Понтрягин Л. С., Болтянский В. Г., Гамкрелидзе Р. В., Мищенко Е. Ф. Математическая теория оптимальных процессов. — М.: Гос. изд-во физ.-мат. лит., 1961. 391 с.
4. Bellman R. E. Dynamic programming. — Princeton, NJ, USA: Princeton University Press, 1957. 340 p.
5. Wonham W. M. Some applications of stochastic differential equations to optimal nonlinear filtering // *SIAM J. Control*, 1965. Vol. 2. No. 3. P. 347–369. doi: 10.1137/0302028.
6. Босов А. В. Управление линейным выходом марковской цепи по квадратичному критерию // *Информатика и её применения*, 2021. Т. 15. Вып. 2. С. 3–11. doi: 10.14357/19922264210201. EDN: EZECOE.
7. Босов А. В. Управление линейным выходом марковской цепи по квадратичному критерию. Случай полной информации // *Информатика и её применения*, 2022. Т. 16. Вып. 2. С. 19–26. doi: 10.14357/19922264220203. EDN: FEQKUN.
8. Босов А. В. Стабилизация и слежение за траекторией линейной системы со скачкообразно изменяющимся дрейфом // *Автоматика и телемеханика*, 2022. № 4. С. 27–46. doi: 10.31857/S0005231022040031.
9. Bregman L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming // *USSR Comp. Math. Math.*, 1967. Vol. 7. No. 3. P. 200–217. doi: 10.1016/0041-5553(67)90040-7. EDN: XLNZWV.
10. Roubíček T. Relaxation in optimization theory and variational calculus. — 2nd ed. — De Gruyter ser. in nonlinear analysis and applications. — Berlin: Walter de Gruyter GmbH, 2020. Vol. 4. 581 p. doi: 10.1515/9783110590852.
11. Fleming W. H., Rishel R. W. Deterministic and stochastic optimal control. — New York, NY, USA: Springer-Verlag, 1975. 222 p. doi: 10.1007/978-1-4612-6380-7.
12. Sager S. Numerical methods for mixed-integer optimal control problems. — Tönning: Der Andere Verlag, 2005. 219 p.
13. Бортакровский А. С. Необходимые условия оптимальности гибридных систем переменной размерности // *Изв. РАН. Теория и системы управления*, 2022. № 1. С. 28–40. doi: 10.31857/S0002338821060056.
14. Фуллер А. Т. Оптимизация релейных систем регулирования по различным критериям качества // *Труды I конгресса ИФАК*. — М.: Изд-во АН СССР, 1961. Т. 2. С. 584–605.
15. Сухинин Б. В., Сурков В. В. Феномен Фуллера в теории и практике оптимального управления // *J. Ad-*

vanced Research Technical Science, 2021. Vol. 2. Iss. 23. P. 94–99. doi: 10.26160/2474-5901-2021-23-94-99. EDN: VBLUGF.

16. Chyba M., Haberkorn T. Autonomous underwater vehicles: Singular extremals and chattering // Systems, control, modeling and optimization / Eds. F. Ceragioli, A. Dontchev, H. Furuta, K. Marti, L. Pandolfi. — IFIP International Federation for Information Process-

ing ser. — Boston, MA, USA: Springer, 2006. Vol. 202. P. 103–113. doi: 10.1007/0-387-33882-9_10.

17. Борисов А. В. Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным наблюдениям II: случай аддитивных шумов // Информатика и её применения, 2020. Т. 14. Вып. 1. С. 17–23. doi: 10.14357/19922264200103. EDN: BDUYGW.

Поступила в редакцию 15.06.2025

Принята к публикации 15.08.2025

STABILIZATION OF THE TRAJECTORY OF A LINEAR SYSTEM WITH JUMPING DRIFT UNDER INTEGER CONTROL CONSTRAINTS

A. V. Bosov

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

Abstract: An integer-valued version of the stabilization problem for a linear stochastic differential system is considered where the drift evolves in a jumping manner determined by a Markov chain. The control objective is formalized via a quadratic cost functional. Depending on the conditions, both the full information case (the state of the chain is known) and the indirect observation case (the system state serves as an indirect observation of the unknown chain state) are possible. A distinguishing feature of the formulation lies in the integer constraints on the admissible control values. Unlike the previously solved unconstrained problem, the existence conditions for a solution are not satisfied in the “integer” formulation; therefore, an ε -optimal solution is investigated. An ε -optimal control can be obtained by discretizing the optimal solution of the unconstrained problem and applying mixed-integer nonlinear programming. However, the stochastic nature of the problem and the large number of switching scenarios prevent the guaranteed computational feasibility of solving it using dynamic programming. For practical implementation, a relaxation method is used: a heuristic approximation is computed as the result of an integer transformation of the ε -optimal control in the unconstrained problem. Three variants of such transformations are proposed. A numerical experiment was conducted using the same applied model as in previous works on unconstrained control (position dynamics of a simple mechanical actuator). The results primarily confirm the applicability of the proposed solutions in terms of the stabilization objective and also allow for a comparison of the nature of the relaxation strategies.

Keywords: stabilization of a linear system; quadratic cost functional; dynamic programming; feedback control; Wonham filter; mixed-integer nonlinear programming; relaxation method; mechanical actuator

DOI: 10.14357/19922264250302

EDN: YZEHCC

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. Belotti, P., C. Kirches, S. Leyffer, J. Linderoth, J. Luedtke, and A. Mahajan. 2013. Mixed integer nonlinear optimization. *Acta Numer.* 22:1–131. doi: 10.1017/S0962492913000032.
2. Sahinidis, N. V. 2019. Mixed-integer nonlinear programming 2018. *Optim. Eng.* 20:301–306. doi: 10.1007/s11081-019-09438-1.
3. Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. 1962. *The mathematical theory of optimal processes*. New York, London: John Wiley & Sons. 360 p.
4. Bellman, R. E. 1957. *Dynamic programming*. Princeton, NJ: Princeton University Press. 342 p.
5. Wonham, W. M. 1965. Some application of stochastic differential equations to optimal nonlinear filtering. *SIAM J. Control* 2(3):347–369. doi: 10.1137/0302028.

6. Bosov, A. V. 2021. Upravlenie lineynym vykhodom markovskoy tsepi po kvadraticnomu kriteriyu [Linear output control of Markov chains by the quadratic criterion]. *Informatika i ee Primeneniya — Inform. Appl.* 15(2):3–11. doi: 10.14357/19922264210201. EDN: EZECO.
7. Bosov, A. V. 2022. Upravlenie lineynym vykhodom markovskoy tsepi po kvadraticnomu kriteriyu. Sluchay polnoy informatsii [Linear output control of Markov chain by square criterion. Complete information case]. *Informatika i ee Primeneniya — Inform. Appl.* 16(2):19–26. doi: 10.14357/19922264220203. EDN: FEQKUN.
8. Bosov, A. V. 2022. Stabilization and tracking of the trajectory of a linear system with jump drift. *Automat. Rem. Contr.* 83(4):520–535. doi: 10.1134/S0005117922040026.
9. Bregman, L. M. 1967. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comp. Math. Math.* 7(3):200–217. doi: 10.1016/0041-5553(67)90040-7. EDN: XLNZWV.
10. Roubíček, T. 2020. *Relaxation in optimization theory and variational calculus*. 2nd ed. De Gruyter ser. in nonlinear analysis and applications. Berlin: Walter de Gruyter GmbH. Vol. 4. 581 p. doi: 10.1515/9783110590852.
11. Fleming, W. H., and R. W. Rishel. 1975. *Deterministic and stochastic optimal control*. New York, NY: Springer-Verlag. 222 p. doi: 10.1007/978-1-4612-6380-7.
12. Sager, S. 2005. *Numerical methods for mixed-integer optimal control problems*. Tönning: Der Andere Verlag. 219 p.
13. Bortakovsky, A. S. 2021. The necessary conditions for optimal hybrid systems of variable dimensions. *J. Comput. Sys. Sc. Int.* 60(6):883–894. doi: 10.1134/S1064230721060058. EDN: YBJIUW.
14. Fuller, A. T. 1960. Relay control systems optimized for various performance criteria. *IFAC Proceedings Volumes* 1(1):520–529. doi: 10.1016/S1474-6670(17)70097-3.
15. Sukhinin, B. V., and V. V. Surkov. 2021. Fenomen Fullera v teorii i praktike optimal'nogo upravleniya [Phenomenon by Fuller in theory and practice of the optimum control]. *J. Advanced Research Technical Science* 2(23):94–99. doi: 10.26160/2474-5901-2021-23-94-99. EDN: VBLUGF.
16. Chyba, M., and T. Haberkorn. 2006. Autonomous underwater vehicles: Singular extremals and chattering. *Systems, control, modeling and optimization*. Eds. F. Ceragioli, A. Dontchev, H. Furuta, K. Marti, and L. Pandolfi. IFIP International Federation for Information Processing ser. Boston, MA: Springer. 202:103–113. doi: 10.1007/0-387-33882-9-10.
17. Borisov, A. V. 2020. Chislennyye skhemy fil'tratsii markovskikh skachkoobraznykh protsessov po diskretizovannym nablyudeniya II: sluchay additivnykh shumov [Numerical schemes of Markov jump process filtering given discretized observations II: Additive noise case]. *Informatika i ee primeneniya — Inform. Appl.* 14(1):17–23. doi: 10.14357/19922264200103. EDN: BDUYGW.

Received June 15, 2025

Accepted August 15, 2025

Contributor

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

ОЦЕНКА РИСКА МЕТОДА БЛОЧНОЙ ПОРОГОВОЙ ОБРАБОТКИ ПРИ РЕШЕНИИ ОБРАТНЫХ СТАТИСТИЧЕСКИХ ЗАДАЧ С ДАННЫМИ, ЗАДАНЫМИ НА СЛУЧАЙНОЙ СЕТКЕ

О. В. Шестаков¹

Аннотация: Методы вейвлет-анализа широко используются при решении обратных статистических задач для обращения линейных однородных операторов. Преимущество этих методов заключается в вычислительной эффективности и возможности адаптации как к виду оператора, так и к локальным особенностям оцениваемой функции. Для подавления шума в наблюдаемых данных применяется пороговая обработка коэффициентов разложения наблюдаемой функции по вейвлет-базису. Одним из наиболее эффективных оказался метод блочной пороговой обработки, в котором коэффициенты разложения обрабатываются группами, что позволяет учитывать информацию о соседних коэффициентах. Иногда природа данных такова, что регистрация наблюдений проводится в случайные моменты времени. Если точки отсчетов образуют вариационный ряд, построенный по выборке из равномерного распределения на отрезке регистрации данных, то использование процедур пороговой обработки оказывается адекватным и не ухудшает качество получаемых оценок. Проведен анализ оценки среднеквадратичного риска метода блочной пороговой обработки и показано, что при определенных условиях данная оценка сильно состоятельна и асимптотически нормальна.

Ключевые слова: линейный однородный оператор; вейвлеты; блочная пороговая обработка; несмещенная оценка риска; случайные отсчеты

DOI: 10.14357/19922264250303

EDN: ZJXMFR

1 Введение

Задача обращения линейного однородного оператора возникает в различных прикладных областях, включая медицину, астрономию, физику плазмы и т. д. При наличии искажений (шума) в наблюдениях необходимо использовать методы регуляризации, и классические подходы к решению этой проблемы основаны на использовании сингулярного разложения и проекционных методах [1]. Однако при этом учитывается лишь вид оператора, участвующего в формировании наблюдаемых данных, и никак не учитываются свойства самого объекта наблюдения. Для линейных однородных операторов эта проблема решается с помощью методов вейвлет-анализа, позволяющих адаптироваться одновременно к виду оператора и локальным особенностям функции, описывающей объект. Подавление шума осуществляется с помощью вейвлет-разложения и процедур пороговой обработки. Чаще всего процедуры пороговой обработки применяются отдельно к каждому коэффициенту вейвлет-разложения. Коэффициент сравнивается с пороговым значением и, если его абсолютная величина оказывается меньше этого значения, обнуляется. Среди таких процедур самыми распространенными стали

процедуры жесткой и мягкой пороговой обработки [2–4]. Однако при использовании этих процедур не удается достичь оптимального (в минимаксном смысле) порядка среднеквадратичного риска. В работе [5] предложен метод пороговой обработки, при котором вейвлет-коэффициенты обрабатываются не отдельно, а блоками, что позволяет учитывать информацию о соседних коэффициентах. При надлежащем выборе размера блоков и порогового значения получаемые оценки имеют оптимальный минимаксный порядок среднеквадратичного риска [6].

Упомянутые методы разработаны для дискретных отсчетов исследуемой функции, заданных на равномерной сетке. В некоторых ситуациях нет возможности (или она сильно затруднена) регистрировать наблюдения через равные интервалы [7]. Иногда природа поступающих данных такова, что регистрация отсчетов производится в случайные моменты времени. В работах [8, 9] рассмотрен случай, когда отсчеты образуют вариационный ряд, построенный по выборке из равномерного распределения на заданном отрезке, и показано, что порядок среднеквадратичного риска остается таким же, как в случае равномерной неслучайной сет-

¹Московский государственный университет имени М. В. Ломоносова, факультет вычислительной математики и кибернетики; Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Московский центр фундаментальной и прикладной математики, oshestakov@cs.msu.ru

ки. Равномерное распределение может возникнуть в ситуации, когда моменты регистрации наблюдений образуют пуассоновский процесс, поскольку условное распределение скачков пуассоновского процесса при их фиксированном числе совпадает с распределением вариационного ряда, построенного по выборке из равномерного распределения. Такие модели могут возникать, например, в астрономии при изучении интенсивности излучения космических объектов.

Для практического анализа погрешности методов обращения линейных однородных операторов можно использовать оценку среднеквадратичного риска, которая зависит только от наблюдаемых данных. В работах [10–12] показано, что для равноотстоящих отсчетов эта оценка сильно состоятельна и асимптотически нормальна. В данной работе рассматривается случайная сетка с равномерным распределением отсчетов и показывается, что и в этой ситуации оценка среднеквадратичного риска сохраняет свои статистические свойства.

2 Обращение линейных однородных операторов

Линейным однородным оператором называется такое линейное преобразование K функции f , при котором

$$K[f(a(x - x_0))] = a^{-\alpha}(Kf)[a(x - x_0)]$$

для любого x_0 и любого $a > 0$. Параметр α называется показателем однородности. Примерами однородных линейных операторов служат операторы дифференцирования и интегрирования, преобразование Гильберта, преобразование Абеля и некоторые операторы свертки. Математические модели с такими операторами используются при решении задач медицины, астрономии, физики плазмы и т. д. Рассмотрим метод обращения оператора K , основанный на свойствах вейвлет-разложений [13–15]. Преимущество этого метода заключается в адаптации не только к свойствам оператора K , но и к свойствам функции f .

Вейвлет-разложение функции Kf имеет вид:

$$Kf = \sum_{j,k \in \mathbb{Z}} \langle Kf, \psi_{j,k} \rangle \psi_{j,k}, \quad (1)$$

где $\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$; ψ — некоторая материнская вейвлет-функция (семейство $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$ образует ортонормированный базис в $L^2(\mathbb{R})$). Индекс j в (1) называется масштабом, а индекс k — сдвигом.

Поскольку оператор K линеен и однороден, существуют такие функции $u_{j,k}$, что $\langle f, u_{j,k} \rangle = \langle Kf, \psi_{j,k} \rangle$. При этом функция f представляется в виде ряда

$$f = \sum_{j,k \in \mathbb{Z}} \beta_j \langle Kf, \psi_{j,k} \rangle u_{j,k}. \quad (2)$$

Здесь $u_{j,k} = K^{-1} \psi_{j,k} / \beta_{j,k}$, где $\beta_j = 2^{\alpha j} \beta_0$, $\beta_0 = \|K^{-1} \psi\|$. Функции $u_{j,k}$ называются вейглетами. По своим свойствам они похожи на вейвлеты и также представляют собой сдвиги и растяжения некоторой материнской функции u . Формула (2) называется вейвлет-вейвлет-разложением и служит основой метода обращения оператора K [15]. При соответствующем выборе вейвлет-функции ψ последовательность $\{u_{j,k}\}$ образует устойчивый базис [10, 14]. Также, если ψ имеет r непрерывных производных и r нулевых моментов, определим при $0 < \gamma < r$ и $1 \leq p, q \leq \infty$ полунорму последовательности вейвлет-коэффициентов выражением

$$|Kf|_{B_{p,q}^\gamma} = \left(\sum_{j=0}^{\infty} \left(2^{sj} \left(\sum_k |\langle Kf, \psi_{j,k} \rangle|^p \right)^{1/p} \right)^q \right)^{1/q},$$

где $s = \gamma + 1/2 - 1/p$. Далее будем считать, что ψ удовлетворяет всем необходимым требованиям, а Kf задана на отрезке $[0, 1]$ и принадлежит пространству Бесова $B_{p,q}^\gamma(A)$ ($A > 0$), т. е. $|Kf|_{B_{p,q}^\gamma} \leq A$ (в частности, пространство Гельдера $\Lambda^\gamma(A) = B_{\infty,\infty}^\gamma(A)$).

3 Модель данных со случайными отсчетами

Предположим, что отсчеты Kf регистрируются в случайные моменты времени и содержат гауссовый белый шум, т. е. рассмотрим следующую модель данных:

$$Y_i = Kf(x_i) + z_i, \quad i = \overline{1, N},$$

где x_i независимы и равномерно распределены на $[0, 1]$; z_i — независимые от x_i и между собой «шумовые» коэффициенты, относительно которых предполагается, что они имеют нормальное распределение с нулевым средним и дисперсией σ^2 .

Пусть $0 \leq x_{(1)} < \dots < x_{(N)} \leq 1$ — вариационный ряд, построенный по выборке x_i , $i = \overline{1, N}$. Тогда, перенумеровав Y_i и z_i , получаем модель

$$Y_i = Kf(x_{(i)}) + \varepsilon_i, \quad i = \overline{1, N}. \quad (3)$$

Наблюдения состоят из пар $(x_{(1)}, Y_1), \dots, (x_{(N)}, Y_N)$, в которых расстояния между отсчетами в общем

случае не равны. Наряду с (3) рассмотрим выборку с равными расстояниями между отсчетами

$$\left(\frac{1}{N+1}, Z_1\right), \dots, \left(\frac{N}{N+1}, Z_N\right), \quad (4)$$

где

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

Для того чтобы можно было пользоваться методами вейвлет-анализа, N должно равняться 2^J при некотором натуральном $J > 0$. Тогда к выборке (4) можно применить дискретное ортогональное вейвлет-преобразование и получить набор эмпирических вейвлет-коэффициентов [11, 12]

$$W_{j,k} = \mu_{j,k} + \xi_{j,k},$$

где $\mu_{j,k} \approx 2^{J/2} \langle Kf, \psi_{j,k} \rangle$; шумовые коэффициенты $\xi_{j,k}$ независимы и имеют такое же распределение, как и ε_i . Далее, не ограничивая общности, будем полагать, что $N = 2^J$.

4 Блочная пороговая обработка коэффициентов разложения

Популярным методом подавления шума стала пороговая обработка эмпирических коэффициентов. Самыми распространенными видами пороговой обработки стали методы жесткой и мягкой пороговой обработки [2–4]. Их суть заключается в удалении достаточно маленьких коэффициентов, которые считаются шумом. При использовании этих методов происходит сравнение абсолютной величины каждого коэффициента с некоторым порогом (например, универсальным порогом, равным $\sigma\sqrt{2\log N}$), и если это значение оказалось ниже порога, то коэффициент считается шумом и обнуляется. Такие методы обрабатывают каждый коэффициент отдельно, не используя информацию о других коэффициентах. Этот недостаток можно устранить, используя блочную пороговую обработку, при которой решение об обнулении применяется одновременно к группе соседних коэффициентов [5].

Пусть $B_{j,1}, \dots, B_{j,M_j}$ — разбиение множества индексов $\{0, \dots, 2^j - 1\}$ на блоки одинаковой длины L (для удобства предположим, что 2^j делится на L). Пусть $S_{j,m}^2 = \sum_{k \in B_{j,m}} W_{j,k}^2$. Оценки коэффициентов $\mu_{j,k}$ вычисляются по правилу:

$$\widehat{W}_{j,k} = \left(1 - \frac{TL\sigma^2}{S_{j,m}^2}\right)_+ W_{j,k}, \quad j = \overline{0, J-1}, \quad k \in B_m,$$

т.е. если величина $\sum_{k \in B_{j,m}} W_{j,k}^2$ меньше порога $TL\sigma^2$, то все коэффициенты в рассматриваемом блоке обнуляются.

На качество оценок, получаемых с помощью блочной пороговой обработки, естественно, влияют размер блока L и значение порога T . В работе [6] показано, что при $L = \log 2^j$ достигается баланс между локальной и глобальной адаптивностью метода блочной пороговой обработки, и если при этом $T^* \approx 4,50524$ (T^* — корень уравнения $T - \log T - 3 = 0$), то среднеквадратичный риск оказывается в минимаксном смысле (почти) оптимальным. В данной работе рассматриваются именно такие значения L и T .

Если применить дискретное вейвлет-преобразование к выборке (3), то получится набор эмпирических коэффициентов

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

Здесь $\nu_{j,k}$ — коэффициенты дискретного вейвлет-преобразования выборки $Kf(x_{(1)}), \dots, Kf(x_{(N)})$. В общем случае $V_{j,k}$ не равны $W_{j,k}$, а $\nu_{j,k}$ не равны $\mu_{j,k}$. Однако к $V_{j,k}$ можно применить ту же процедуру, что и к коэффициентам $W_{j,k}$, и получить оценки $\widehat{V}_{j,k}$. В следующих разделах обсуждаются свойства этих оценок.

5 Среднеквадратичный риск блочной пороговой обработки

Среднеквадратичный риск пороговой обработки для выборки со случайными отсчетами определим как

$$R_\nu = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_j^2 \mathbb{E}(\widehat{V}_{j,k} - \mu_{j,k})^2.$$

Также определим среднеквадратичный риск для выборки с равными расстояниями между отсчетами:

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

Используя результаты работ [6, 9], можно убедиться, что при блочной пороговой обработке величины R_μ и R_ν имеют минимаксный порядок.

Теорема 1. Пусть $Kf \in B_{p,q}^\gamma(A)$ и задана на отрезке $[0, 1]$. Пусть вейвлет-функция ψ имеет r непрерывных производных и r нулевых моментов. Если при этом $2 \leq p \leq \infty$, $1 \leq q \leq \infty$ и $\alpha < \gamma < r$, тогда

$$R_\mu \leq C 2^{(2\alpha+1)J/(2\gamma+1)},$$

где C — некоторая положительная константа.

Теорема 2. Пусть $Kf \in B_{p,q}^\gamma(A) \cap \Lambda^{2\gamma/(2\gamma+1)}(A)$ и задана на отрезке $[0, 1]$. Пусть вейвлет-функция ψ

имеет r непрерывных производных и r нулевых моментов. Если при этом $2 \leq p \leq \infty$, $1 \leq q \leq \infty$ и $\max(\alpha, 1/2) < \gamma < r$, тогда

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

где C — некоторая положительная константа.

Таким образом, замена равноотстоящих отсчетов на случайные не ухудшает оценку порядка среднеквадратичного риска.

6 Асимптотические свойства несмещенной оценки риска

Вычислить значение среднеквадратичного риска на практике нельзя, поскольку оно зависит от ненаблюдаемых «чистых» коэффициентов. Однако можно построить его оценку, используя только наблюдаемые данные. По аналогии с [6] в качестве оценки риска будем использовать величину

$$\widehat{R}_\nu = \sum_{j=0}^{J-1} \beta_j^2 \sum_{m=1}^{M_j} F_{j,m}, \quad (5)$$

где

$$F_{j,m} = \left[L\sigma^2 + \frac{T^2 L^2 \sigma^4 - 2TL\sigma^4(L-2)}{\tilde{S}_{j,m}^2} \mathbf{1}(\tilde{S}_{j,m}^2 > TL\sigma^2) + \left(\tilde{S}_{j,m}^2 - 2L\sigma^2 \right) \mathbf{1}(\tilde{S}_{j,m}^2 \leq TL\sigma^2) \right],$$

$$\tilde{S}_{j,m}^2 = \sum_{k \in B_{j,m}} V_{j,k}^2.$$

В случае использования сетки с равноотстоящими отсчетами такая оценка риска не смещена.

Оценка риска (5) дает возможность получить представление о погрешности, с которой оценивается функция сигнала, используя только наблюдаемые данные. Покажем, что оценка (5) асимптотически нормальна. Это свойство, в частности, позволяет строить асимптотические доверительные интервалы для R_ν .

Теорема 3. Пусть $Kf \in B_{p,q}^\gamma(A) \cap \Lambda^{2\gamma/(2\gamma+1)}(A)$ и задана на отрезке $[0, 1]$. Пусть вейвлет-функция ψ имеет r непрерывных производных и r нулевых моментов. Если при этом $2 \leq p \leq \infty$, $1 \leq q \leq \infty$ и $\max(\alpha, 1/2) < \gamma < r$, тогда при $J \rightarrow \infty$

$$\mathbb{P} \left(\frac{\widehat{R}_\nu - R_\nu}{D_J} < x \right) \rightarrow \Phi(x),$$

где

$$D_J^2 = \frac{2\sigma^4 \beta_0^4}{2^{4\alpha+1} - 1} 2^{(4\alpha+1)J};$$

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

Доказательство. Наряду с \widehat{R}_ν рассмотрим

$$\widehat{R}_\mu = \sum_{j=0}^{J-1} \beta_j^2 \sum_{m=1}^{M_j} G_{j,m},$$

где

$$G_{j,m} = \left[L\sigma^2 + \frac{T^2 L^2 \sigma^4 - 2TL\sigma^4(L-2)}{S_{j,m}^2} \mathbf{1}(S_{j,m}^2 > TL\sigma^2) + (S_{j,m}^2 - 2L\sigma^2) \mathbf{1}(S_{j,m}^2 \leq TL\sigma^2) \right],$$

и запишем разность $\widehat{R}_\nu - R_\nu$ в виде

$$\widehat{R}_\nu - R_\nu = \widehat{R}_\mu - R_\mu + \widetilde{R},$$

где

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

В [16] показано, что

$$\mathbb{P} \left(\frac{\widehat{R}_\mu - R_\mu}{D_J} < x \right) \rightarrow \Phi(x) \text{ при } J \rightarrow \infty.$$

Следовательно, для доказательства теоремы достаточно показать, что

$$\frac{\widetilde{R}}{2^{J(1/2+2\alpha)}} \xrightarrow{\mathbb{P}} 0 \text{ при } J \rightarrow \infty.$$

В силу теорем 1 и 2

$$\frac{R_\nu(f, T_\gamma) - R_\mu(f, T_\gamma)}{2^{J(1/2+2\alpha)}} \rightarrow 0 \text{ при } J \rightarrow \infty.$$

Пусть $j_0 \approx cJ/(2\gamma+1)$. Представим $\widehat{R}_\nu - \widehat{R}_\mu$ в виде

$$\widehat{R}_\nu - \widehat{R}_\mu = S_1 + S_2,$$

где

$$S_1 = \sum_{j=0}^{j_0-1} \beta_j^2 \sum_{m=1}^{M_j} (F_{j,m} - G_{j,m});$$

$$S_2 = \sum_{j=j_0}^{J-1} \beta_j^2 \sum_{m=1}^{M_j} (F_{j,m} - G_{j,m}).$$

Поскольку для некоторой константы $C_F > 0$ [6]

$$|F_{j,m} - G_{j,m}| \leq C_F T^* L \text{ п. в.}, \quad (6)$$

то

$$\frac{S_1}{2^{J(1/2+2\alpha)}} \rightarrow 0 \text{ п. в. при } J \rightarrow \infty.$$

Далее

$$\begin{aligned} S_2 &= \sum_{j=j_0}^{J-1} \beta_j^2 \sum_{m=1}^{M_j} (F_{j,m} - G_{j,m}) = \\ &= \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \beta_j^2 (V_{j,k}^2 - W_{j,k}^2) + \\ &+ \sum_{j=j_0}^{J-1} \beta_j^2 \sum_{m=1}^{M_j} (U(\tilde{S}_{j,m}^2) - U(S_{j,m}^2)), \quad (7) \end{aligned}$$

где

$$U(S^2) = \left[\frac{T^2 L^2 \sigma^4 - 2TL\sigma^4(L-2)}{S^2} - S^2 + 2L\sigma^2 \right] \mathbf{1}(S^2 > TL\sigma^2).$$

Рассмотрим первую сумму:

$$\begin{aligned} \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \beta_j^2 (V_{j,k}^2 - W_{j,k}^2) &= \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \beta_j^2 (\nu_{j,k}^2 - \mu_{j,k}^2) + \\ &+ 2 \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \beta_j^2 \xi_{j,k} (\nu_{j,k} - \mu_{j,k}). \end{aligned}$$

Условное распределение этой суммы при фиксированных x_i нормально с математическим ожиданием $\sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \beta_j^2 (\nu_{j,k}^2 - \mu_{j,k}^2)$ и дисперсией $4\sigma^2 \sum_{j=j_0}^{J-1} \beta_j^4 \sum_{k=0}^{2^j-1} (\nu_{j,k} - \mu_{j,k})^2$.

Так как $f \in B_{p,q}^\gamma(A) \cap \Lambda^{2\gamma/(2\gamma+1)}(A)$, то, повторяя рассуждения из работ [17–19] с учетом леммы 3 из [8] и леммы 4 из [9], можно показать, что

$$\begin{aligned} \frac{1}{2^{J(1/2+2\alpha)}} \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \beta_j^2 (\nu_{j,k}^2 - \mu_{j,k}^2) &\xrightarrow{P} 0; \\ \frac{1}{2^{J(1/2+2\alpha)}} \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \beta_j^2 (\nu_{j,k} - \mu_{j,k})^2 &\xrightarrow{P} 0 \end{aligned}$$

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

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

Во второй сумме в (7) содержатся индикаторы, в которых либо $\tilde{S}_{j,m}^2 > TL\sigma^2$, либо $S_{j,m}^2 > TL\sigma^2$,

причем $\sum_{k=0}^{2^j-1} \mu_{j,k}^2 \rightarrow 0$ при $J \rightarrow \infty$ для всех $j > j_0$ [16]. Повторяя рассуждения из работы [20] с использованием леммы 2 из работы [6], можно показать, что эти суммы при делении на $2^{J(1/2+2\alpha)}$ также сходятся к нулю по вероятности. Теорема доказана.

Помимо асимптотической нормальности оценка (5) также обладает свойством сильной состоятельности.

Теорема 4. Пусть выполнены условия теоремы 3. Тогда при любом $\lambda > 1/2 + 2\alpha$

$$\frac{\hat{R}_\nu - R_\nu}{2^{\lambda J}} \rightarrow 0 \text{ п. в. при } J \rightarrow \infty.$$

Поскольку выполнено (6) и при фиксированных x_i слагаемые в (5) условно независимы, доказательство этой теоремы аналогично доказательству соответствующего утверждения из работы [21].

Теоремы 3 и 4 дают теоретическое обоснование использования значения \hat{R}_ν в качестве оценки неизвестной величины риска (погрешности) R_ν , а также дают возможность строить асимптотические доверительные интервалы для R_ν .

Литература

1. Engl H. W., Hanke M., Neubauer A. Regularization of inverse problems. — Mathematics and its applications ser. — Dordrecht: Kluwer Academic Publs., 1996. Vol. 375. 330 p.
2. Donoho D., Johnstone I. M. Ideal spatial adaptation via wavelet shrinkage // Biometrika, 1994. Vol. 81. No. 3. P. 425–455. doi: 10.1093/biomet/81.3.425.
3. Donoho D., Johnstone I. M. Adapting to unknown smoothness via wavelet shrinkage // J. Am. Stat. Assoc., 1995. Vol. 90. P. 1200–1224. doi: 10.1080/01621459.1995.10476626.
4. Donoho D., Johnstone I. M. Minimax estimation via wavelet shrinkage // Ann. Stat., 1998. Vol. 26. No. 3. P. 879–921. doi: 10.1214/aos/1024691081.
5. Hall P., Kerkycharian G., Picard D. On the minimax optimality of block thresholded wavelet estimators // Stat. Sinica, 1999. Vol. 9. P. 33–49.
6. Cai T. Adaptive wavelet estimation: A block thresholding and oracle inequality approach // Ann. Stat., 1999. Vol. 28. No. 3. P. 898–924. doi: 10.1214/aos/1018031262.
7. Cai T., Brown L. Wavelet shrinkage for nonequispaced samples // Ann. Stat., 1998. Vol. 26. No. 5. P. 1783–1799. doi: 10.1214/aos/1024691357.
8. Cai T., Brown L. Wavelet estimation for samples with random uniform design // Stat. Probabil. Lett., 1999. Vol. 42. P. 313–321. doi: 10.1016/S0167-7152(98)00223-5.

9. *Chicken E.* Block thresholding and wavelet estimation for nonequispaced samples // *J. Stat. Plan. Infer.*, 2003. Vol. 116. No. 1. P. 113–129. doi: 10.1016/S0378-3758(02)00238-0.
10. *Кудрявцев А. А., Шестаков О. В.* Асимптотика оценки риска при вейвлет-вейвлет разложении наблюдаемого сигнала // *T-Comm — Телекоммуникации и Транспорт*, 2011. № 2. С. 54–57. EDN: ОРНУQN.
11. *Ерошенко А. А., Шестаков О. В.* Асимптотическая нормальность оценки риска при вейвлет-вейвлет-разложении функции сигнала в модели с коррелированным шумом // *Вестн. Моск. ун-та. Сер. 15: Вычисл. матем. и киберн.*, 2014. № 3. С. 23–30. EDN: QPRMZD.
12. *Ерошенко А. А., Кудрявцев А. А., Шестаков О. В.* Предельное распределение оценки риска метода вейвлет-вейвлет-разложения сигнала в модели с коррелированным шумом // *Вестн. Моск. ун-та. Сер. 15: Вычисл. матем. и киберн.*, 2015. № 1. С. 12–18. EDN: TGGPDF.
13. *Donoho D.* Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition // *Appl. Comput. Harmon. A.*, 1995. Vol. 2. P. 101–126. doi: 10.1006/acha.1995.1008.
14. *Lee N.* Wavelet-vaguelette decompositions and homogeneous equations: PhD Thesis. — West Lafayette, IN, USA: Purdue University, 1997. 93 p.
15. *Abramovich F., Silverman B. W.* Wavelet decomposition approaches to statistical inverse problems // *Biometrika*, 1998. Vol. 85. No. 1. P. 115–129. doi: 10.1093/biomet/85.1.115.
16. *Шестаков О. В., Степанов Е. П.* Нелинейная регуляризация обращения линейных однородных операторов с помощью метода блочной пороговой обработки // *Информатика и её применения*, 2023. Т. 17. Вып. 4. С. 2–8. doi: 10.14357/19922264230401. EDN: PGKKYE.
17. *Шестаков О. В.* Свойства вейвлет-оценок сигналов, регистрируемых в случайные моменты времени // *Информатика и её применения*, 2019. Т. 13. Вып. 2. С. 16–21. doi: 10.14357/19922264190203. EDN: SEYEXW.
18. *Шестаков О. В.* Асимптотическая регулярность вейвлет-методов обращения линейных однородных операторов по наблюдениям, регистрируемым в случайные моменты времени // *Информатика и её применения*, 2020. Т. 14. Вып. 1. С. 3–9. doi: 10.14357/19922264200101. EDN: GJHMNK.
19. *Шестаков О. В.* Статистические свойства оценки среднеквадратичного риска метода блочной пороговой обработки в задачах непараметрической регрессии со случайной сеткой // *Информатика и её применения*, 2024. Т. 18. Вып. 4. С. 26–33. doi: 10.14357/19922264240404. EDN: WHZQPX.
20. *Шестаков О. В.* Аппроксимация распределения оценки риска пороговой обработки вейвлет-коэффициентов нормальным распределением при использовании выборочной дисперсии // *Информатика и её применения*, 2010. Т. 4. Вып. 4. С. 72–79. EDN: NULNZH.
21. *Шестаков О. В.* Решение обратных статистических задач с помощью методов пороговой обработки, допускающих построение несмещенной оценки среднеквадратичного риска // *Информатика и её применения*, 2025. Т. 19. Вып. 1. С. 74–81. doi: 10.14357/19922264250110. EDN: ZSIBDL.

Поступила в редакцию 22.06.2025

Принята к публикации 15.08.2025

RISK ESTIMATE FOR THE BLOCK THRESHOLDING METHOD OF SOLVING INVERSE STATISTICAL PROBLEMS WITH DATA ON A RANDOM GRID

O. V. Shestakov^{1,2,3}

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

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

³Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

Abstract: Wavelet analysis methods are widely used to solve inverse statistical problems for inverting linear homogeneous operators. The advantage of these methods is their computational efficiency and the ability to adapt to both the operator type and local features of the estimated function. To suppress the noise in the observed data,

threshold processing of the expansion coefficients of the observed function over the wavelet basis is used. One of the most effective is the block thresholding method in which the expansion coefficients are processed in groups that allows taking into account information about neighboring coefficients. Sometimes, the nature of the data is such that observations are recorded at random times. If the sample points form a variation series constructed from a uniform distribution sample over the data recording interval, then the use of threshold processing procedures is adequate and does not worsen the quality of the estimates obtained. The paper analyzes the estimate of the mean square risk of the block thresholding method and shows that under certain conditions, this estimate is strongly consistent and asymptotically normal.

Keywords: linear homogeneous operator; wavelets; block thresholding; unbiased risk estimate; random samples

DOI: 10.14357/19922264250303

EDN: ZJXMFR

References

- Engl, H. W., M. Hanke, and A. Neubauer. 1996. *Regularization of inverse problems*. Mathematics and its applications ser. Dordrecht: Kluwer Academic Publs. Vol. 375. 330 p.
- Donoho, D., and I. M. Johnstone. 1994. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81(3):425–455. doi: 10.1093/biomet/81.3.425.
- Donoho, D., and I. M. Johnstone. 1995. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* 90:1200–1224. doi: 10.1080/01621459.1995.10476626.
- Donoho, D., and I. M. Johnstone. 1998. Minimax estimation via wavelet shrinkage. *Ann. Stat.* 26(3):879–921. doi: 10.1214/aos/1024691081.
- Hall, P., G. Kerkycharian, and D. Picard. 1999. On the minimax optimality of block thresholded wavelet estimators. *Stat. Sinica* 9:33–49.
- Cai, T. 1999. Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Stat.* 28(3):898–924. doi: 10.1214/aos/1018031262.
- Cai, T., and L. Brown. 1998. Wavelet shrinkage for nonequispaced samples. *Ann. Stat.* 26(5):1783–1799. doi: 10.1214/aos/1024691357.
- Cai, T., and L. Brown. 1999. Wavelet estimation for samples with random uniform design. *Stat. Probabil. Lett.* 42:313–321. doi: 10.1016/S0167-7152(98)00223-5.
- Chicken, E. 2003. Block thresholding and wavelet estimation for nonequispaced samples. *J. Stat. Plan. Infer.* 116(1):113–129. doi: 10.1016/S0378-3758(02)00238-0.
- Kudryavtsev, A. A., and O. V. Shestakov. 2011. Asimptotika otsenki riska pri veyglet-veyvlet razlozhenii nablyudemogo signala [The average risk assessment of the wavelet decomposition of the signal]. *T-Comm: Telekommunikatsii i transport* [T-Comm: Telecommunications and Transport] 5(2):54–57. EDN: OPHYQN.
- Eroshenko, A. A., and O. V. Shestakov. 2014. Asymptotic normality of estimating risk upon the wavelet-vaguelette decomposition of a signal function in a model with correlated noise. *Moscow University Computational Mathematics Cybernetics* 38(3):110–117. doi: 10.3103/S0278641914030042. EDN: UEUPKZ.
- Eroshenko, A. A., A. A. Kudryavtsev, and O. V. Shestakov. 2015. Limit distribution of a risk estimate using the vaguelette-wavelet decomposition of signals in a model with correlated noise. *Moscow University Computational Mathematics Cybernetics* 39(1):6–13. doi: 10.3103/S0278641915010021. EDN: UFLCXN.
- Donoho, D. 1995. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. A.* 2(2):101–126. doi: 10.1006/acha.1995.1008.
- Lee, N. 1997. Wavelet-vaguelette decompositions and homogeneous equations. West Lafayette, IN: Purdue University. PhD Thesis. 93 p.
- Abramovich, F., and B. W. Silverman. 1998. Wavelet decomposition approaches to statistical inverse problems. *Biometrika* 85(1):115–129. doi: 10.1093/biomet/85.1.115.
- Shestakov, O. V., and E. P. Stepanov. 2023. Nelineynaya regularizatsiya obrashcheniya lineynykh odnorodnykh operatorov s pomoshch'yu metoda blochnoy porogovoy obrabotki [Nonlinear regularization of the inversion of linear homogeneous operators using the block thresholding method]. *Informatika i ee Primeneniya — Inform. Appl.* 17(4):2–8. doi: 10.14357/19922264230401. EDN: PGKKEY.
- Shestakov, O. V. 2019. Svoystva veyvlet-otsenok signalov, registriruemykh v sluchaynye momenty vremeni [Properties of wavelet estimates of signals recorded at random time points]. *Informatika i ee Primeneniya — Inform. Appl.* 13(2):16–21. doi: 10.14357/19922264190203. EDN: SEYEXW.
- Shestakov, O. V. 2020. Asimptoticheskaya regulyarnost' veyvlet-metodov obrashcheniya lineynykh odnorodnykh operatorov po nablyudeniyam, registriruemykh v sluchaynye momenty vremeni [Asymptotic regularity of the wavelet methods of inverting linear homogeneous operators from observations recorded at random times]. *Informatika i ee Primeneniya — Inform. Appl.* 14(1):3–9. doi: 10.14357/19922264200101. EDN: GJHMNK.
- Shestakov, O. V. 2024. Statisticheskie svoystva otsenki srednekvadratsionnogo riska metoda blochnoy porogovoy obrabotki v zadachakh neparametricheskoy regressii so sluchaynoy setkoy [Statistical properties of the mean-square risk estimate for the block threshold processing method in nonparametric regression problems with a random grid]. *Informatika i ee Primeneniya — Inform. Appl.* 18(4):26–33. doi: 10.14357/19922264240404. EDN: WHZQPX.
- Shestakov, O. V. 2010. Approksimatsiya raspredeleniya otsenki riska porogovoy obrabotki veyvlet-koeffitsientov

normal'nym raspredeleniem pri ispol'zovanii vyborochnoy dispersii [Normal approximation for distribution of risk estimate for wavelet coefficients thresholding when using sample variance]. *Informatika i ee Primeneniya — Inform. Appl.* 4(4):72–79. EDN: NULNZH.

21. Shestakov, O. V. 2025. Reshenie obratnykh statisticheskikh zadach s pomoshch'yu metodov porogovoy ob-

rabotki, dopuskayushchikh postroenie nesmeshchennoy otsenki srednekvadrachnogo riska [Solving inverse statistical problems using threshold processing methods that allow the construction of an unbiased estimate of the mean-square risk]. *Informatika i ee Primeneniya — Inform. Appl.* 19(1):74–81. doi: 10.14357/19922264250110. EDN: ZSIBDL.

Received June 22, 2025

Accepted August 15, 2025

Contributor

Shestakov Oleg V. (b. 1976) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, 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; leading scientist, Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; oshestakov@cs.msu.su

ОПТИМИЗАЦИЯ СКОРОСТНОГО РЕЖИМА ГРУЗОВОГО ПОЕЗДА ПО КРИТЕРИЮ СРЕДНЕГО УЩЕРБА ОТ АВАРИЙ*

А. В. Борисов¹, А. Н. Игнатов², В. А. Борисов³

Аннотация: Исследована задача определения скоростного режима поезда, минимизирующего ожидаемый ущерб, понесенный от аварий железнодорожного транспорта. Ущерб складывается из собственных потерь состава и возможных сопутствующих потерь составов на соседних путях. Функции вероятностей аварий различного типа параметризованы характеристиками топологии маршрута и показателями профиля — уклонами и кривизной пути. Вероятности также зависят от управления — скорости состава. От нее же зависят и функции ущерба — средние финансовые потери при повреждении одного вагона. Скоростной режим определяется решением задачи математического программирования с ограничениями. Он представляет собой функцию от расстояния до начала маршрута, сохраняющую постоянное значение на отрезках пути с постоянным профилем. Функция также удовлетворяет мгновенным геометрическим и интегральному временному ограничениям. Кусочно-постоянная скорость — идеализация, поэтому в работе дан алгоритм ее преобразования к физически реализуемому движению с равноускоренными переходами. В численном примере проанализировано влияние вида функций ущерба и временных ограничений на выбор оптимального скоростного режима.

Ключевые слова: кусочно-постоянное управление; скоростной режим; средний ущерб; нелинейная оптимизация

DOI: 10.14357/19922264250304

EDN: BJFFCD

1 Введение

Исследования в области оптимизации транспортных потоков посвящены по большей части поиску эффективного расписания или маршрутов движения [1–4]. Построение оптимального расписания на железнодорожном транспорте представляет собой отдельный класс задач [5–7]. Расчету надежности такого движения посвящено меньшее число работ, хотя задачи оценивания и управления надежностью в технических и коммуникационных системах не теряют свой актуальности [8, 9].

Закон движения поезда можно описать системой дифференциальных уравнений [10]. Управляющие переменные — тяга и торможение — были получены решением задачи оптимального управления, где в качестве критерия выступал общий расход энергии. Существует и другой критерий оптимальности перевозок — интегральный риск возникновения аварии на маршруте.

Впервые интегральный риск в качестве критерия оптимальности перевозок был использован в [11, 12]. Целью этих работ ставилось оптимальное расположение вагонов с опасными грузами, минимизирующее этот риск. Функция риска зависит

от скорости состава, максимально допустимой на участке маршрута, номера первого сошедшего вагона и причины схода. В [13] функция интегрального риска была усовершенствована и стала зависеть от реальной скорости состава, однако задача оптимизации скоростного режима поставлена не была.

Цель предлагаемой работы заключается в определении скоростного режима состава, минимизирующего функцию интегрального риска. Статья имеет следующую структуру. Раздел 2 представляет стохастическую модель ущерба, наносимого аварией состава, и постановку соответствующей задачи оптимизации скоростного режима. В задаче учитывается допустимый интервал скорости состава, а также ограничение сверху на время прохождения маршрута. Скорость состава постоянна на отрезках маршрута с постоянным профилем, т.е. кривизной и уклоном. В разделе также приведена процедура трансформации идеализированного кусочно-постоянного режима в физически реализуемый режим с равноускоренными переходами. В разд. 3 представлены результаты сравнительного численного анализа решения поставленной задачи оптимизации. В качестве функций потерь рассмотрены линейная и квадратичная функции, а также

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

¹ Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Московский центр фундаментальной и прикладной математики, aborisov@frccsc.ru

² Московский авиационный институт, alexei.ignatov1@gmail.com

³ Московский авиационный институт, forexample3423@gmail.com

их версии с насыщением. Исследовано влияние выбора этих функций, а также ограничения на время прохождения маршрута на вид оптимального скоростного режима. В разд. 4 предложены направления дальнейших исследований.

2 Модель ущерба от аварий и задача оптимизации скоростного режима

В работе предлагается описывать ущерб C от аварии грузового поезда как семейство неотрицательных случайных величин $C_E(\omega, V)$, где ω — случайное событие; V — вектор переменных программного управления (в данной работе представляет собой искомый скоростной режим прохождения маршрута); $E \triangleq \text{col}(\Theta, \Pi)$ — известный детерминированный вектор параметров, содержащий информацию Θ о составе поезда (количестве и длине вагонов и локомотивов, степени загруженности вагонов и пр.) и профиле маршрута Π (описании отрезков пути с одинаковыми уклоном и кривизной, наличии стрелок, переездов, параллельных путей и пр.).

Задача оптимизации скоростного режима прохождения поездом маршрута имеет вид:

$$R(V) = M \{C_E(\omega, V)\} \rightarrow \min_{V \in \mathcal{V}}, \quad (1)$$

где \mathcal{V} — множество допустимых управлений. В данной работе в их качестве выступает интервал допустимых скоростей, ограничение сверху на время прохождения маршрута, а также дополнительное ограничение на непрерывность изменения скорости.

Для решения этой задачи следует уточнить описание ущерба $C_E(\omega, V)$ и способа вычисления его среднего значения. Детальное описание всех характеристик и их значений приведено в [13, 14]. Прежде всего,

$$C_E(\omega, V) = C_E^0(\omega, V) + C_E^1(\omega, V),$$

где $C_E^0(\omega, V) \geq 0$ — ущерб, связанный с повреждением данного поезда; $C_E^1(\omega, V) \geq 0$ — возможный сопутствующий ущерб, связанный с аварией этого поезда, приведшей к аварии поезда на соседних путях.

Для вычисления $M \{C_E(\omega, V)\}$ воспользуемся формулой полного условного математического ожидания, разбив вероятностное пространство Ω на события $\{A_n\}_{n=0, \overline{N}}$, где A_0 — успешное прохождение поездом маршрута, а A_n , $n \geq 1$, — аварии, произошедшие на всех возможных шагах маршрута,

вызванные различными причинами и приведшие к сходу q_n единиц подвижного состава. Рассмотрим также события $\{B_n\}_{n=0, \overline{N}}$, заключающиеся в авариях поездов на соседних путях, вызванных аварией данного поезда. Отсутствие аварии влечет нулевой ущерб, и среднее значение принимает вид:

$$M \{C_E(\omega, V)\} = \sum_{n=1}^N (M \{C_E^0(\omega, V)|A_n\} + M \{C_E^1(\omega, V)|A_n B_n\} P \{B_n|A_n\}) P \{A_n\}.$$

Пусть функция $v = v_n(V)$ «выбирает» из вектора скоростного режима V скорость поезда в момент аварии в зависимости от n , а функция $\pi = \pi_n(\Pi)$ «выбирает» из вектора Π профиль пути в момент аварии в зависимости от n . Тогда отдельные элементы последнего выражения и их зависимости от параметров могут быть уточнены:

$$P \{A_n\} = P_n(V, E);$$

$$P \{B_n|A_n\} = p(v_n(V), \pi_n(\Pi), \Theta);$$

$$M \{C_E^0(\omega, V)|A_n\} = c^0(v_n(V)) q_n;$$

$$M \{C_E^1(\omega, V)|A_n B_n\} = c^1.$$

Это означает, что

- вероятность $p(v, \pi, \Theta)$ возникновения дополнительных аварий, вызванных данной, зависит только от скорости рассматриваемого поезда v и профиля места аварии π , а дополнительный средний ущерб постоянен;
- средний ущерб от аварии A_n рассматриваемого поезда пропорционален числу q_n сошедших вагонов с коэффициентом пропорциональности $c^0(v)$ — средним ущербом, причиненным одному вагону, движущемуся со скоростью v .

Таким образом, критерий $R(V)$ приобретает вид:

$$R(V) = \sum_{n=1}^N (c^0(v_n(V)) q_n + c^1 p(v_n(V), \pi_n(\Pi), \Theta)) P_n(V, \Pi).$$

Пусть весь маршрут S протяженностью S метров разбит на протяженные участки $\{S_\ell\}_{\ell=1, \overline{L}}$, $L \ll S$, имеющие постоянный профиль. Это означает, что профиль пути $\pi(s)$ имеет кусочно-постоянный вид:

$$\pi(s) = \sum_{\ell=1}^L \mathbf{I}_{S_\ell}(s) \bar{\pi}_\ell$$

и набор пар $\{(S_\ell, \bar{\pi}_\ell)\}_{\ell=1, \overline{L}}$ образует карту профиля маршрута.

Множество \mathcal{V} допустимых скоростных режимов поезда $V = \text{col}(v(1), \dots, v(S))$ удовлетворяет следующим ограничениям:

$$\begin{aligned} \text{I } v(s) &= \sum_{\ell=1}^L \mathbf{I}_{S_\ell}(s) \bar{v}_\ell; \\ \text{II } \underline{V} &\leq \bar{v}_\ell \leq \bar{V}; \\ \text{III } \sum_{\ell=1}^L \frac{|S_\ell|}{\bar{v}_\ell} &\leq T. \end{aligned}$$

Первое ограничение постулирует постоянную скорость прохождения поездом участка маршрута с постоянным профилем. Второе ограничение определяется допустимым интервалом скоростей $[\underline{V}, \bar{V}]$. Третье ограничение определяется максимальным временем T прохождения данного маршрута.

Задача поиска оптимального кусочно-постоянного скоростного режима имеет вид:

$$\begin{aligned} R(V) &= \sum_{n=1}^N (c^0(v_n(V)) q_n + \\ &+ c^1 p(v_n(V), \pi_n(\Pi), \Theta)) P_n(V, \Pi) \rightarrow \min_V \quad (2) \end{aligned}$$

при ограничениях I-III. Следует отметить, что функция потерь $c^0(v)$ должна удовлетворять двум очевидным свойствам: $c^0(0) = 0$ и неубыванию $c^0(v)$ по v .

Задача оптимизации (2) при ограничениях I–III имеет ряд положительных качеств. Во-первых, стратегия кусочно-постоянного скоростного режима интуитивно понятна: машинист выбирает и поддерживает постоянную скорость, соответствующую «степени опасности» профиля текущего участка. Во-вторых, решение оптимизационной задачи существенно упрощается, так как размерность L оптимизируемого вектора значительно меньше размерности исходного вектора $\text{col}(v(1), \dots, v(S))$ скоростного режима, т. е. длины S маршрута в метрах. Однако у кусочно-постоянного скоростного режима имеется существенный недостаток: он нереализуем на практике из-за невозможности мгновенного изменения скорости поезда. Тем не менее в случае, если отрезки постоянства скорости имеют достаточную протяженность, кусочно-постоянный режим может быть приведен к непрерывному с помощью простого рекуррентного преобразования, представленного ниже.

Зададим $\bar{v}_\ell \stackrel{\text{def}}{=} v(s_\ell)$, $\ell = \overline{1, L}$. Предположим, что состав может изменять скорость с максимальным ускорением a .

Скоростной режим V в задаче определяется как функция расстояния s от начала маршрута. Пусть $h > 0$ — это шаг по пройденному пути, с которым изменяется скоростной режим; в данной задаче $h = 1$ м. Согласно элементарным кинематическим соотношениям для равноускоренного движения, если на s -м шаге скорость поезда равна $v(s)$,

то при равноускоренном и равнозамедленном движении скорость на следующем шаге определяется формулой $v(s+1) = \sqrt{v^2(s) \pm 2ah}$.

Достаточная протяженность участков S_ℓ означает, что их длины достаточно, чтобы изменить скорость нужным образом за время прохождения участка. Алгоритм преобразования кусочно-постоянного скоростного режима V в непрерывный $\tilde{V} = \text{col}(\tilde{v}(1), \dots, \tilde{v}(S))$ имеет следующий вид.

Шаг 0. задается $a > 0$.

Шаг 1. $\ell := 2$; $\tilde{V} := V$.

Шаг 2. Если $\ell > L$, то закончить алгоритм.

Шаг 3. Если $\bar{v}_\ell > \bar{v}_{\ell-1}$, то перейти к шагу 7.

Шаг 4. $\sigma := s_\ell$.

Шаг 5. $\tilde{v}(\sigma) = \max(\sqrt{\bar{v}^2(\sigma) - 2ah}, \bar{v}_{\ell-1})$; $\sigma := \sigma + 1$.

Шаг 6. Если $\sqrt{\bar{v}^2(\sigma) - 2ah} \geq \bar{v}_{\ell-1}$, то перейти к шагу 5, иначе к шагу 10.

Шаг 7. $\sigma := s_\ell - 1$.

Шаг 8. $\tilde{v}(\sigma) = \max(\sqrt{\bar{v}^2(\sigma) - 2ah}, \bar{v}_\ell)$; $\sigma := \sigma - 1$.

Шаг 9. Если $\sqrt{\bar{v}^2(\sigma) - 2ah} \geq \bar{v}_\ell$, то перейти к шагу 8, иначе к шагу 10.

Шаг 10. $\ell := \ell + 1$; перейти к шагу 2.

Смысл преобразования следующий. Участки с постоянной скоростью преобразуются последовательно. Если скорость на предыдущем участке меньше скорости на последующем, то несколько последних интервалов предыдущего участка состав должен пройти равноускоренно с ускорением a , кроме, возможно, одного, ускорение на котором будет меньше a . Если скорость на предыдущем участке выше скорости на последующем, то несколько первых интервалов последующего участка состав также должен пройти равнозамедленно с ускорением a , кроме, возможно, одного, ускорение на котором будет меньше a . Данный алгоритм позволяет не только привести кусочно-постоянный скоростной режим к физически реализуемому непрерывному, но и сохранить выполнение ограничения на время прохождения маршрута $h \sum_{s=1}^S 1/\tilde{v}(s) \leq T$.

3 Сравнительный численный анализ оптимальных скоростных режимов при различных функциях потерь

Данный раздел содержит результаты комплекса численных экспериментов. Его цели:

- анализ зависимости оптимальных скоростных режимов от выбора пары $(c^0(v), c^1)$;
- исследование изменения среднего ущерба при переходе от оптимального кусочно-постоянного к непрерывному скоростному режиму.

Параметры вычислительных экспериментов совпадают с использованными в [13], за исключением вспомогательной функции $g_2(\cdot)$, взятой из [15].

- Маршрут поезда составляет 250 км. Поезд включает в себя 2 локомотива по 20 м и 58 вагонов по 14 м, т. е. общая длина поезда составляет 852 м. Движение проходит по двупутной дороге без переездов; имеются стрелочные переводы. Аварии в данном примере порождаются следующими причинами: неисправностью подвижного состава, полотна пути или сходом на стрелочном переводе. При этом возможно нанесение ущерба как самому составу, так и поездам, следующим по соседнему пути, т. е. при расчете критерия задействованы оба элемента пары $(c^0(v), c^1)$.
- Зависимость уклона и кривизны путей от расстояния до начала маршрута приведена в табл. 1.

Таблица 1 Карта уклонов и кривизны

ℓ	Положение	Кривизна κ	Уклон γ
0	$[-852, 0]$	0	0
1	$[50001, 60000]$	1/1500	-0,009
2	$[60001, 70000]$	1/1500	-0,003
3	$[70001, 100000]$	0	0,01
4	$[100001, 110000]$	1/800	-0,004
5	$[1, 50000]$	1/2000	-0,0001
6	$[110001, 150000]$	1/1000	0,008
7	$[150001, 200000]$	0	0
8	$[200001, 205000]$	1/600	0,005
9	$[205001, 220000]$	0	-0,01
10	$[220001, 250000]$	0	0,005

- Вероятности аварий различного вида $P_n(V, \Pi)$ и $p(v, \pi)$ вычисляются по формулам, представленным в [13].
- Допустимый интервал скоростей состава — от 5 до 25 м/с, максимально допустимое ускорение $a = 0,1$ м/с². Оптимизация скорости выполняется для двух ограничений по времени: 15 000 и 20 000 с.
- Сопутствующие средние потери от аварий на соседних путях $c^1 = 1,6 \cdot 10^6$.

В качестве функций потерь от аварии состава выбраны линейная функция $c_\ell^0(v)$; квадратичная функция $c_q^0(v)$; линейная функция с насыщением $c_{\ell s}^0(v)$; квадратичная функция с насыщением, вариант № 1 $c_{qs1}^0(v)$; квадратичная

функция с насыщением, вариант № 2 $c_{qs2}^0(v)$. Были выбраны два набора функций:

(а) первый набор:

$$c_\ell^0(v) = \frac{9}{200} v; \quad c_q^0(v) = \frac{81}{31\,000} v^2;$$

$$c_{\ell s}^0(v) = \frac{2,358}{1 + e^{-0,15v}} - 1,179;$$

$$c_{qs1}^0(v) = 1,148 (1 - e^{-0,3v} (0,045v^2 + 0,03v + 1));$$

$$c_{qs2}^0(v) = \frac{0,0018 (v^2 e^{-0,5(v-19)} + 625)}{1 + e^{-0,5(v-19)}};$$

(б) второй набор:

$$c_\ell^0(v) = \frac{9}{200} v; \quad c_q^0(v) = 0,0018v^2;$$

$$c_{\ell s}^0(v) = \frac{1,782}{1 + e^{-0,15v}} - 0,891;$$

$$c_{qs1}^0(v) = 0,92 (1 - e^{-0,3v} (0,045v^2 + 0,03v + 1));$$

$$c_{qs2}^0(v) = \frac{0,0022 (v^2 e^{-0,5(v-19)} + 625)}{1 + e^{-0,5(v-19)}}.$$

Выбор типов функций объясняется следующим. Линейная функция представляет простейшую зависимость между величинами. Линейная функция с насыщением позволяет учитывать ограничение сверху на ущерб: прямой ущерб от аварии вагона не может превышать его стоимость. Квадратичная функция от скорости предполагает модель, когда ущерб зависит от кинетической энергии вагона в момент аварии, которая, в свою очередь, пропорциональна квадрату скорости. Различные квадратичные функции с насыщением учитывают ограничение сверху на величину ущерба и отличаются близостью к исходной параболе и скоростью насыщения.

Наборы функций потерь отличаются способом их «нормировки». Дело в том, что критерии должны быть сравнимыми. Первый набор предполагает равенство функций при максимально допустимой скорости: $c_i^0(25) \equiv \text{Const}$ для всех видов критерия. Второй набор предполагает одинаковые усредненные потери:

$$\frac{1}{25 - 5} \int_5^{25} c_i^0(v) dv \equiv \text{Const}.$$

Графики функций первого набора представлены на рис. 1, а, второго — на рис. 1, б.

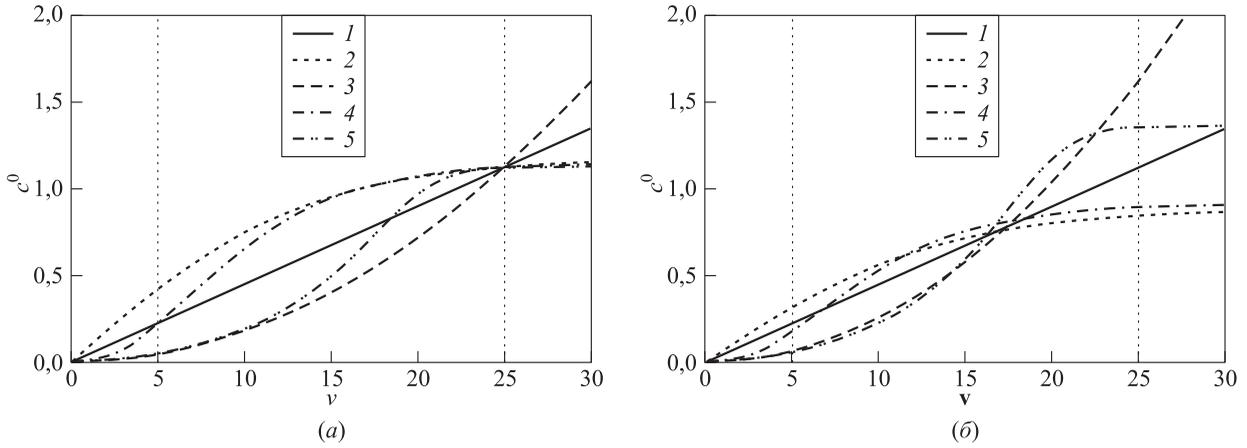


Рис. 1 Два набора функций потерь: (а) первый набор ($1 - c_{\ell}^0$; $2 - c_{\ell s}^0$; $3 - c_q^0$; $4 - c_{qs1}^0$; $5 - c_{qs2}^0$); (б) второй набор ($1 - c_{\ell}^0$; $2 - c_{\ell s}^0$; $3 - c_q^0$; $4 - c_{qs1}^0$; $5 - c_{qs2}^0$)

Вычисления выполнялись на высокопроизводительном комплексе *Huawei Taishan 200* архитектуры ARM с двумя процессорами *Huawei Kunpeng 920* (2,6 ГГц, 64 ядра каждый) и 512 ГБ ОЗУ под управлением операционной системы *CentOS* с использованием системы управления заданиями *SLURM*. Программный комплекс написан на языке *Julia 1.11.1* с привлечением библиотек *JuMP* и *IPOPT*, предназначенных для математического моделирования задач оптимизации и их численного решения. Был использован метод внутренней точки [16].

На рис. 2 представлены преобразованные непрерывные скоростные режимы:

- рис. 2, а — скорости $v_{\ell}^{15}(s)$ и $v_{\ell}^{20}(s)$, вычисленные при линейных потерях и временных ограничениях 15 000 и 20 000 с, а также соответствующие скорости $v_{\ell s}^{15}(s)$ и $v_{\ell s}^{20}(s)$ для функций линейных потерь с насыщением (первый набор функций потерь);
- рис. 2, б — скорости $v_q^{15}(s)$ и $v_q^{20}(s)$, вычисленные при квадратичных потерях и временных ограничениях 15 000 и 20 000 с, и $v_{qs1}^{15}(s)$, $v_{qs1}^{20}(s)$, $v_{qs2}^{15}(s)$ и $v_{qs2}^{20}(s)$ для функций квадратичных потерь с насыщением обоих типов (первый набор);
- рис. 2, в — скорости $v_{\ell}^{15}(s)$ и $v_{\ell}^{20}(s)$, вычисленные при линейных потерях и временных ограничениях 15 000 и 20 000 с, и $v_{\ell s}^{15}(s)$ и $v_{\ell s}^{20}(s)$ для функций линейных потерь с насыщением (второй набор);
- рис. 2, г — скорости $v_q^{15}(s)$ и $v_q^{20}(s)$, вычисленные при квадратичных потерях и временных ограничениях 15 000 и 20 000 с, и $v_{qs1}^{15}(s)$, $v_{qs1}^{20}(s)$, $v_{qs2}^{15}(s)$ и $v_{qs2}^{20}(s)$ для функций квадратичных потерь с насыщением обоих типов (второй набор).

Результаты вычислений представлены в табл. 2. В ней для обоих наборов функций потерь и ограничений по времени 15 000 и 20 000 с собраны

- ущерб $R(v)$, соответствующий идеализированному оптимальному кусочно-постоянному скоростному режиму движения состава;
- ущерб $R(\tilde{v})$, соответствующий субоптимальному непрерывному режиму;
- процент потерь в критерии при переходе от кусочно-постоянного к непрерывному режиму

$$\text{Loss} = \frac{R(\tilde{v}) - R(v)}{R(v)} \cdot 100\%;$$

- время прохождения маршрута при непрерывном скоростном режиме $T(\tilde{v})$;
- число итераций алгоритма оптимизации.

По результатам вычислений можно сделать следующие выводы.

1. Во всех расчетах оптимум достигался за умеренное число итераций: не более 55.
2. Переход от кусочно-постоянного к непрерывному скоростному режиму увеличивает потери не существенно: не более чем на 1,91%. При этом ограничение на максимальное время прохождения маршрута не нарушается.
3. Ужесточение временного ограничения с 20 000 до 15 000 с может увеличить средний ущерб более чем в 2 раза.
4. При ограничении времени 20 000 с сами скоростные режимы близки, чего нельзя сказать о значении среднего ущерба.

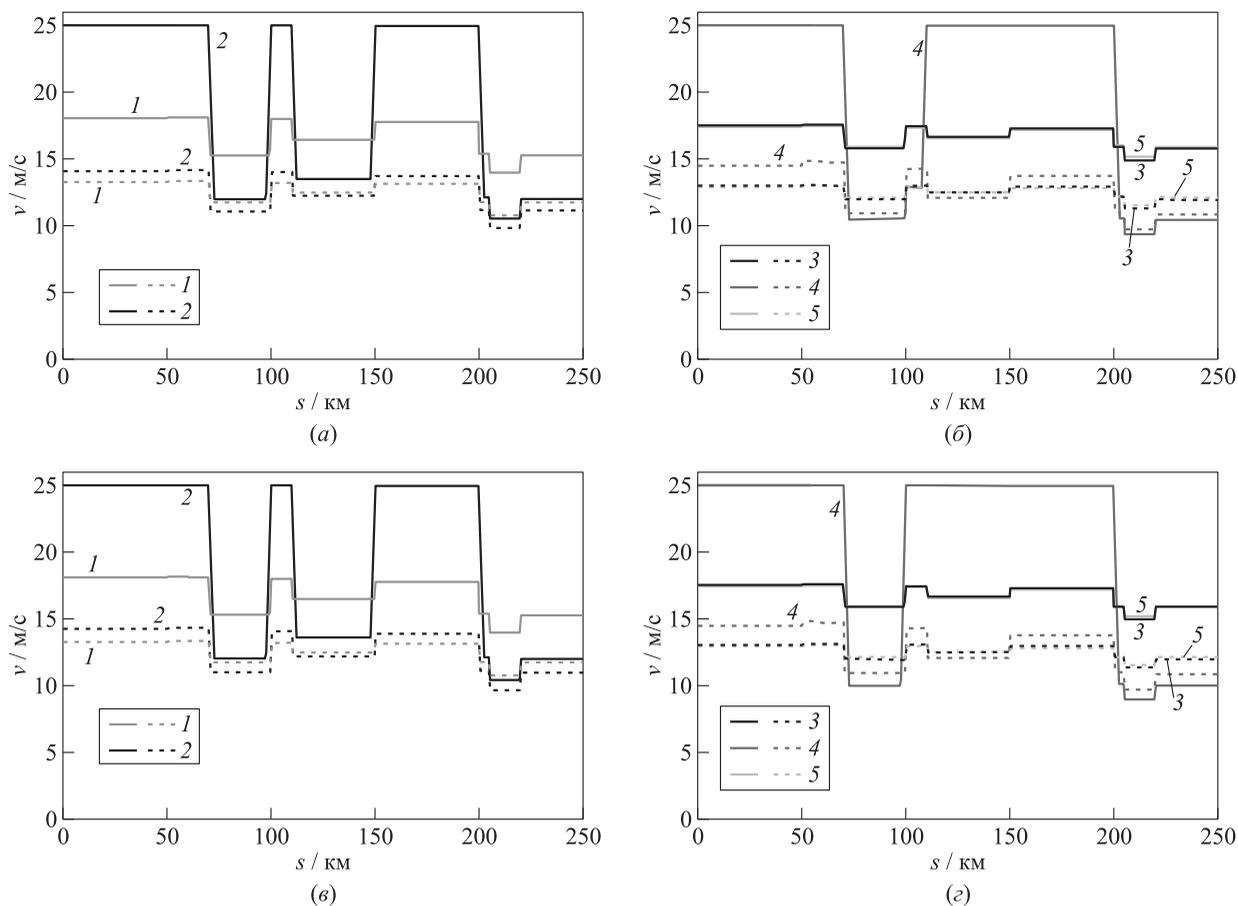


Рис. 2 Непрерывные скоростные режимы, рассчитанные для разных функций потерь (верхний ряд — первый набор; нижний ряд — второй набор) и временных ограничений (сплошные линии — 15 000 с; пунктирные линии — 20 000 с): 1 — v_ℓ ; 2 — $v_{\ell s}$; 3 — v_q ; 4 — v_{qs1} ; 5 — v_{qs2}

Таблица 2 Сводная таблица результатов эксперимента

№ набора	T	Параметр	c_ℓ^0	$c_{\ell s}^0$	c_q^0	c_{qs1}^0	c_{qs2}^0
1	15 000	$R(v)$	279,14	358,12	186,78	352,06	241,38
		$R(\tilde{v})$	279,37	363,86	186,83	357,07	241,5
		Loss, %	0,08	1,6	0,02	1,4	0,05
		$T(\tilde{v})$	14991	14691	14997	14739	14997
		Iтер	28	33	38	40	39
	20 000	$R(v)$	188,35	288,07	95,09	275,29	105,59
		$R(\tilde{v})$	188,4	288,28	95,11	275,65	105,61
		Loss, %	0,03	0,07	0,02	0,1	0,02
		$T(\tilde{v})$	19996	19986	19998	19980	19999
		Iтер	42	27	30	46	38
2	15 000	$R(v)$	279,14	411,66	270,25	282,16	290,05
		$R(\tilde{v})$	279,37	417,01	270,37	287,55	290,19
		Loss, %	0,08	1,3	0,04	1,91	0,05
		$T(\tilde{v})$	14990	14693	14997	14660	14997
		Iтер	39	22	35	22	55
	20 000	$R(v)$	188,35	347,83	137,15	220,89	126,66
		$R(\tilde{v})$	188,4	348,07	137,18	221,17	126,68
		Loss, %	0,02	0,07	0,02	0,13	0,02
		$T(\tilde{v})$	19996	19983	19998	19980	19999
		Iтер	39	32	42	42	43

5. Скоростные режимы, вычисленные для одинаковых функций потерь разных наборов, очень близки, но значения критериев отличаются.
6. В обоих наборах скоростные режимы, построенные для линейной с насыщением и квадратичной с насыщением (первый вариант) функций потерь, предполагают прохождение части участков с максимально допустимой скоростью. Данные критерии растут очень быстро и затем стабилизируются, поэтому рациональнее проходить часть относительно безопасных участков максимально быстро, чтобы двигаться на опасных участках с умеренной скоростью.

4 Заключение

В работе решена задача выбора оптимального программного скоростного режима движения грузового состава, минимизирующего средний интегральный ущерб от аварий, при мгновенных геометрических и интегральном временном ограничениях.

Представленные исследования могут быть продолжены по следующим направлениям.

Во-первых, перспективным представляется построение модели момента наступления аварии и ее последствий, обладающей марковским свойством. Она позволила бы существенно упростить как саму задачу оптимизации, так и идентификацию ее вероятностных параметров по реальным статистическим данным.

Во-вторых, ущерб от аварий, нанесенных подвижному составу, не единственная статья расходов на железнодорожном транспорте. В критерий оптимальности следует включить затраты энергии на перевозку, амортизацию оборудования, страховые отчисления и пр.

В-третьих, интерес представляет рассмотрение аналога этой задачи в непрерывном времени и использование для ее решения аппарата стохастического управления скачкообразными процессами.

Литература

1. *Voctor F. F., Laporte G., Renaud J.* Heuristics for the traveling purchaser problem // *Comput. Oper. Res.*, 2003. Vol. 30. No. 4. P. 491–504. doi: 10.1016/S0305-0548(02)00020-5.
2. *Хачай М. Ю., Огородников Ю. Ю.* Полиномиальная приближенная схема для задачи маршрутизации транспортных средств с ограничениями на грузоподъемность и временные промежутки обслуживания // *Труды Института математики и механики УрО РАН*, 2018. Т. 24. Вып. 3. С. 233–246. doi: 10.21538/0134-4889-2018-24-3-233-246.
3. *Fluschnik T., Morik M., Sorge M.* The complexity of routing with collision avoidance // *J. Comput. Syst. Sci.*, 2019. Vol. 102. P. 69–86. doi: 10.1016/j.jcss.2019.01.001.
4. *Mor A., Speranza M. G.* Vehicle routing problems over time: A survey // *4OR — Q. J. Oper. Res.*, 2020. Vol. 18. No. 2. P. 129–149. doi: 10.1007/s10288-020-00433-2.
5. *Лазарев А. А., Мусатова Е. Г.* Целочисленные постановки задачи формирования железнодорожных составов и расписания их движения // *Управление большими системами*, 2012. Т. 38. С. 161–169. doi: 10.1134/S0005117913120084. EDN: PBXQQD.
6. *Forsgren M., Aronsson M., Gestrelius S.* Maintaining tracks and traffic flow at the same time // *J. Rail Transport Planning Management*, 2013. Vol. 3. No. 3. P. 111–123. doi: 10.1016/j.jrtpm.2013.11.001.
7. *Босов А. В., Игнатов А. Н., Наумов А. В.* Алгоритмы приближенного решения задачи назначения «технологического окна» на участках железнодорожной сети // *Информатика и её применения*, 2021. Т. 15. Вып. 4. С. 3–11. doi: 10.14357/19922264210401. EDN: OPTRUD.
8. *Соколов И. А., Степченков Ю. А., Дьяченко Ю. Г., Рожественский Ю. В.* Оценка надежности синхронного и самосинхронного конвейеров // *Информатика и её применения*, 2022. Т. 16. Вып. 4. С. 2–7. doi: 10.14357/19922264220401. EDN: GWXJHM.
9. *Малашенко Ю. Е., Назарова И. А.* Гарантированные оценки показателей работоспособности многопользовательской сети при повреждениях // *Информатика и её применения*, 2025. Т. 19. Вып. 1. С. 16–24. doi: 10.14357/19922264250103. EDN: MYNICH.
10. Оптимизация управления движением поездов / Под ред. Л. А. Баранова. — М.: МИИТ, 2011. 164 с. EDN: QNYRIV.
11. *Bagheri M., Saccomanno F., Fu L.* Effective placement of dangerous goods cars in rail yard marshaling operation // *Can. J. Civil Eng.*, 2010. Vol. 37. P. 753–762. doi: 10.1139/L10-015.
12. *Rahbar M., Bagheri M.* Risk assessment framework for the rail transport of hazardous materials // *Transp. Res. Rec.*, 2014. Vol. 2411. No. 1. P. 90–95. doi: 10.3141/2411-11.
13. *Босов А. В., Игнатов А. Н.* О задаче оценки и анализа риска транспортных происшествий на рельсовом транспорте // *Информатика и её применения*, 2023. Т. 17. Вып. 1. С. 73–82. doi: 10.14357/19922264230110. EDN: SSMEEK.
14. *Замышляев А. М., Игнатов А. Н., Кибзун А. И., Новожилов Е. О.* О нарушении безопасности движения, связанном с выходом в габарит соседнего пути подвижных единиц грузового поезда, сошедших с рельсов // *Надежность*, 2018. Т. 18. Вып. 3. С. 39–45. doi: 10.21683/1729-2646-2018-18-3-39-45. EDN: YATOLJ.
15. *Замышляев А. М., Игнатов А. Н., Кибзун А. И., Новожилов Е. О.* Функциональная зависимость между количеством вагонов в сходе из-за неисправностей

вагонов или пути и факторами движения // Надежность, 2018. Т. 18. Вып. 1. С. 53–60. doi: 10.21683/1729-2646-2018-18-1-53-60.

16. Nesterov Y., Nemirovski A. Interior-point polynomial algorithms in convex programming. — Philadelphia, PA, USA: SIAM, 1994. 405 p.

Поступила в редакцию 29.05.2025

Принята к публикации 15.08.2025

OPTIMIZATION OF A TRAIN SPEED PROFILE BASED ON THE EXPECTED ACCIDENT DAMAGE CRITERION

A. V. Borisov^{1,2}, A. N. Ignatov³, and V. A. Borisov³

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

²Moscow Center for Fundamental and Applied Mathematics, M.V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

³Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation

Abstract: The paper focuses on designing a freight train speed profile that minimizes the expected damage from various types of railway accidents. Total losses include both damage to a considered train and potential harm to trains on adjacent tracks. The probability functions for all accident types are parameterized by the route’s topology and profile, i. e., its local slope and curvature. These probability functions, along with those describing the average financial loss per derailed car, treat train speed as a control variable. The speed profile is a solution to the constrained mathematical programming problem. It is represented as a piecewise constant function, remaining constant over each segment of the route with uniform slope or curvature. This profile satisfies both instantaneous geometric and integral time constraints. Since a piecewise constant speed profile looks physically unrealistic, the paper also proposes a method for transforming it into a profile with uniformly accelerated transitions. A numerical example is provided to illustrate how different loss functions and time constraints affect the choice of an optimal speed profile.

Keywords: piecewise constant control; speed profile; expected damage; nonlinear optimization

DOI: 10.14357/19922264250304

EDN: BJFFCD

Acknowledgments

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

References

1. Boctor, F. F., G. Laporte, and J. Renaud. 2003. Heuristics for the traveling purchaser problem. *Comput. Oper. Res.* 30(4):491–504. doi: 10.1016/S0305-0548(02)00020-5.
2. Khachai, M. Yu., and Yu. Yu. Ogorodnikov. 2019. Polynomial-time approximation scheme for the capacitated vehicle routing problem with time windows. *P. Steklov I. Math.* 307(S1):51–63. doi: 10.1134/S0081543819070058. EDN: MWXBYZ.
3. Fluschnik, T., M. Morik, and M. Sorge. 2019. The complexity of routing with collision avoidance. *J. Comput. Syst. Sci.* 102:69–86. doi: 10.1016/j.jcss.2019.01.001.
4. Mor, A., and M. G. Speranza. 2020. Vehicle routing problems over time: A survey. *4OR—Q. J. Oper. Res.* 18(2):129–149. doi: 10.1007/s10288-020-00433-2.
5. Lazarev, A. A., and E. G. Musatova. 2012. Tselochislennye postanovki zadachi formirovaniya zheleznodorozhnykh sostavov i raspisaniya ikh dvizheniya [Integer formulations of freight train design and scheduling problems]. *Upravlenie bol'shimi sistemami* [Large Scale Systems Control] 38:161–169. doi: 10.1134/S0005117913120084. EDN: PBXQQD.
6. Forsgren, M., M. Aronsson, and S. Gestrelus. 2013. Maintaining tracks and traffic flow at the same time. *J. Rail Transport Planning Management* 3(3):111–123. doi: 10.1016/j.jrtpm.2013.11.001.
7. Bosov, A. V., A. N. Ignatov, and A. V. Naumov. Algoritmy priblizhennogo resheniya zadachi naznacheniya “tekhnologicheskogo okna” na uchastkakh zheleznodorozhnoy seti [Algorithms for an approximate solution of the track possession problem on the railway network segment]. *Informatika i ee Primeneniya — Inform. Appl.* 15(4):3–11. doi: 10.14357/19922264210401. EDN: OPTRUD.
8. Sokolov, I. A., Yu. A. Stepchenkov, Yu. G. Diachenko, and Yu. V. Rogdestvenski. 2022. Otsenka nadezhnosti sinkhronnogo i samosinkhronnogo konveyerov [Synchronous and self-timed pipeline’s reliability estimation].

- Informatika i ee Primeneniya — Inform. Appl.* 16(4):2–7. doi: 10.14357/19922264220401. EDN: GWXJHM.
9. Malashenko, Yu. E., and I. A. Nazarova. 2025. Garantirovannye otsenki pokazateley rabotosposobnosti mnogopol'zovatel'skoy seti pri povrezhdeniyakh [Guaranteed estimates of the multiuser network performance indicators in case of damage]. *Informatika i ee Primeneniya — Inform. Appl.* 19(1):16–24. doi: 10.14357/19922264250103. EDN: MYHICH.
 10. Baranov, L. A., ed. 2011. *Optimizatsiya upravleniya dvizheniem poezdov* [Control optimization of train transportation]. Moscow: MIIT. 164 p. EDN: QNYRIV.
 11. Bagheri, M., F. Saccomanno, and L. Fu. 2010. Effective placement of dangerous goods cars in rail yard marshaling operation. *Can. J. Civil Eng.* 37(5):753–762. doi: 10.1139/L10-015.
 12. Rahbar, M., and M. Bagheri. 2014. Risk assessment framework for the rail transport of hazardous materials: Formulation and solution. *Transp. Res. Rec.* 2411(1):90–95. doi: 10.3141/2411-11.
 13. Bosov, A. V., and A. N. Ignatov. 2023. O zadache otsenki i analiza riska transportnykh proisshestviy na rel'sovom transporte [On the problem of assessing and analyzing traffic accidents risk on the rail transport]. *Informatika i ee Primeneniya — Inform. Appl.* 17(1):73–82. doi: 10.14357/19922264230110. EDN: SSMEEK.
 14. Zamyshliaev, A. M., A. N. Ignatov, A. I. Kibzun, and E. O. Novozhilov. 2018. O narushenii bezopasnosti dvizheniya, svyazannom s vykhodom v gabarit sosednego puti podvizhnykh edinit grazhkovogo poezda, soshedshikh s rel'sov [On traffic safety incidents caused by intrusion of derailed freight cars into the operational space of an adjacent track]. *Nadezhnost' [Dependability]* 18(3):39–45. doi: 10.21683/1729-2646-2018-18-3-39-45. EDN: YATOLJ.
 15. Zamyshliaev, A. M., A. N. Ignatov, A. I. Kibzun, and E. O. Novozhilov. 2018. Funktsional'naya zavisimost' mezhdru kolichestvom vagonov v skhode iz-za neispravnostey vagonov ili puti i faktorami dvizheniya [Functional dependency between the number of wagons derailed due to wagon or track defects and the traffic factors]. *Nadezhnost' [Dependability]* 18(1):53–60. doi: 10.21683/1729-2646-2018-18-1-53-60.
 16. Nesterov, Y., and A. Nemirovski. 1994. *Interior-point polynomial algorithms in convex programming*. Philadelphia, PA: SIAM. 405 p.

Received May 29, 2025

Accepted August 15, 2025

Contributors

Borisov Andrey V. (b. 1965) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; aborisov@frcsc.ru

Ignatov Aleksei N. (b. 1991) — Candidate of Science (PhD) in physics and mathematics, assistant professor, Department of Probability Theory and Computational Modeling, Faculty of Computer Science and Applied Mathematics, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; alexei.ignatov1@gmail.com

Borisov Vasily A. (b. 2002) — master student, Department of Mathematical Cybernetics, Faculty of Computer Science and Applied Mathematics, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; forexample3423@gmail.com

МОДЕЛЬ РАЗВИТИЯ МНОГОПОЛЬЗОВАТЕЛЬСКОЙ СИСТЕМЫ СВЯЗИ ПРИ НАРАЩИВАНИИ ПРОПУСКНОЙ СПОСОБНОСТИ СЕТИ

Ю. Е. Малащенко¹, И. А. Назарова², М. В. Козлов³

Аннотация: Исследуются показатели функциональных возможностей многопользовательской системы связи при наращивании пропускной способности сети. В рамках вычислительного эксперимента анализируется влияние увеличения пропускной способности ребер вдоль маршрутов передачи максимальных потоков. Для каждой пары узлов-корреспондентов независимо определяется величина предельно допустимого межузлового потока при монопольном режиме управления. Полученные максимальные значения используются для построения и сравнения векторов равных межузловых потоков всех видов, которые могут одновременно передаваться в сети. Вводится понятие вектор-отклика системы на наращивание пропускной способности ребер вдоль маршрута передачи. Для каждого варианта проекта реконструкции и каждой пары узлов-корреспондентов вычисляется отношение пропускной способности к приросту максимального потока, найденные значения переупорядочиваются по максиминному правилу. На основании вектор-откликов формируется набор гарантированных оценок предельно возможных изменений эксплуатационных параметров. Анализируются результаты вычислительных экспериментов для сетей с различными структурными особенностями

Ключевые слова: потоковая модель сети связи; гарантированные оценки при наращивании пропускной способности; маршруты передачи максимального потока

DOI: 10.14357/19922264250305

EDN: EBFVDL

1 Введение

В рамках вычислительных экспериментов на математической модели передачи многопродуктового потока [1] анализируется изменение показателей функционирования многопользовательской системы связи [2].

В монопольном режиме управления для каждой пары узлов независимо вычисляется максимальный однопродуктовый поток и создается список (набор) ребер маршрутов его передачи. При моделировании каждый полученный набор рассматривается как вариант *проекта* модернизации сети путем увеличения пропускной способности ребер указанных маршрутов.

Предельные значения однопродуктового потока определяют угловые точки множества достижимых межузловых потоков. С помощью векторов, соответствующих угловым точкам, формируется вектор-луч равных межузловых потоков различных видов, которые могут передаваться одновременно в исходной сети.

Увеличение пропускной способности ребер в процессе модернизации сети приводит к росту максимальных однопродуктовых потоков, что ве-

дет к изменению координат угловых точек. При этом вектор-луч равных межузловых потоков сохраняет ориентацию относительно осей координат, что позволяет оценить относительные изменения характеристик сети при увеличении пропускной способности различных вариантов проектов.

В ходе эксперимента в качестве эталонных показателей для каждой пары узлов-корреспондентов рассматривается максимально допустимый однопродуктовый межузловой поток, передаваемый в монопольном режиме [3] в исходной сети. Для каждого варианта *проекта* при увеличении пропускной способности ребер вдоль маршрутов передачи рассчитываются новые значения максимальных потоков и соответствующие компоненты вектора равных потоков [4] для всех пар в модернизированной сети. На последующих этапах экспериментов векторы используются для построения гарантированных оценок функциональных характеристик каждой реконструируемой сети.

Полученные массивы данных сохраняются для каждого варианта *проекта* и подсчитываются относительные изменения максимальных потоков. С геометрической точки зрения оценивается деформация множества достижимых потоков — отно-

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, malash09@ccas.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, irina-nazar@yandex.ru

³Федеральный исследовательский центр «Информатика и управление» Российской академии наук, kztmp2@ccas.ru

сительный сдвиг угловых точек по координатным осям.

В модели вводится понятие вектор-отклика на рост пропускной способности сети. Для каждой пары узлов и каждого варианта *проекта* вычисляется отношение увеличения пропускной способности к приросту максимального потока в обновленной (модернизированной) сети.

Для анализа изменения технических возможностей сети компоненты вектор-откликов перепорядочиваются по максиминному правилу. Полученная лексикографически упорядоченная последовательность представляет собой набор гарантированных оценок предельно возможных изменений агрегированных характеристик для различных проектов развития.

Данная работа продолжает исследования функциональных характеристик многопользовательских систем сетей связи при изменении пропускной способности сети [1]. Вычислительные эксперименты проводились в рамках методологии исследования операций [5, 6] на основе методов потокового программирования [7–9].

2 Математическая модель

Для описания многопользовательской сетевой системы связи воспользуемся следующей математической записью модели передачи многопродуктового потока (далее — М-модель). Сеть G задается множествами $\langle V, R, U, P \rangle$: узлов (вершин) сети $V = \{v_1, v_2, \dots, v_n, \dots, v_N\}$; неориентированных ребер $R = \{r_1, r_2, \dots, r_k, \dots, r_E\}$; ориентированных дуг $U = \{u_1, u_2, \dots, u_k, \dots, u_{2E}\}$; пар узлов-корреспондентов $P = \{p_1, p_2, \dots, p_M\}$. Предполагается, что в сети отсутствуют петли и двоянные ребра.

Ребро $r_k \in R$ соединяет *смежные* вершины v_{n_k}, v_{j_k} . Каждому ребру r_k ставятся в соответствие две ориентированные дуги u_k, u_{k+E} из множества U . Дуги $\{u_k, u_{k+E}\}$ определяют прямое и обратное направление передачи потока по ребру r_k между концевыми вершинами v_{n_k}, v_{j_k} . Для каждой вершины v_n формируется список $K(n)$ номеров инцидентных ей ребер: $K(n) = \{k_1, k_2, \dots, k_{a(n)}\}$, где $a(n)$ — число инцидентных ребер для v_n .

В многопользовательской сети G рассматривается $M = N(N - 1)$ независимых, невазимоменяемых и равноправных межузловых потоков различных видов. Каждой паре узлов-корреспондентов p_m из множества P соответствуют: вершина-источник с номером s_m , из s_m входной поток m -го вида поступает в сеть; вершина-приемник с номером t_m , из t_m поток m -го вида покидает сеть. Для каждой вершины $v_n \in V, n = \overline{1, N}$, в подмножество $P(v_n)$

входят все пары-корреспонденты, для которых v_n служит узлом-источником:

$$P(v_n) = \{p_m | s_m = n, t_m \neq n, t_m = \overline{1, N}\},$$

а для каждого $P(v_n)$ — список номеров $M(n)$ пар p_m , входящих в подмножество $P(v_n)$:

$$M(n) = \{m_1(n), m_2(n), \dots, m_{N-1}(n)\}.$$

Обозначим через z_m величину *межузлового* потока m -го вида, поступающего в сеть через узел с номером s_m и покидающего сеть из узла с номером t_m ; $x_{mk}, x_{m(k+E)}$ — поток m -го вида, который передается по дугам u_k и u_{k+E} согласно направлению передачи, $x_{mk} \geq 0, x_{m(k+E)} \geq 0, m = \overline{1, M}, k = \overline{1, E}$; $S(v_n)$ — множество номеров исходящих дуг, по ним поток покидает узел v_n ; $T(v_n)$ — множество номеров входящих дуг, по ним поток поступает в узел v_n . Состав множеств $S(v_n), T(v_n)$ однозначно определяется в ходе выполнения следующей процедуры.

Пусть некоторое ребро $r_k \in R$ соединяет вершины с номерами n и j , такими что $n < j$. Тогда ориентированная дуга $u_k = (v_n, v_j)$, направленная из вершины v_n в v_j , считается *исходящей* из вершины v_n и ее номер k заносится в множество $S(v_n)$, а дуга u_{k+E} , направленная из v_j в v_n , — *входящей* для v_n и ее номер $k + E$ помещается в список $T(v_n)$. Дуга u_k является *входящей* для v_j , и ее номер k попадает в $T(v_j)$, а дуга u_{k+E} — *исходящей*, и номер $k + E$ вносится в список исходящих дуг $S(v_j)$.

Во всех узлах сети $v_n \in V, n = \overline{1, N}$, для каждого вида потока должны выполняться условия сохранения потоков:

$$\sum_{i \in S(v_n)} x_{mi} - \sum_{i \in T(v_n)} x_{mi} = \begin{cases} z_m, & \text{если } v_n = v_{s_m}; \\ -z_m, & \text{если } v_n = v_{t_m}; \\ 0 & \text{в остальных случаях,} \end{cases} \quad (1)$$

$$n = \overline{1, N}, m = \overline{1, M}, x_{mi} \geq 0, z_m \geq 0.$$

Величина z_m равна входному межузловому потоку m -го вида, проходящему от источника s_m к приемнику t_m пары p_m при распределении потоков $\{x_{mi}\} = \mathbf{x}$ по дугам сети.

Каждому ребру $r_k \in R$ приписывается неотрицательное число d_k — суммарный предельно допустимый поток, который можно передать по ребру r_k в обоих направлениях. В исходной сети компоненты вектора пропускных способностей $\mathbf{d} = (d_1, d_2, \dots, d_k, \dots, d_E)$ — положительные числа d_k . Вектор \mathbf{d} ограничивает сумму потоков всех видов, передаваемых по ребру r_k одновременно:

$$\sum_{m=1}^M (x_{mk} + x_{m(k+E)}) \leq d_k, \quad x_{mk} \geq 0, \\ x_{m(k+E)} \geq 0, \quad k = \overline{1, E}. \quad (2)$$

Условия (1), (2) задают множество допустимых значений вектора межузловых потоков $\mathbf{z} = (z_1, z_2, \dots, z_m, \dots, z_M)$:

$$\mathcal{Z}(\mathbf{d}) = \{\mathbf{z} \geq 0 \mid \exists \mathbf{x} \geq 0 : (\mathbf{z}, \mathbf{x}) \text{ удовлетворяют (1), (2)}\}.$$

3 Распределение дуговых потоков при монопольных режимах

На начальном этапе вычислительного эксперимента задается исходная сеть $G(0)$, пропускные способности $d_k(0)$, $k = \overline{1, E}$, и предполагается, что между парой узлов $p_m \in P$ передается поток m -го вида. В исходной сети $G(0)$ независимо и последовательно определяется максимальный поток z_m , $m = \overline{1, M}$, передаваемый в монопольном режиме из s_m в t_m .

Задача 1. Для некоторого $p_m \in P$ найти

$$z_m^0(1) = \max_{\mathcal{Z}(\mathbf{d}), \mathbf{x}} z_m$$

при условиях (1), (2), $z_i = 0$ для $i \neq m$, $p_i \in P$.

Решением задачи 1 являются: величина максимального потока $z_m^0(1)$, $m = \overline{1, M}$, соответствующие значения дуговых потоков $x_{mk}^0(1) > 0$, $m = \overline{1, M}$, $k = \overline{1, 2E}$. Формируется множество номеров ребер, для которых дуговые потоки положительны:

$$L(m) = \{k \mid x_{mk}^0(1) + x_{m(k+E)}^0(1) > 0, k = \overline{1, E}\}.$$

После решения последовательности задач 1 для всех $p_m \in P$ устанавливаются нормирующие коэффициенты

$$\omega_m^0(1) = \frac{1}{z_m^0(1)}, \quad z_m^0(1) \neq 0, \quad m = \overline{1, M},$$

и нормированные дуговые потоки

$$x_{mk}^0 = \omega_m^0(1)x_{mk}^0(1), \quad m = \overline{1, M}, \quad k = \overline{1, 2E}. \quad (3)$$

Согласно условиям *нормировки* для дуговых потоков (3) при передаче всех x_{mk}^0 по ребрам сети межузловой поток из узла s_m в узел t_m равен единице для всех $p_m \in P$.

Полученные при решении задачи 1 предельно допустимые значения межузловых потоков $z_m^0(1)$

отвечают точкам пересечения m -й координатной оси с границей множества допустимых потоков $\mathcal{Z}(\mathbf{d})$. Таким образом каждый вектор $\vec{z}(m) = \langle 0, 0, \dots, z_m^0(1), \dots, 0, 0 \rangle$, $m = \overline{1, M}$, определяет угловую точку множества допустимых значений межузловых потоков $\mathcal{Z}(\mathbf{d})$.

Рассмотрим вектор $\vec{\beta}(1)$ с равными значениями межузловых потоков

$$\vec{\beta}(1) = \langle \beta(1), \beta(1), \dots, \beta(1), \beta(1) \rangle,$$

который построен с помощью векторов набора $\vec{z}(m)$, $m = \overline{1, M}$. Вектор $\vec{\beta}(1)$ принадлежит множеству $\mathcal{Z}(\mathbf{d})$ как выпуклая комбинация векторов, соответствующих угловым точкам $\vec{z}(m)$. По определению,

$$\vec{\beta}(1) = \sum_{m=1}^M \gamma_m \vec{z}(m), \quad \sum_{m=1}^M \gamma_m = 1.$$

По построению, m -я компонента вектора $\vec{\beta}(1)$ равна

$$\beta_m(1) = \beta(1) = \gamma_m z_m^0(1),$$

откуда

$$\gamma_m = \frac{\beta(1)}{z_m^0(1)}, \quad m = \overline{1, M}.$$

Поскольку

$$\sum_{m=1}^M \gamma_m = 1,$$

то

$$\beta(1) \left[\sum_{m=1}^M \frac{1}{z_m^0(1)} \right] = 1$$

и

$$\beta(1) = \left[\sum_{m=1}^M \frac{1}{z_m^0(1)} \right]^{-1}. \quad (4)$$

Вектор $\vec{\beta}(1)$ с компонентами, которые определены согласно (4), представлен через угловые точки, отвечающие максимальным потокам $z_m^0(1)$ — решениям задачи 1. Указанный вектор служит гарантированным вектором равных межузловых потоков, передаваемых в сети одновременно. $\vec{\beta}(1)$ — вектор-нормаль секущей гиперплоскости с заданной суммой межузловых потоков $M\beta$. При увеличении пропускной способности ребер сети координаты угловых точек множества $\mathcal{Z}(\mathbf{d})$ изменяются, однако вектор-луч равных межузловых потоков сохраняет направление. Направляющий вектор $\vec{\beta}(1)$ позволяет проводить оценки относительных изменений характеристик сети при деформации множества достижимости $\mathcal{Z}(\mathbf{d})$.

4 Вычислительный эксперимент

Вычислительный эксперимент проводился на моделях сетевых систем, представленных на рис. 1. В каждой сети имеются 69 узлов. Пропускные способности ребер $d_k(0)$ равномерно распределены на отрезке $[900, 999]$. В ходе исследования была сделана нормировка, и потому суммарная пропускная способность в обеих сетях одинакова:

$$\sum_{k=1}^E d_k(0) = D(0) = 68\,256.$$

На начальном этапе с помощью решения задачи 1 для некоторой произвольной фиксированной пары узлов $p_j \in P$ в сети $G(0)$ был определен максимальный однопродуктовый поток $z_j^0(1)$ из узла s_j в узел t_j с соответствующими значениями потоков $\{x_{jk}^0(1), x_{j(k+E)}^0(1)\}$, $k = \overline{1, E}$, на ребрах сети. Для каждого j -го вида потока рассмотрим набор ребер (маршруты) R_j , по которым межузловой поток $z_j^0(1)$ передается из узла s_j в узел t_j :

$$R_j = \left\{ r_k \mid k \in L(j), x_{jk}^0(1) + x_{j(k+E)}^0(1) > 0 \right\}.$$

Для оценки последствий роста пропускной способности ребер набора R_j на основе исходной сети $G(0)$ рассматривается проект модернизированной сети $G(R_j)$, в которой

$$d_k(R_j) = \begin{cases} 2d_k(0), & \text{если } k \in L(j); \\ d_k(0) & \text{в остальных случаях.} \end{cases} \quad (5)$$

Для проекта $G(R_j)$ при заданных значениях пропускной способности (5) ребер из набора R_j для каждой пары узлов-корреспондентов $p_m \in P$ последовательно для каждого $m = \overline{1, M}$ решается задача 1. В результате вычисляются новые значения предельно допустимых межузловых потоков $z_m^0(R_j)$ и по формуле (4) рассчитывается величина $\beta(R_j)$.

Обозначим через $\lambda(j)$ относительное изменение исходного вектора равных значений

$$\lambda(j) = \frac{\beta(R_j)}{\beta(1)} - 1, \quad j = \overline{1, M}.$$

Значения $\lambda(j)$ переупорядочиваются по величине по невозрастанию и перенумеровываются согласно порядку

$$\{\lambda^*(i)\}: \lambda^*(i) \geq \lambda^*(i+1), \quad i = \overline{1, M-1}.$$

На рис. 2 представлены диаграммы $\lambda^*(\cdot)$ для базовой и кольцевой сетей. По вертикальной оси откладываются значения $\lambda^*(\cdot)$, а по горизонтальной — относительные номера (i/M) , $i = \overline{1, M}$. Из диаграмм для базовой сети на рис. 2, а следует, что более 20% проектов $G(R_j)$ дают возможность увеличить межузловые потоки на 15%–20%. Более половины проектов модернизации приводят к росту показателя «равных» потоков менее чем на 6%. Однако 20% проектов позволяют получить не среднюю, а гарантированную оценку увеличения на 15% всех межузловых потоков при одновременной передаче. Для кольцевой сети на рис. 2, б только для 10% проектов $G(R_j)$ достигается рост показателя $\lambda^*(\cdot)$ более чем на 10%. Дело в том, что в кольцевой сети больше ребер и больше исходные предельные показатели. Однако до 40% проектов также достигает увеличения на те 5%, что и в базовой сети.

Для каждого набора R_j вводится индикаторная функция

$$\mu_m^j = \begin{cases} 1, & \text{если } \frac{z_m^0(R_j)}{z_m^0(1)} \geq 1,3; \\ 0 & \text{в остальных случаях.} \end{cases}$$

Таким образом, если для проекта $G(R_j)$ максимальный поток m -го вида превышает значения $z_m^0(1)$ в исходной сети $G(0)$ более чем на 30%, то $\mu_m^j = 1$.

Для каждого набора R_j , $j = \overline{1, M}$, определяется доля пар-корреспондентов, для которых максимальные потоки возрастают более чем на 30%:

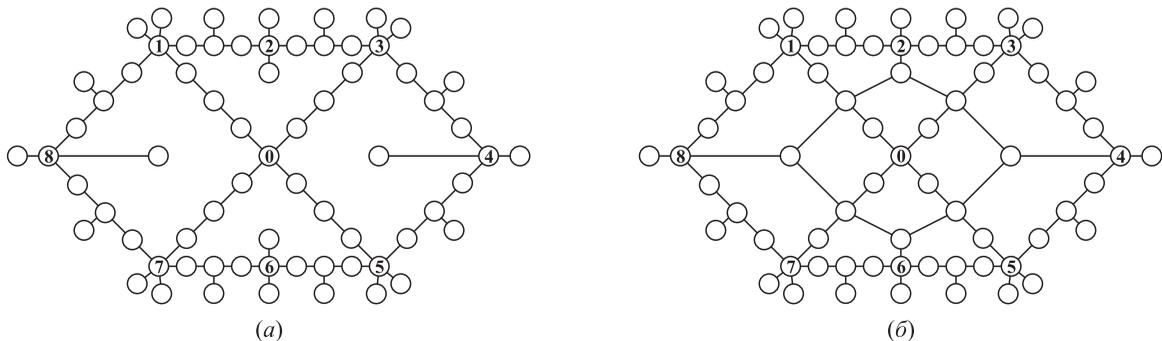


Рис. 1 Базовая (а) и кольцевая (б) сети

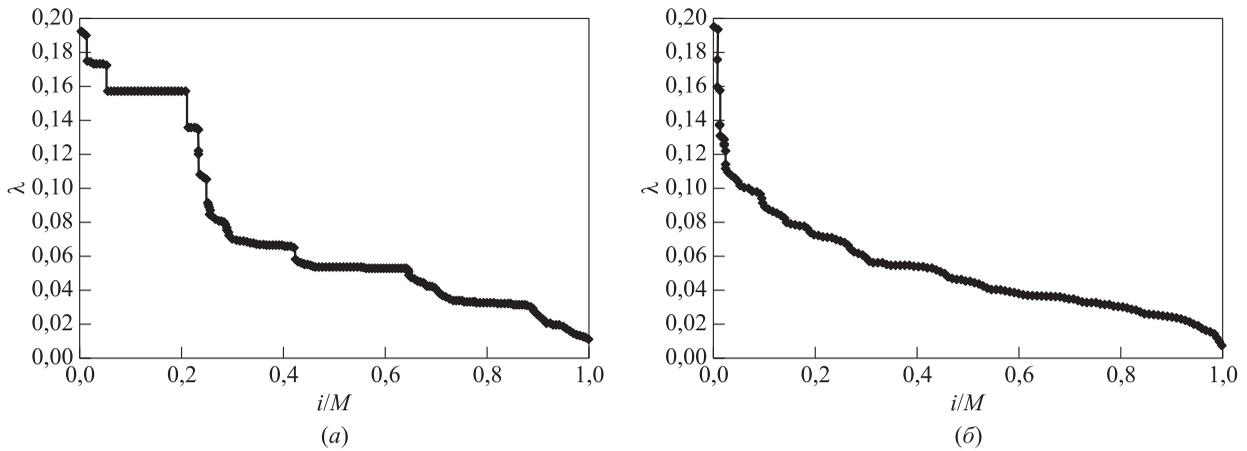


Рис. 2 Равные межузловые потоки в базовой (а) и кольцевой (б) сетях

$$\mu(j) = \frac{\sum_{m=1}^M \mu_m^j}{M}.$$

Значения $\mu(j)$ переупорядочиваются по величине по невозрастанию и перенумеровываются согласно порядку

$$\{\mu^*(i)\}: \mu^*(i) \geq \mu^*(i + 1), i = \overline{1, M - 1}.$$

Диаграммы $\mu^*(\cdot)$ для базовой и кольцевой сетей представлены на рис. 3. По вертикальной оси откладываются значения $\mu^*(\cdot)$, а по горизонтальной — относительные номера (i/M) , $i = \overline{1, M}$. Из диаграммы для базовой сети на рис. 3, а следует, что существуют около 900 наборов R_j (почти 20%) при увеличении пропускной способности которых около 40% узлов-корреспондентов смогут увеличить межузловые потоки более чем на 30%. «Длинные ступени» равных значений $\mu^*(\cdot)$ на рис. 3, а соответствуют наборам R_j , увеличение пропускной

способности которых одинаково влияет на предельные показатели передаваемых потоков. Диаграмма для кольцевой сети на рис. 3, б состоит из большого числа «коротких» ступенек. Следовательно, увеличение пропускной способности ребер наборов R_j в кольцевой сети затрагивает меньшее число транзитных маршрутов.

В базовой сети (см. рис. 3, а) значение $\mu^* = 0,422$ соответствует 750 проектам $G(R_j)$. При проведении вычислительного эксперимента для сортировки вводится большое пороговое значение при подсчете $\mu^*(\cdot)$, которое достигается для различных проектов. В кольцевой сети более 20% проектов $G(R_j)$, модернизирующих сеть $G(0)$, приводят к значительному росту максимального потока (более чем на 30%) для более чем 40% пар-корреспондентов.

Формально пара $\{\lambda(j), \mu(j)\}$ описывает изменение характеристик множества достижимости при увеличении пропускной способности ребер наборов

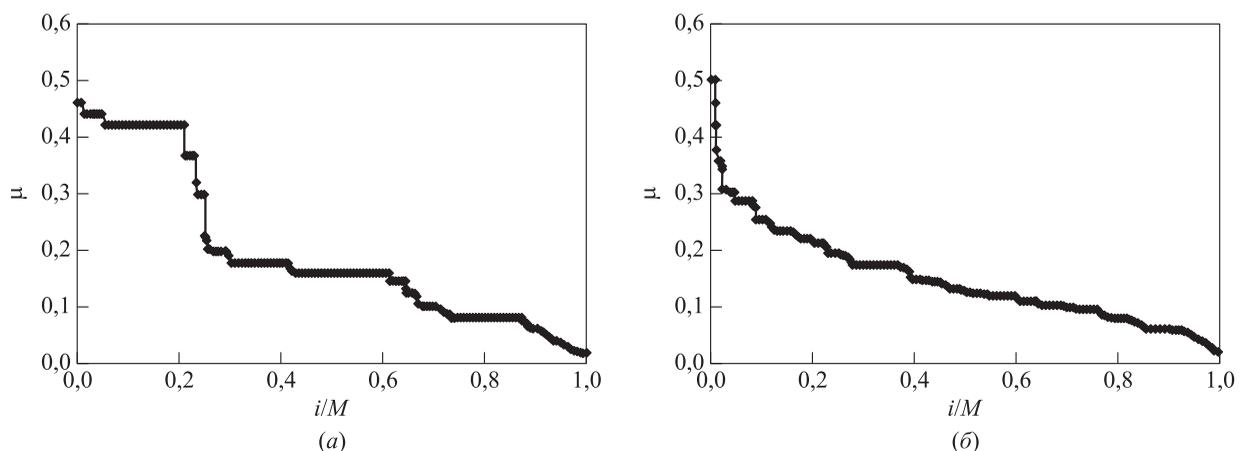


Рис. 3 Максимальные однопродуктовые потоки в базовой (а) и кольцевой (б) сетях

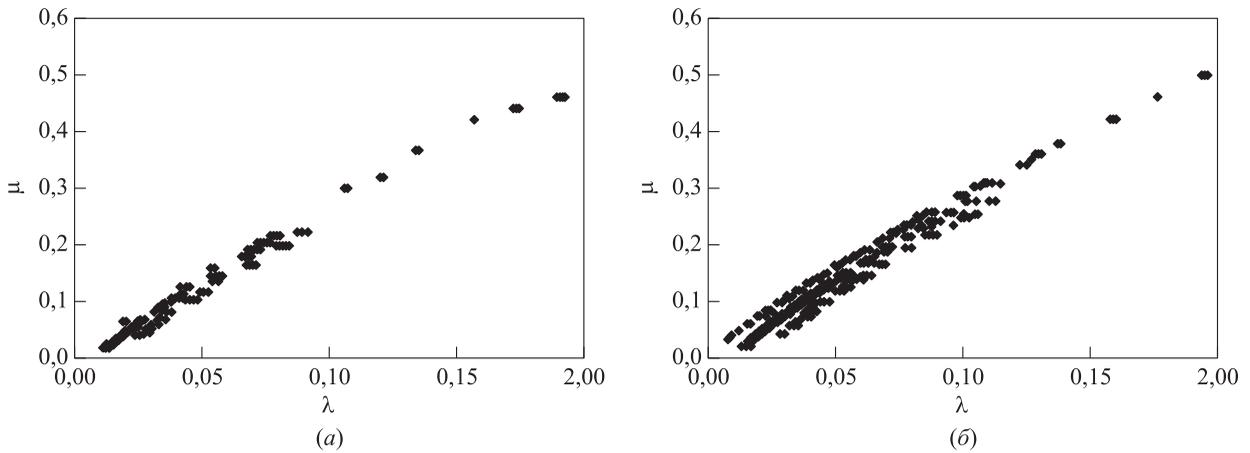


Рис. 4 Оценки проектов $G(R_j)$ в базовой (а) и кольцевой (б) сетях

ра R_j для каждого заданного j . На диаграммах рис. 4 по вертикальной оси откладываются значения $\mu(j)$, а по горизонтальной — $\lambda(j)$ для каждого R_j , $j = \overline{1, M}$. Каждая отдельная точка-ромб на рис. 4 соответствует проекту $G(R_j)$.

Менее 10 проектов для базовой сети имеют значения $\lambda(j) \geq 0,15$, $\mu(j) \geq 0,4$, из них для четырех проектов $\lambda(j) \geq 0,18$, $\mu(j) \geq 0,43$. Для более 90% проектов $G(R_j)$ показатели $\lambda(j) \leq 0,1$, $\mu(j) \leq 0,3$. Для кольцевой сети — схожие результаты: для двух проектов удается получить $\mu(j) \geq 0,5$, а значение $\lambda(j)$ — почти 0,2. При этом для 90% проектов показатели оказываются ниже: $\lambda(j) \leq 0,1$, $\mu(j) \leq 0,3$.

5 Отклик системы на увеличение пропускной способности

Для получения гарантированных оценок характеристик и затрат ресурсов рассчитывается относительное изменение величины максимального потока $z_m^0(R_j)$ при изменении пропускной способности ребер из $R_j \in R$:

$$\xi_m^j = \frac{z_m^0(R_j)}{z_m^0(1)} - 1, \quad m = \overline{1, M}.$$

Для проекта $G(R_j)$ вычисляется относительное изменение пропускной способности

$$\rho_j = \frac{\sum_{k \in L(j)} [d_k(R_j) - d_k(0)]}{D(0)},$$

где $L(j)$ — множество номеров ребер, входящих в набор R_j .

Для определения количественных оценок работоспособности сети $G(R_j)$ подсчитывается относи-

тельное изменение предельных значений функциональных показателей системы:

$$\theta_m^j = \frac{\xi_m^j}{\rho_j}, \quad m = \overline{1, M}.$$

Фактически если значения ξ_m^j и ρ_j записать в процентах, то величина θ_m^j покажет, на сколько процентов увеличится максимальный поток для пары $p_m \in P$ при росте пропускной способности сети на 1% за счет ребер из набора R_j .

Значения θ_m^j упорядочиваются по величине от большего к меньшему (по невозрастанию) и перенумеровываются согласно порядку

$$\{\theta_i^j\} : \theta_i^j \geq \theta_{i+1}^j, \quad i = \overline{1, M-1}.$$

Полученные лексикографически упорядоченные величины $\{\theta_i^j\}$ служат i -ми компонентами вектор-отклика $\theta(j)$ при изменении пропускной способности ребер из набора R_j .

Для получения гарантированных оценок вектор-откликов $\theta(\cdot)$ для каждого проекта $G(R_j)$, $j = \overline{1, M}$, формируется вектор-отклик $\theta(j)$. Для каждой фиксированной i -й компоненты вектор-отклика $\theta(j)$ во всех сетях $G(R_j)$, $j = \overline{1, M}$, определяются

$$\Theta_i^* = \max_j \theta_i^j; \quad \Theta_i^{**} = \min_j \theta_i^j, \quad i = \overline{1, M}.$$

По построению, $\Theta_i^* \geq \theta_i^j$ для любого $j = \overline{1, M}$.

Набор $\{\Theta_i^*\}$, $i = \overline{1, M}$, представляет собой вектор гарантированных оценок компонент вектор-откликов при увеличении пропускной способности ребер из каждого отдельного набора R_j .

На рис. 5 представлены Θ^* -диаграммы для базовой и кольцевой сетей. По вертикальной оси откладываются значения $\{\Theta_i^*\}$, а по горизонтальной — относительные номера наборов $R_j - (i/M)$. Резкий скачок на диаграмме как для кольцевой, так и для базовой сетей относится к смежным парам-

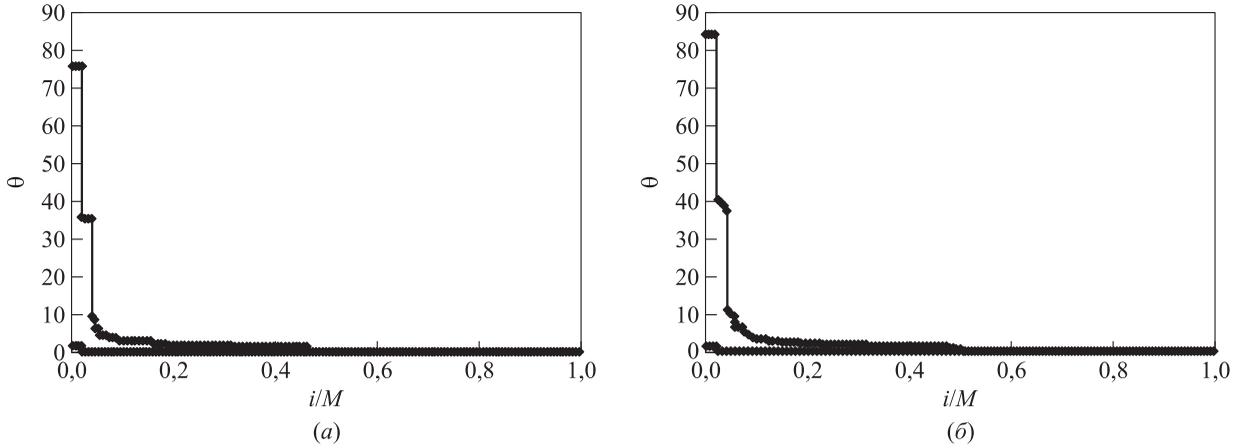


Рис. 5 Относительные оценки максимальных потоков в базовой (а) и кольцевой (б) сетях

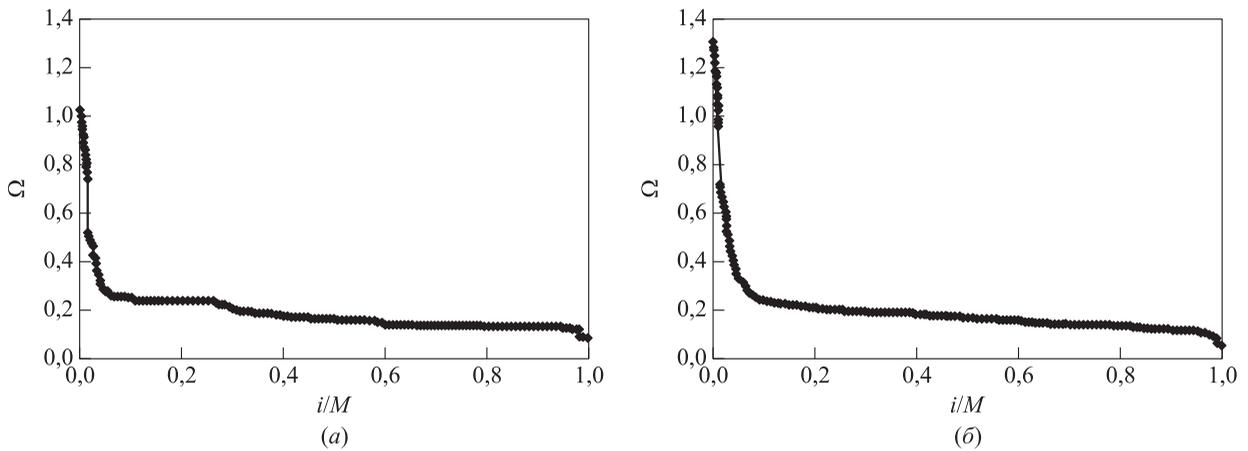


Рис. 6 Относительные оценки равных межузловых потоков в базовой (а) и кольцевой (б) сетях

корреспондентам на ребрах к всеячим вершинам. По построению, исходные значения Θ_i^* показывают относительное изменение максимального потока при увеличении пропускной способности всех ребер набора R_j . Увеличение пропускной способности только одного ребра к всеячей вершине дает возможность нарастить вдвое максимальный поток между указанными смежными вершинами.

Для оценки относительных затрат ресурсов при увеличении значений $\beta(R_j)$ вследствие изменения пропускной способности ребер наборов R_j вычисляется

$$\Omega_j = \frac{\lambda(j)}{\rho_j}, \quad j = \overline{1, M}.$$

Значения Ω_j переупорядочиваются по величине от большего к меньшему (по невозрастанию) и перенумеровываются согласно порядку

$$\{\Omega_i^*\} : \Omega_i^* \geq \Omega_{i+1}^*, \quad i = \overline{1, M-1}.$$

На рис. 6 представлены Ω^* -диаграммы для базовой и кольцевой сетей. По вертикальной оси откла-

дываются значения $\{\Omega_i^*\}$, а по горизонтальной — относительные номера наборов $R_j - (i/M)$.

Величина Ω_j характеризует относительное изменение вектора равных межузловых потоков при увеличении пропускной способности ребер набора R_j . Диаграммы на рис. 6 как для базовой, так и для кольцевой сетей показывают, что существуют не более 10–15 проектов $G(R_j)$ — менее 1% от их общего числа, реализация которых приводит к росту $\beta(R_j)$ более чем на 1% при увеличении пропускной способности ребер набора R_j на 1%.

В кольцевой сети существует уникальный проект R_j^* , для всех пар-корреспондентов которого удастся увеличить $\beta(R_j^*)$ на 1,3%. Диаграммы на рис. 6 показывают, что при увеличении пропускной способности для 80% проектов $G(R_j)$ значение Ω_j не превышает 0,2%. Однако удастся находить уникальные проекты, для которых эффективность использования вложенных ресурсов в 5–6 раз выше средних показателей.

6 Заключение

Перекрестный анализ структурных и потоковых характеристик, их взаимосвязи и взаимовлияния может быть использован на предпроектном этапе создания сетей связи различного назначения. Специализированные модели [10–13] позволяют получать различные оценки показателей работоспособности многопользовательских сетей связи. В ряде исследований [3, 4] оценка функционирования и устойчивости современных сетевых систем основывается на использовании многопараметрических векторов, изучается их зависимость от схем управления перераспределением ресурсов при диспетчеризации.

Предложенный в настоящей работе подход к вычислению предельных значений при увеличении пропускной способности сети может быть применен как для априорного анализа проектов модернизации $G(R_j)$ и развития сетей, так и для оценки параметров реальных сетевых структур. Данный метод дает возможность выявлять доминирующие группы или наборы ребер R_j , входящие в состав маршрутов при передаче максимального потока в монопольном режиме, а также строить гарантированные оценки изменения межузловых потоков при увеличении пропускной способности ребер вдоль указанных маршрутов.

Результаты экспериментов подтверждают эффективность подхода: установлено, что увеличение пропускной способности определенных ребер может привести к существенному росту части максимальных потоков, особенно в рамках конкретных проектов модернизации. Описанная схема эксперимента имеет полиномиальную оценку сложности [14] и может применяться для расчета сетей с десятками тысяч узлов за счет использования современных программных реализаций с минимальными вычислительными затратами.

Литература

1. Малашенко Ю. Е., Назарова И. А. Гарантированные оценки показателей работоспособности многопользовательской сети при повреждении // Информатика и её применения, 2025. Т. 19. Вып. 1. С. 16–24. doi: 10.14357/19922264250103. EDN: MYNICH.
2. Пехтерев С. В., Макаренко С. И., Ковальский А. А. Описательная модель системы спутниковой связи Starlink // Системы управления, связи и безопасности, 2022. № 4. С. 190–255. doi: 10.24412/2410-9916-2022-4-190-255. EDN: QMOLDV.
3. Малашенко Ю. Е., Назарова И. А. Сравнительный анализ узловых мультипотоков в многопользовательской сетевой системе // Информатика и её применения, 2024. Т. 18. Вып. 1. С. 40–45. doi: 10.14357/19922264240106. EDN: AKCMCQ.
4. Малашенко Ю. Е., Назарова И. А. Оценки распределения ресурсов в многопользовательской сети при равных межузловых нагрузках // Информатика и её применения, 2023. Т. 17. Вып. 1. С. 83–88. doi: 10.14357/19922264230111. EDN: BUKVGV.
5. Данскин Дж. М. Теория максимина и ее приложение к задачам распределения вооружения / Пер. с англ. — М.: Сов. радио, 1970. 200 с. (*Danskin J. M. The theory of Max-Min and its application to weapons allocation problems.* — Berlin: Springer-Verlag, 1970. 128 p. doi: 10.1007/978-3-642-46092-0.)
6. Гермейер Ю. Б. Введение в теорию исследования операций. — М.: Наука, 1971. 383 с.
7. Йенсен П., Барнес Д. Потокное программирование / Пер. с англ. — М.: Радио и связь, 1984. 392 с. (*Jensen P. A., Barnes J. W. Network flow programming.* — New York, NY, USA: Wiley, 1980. 408 p.)
8. Лотов А. В., Поспелова И. И. Многокритериальные задачи принятия решений. — М.: Макс Пресс, 2008. 197 с.
9. Ogryczak W., Luss H., Pioro M., Nace D., Tomaszewski A. Fair optimization and networks: A survey // J. Appl. Math., 2014. Vol. 2014. Iss. 1. Art. 612018. 25 p. doi: 10.1155/2014/612018.
10. Соколов Н. А. Задачи планирования сетей электросвязи. — СПб.: Техника связи, 2012. 432 с. EDN: RPYWOL.
11. Бесчастный В. А., Острикова Д. Ю., Шоргин С. Я., Молчанов Д. А., Гайдамака Ю. В. Анализ плотности базовых станций 5G NR для предоставления услуг виртуальной и дополненной реальности // Информатика и её применения, 2022. Т. 16. Вып. 2. С. 102–108. doi: 10.14357/19922264220213. EDN: VPIRYN.
12. Мачнев Е. А., Бесчастный В. А., Острикова Д. Ю., Шоргин С. Я., Гайдамака Ю. В. Об оптимальном расположении антенн для V2X-соединений в субтерагерцевом диапазоне // Информатика и её применения, 2022. Т. 16. Вып. 4. С. 42–50. doi: 10.14357/19922264220407. EDN: WFMJTI.
13. Самуйлов А. К., Платонова А. А., Шоргин В. С., Гайдамака Ю. В. К моделированию эффектов обслуживания многоадресного трафика в сетях 5G NR // Информатика и её применения, 2023. Т. 17. Вып. 2. С. 71–77. doi: 10.14357/19922264230210. EDN: SLMGZU.
14. Кормен Т. Х., Лейзерсон Ч. И., Ривест Р. Л., Штайн К. Алгоритмы: построение и анализ / Пер. с англ. — 3-е изд. — М.: Вильямс, 2013. 1328 с. (*Cormen T. H., Leiserson C. I., Rivest R. L., Stein C. Introduction to algorithms.* — 3rd ed. — Cambridge, MA, USA: The MIT Press, 2009. 1313 p.)

Поступила в редакцию 17.06.2025

Принята к публикации 15.08.2025

MULTICOMMODITY FLOW MODEL FOR ANALYSIS PROJECTS OF COMMUNICATION NETWORK CAPACITY SCALING

Yu. E. Malashenko, I. A. Nazarova, and M. V. Kozlov

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

Abstract: The performance capabilities of a multiuser communication system under network capacity scaling is investigated. Within the framework of computational experiments, the impact of increasing edge capacities along flow paths is analyzed in terms of transmitting maximum flows. For each pair of communicating nodes, the value of maximum allowable internodal flow is determined independently under monopoly control modes. The obtained maximum flow values are used to construct and compare vectors of uniform internodal flows of all types that can be simultaneously transmitted through the network. The concept of a vector-response of the system to increasing edge capacities along transmission routes is introduced. For each reconstruction project option and for every pair of communicating nodes, the ratio of capacity increase to the growth of the maximum flow is calculated. The obtained values are reordered according to the max-min rule. Based on the vectors-response, a set of guaranteed estimates of the maximum possible changes in operational parameters is formed. The results of the computational experiments for networks with different structural characteristics are analyzed.

Keywords: multicommodity model of the communication network; guaranteed estimate in case of network capacity scaling; maximum flow transmission routes

DOI: 10.14357/19922264250305

EDN: EBFVDL

References

1. Malashenko, Yu. E., and I. A. Nazarova. 2025. Garantirovannye otsenki pokazateley rabotosposobnosti mnogopol'zovatel'skoy seti pri povrezhdeniyakh [Guaranteed estimates of the multiuser network performance indicators in case of damage]. *Informatika i ee Primeneniya — Inform. Appl.* 19(1):16–24. doi: 10.14357/19922264250103. EDN: MYHICH.
2. Pehterev, S. V., S. I. Makarenko, and A. A. Kovalsky. 2022. Opisatel'naya model' sistemy sputnikovoy svyazi Starlink [Descriptive model of Starlink satellite communication system]. *Sistemy upravleniya, svyazi i bezopasnosti* [Systems of Control, Communication and Security] 4:190–255. doi: 10.24412/2410-9916-2022-4-190-255. EDN: QMOLDV.
3. Malashenko, Yu. E., and I. A. Nazarova. 2024. Sravnitel'nyy analiz uzlovykh mul'tipotokov v mnogopol'zovatel'skoy setevoy sisteme [Analysis of node multiflows in a multiuser network system]. *Informatika i ee Primeneniya — Inform. Appl.* 18(1):40–45. doi: 10.14357/19922264240106. EDN: AKCMCQ.
4. Malashenko, Yu. E., and I. A. Nazarova. 2023. Otsenki raspredeleniya resursov v mnogopol'zovatel'skoy seti pri ravnykh mezhuzlovykh nagruzkakh [Estimates of the resource distribution in the multiuser network with equal internodal loads]. *Informatika i ee Primeneniya — Inform. Appl.* 17(1):83–88. doi: 10.14357/19922264230111. EDN: BUKVGV.
5. Danskin, J. M. 1970. *The theory of Max-Min and its application to weapons allocation problems*. Berlin: Springer-Verlag. 128 p. doi: 10.1007/978-3-642-46092-0.
6. Germeyer, Yu. B. 1971. *Vvedenie v teoriyu issledovaniya operatsiy* [Introduction to operations research theory]. Moscow: Nauka. 383 p.
7. Jensen, P. A., and J. W. Barnes. 1980. *Network flow programming*. New York, NY: Wiley. 408 p.
8. Lotov, A. V., and I. I. Pospelova. 2008. *Mnogokriterial'nye zadachi prinyatiya resheniy* [Multicriteria decision making tasks]. Moscow: Maks Press. 197 p.
9. Ogryczak, W., H. Luss, M. Pioro, D. Nace, and A. Tomaszewski. 2014. Fair optimization and networks: A survey. *J. Appl. Math.* 2014(1):612018. 25 p. doi: 10.1155/2014/612018.
10. Sokolov, N. A. 2012. *Zadachi planirovaniya setey elektrosvyazi* [Tasks of telecommunications network planning]. St. Petersburg: Tekhnika svyazi. 432 p. EDN: RPYWOL.
11. Beschastnyi, V. A., D. Yu. Ostrikova, S. Ya. Shorgin, D. A. Moltchanov, and Yu. V. Gaidamaka. 2022. Analiz plotnosti bazovykh stantsiy 5G NR dlya predostavleniya uslug virtual'noy i dopolnennoy real'nosti [Density analysis of mmWave NR deployments for delivering scalable AR/VR video services]. *Informatika i ee Primeneniya — Inform. Appl.* 16(2):102–108. doi: 10.14357/19922264220213. EDN: VPIRYN.
12. Machnev, E. A., V. A. Beschastnyi, D. Yu. Ostrikova, Yu. V. Gaidamaka, and S. Ya. Shorgin. 2022. Ob optimal'nom raspolozhenii antenn dlya V2X-soedineniy v subteragertsevom diapazone [On the optimal antenna deployment for subterahertz V2X communications]. *Informatika i ee Primeneniya — Inform. Appl.*

- 16(4):42–50. doi: 10.14357/19922264220407. EDN: WFMJTI.
13. Samoylov, A. K., A. A. Platonova, V. S. Shorgin, and Yu. V. Gaidamaka. 2023. K modelirovaniyu effektivov obsluzhivaniya mnogoadresnogo trafika v setyakh 5G NR [On modeling the effects of multicast traffic servicing in 5G NR networks]. *Informatika i ee Primeneniya — Inform. Appl.* 17(2):71–77. doi: 10.14357/19922264230210. EDN: SLMGZU.
14. Cormen, T. H., C. I. Leiserson, R. L. Rivest, and C. Stein. 2009. *Introduction to algorithms*. 3rd ed. Cambridge, MA: The MIT Press. 1313 p.

Received June 17, 2025

Accepted August 15, 2025

Contributors

Malashenko Yuri E. (b. 1946) — Doctor of Science in physics and mathematics, senior scientist, Federal Research Center “Computer Science and Control of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; malash09@ccas.ru

Nazarova Irina A. (b. 1966) — Candidate of Science (PhD) in physics and mathematics, scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; irina-nazar@yandex.ru

Kozlov Mihail V. (b. 1958) — senior mathematician, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; kztmp2@ccas.ru

РАЗРАБОТКА МОДЕЛИ ДЛЯ РАСЧЕТА ПРОИЗВОДИТЕЛЬНОСТИ ПЛАНИРОВАНИЯ ПЕРЕДАЧ В СЕТЯХ 5G NR С УЧЕТОМ СИГНАЛЬНОГО ТРАФИКА*

Э. С. Сопин¹, А. И. Назарьин², С. Я. Шоргин³

Аннотация: Цель исследования — разработка математической модели для анализа компромисса между ресурсами физического канала управления нисходящего канала связи (PDCCH — Physical Downlink Control Channel) и канала передачи данных (PDSCH — Physical Data Shared Channel) в 5G NR (New Radio, «Новое радио») при увеличении числа абонентов. Разработаны модели стратегий назначения первичных блоков ресурсов (PRB — Primary Resource Block). Затем построена ресурсная модель массового обслуживания, учитывающая потребности абонентов в обоих каналах. Численные результаты показывают, что при последовательном назначении PRB узким местом становится размер PDCCH, а при последовательном — недостаток ресурсов PDSCH приводит к упущенным возможностям планирования. Непоследовательное назначение ресурсов потенциально позволяет увеличить коэффициент использования PDSCH, но требует тщательной настройки соотношения ресурсов PDCCH и PDSCH. Предложенная модель позволяет определять объем ресурсов, выделяемых для PDCCH/PDSCH, в зависимости от потребностей пользователей, минимизируя упущенные возможности планирования и максимизируя использование ресурсов. Проведенное численное исследование показало, что последовательное назначение ресурсов в PDSCH приводит к значительному ухудшению производительности системы. При этом эффективное использование метода последовательного назначения ресурсов требует динамического изменения объема ресурсов, назначаемых каналу PDCCH.

Ключевые слова: 5G; миллиметровый диапазон частот; планирование ресурсов; ресурсная система массового обслуживания

DOI: 10.14357/19922264250306

EDN: EMAQVK

1 Введение

Системы 5G «Новое радио» были разработаны для приложений, характеризующихся высокими требованиями к скорости передачи. С этой целью в дополнение к диапазону частот ниже 6 ГГц они используют диапазон миллиметровых волн (mmWave, 30–100 ГГц), где каналы шириной 50 МГц могут быть объединены для обеспечения непрерывной полосы пропускания до 400 МГц [1].

Планирование передачи пользовательского трафика — один из наиболее сложных аспектов проектирования сотовых систем. Важным вопросом при разработке планировщиков становится объем ресурсов, доступных для физического канала управления нисходящего канала связи, который служит для информирования пользователей об использовании первичных ресурсных блоков для нисходящего канала передачи данных. Для увеличения спектральной эффективности использование направленных антенн в частотно-селективном много-

лучевом канале с замираниями требует детального планирования для каждого PRB в сетке ортогонального частотного мультиплексирования (OFDMA — Orthogonal Frequency-Division Multiple Access). Однако это приводит к высоким требованиям к пропускной способности PDCCH, так как требует передачи номеров отдельных ресурсных блоков.

Альтернативой может стать использование ресурсоэффективного последовательного выделения PRB, что снизит требования к ресурсам PDCCH, но может увеличить использование ресурсов для PDSCH из-за снижения спектральной эффективности. Таким образом, в зависимости от схемы кодирования PDCCH ресурсы PDCCH или PDSCH могут использоваться неэффективно. Насколько известно авторам, эти аспекты не были исследованы в литературе.

Отметим, что планирование ресурсов охватывает функции, связанные с выделением PRB пользователям в PDSCH и восходящем канале передачи данных (PUSCH — Physical Uplink Shared Channel),

* Исследование выполнено за счет гранта Российского научного фонда № 22-79-10128, <https://rscf.ru/project/22-79-10128/>.

¹ Российский университет дружбы народов имени Патриса Лумумбы; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sopin-es@rudn.ru

² Российский университет дружбы народов имени Патриса Лумумбы, nazaryin-ai@rudn.ru

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

а также уведомлением пользовательского устройства (ПУ) о выделенных PRB в PDCCH. В контексте планирования PRB большая часть литературы сосредоточена на вопросах справедливости распределения ресурсов на абонентском участке доступа с учетом новых возможностей сетей 5G NR, таких как, например, мультисвязность, агрегирование частот (Carrier Aggregation), различные режимы MIMO (Multiple-in Multiple-Out), нарезка ресурсов сети (Network Slicing) [2, 3]. В то же время авторы в [4] и [5] подчеркнули, что PDCCH может стать узким местом на практике. В данном исследовании эта проблема решается путем анализа производительности базовой станции (БС) 5G NR в условиях ограниченных ресурсов, доступных для PDCCH и PDSCH.

Целью данного исследования ставится разработка модели, учитывающей процесс обслуживания сессий как в PDSCH, так и в PDCCH. С этой целью в разд. 2 определены требования к ресурсам, предъявляемые к PDCCH и PDSCH. В разд. 3 предложены модель и метод расчета характеристик

обслуживания сессий. Численный пример и заключение завершают работу.

2 Методы назначения ресурсов

В данном разделе определено распределение требований к ресурсам для PDCCH и PDSCH, т.е. случайного вектора $\mathbf{r} = (r_1, r_2)$. Отметим, что r_1 определяет объем ресурсов, необходимых для обслуживания сессий в сигнальном PDCCH, r_2 — объем ресурсов, необходимых для обслуживания сессий в канале передачи данных (PDSCH). Для сигнального канала схема модуляции и кодирования неизменна и не зависит от текущего значения отношения сигнал/шум (ОСШ) на устройстве. Это необходимо для того, чтобы все устройства могли принимать PDCCH полностью. Поэтому объем ресурсов, требуемых в PDCCH, зависит только от размера передаваемых данных и для выбранной схемы планирования ресурсов остается константой. Однако метод последовательного назначения ресурсов

Таблица 1 Список обозначений

Переменная	Значение
M	Число схем модуляции и кодирования
s_m	Нижняя граница ОСШ на схеме модуляции и кодирования m
C	Требуемая скорость передачи данных
e_m	Спектральная эффективность на схеме модуляции и кодирования m
w_p	Размер первичного ресурсного блока
$\phi_{n,m}$	Вероятность того, что при выборе n наилучших по качеству радиоканала PRB значение ОСШ наилучшего из них находится в диапазоне (s_m, s_{m+1})
r_C	Радиус области покрытия БС
$r_{C,k}$	Внешний радиус кольца k
μ_k	Среднее значение потерь распространения на расстоянии $(r_{C,k} + r_{C,k+1})/2$ от БС
c_k	Вероятность того, что запрос на сессию поступает из области k
σ	Среднеквадратичное отклонение распределения теневого замирания
$f_{k,i}(x)$	Плотность распределения i -го наилучшего PRB по значению ОСШ
R_1	Емкость PDCCH
R_2	Емкость PDSCH
r_1	Случайная величина требуемых PRB в PDCCH
r_2	Случайная величина требуемых PRB в PDSCH
N	Число приборов в РеСМО
λ	Интенсивность поступления сессий
b	Среднее время обслуживания сессии
$q_n(\mathbf{r})$	Совместное стационарное распределение числа заявок в системе и суммарного занятого ими объема ресурса
π	Общая вероятность потери заявки
π_i	Вероятность потери заявки, вызванная недостаточностью типа ресурса i
\mathbf{R}	Средний объем занятого ресурса
\mathbf{R}_{util}	Коэффициент использования ресурсов
$L(y)$	Показатель затухания волны на расстоянии y
f_C	Несущая частота
P_U	Мощность передачи ПУ
G_A, G_U	Коэффициент усиления антенной решетки на БС/ПУ

характеризуется меньшим количеством информации, требуемой для определения выделяемых PRB, поскольку позволяет указывать только начальный номер PRB и количество выделяемых PRB. Напротив, метод непоследовательного выделения ресурсов требует передачи полной маски ресурсной сетки, где 1 указывает на PRB, выделяемых пользователю, и, следовательно, требует большего объема ресурсов для каждого пользователя в PDCCH [6]. В данной работе предполагается, что используются одна единица ресурса для последовательного назначения PRB и две единицы ресурса — для непоследовательного. На практике это соотношение может зависеть от дополнительных факторов, таких как, например, полоса частот, выделенная системе связи.

Объем ресурсов, необходимых для обслуживания сессии в PDSCH, зависит от текущего значения ОСШ на приемнике и индивидуален для каждого пользователя. Поэтому компонент r_2 представляет собой невырожденную дискретную случайную величину. Перейдем к выводу распределений для второй компоненты r_2 .

Для облегчения понимания введем табл. 1 с используемыми обозначениями.

2.1 Метод непоследовательного назначения PRB

Предполагается, что PDSCH в сетях 5G NR использует M схем модуляции и кодирования (MCS — Modulation and Coding Scheme). Обозначим за s_m ОСШ, а за e_m — спектральную эффективность, соответствующие m -й схеме MCS, $m = \overline{1, M}$. Для достижения требуемой скорости C при уровне спектральной эффективности e_m требуется $n_m = \lceil C / (e_m w_p) \rceil$ PRB, где w_p — размер PRB, соответствующий используемой нумерологии 5G NR; $\lceil \cdot \rceil$ — операция округления до наибольшего ближайшего целого числа. Пользовательскому устройству необходимо выделить n_m PRB, если наихудшее значение ОСШ, полученное на одном из этих PRB, находится в диапазоне (s_m, s_{m+1}) , в то время как остальные PRB имеют значение ОСШ не менее s_m . Таким образом, чтобы вычислить объем требований к ресурсу физического PDSCH при непоследовательном назначении PRB, необходимо определить вероятности $\phi_{n_m, m}$, $m = \overline{1, M}$, того, что при выборе n_m наилучших по качеству радиоканала PRB значение ОСШ наихудшего из них находится в диапазоне (s_m, s_{m+1}) .

Зона покрытия БС, представляющая собой окружность радиусом r_C , делится на K областей с помощью $K - 1$ концентрических окружностей.

Радиусы окружностей $r_{C,k}$ задаются с определенным шагом l , так что $r_{C,k} = r_C - (k - 1)l$, $k = \overline{1, K}$. Каждый PRB в области k испытывает ОСШ, распределенный в соответствии с нормальным распределением $\mathcal{N}(\mu_k, \sigma^2)$, где μ_k — среднее значение потерь распространения на расстоянии $(r_{C,k} + r_{C,k+1})/2$ от БС; σ — среднеквадратичное отклонение распределения теневого замирания, полученное из спецификации TR 38.901 [7]. Несмотря на то что ОСШ соседних PRB являются зависимыми случайными величинами, для рассматриваемых в данной работе систем 5G NR, использующих миллиметровые диапазоны частот с широкими каналами связи (> 50 МГц), выбранные PRB при непоследовательном назначении, как правило, независимы.

Сделаем следующее упрощающее предположение о поведении канала. Если объем ресурсов достаточно большой, непоследовательный выбор PRB приводит к тому, что выбранные PRB будут находиться в различных областях ресурсной сетки. При этом даже при больших значениях коэффициента корреляции ν число соседних PRB, качество которых окажется зависимым, не превысит нескольких единиц. Таким образом, можно предположить, что значения ОСШ выбранных PRB будут независимыми случайными величинами. Данное приближение становится лучше, когда объем ресурсов системы увеличивается, коэффициент ν не превышает 0,9 и скорость сессии не слишком велика относительно доступной полосы частот.

Используя результаты теории порядковых статистик [8], плотность распределения i -го наихудшего по значению ОСШ PRB определим по формуле:

$$f_{k,i}(x) = \frac{R_2!}{(i-1)!(R_2-i)!} [F_k(x)]^{i-1} [1-F_k(x)]^{R_2-i} f_k(x),$$

где R_2 — число PRB в PDSCH; $F_k(x)$ и $f_k(x)$ — функция и плотность распределения ОСШ в k -й области. Далее следует определить совместное распределение n_i -х наилучших PRB, однако в таком случае вычислительная сложность расчета числа требуемых PRB в канале PDSCH r_2 станет неприемлемой. Чтобы этого избежать, воспользуемся аппроксимацией, используя ОСШ наихудшего из выбранных PRB для расчета вероятностей $\phi_{i,m}$ следующим образом:

$$\phi_{i,m} = \sum_{k=1}^K c_k (F_{k,R_2-i+1}(s_{m+1}) - F_{k,R_2-i+1}(s_m)),$$

где c_k — вероятность того, что запрос на сессию поступает из области k . Из предположения о том, что ПУ расположены на плоскости в соответствии

с пуассоновским точечным процессом, следует пропорциональность вероятности c_k площади соответствующей k -й области, т. е.

$$c_k = \frac{r_{C,k+1}^2 - r_{C,k}^2}{r_C^2}.$$

Проведя указанные вычисления для всех $i \in \{n_1, n_2, \dots, n_M\}$, получим вероятностный ряд $\{\phi_{n,m,m}\}$, $m = \overline{1, M}$, случайной величины r_2 , соответствующий методу непоследовательного назначения PRB.

2.2 Метод последовательного назначения PRB

Метод последовательного назначения PRB предполагает, что каждое ПУ получает свои i PRB последовательно, причем не обязательно те PRB, которые характеризуются наилучшим качеством канала. Отметим, что в незагруженной системе планировщик может найти наилучшие непересекающиеся последовательности из n PRB для каждого ПУ. Однако в случае, когда система работает в условиях, близких к перегрузке, что представляет интерес для данного исследования, задача может не иметь решения даже при наличии свободных ресурсов. При учете временных ограничений на планирование используются различные эвристические алгоритмы (см., например, [9]). В данной работе предполагается, что для каждого ПУ планировщик выбирает PRB, характеризующийся «средними» условиями канала, распределенными в соответствии с $\mathcal{N}(\mu_k, \sigma^2)$, и начинает выделение ресурсов с этого PRB. Качество канала последующих PRB коррелирует с первым выбранным, причем коэффициент корреляции представляет геометрическую прогрессию ν^i .

При последовательном назначении PRB аналогично предыдущему разделу представляет интерес вероятность $\phi_{n,m}$ того, что при назначении n последовательных PRB ОСШ худшего из них будет в промежутке (s_m, s_{m+1}) . Плотность совместного распределения вероятности n нормально распределенных случайных величин, задающих условия канала, определяется выражением:

$$f_S(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\Delta}} e^{-\sum_{i=1}^n \sum_{j=1}^n k_{ij}^{(-1)} (x_i - m_i)(x_j - m_j)/2}, \quad (1)$$

где m_i — среднее значение случайной величины S_i , $i = \overline{1, n}$; Δ — определитель ковариационной матрицы \mathbf{K} системы (S_1, S_2, \dots, S_n) ; $k_{ij}^{(-1)}$ — элементы

обратной матрицы \mathbf{K}^{-1} . Заметим, что в рассматриваемом случае $m_i = m_j$ для любых $i, j = \overline{1, n}$, при этом $\mathbf{K}_{ii} = \sigma^2$. В силу корреляции качества канала в соседних PRB недиагональные элементы матрицы \mathbf{K} связаны с диагональными соотношениями $\mathbf{K}_{ij} = \mathbf{K}_{ii} \nu^{|i-j|}$. Теперь можно оценить искомые вероятности $\phi_{i,m}$, используя формулу (1).

В методе непоследовательного назначения PRB искомая вероятность выражается в терминах одномерной функции распределения. Однако в методе последовательного назначения PRB необходимо использовать совместную функцию распределения (1), что приводит к вычислительным сложностям. Вместо этого можно воспользоваться известной задачей о наилучшем линейном прогнозе [10]. Так, если даны две случайные величины ξ_1 и ξ_2 с математическими ожиданиями $\mathcal{M}[\xi_1]$ и $\mathcal{M}[\xi_2]$, дисперсиями $D[\xi_1]$ и $D[\xi_2]$ и ковариацией $\text{cov}(\xi_1, \xi_2)$, то наилучшее линейное приближение случайной величины $\xi_2 = x\xi_1 + a$ определяется выражениями:

$$x = \frac{\text{cov}(\xi_1, \xi_2)}{D\xi_2}; \quad a = (1 - x)\mathcal{M}\xi_2.$$

В нашем случае S_2, S_3, \dots, S_n можно выразить через одну случайную величину S_1 . Наилучшее линейное приближение случайной величины S_i через S_1 достигается при

$$x_i = \frac{\text{cov}(S_1, S_i)}{DS_k} = \frac{\nu^{i-1}\sigma^2}{\sigma^2} = \nu^{i-1};$$

$$a_i = x_i \mathcal{M}S_i = (1 - \nu^{i-1})\mu_i.$$

Таким образом, нужно найти вероятности

$$\begin{aligned} & \mathbf{P}\{\min(S_1, \dots, S_n) \in (s_k, s_{k+1})\} = \\ & = \mathbf{P}\{\min(S_1, \dots, S_n) < s_{k+1}\} - \\ & - \mathbf{P}\{\min(S_1, \dots, S_n) < s_k\} = \\ & = \mathbf{P}\{\min(S_1, \dots, S_n) > s_k\} - \\ & - \mathbf{P}\{\min(S_1, \dots, S_n) > s_{k+1}\}. \quad (2) \end{aligned}$$

Выражая все случайные величины через S_1 , получим:

$$\begin{aligned} & \mathbf{P}\{\min(S_1, \dots, S_n) > s_k\} = \\ & = \mathbf{P}\{S_1 > s_k, \dots, S_n > s_k\} = \\ & = \mathbf{P}\{S_1 > s_k, \nu S_1 + (1 - \nu)\mu_k > s_k, \dots, \\ & \dots, \nu^{n-1} S_1 + (1 - \nu^{n-1})\mu_k > s_k\} = \\ & = \mathbf{P}\left\{S_1 > \max\left(\frac{s_k - (1 - \nu^{i-1})\mu_k}{\nu^{i-1}}, i = \overline{1, n}\right)\right\}. \end{aligned}$$

Теперь (2) можно записать в терминах одномерной функции распределения. Принимая во внимание разделение зоны покрытия на K областей,

получим следующее выражение для искомых вероятностей:

$$\phi_{n,m} = \sum_{k=1}^K c_k \left[F_k \left(\max \left(\frac{s_{m+1} - (1 - \nu^{i-1})\mu_k}{\nu^{i-1}}, \right. \right. \right. \\ \left. \left. \left. i = \overline{1, n} \right) \right) - \right. \\ \left. - F_k \left(\max \left(\frac{s_m - (1 - \nu^{i-1})\mu_k}{\nu^{i-1}}, i = \overline{1, n} \right) \right) \right].$$

По аналогии с непоследовательным методом назначения PRB вероятности $\phi_{n,m}$, $m = \overline{1, M}$, определяют распределение r_2 .

3 Оценка характеристик обслуживания

Для оценки характеристик обслуживания при различных методах выделения ресурсов применяется ресурсная система массового обслуживания (РеСМО) с двумя типами ресурсов, которые отражают число PRB для PDCCH и PDSCH соответственно.

Рассматривается РеСМО с N приборами и двумя типами ресурсов, объем которых задается вектором $\mathbf{R} = (R_1, R_2)$. На систему поступает пуассоновский поток заявок с интенсивностью λ , соответствующих поступлению запросов на установление сессий передачи данных. Для обслуживания заявки требуется некоторое случайное число ресурсных единиц в каждом из ресурсных пулов, распределение требований задано вероятностным рядом $\{p_{\mathbf{r}}\}$, $\mathbf{0} \leq \mathbf{r} \leq \mathbf{R}$. Если свободного ресурса каждого типа достаточно для удовлетворения требований заявки, то она занимает прибор и требуемый объем ресурса на случайное время обслуживания со средним $b < \infty$. В противном случае заявка теряется.

Функционирование системы описывается случайным процессом $X(t) = (\xi(t), \delta(t))$, где $\xi(t)$ — число заявок в системе в момент t ; $\delta(t)$ — суммарный объем занятого ими ресурса. В работе [11] было показано, что совместное стационарное распределение числа заявок в системе и суммарного занятого ими объема ресурса имеет вид:

$$q_n(\mathbf{r}) = \lim_{t \rightarrow \infty} P\{\xi(t) = k, \delta(t) = \mathbf{r}\} = q_0 p_{\mathbf{r}}^{(n)} \frac{\rho^n}{n!}, \\ q_0 = \left(1 + \sum_{n=1}^N \frac{\rho^n}{n!} \sum_{\mathbf{0} \leq \mathbf{r} \leq \mathbf{R}} p_{\mathbf{r}}^{(n)} \right)^{-1}, \quad (3)$$

где $\rho = \lambda b$ — предложенная нагрузка; $p_{\mathbf{r}}^{(n)}$ — вероятность того, что суммарное требование n заявок равно \mathbf{r} .

Стоит отметить, что формулы для стационарного распределения (3) справедливы не только для экспоненциального времени обслуживания, но и для произвольного распределения с конечным средним [12].

Получив стационарное распределение, можно перейти к вероятностно-временным характеристикам модели. Во-первых, это вероятность потери заявки π :

$$\pi = q_0 \sum_{k=0}^{N-1} \frac{\rho^k}{k!} \sum_{\mathbf{0} \leq \mathbf{r} \leq \mathbf{R}} p_{\mathbf{r}}^{(k)} \left(1 - \sum_{\mathbf{0} \leq \mathbf{j} \leq \mathbf{R} - \mathbf{r}} p_{\mathbf{j}} \right) + \\ + q_0 \frac{\rho^N}{N!} \sum_{\mathbf{0} \leq \mathbf{r} \leq \mathbf{R}} p_{\mathbf{r}}^{(N)}.$$

Помимо общей вероятности потери интерес представляет вероятность потери π_i , вызванная недостаточностью i -го типа ресурса. Эти вероятности можно интерпретировать как вероятности потери, вызванной нехваткой емкости контрольного PDCCH или физического PDSCH:

$$\pi_i = q_0 \sum_{k=0}^{N-1} \frac{\rho^k}{k!} \sum_{\mathbf{0} \leq \mathbf{r} \leq \mathbf{R}} p_{\mathbf{r}}^{(k)} \sum_{\mathbf{j}: \mathbf{r}_i + \mathbf{j}_i > \mathbf{R}_i} p_{\mathbf{j}}, \quad i = \{1, 2\}.$$

Средний объем занятого ресурса вычисляется по формуле:

$$\bar{\mathbf{R}} = q_0 \sum_{k=1}^N \frac{\rho^k}{k!} \sum_{\mathbf{0} \leq \mathbf{r} \leq \mathbf{R}} \mathbf{r} p_{\mathbf{r}}^{(k)},$$

что, в свою очередь, позволяет получить коэффициент использования ресурсов:

$$\mathbf{R}_{\text{util}} = \left(\frac{\bar{R}_1}{R_1}, \frac{\bar{R}_2}{R_2} \right).$$

4 Численные результаты

В данном разделе приводится численный пример расчета искомых характеристик. Исходные данные, использованные для получения графиков и соответствующие потоковому трафику с постоянной скоростью (CBR — Constant Bitrate), представлены в табл. 2. Для расчета средних значений ОСШ μ_k , $k = \overline{1, K}$, используется модель распространения по сценарию городского каньона UMi (Street Canyon) из технической спецификации 3GPP TS 38.901 [7]. Так, затухание волны $L(y)$ на расстоянии y при распространении в пространстве в линейном масштабе имеет вид:

$$L(y) = 10^{2 \lg f_c + 3.24} y^{-\zeta},$$

Таблица 2 Параметры для численного примера

Обозначение	Описание	Значение по умолчанию
s_1	Нижняя граница ОСШ	-9,47 дБ
σ	Среднеквадратичное отклонение распределения теневого замирания	4 дБ
C	Требуемая скорость	15 Мбит/с
B	Объем выделенного ресурса	[100–400] МГц
r_C	Радиус зоны покрытия БС	480 м
N	Число приборов	100
R_1	Емкость PDCCH для одного CCE	15 PRB
w_p	Размер PRB	1,44 МГц
λ	Интенсивность поступления сессий	0,5 сессий/с
b	Среднее время обслуживания сессии	10 с
P_U	Мощность передачи ПУ	33 дБм
f_C	Несущая частота	28 ГГц
G_A, G_U	Коэффициент усиления антенной решетки на БС/ПУ	4 дБи, 16 дБи
N_0	Спектральная мощность плотности шума	-82 дБм/Гц

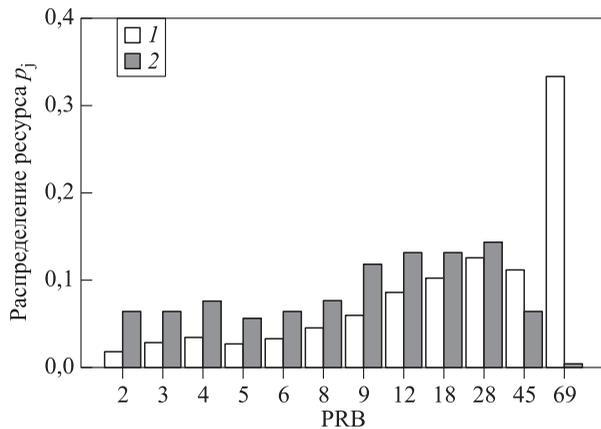


Рис. 1 Ряды распределения количества запрашиваемых PRB: 1 — последовательное; 2 — непоследовательное

где f_C — несущая частота, ГГц; $\zeta = 3,19$ — коэффициент затухания. Среднее значение ОСШ на приемнике μ_k на расстоянии x составляет

$$\mu_k = \frac{P_U G_A G_U}{N_0 L(x_k)} \text{ [дБ]},$$

где P_U — мощность передачи ПУ; G_A — коэффициент усиления антенной решетки на БС; G_U — коэффициент усиления антенной решетки на ПУ; N_0 — спектральная мощность плотности шума; $x_k = (r_{C,k} + r_{C,k+1})/2$ — расстояние от середины k -й области до БС.

Графики, представленные на рис. 1, иллюстрируют ряды распределения ресурсов, необходимых для обслуживания сессии для последовательного и непоследовательного назначения ресурсов планировщиком. Отметим как качественную, так и количественную разницу в структуре рядов распределения. В частности, последовательный выбор PRB

для сессий приводит к тому, что вероятность запроса максимального объема ресурсов для сессии в несколько раз превышает аналогичное значение для непоследовательного назначения PRB. При этом среднее значение для непоследовательного назначения в 2,52 раза больше, чем для последовательного.

Рассмотрим вероятности переполнения PDSCH и PDCCH, представленные на рис. 2. Данные вероятности были подсчитаны при различных значениях числа контрольных PDCCH, которые задаются параметром элемента контрольного канала (CCE — Control Channel Element). Дополнительно рис. 3 иллюстрирует вероятности переполнения индивидуальных каналов для одного элемента CCE, включающего в себя 15 PRB. Отметим, что для последовательного назначения количества PRB общая вероятность переполнения каналов не меняется с увеличением объема ресурса (CCE) в сигнальном канале, что объясняется тем, что узким местом при таком методе выделения ресурсов оказывается канал PDSCH.

Для метода непоследовательного назначения PRB использование дополнительных ресурсов для канала PDCCH (увеличение CCE) приводит к значительному снижению вероятности переполнения сигнального канала PDCCH. При небольшом объеме выделенного ресурса (100–150 МГц) узким местом системы все еще остается канал передачи данных (PDSCH). При увеличении объема ресурсов системы вероятность переполнения PDSCH падает, что приводит к росту принятой нагрузки и, следовательно, к некоторому увеличению вероятности переполнения контрольного канала. Дальнейшее увеличение объема ресурсов системы приводит к тому, что вероятность переполнения PDSCH становится

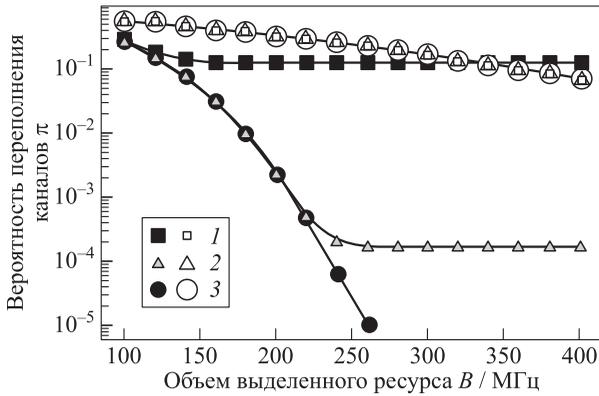


Рис. 2 Вероятность блокировки каналов как функции от количества ресурсов системы: залитые и пустые значки — непоследовательное и последовательное назначения PRB соответственно; 1 — 1 CCE; 2 — 2 CCE; 3 — 3 CCE

крайне малой, а узким местом системы становится сигнальный PDCCH. Так как объем требований к PDCCH фиксирован для каждого метода выделения ресурса, то при прекращении роста принятой нагрузки вероятность переполнения сигнального PDCCH тоже перестает расти.

Проанализируем коэффициент использования ресурсов PDCCH и PDSCH, представленный на рис. 4 для одного CCE. Отметим, что при использовании последовательного назначения PRB канал управления нисходящего канала связи (PDCCH) остается недогруженным при всех рассматриваемых значениях объема ресурсов. Напротив, при непоследовательном распределении коэффициент использования ресурсов PDSCH уменьшается с увеличением объема ресурсов системы. Оба эффекта обусловлены поведением рядов распреде-

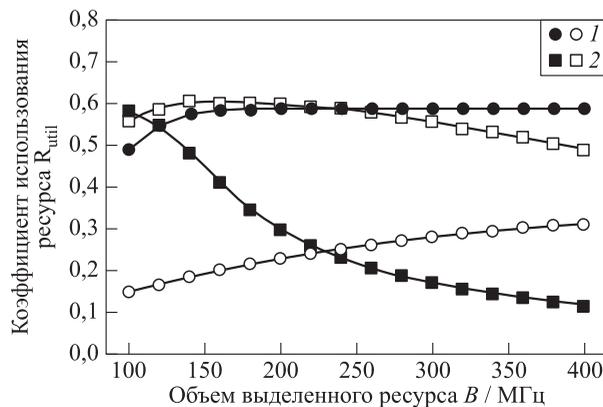


Рис. 4 Коэффициент использования ресурсов каналов: залитые и пустые значки — непоследовательное и последовательное назначения PRB соответственно; 1 — R1, PDCCH; 2 — R2, PDSCH

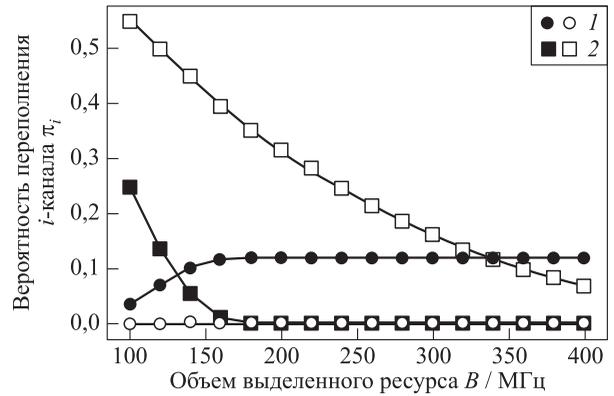


Рис. 3 Вероятность блокировки канала i -типа как функции от количества ресурсов системы: залитые и пустые значки — непоследовательное и последовательное назначения PRB соответственно; 1 — PDCCH; 2 — PDSCH

ления ресурсов, необходимых для обслуживания сессии, представленных на рис. 1. При обоих типах назначения ресурсов коэффициент использования любого типа канала не превышает 0,6, что также следует из структуры рядов распределения ресурсов, необходимых для обслуживания сессии.

5 Заключение

Процесс планирования ресурсов в системе 5G NR — сложная процедура, производительность которой зависит от взаимодействия между объемом доступных ресурсов в каналах данных и управления, а также от типов алгоритмов распределения PRB. В данном исследовании разработана модель, которая учитывает процесс обслуживания сессии как в канале PDSCH, так и в PDCCH, а также различные типы назначения PRB.

Проведенные численные исследования показывают, что непоследовательное назначение PRB превосходит последовательное во всех режимах работы системы. При этом эффективное использование системных ресурсов может требовать динамического назначения ресурсов канала PDCCH, которое зависит от полосы пропускания системы.

Литература

1. Moltchanov D., Sopin E., Begishev V., Samuylov A., Koucheryavy Y., Samuylov 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.

2. Сопин Э. С., Маслов А. Р., Шоргин В. С., Бегишев В. О. Моделирование настойчивого поведения пользователей в сетях 5G NR с адаптацией скорости и блокировками // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 25–32. doi: 10.14357/19922264230304.
3. Иванова Д. В., Маркова Е. В., Шоргин С. Я., Гайдамака Ю. В. Модели совместного обслуживания трафика eMBB и URLLC на основе приоритетов в промышленных развертываниях 5G NR // Информатика и её применения, 2023. Т. 17. Вып. 4. С. 64–70. doi: 10.14357/19922264230409.
4. Hamidi-Selehr F., Kwak Y., Chatterjee D. 5G NR PDCCH: Design and performance // IEEE 5G World Forum Proceedings. — IEEE, 2018. P. 250–255. doi: 10.1109/5GWF.2018.8517070.
5. Reddy M. P., Kumar D. H., Amuru S., Kuchi K. Removing the PDCCH bottleneck and enhancing the capacity of 4G massive MIMO systems // Conference (International) on Communication Systems & Networks Proceedings. — IEEE, 2020. P. 237–244. doi: 10.1109/COMSNETS48256.2020.9027449.
6. 3GPP Technical Specification 38.213: NR; Physical layer procedures for control (Release 15), 2022. https://www.3gpp.org/ftp/Specs/archive/38_series/38.213/38213-001.zip.
7. 3GPP Technical Specification 38.901: Study on channel model for frequencies from 0.5 to 100 GHz (Release 14), 2017. https://www.3gpp.org/ftp/Specs/archive/38_series/38.901/38901-100.zip.
8. David H. A., Nagaraja H. N. Order statistics. — Hoboken, NJ, USA: John Wiley & Sons, 2004. 488 p.
9. Capozzi F., Piro G., Grieco L. A., Boggia G., Camarda P. Downlink packet scheduling in LTE cellular networks: Key design issues and a survey // IEEE Commun. Surv. Tut., 2012. Vol. 15. Iss. 2. P. 678–700. doi: 10.1109/SURV.2012.060912.00100.
10. Бочаров П. П., Печинкин А. В. Теория вероятностей. Математическая статистика. — М.: Физматлит, 2005. 296 с.
11. Горбунова А. В., Наумов В. А., Гайдамака Ю. В., Самуйлов К. Е. Ресурсные системы массового обслуживания как модели беспроводных систем связи // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 48–55. doi: 10.14357/19922264180307. EDN: YAMDIL.
12. Горбунова А. В., Наумов В. А., Гайдамака Ю. В., Самуйлов К. Е. Ресурсные системы массового обслуживания с произвольным обслуживанием // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 99–107. doi: 10.14357/19922264190114. EDN: ZASZJJ.

Поступила в редакцию 30.06.2025

Принята к публикации 15.08.2025

MODEL DEVELOPMENT FOR PERFORMANCE OF TRANSMISSION SCHEDULING IN 5G NEW RADIO NETWORKS UNDER SIGNALING AND USER TRAFFIC CONSTRAINTS

E. S. Sopin^{1,2}, A. I. Nazarin¹, and S. Ya. Shorgin²

¹Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

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

Abstract: The goal of the study is to analyze the trade-off between Physical Downlink Control Channel (PDCCH) and Physical Downlink Shared Channel (PDSCH) resources in 5G NR (New Radio) as the number of subscribers increases. Using stochastic geometry and probability theory methods, the models for Primary Resource Block (PRB) assignment strategies were developed. Subsequently, by applying queueing theory, a model jointly considering subscriber demands for both channels was formulated. Numerical results indicate that with nonsequential PRB assignment, the PDCCH size becomes the bottleneck, while with sequential assignment, a lack of PDCCH resources leads to missed scheduling opportunities. Nonsequential resource assignment potentially allows for increased PDSCH utilization but requires careful tuning of the PDCCH and PDSCH resource ratio. The proposed model enables determining the volume of resources allocated for PDCCH/PDSCH based on user needs, minimizing missed scheduling opportunities and maximizing resource utilization. The numerical investigation revealed that sequential resource assignment in the PDSCH channel leads to a significant degradation in system performance. Furthermore, for effective utilization of the nonsequential resource assignment method, dynamic adjustment of the resources allocated to the PDCCH is required.

Keywords: 5G; millimeter wave; scheduling; resource loss system

DOI: 10.14357/19922264250306

EDN: EMAQVK

Acknowledgments

The research was funded by the Russian Science Foundation, project No.22-79-10128, <https://rscf.ru/en/project/22-79-10128/>.

References

1. Moltchanov, D., E. Sopin, V. Begishev, A. Samuylov, Y. Koucheryavy, and K. Samouylov. 2018. 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.
2. Sopin, E. S., A. R. Maslov, V. S. Shorgin, and V. O. Begishev. 2023. Modelirovanie nastoychivogo povedeniya pol'zovateley v setyakh 5G NR s adaptatsiyey skorosti i blokirovkami [Modeling insistent user behavior in 5G New Radio networks with rate adaptation and blockage]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):25–32. doi: 10.14357/19922264230304.
3. Ivanova, D. V., E. V. Markova, S. Ya. Shorgin, and Yu. V. Gaidamaka. 2023. Modeli sovместnogo obsluzhivaniya trafika eMBB i URLLC na osnove prioritetov v promyshlennykh razvertyvaniyakh 5G NR [Priority-based eMBB and URLLC traffic coexistence models in 5G NR industrial deployments]. *Informatika i ee Primeneniya — Inform. Appl.* 17(4):64–70. doi: 10.14357/19922264230409.
4. Hamidi-Sepehr, F., Y. Kwak, and D. Chatterjee. 2018. 5G NR PDCCH: Design and performance. *IEEE 5G World Forum Proceedings*. IEEE. 250–255. doi: 10.1109/5GWF.2018.8517070.
5. Reddy, M. P., D. H. Kumar, S. Amuru, and K. Kuchi. 2020. Removing the PDCCH bottleneck and enhancing the capacity of 4G massive MIMO systems. *Conference (International) on Communication Systems & Networks Proceedings*. IEEE. 237–244. doi:10.1109/COMSNETS48256.2020.9027449.
6. 3GPP Technical Specification 38.213: NR; Physical layer procedures for control (Release 15). 2022. Available at: https://www.3gpp.org/ftp/Specs/archive/38_series/38.213/38213-001.zip (accessed June 28, 2025).
7. 3GPP Technical Specification 38.901: Study on channel model for frequencies from 0.5 to 100 GHz. 2017. Available at: https://www.3gpp.org/ftp/Specs/archive/38_series/38.901/38901-100.zip (accessed June 28, 2025).
8. David, H. A., and H. N. Nagaraja. 2004. *Order statistics*. Hoboken, NJ: John Wiley & Sons. 488 p.
9. Capozzi, F., G. Piro, L. A. Grieco, G. Boggia, and P. Camarda. 2012. Downlink packet scheduling in LTE cellular networks: Key design issues and a survey. *IEEE Commun. Surv. Tut.* 15(2):678–700. doi: 10.1109/SURV.2012.060912.00100.
10. Bocharov, P. P., and A. V. Pechinkin. 2005. *Teoriya veroyatnostey. Matematicheskaya statistika* [Probability theory. Mathematical statistics]. Moscow: Fizmatlit. 296 p.
11. Gorbunova, A. V., V. A. Naumov, Yu. V. Gaidamaka, and K. E. Samouylov. 2018. Resursnye sistemy massovogo obsluzhivaniya kak modeli besprovodnykh sistem svyazi [Resource queuing systems as models of wireless communication systems]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):48–55. doi: 10.14357/19922264180307. EDN: YAMDIL.
12. Gorbunova, A. V., V. A. Naumov, Yu. V. Gaidamaka, and K. E. Samouylov. 2019. Resursnye sistemy massovogo obsluzhivaniya s proizvol'nym obsluzhivaniem [Resource queuing systems with general service discipline]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):99–107. doi: 10.14357/19922264190114. EDN: ZASZJJ.

Received June 30, 2025

Accepted August 15, 2025

Contributors

Sopin Eduard S. (b. 1987) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Probability Theory and Cyber Security, Peoples' Friendship University of Russia (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; sopin-es@rudn.ru

Nazarin Artem I. (b. 2000) — master student, Department of Probability Theory and Cyber Security, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; nazaryin_ai@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 119133, Russian Federation; sshorgin@ipiran.u

МНОГОВАРИАНТНОЕ РЕЗЕРВИРОВАНИЕ С УЧЕТОМ ЛОГИКО-ТОПОЛОГИЧЕСКИХ ОСОБЕННОСТЕЙ ТРАНЗИСТОРНЫХ СХЕМ

С. Ф. Тюрин¹, М. С. Никитин², Ю. А. Степченков³, Ю. Г. Дьяченко⁴

Аннотация: Рассматривается пассивная отказо- и сбоеустойчивость цифровых элементов и устройств с использованием многовариантной избыточности с учетом топологических особенностей резервирования транзисторов. Строится модель, включающая мажоритарное резервирование каналов с резервированием самих мажоритаров, допускающих «развал» каналов при диагностировании, глубокое резервирование с избыточностью на уровне слоев отдельных каналов со специальными мажоритарными, обеспечивающими конфигурирование слоев в каналы. Известные методы комбинируются в соотношении, оптимизирующем заданную целевую функцию с требуемыми ограничениями. Кроме того, применяется резервирование на уровне отдельных транзисторов с разной степенью парирования отказов. Исследуются топологические особенности такого резервирования путем построения различных вариантов схем на основе дизъюнктивной нормальной формы (ДНФ), конъюнктивной нормальной формы (КНФ) и промежуточных форм. Устанавливается мощность множества таких вариантов. Предлагается метод поиска топологически лучшего варианта при большой размерности устройства. Путем топологического моделирования устанавливается предпочтительный вариант резервирования по показателю произведения потребляемой мощности на задержку переключения. Приводятся примеры параметров созданных топологий.

Ключевые слова: отказо- и сбоеустойчивость; резервирование; мажоритарный элемент; топологическое моделирование

DOI: 10.14357/19922264250307

EDN: TXXPPO

1 Введение

Несмотря на впечатляющие успехи интегральных технологий в области повышения показателей качества интегральных микросхем: у ведущих производителей выход годных чипов, по некоторым оценкам, — до 70% и выше [1], обеспечение сбое- и отказоустойчивости [2–4] цифровых устройств, содержащих десятки миллиардов транзисторов, в условиях воздействия дестабилизирующих факторов [5] остается крайне актуальной задачей, особенно для вычислительных систем критического применения, в том числе для военной техники и аппаратуры.

Микроэлектроника особой надежности создается для работы в условиях ионизирующего излучения и бомбардировки тяжелыми заряженными частицами [5], способна выдерживать ускорения свыше 3G и работать в расширенном диапазоне температур от -273 до $+400$ °C. Хотя качество и надежность микросхем непрерывно возрастают, существующие модели и методы обеспечения сбое- и отказоустойчивости цифровых элементов вычислительных систем по-прежнему основываются на

принципах структурного резервирования, компенсирующих возможную ненадежность единичных экземпляров [2, 6–8]. Интенсивность отказов транзисторов уменьшается до 10^{-12} 1/ч в особо качественных вариантах исполнения, хотя, конечно, такие показатели вряд ли могут быть в полной мере подтверждены ускоренными испытаниями, что справедливо и для температурных испытаний.

При уменьшении проектных норм (с 10 до 2 нм) резко возрастают токи утечки. Этот недостаток устраняется новыми технологиями изготовления транзисторов [9–11]: это и оригинальный химический состав затворов, и специальные «объемные» транзисторы типа Tri-Gate или FinFET, GAA, MBC-FETs и др., позволяющие уменьшить потребляемую мощность или снизить задержку, а также улучшить показатели надежности.

Тем не менее обеспечение сбое- и отказоустойчивости цифровых устройств, вычислительных систем, цифровой аппаратуры критического применения, работающих под воздействием жестких дестабилизирующих воздействий, особенно в военной технике, все так же очень актуально [12–15].

¹Пермский национальный исследовательский политехнический университет; Пермский государственный национальный исследовательский университет, tyurinsergfe@yandex.ru

²Пермский национальный исследовательский политехнический университет, mann1k@yandex.ru

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

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

Далее в работе используется обобщенный термин «отказоустойчивость». В этой области применяются экспоненциальные модели распределения Вейбулла [16] времени наработки на отказ основных вариантов резервирования. При этом особенности топологий резервированных структур учитываются не в полной мере в моделях обеспечения отказоустойчивости [2, 3, 6, 8, 17–19]. Поэтому предлагается новый подход к многовариантному резервированию с выбором наиболее предпочтительного варианта, учитывающего и топологические особенности избыточности на уровне транзисторов. Под топологией понимается «пространственно-геометрическое расположение элементов», в данном случае — транзисторов и «связей между ними на кристалле кремния», выполненное в системе топологического моделирования. При этом логическая функция и вероятность безотказной работы (ВБР) всех вариантов считается одинаковой. Логические элементы могут строиться путем соединения таких транзисторных структур.

2 Постановка задачи

Предметом исследования ставится обеспечение отказоустойчивости цифровых элементов и устройств путем многовариантного резервирования с учетом топологически предпочтительных вариантов транзисторного резервирования. Вероятностные характеристики определяются структурной схемой надежности и соответствующими логическими функциями. Инструментом для изменения этих характеристик служит комбинирование различных известных методов резервирования в цифровом устройстве, содержащем несколько элементов (устройств), причем частичное транзисторное резервирование применяется только для наиболее важных участков, для которых выполняется поиск наилучшего варианта резервирования вместе с поиском наилучшего варианта относительно всего устройства.

При этом строится следующая модель. При использовании мажоритирования мажоритаров [20, 21], представляющих собой схему голосования по большинству голосов, и интенсивности отказов одного мажоритару $\lambda_{мэ}$ для одного выхода канала резервирования по принципу « $\geq n$ », где n — число работоспособных каналов из общего количества $2n - 1$, $n \geq 2$, ВБР рассчитывается по формуле:

$$P_{\geq n}^* = \left(e^{-\lambda t^\alpha (2n-1)} + \sum_{i=1}^{\lfloor (2n-1)/2 \rfloor} \left\{ C_{2n-1}^i \left[e^{-\lambda t^\alpha (2n-1-i)} \left(1 - e^{-\lambda t^\alpha} \right)^i \right] \right\} \right) \times$$

$$\times \left(e^{-\lambda_{мэ} t^\alpha (2n-1)} + \sum_{i=1}^{\lfloor (2n-1)/2 \rfloor} \left\{ C_{2n-1}^i \times \left[e^{-\lambda_{мэ} t^\alpha (2n-1-i)} \left(1 - e^{-\lambda_{мэ} t^\alpha} \right)^i \right] \right\} \right), \quad (1)$$

где $\lfloor \cdot \rfloor$ — округление в меньшую сторону; λ — интенсивность отказов (сбоев) одного канала; C_{2n-1}^i — число сочетаний из $2n - 1$ по i ; $1 \leq \alpha < 2$ — коэффициент распределения Вейбулла [16]; t — время работы. Избыточность в каналах равна $2n - 2$.

Вероятность безотказной работы при использовании так называемого глубокого мажоритирования, когда выделяются k условных слоев, или подканалов, в одновыходных каналах с интенсивностью отказов λ , описывается формулой:

$$P_{\geq n \text{ ГМ}}^* = \left(e^{-(\lambda/k)t^\alpha (2n-1)} + \sum_{i=1}^{\lfloor (2n-1)/2 \rfloor} \left\{ C_{2n-1}^i \times \left[e^{-(\lambda/k)t^\alpha (2n-1-i)} \left(1 - e^{-(\lambda/k)t^\alpha} \right)^i \right] \right\} \right)^k \times$$

$$\times \left(e^{-\lambda_{мэ} t^\alpha (2n-1)} + \sum_{i=1}^{\lfloor (2n-1)/2 \rfloor} \left\{ C_{2n-1}^i \times \left[e^{-\lambda_{мэ} t^\alpha (2n-1-i)} \left(1 - e^{-\lambda_{мэ} t^\alpha} \right)^i \right] \right\} \right)^k. \quad (2)$$

Вероятность безотказной работы при резервировании отдельных транзисторов [21, 22], обозначаемом TrR, описывается выражением:

$$P_{(r+1)^2}(t)_{\text{TrR}} = \left(\sum_{i=0}^r C_{(r+1)^2}^i \times \left\{ e^{-[(r+1)^2-i]\lambda_1 t^\alpha} \left(1 - e^{-\lambda_1 t^\alpha} \right)^i \right\} \right)^w, \quad (3)$$

где λ_1 — интенсивность отказов одного транзистора; r — количество парируемых отказов в группе резервированного транзистора; $(r + 1)^2 - 1$ — избыточность числа транзисторов; w — число транзисторов в устройстве до резервирования; $(r + 1)^2 w$ — всего транзисторов после резервирования. Если же проводилась декомпозиция исходного устройства, то число транзисторов увеличивается для выполнения ограничения Мида–Конвей e на число последовательно соединенных транзисторов, которых

в технологиях с проектными нормами 65 нм и ниже должно быть не более четырех [23]. При этом для $r = 1$, числа транзисторов n и выходов m троирование хуже «расчетверения» (каждый транзистор заменяется четырьмя репликами), если $12m/n \geq 1$ [22]. Тогда исходная модель, комбинирующая произведение (1), (2), (3), описывается выражением $P(t, \lambda, L, W, T, S)$, где t — время работы; λ — интенсивность отказов (сбоев); W — потребляемая мощность; T — задержка переключения; S — площадь топологической реализации схемы на кристалле; L — число транзисторов. В каждом из рассмотренных вариантов резервирования (1), (2), (3) величины n, t, λ, W, T, S и L имеют свое значение. Рассматривается пассивная отказоустойчивость, самая «быстрая», но и самая дорогая. Предлагаемая модель использует комбинирование с использованием логико-топологической модификации транзисторного резервирования $P_{(r+1)^2TrR^*}$. Задача заключается:

(1) в максимизации общей ВБР P^* :

$$P_{\geq n}^*(t, \lambda_{\geq n}, W_{\geq n}, T_{\max, \geq n}, S_{\geq n}, L_{\geq n}) \times P_{\geq n \text{ ГМ}}^*(t, \lambda_{\geq n \text{ ГМ}}, W_{\geq n \text{ ГМ}}, T_{\max, \geq n \text{ ГМ}}, S_{\geq n \text{ ГМ}}, L_{\geq n \text{ ГМ}}) \times P_{(r+1)^2TrR}(t, \lambda_{\geq n \text{ TrR}}, W_{\geq n \text{ TrR}}, T_{\max, \geq n \text{ TrR}}, S_{\geq n \text{ TrR}}, L_{\geq n \text{ TrR}}) \rightarrow \max$$

при заданных ограничениях $WT \leq (WT)_{\text{треб}}$, $S \leq S_{\text{треб}}$, $L \leq L_{\text{треб}}$;

(2) минимизации произведения WT :

$$WT \rightarrow \min$$

при заданных ограничениях

$$(P_{\geq n}^*(t, \lambda_{\geq n}, W_{\geq n}, T_{\max, \geq n}, S_{\geq n}, L_{\geq n}) \times P_{\geq n \text{ ГМ}}^*(t, \lambda_{\geq n \text{ ГМ}}, W_{\geq n \text{ ГМ}}, T_{\max, \geq n \text{ ГМ}}, S_{\geq n \text{ ГМ}}, L_{\geq n \text{ ГМ}}) \times P_{(r+1)^2TrR}(t, \lambda_{\geq n \text{ TrR}}, W_{\geq n \text{ TrR}}, T_{\max, \geq n \text{ TrR}}, S_{\geq n \text{ TrR}}, L_{\geq n \text{ TrR}}) \leq P_{\text{треб}}) \& (S \leq S_{\text{треб}}) \& (L \leq L_{\text{треб}}).$$

3 Логико-топологическая модификация транзисторного резервирования

Резервирование транзистора, управляемого сигналом x , описывается логическими выражениями ДНФ $f_{\vee}^r = \bigvee_{i=1}^{r+1} \bigwedge_{j=1}^{r+1} x_{(i-1)(r+1)+j}$ или КНФ $f_{\wedge}^r = \bigwedge_{i=1}^{r+1} \bigvee_{j=1}^{r+1} x_{(i-1)(r+1)+j}$, которые при $r = 1$ (парировается один отказ, сбой) имеют вид

$$f_{\vee}^1 = x_1 x_2 \vee x_3 x_4; f_{\wedge}^1 = (x_1 \vee x_2)(x_3 \vee x_4) \quad (4)$$

Таблица 1 Функции парирования отказов (сбоев) одного транзистора в ДНФ и КНФ при $r = 1$

№ набора	Переменные				ДНФ/КНФ
	x_4	x_3	x_2	x_1	
0	0	0	0	0	0
1	0	0	0	1	0
2	0	0	1	0	0
3	0	0	1	1	1/0
4	0	1	0	0	0
5	0	1	0	1	0
6	0	1	1	0	0
7	0	1	1	1	1
8	1	0	0	0	0
9	1	0	0	1	0
10	1	0	1	0	0
11	1	0	1	1	1
12	1	1	0	0	1/0
13	1	1	0	1	1
14	1	1	1	0	1
15	1	1	1	1	1

и реализуют «расчетверение» транзистора, т. е. используются четыре транзистора вместо одного по логическим функциям (4). В табл. 1 показаны функции парирования отказов одного транзистора в ДНФ и КНФ (4) при $r = 1$.

На рис. 1 показаны варианты схемы «расчетверения» по ДНФ (4) и по КНФ (4), а также примеры отказов в виде постоянно выключенного транзистора — обрыва, постоянно включенного транзистора — закорачивания под действием тяжелой заряженной частицы. Во всех случаях функция структуры (x_1) сохраняется, но происходит рост сложности в $(r + 1)^2$ раз и увеличение временной задержки переключения более чем в r раз.

Таким образом, имеются всего два логически и вероятностно эквивалентных варианта. «Удевятирование» предполагает использование девяти транзисторов вместо одного:

$$\text{ДНФ: } x_{1.1}x_{1.2}x_{1.3} \vee x_{1.4}x_{1.5}x_{1.6} \vee x_{1.7}x_{1.8}x_{1.9};$$

$$\text{КНФ: } (x_{1.1} \vee x_{1.4} \vee x_{1.7})(x_{1.2} \vee x_{1.5} \vee x_{1.8})(x_{1.3} \vee x_{1.6} \vee x_{1.9}).$$

Оно позволяет парировать два отказа или сбоя, как показано на рис. 2.

Шестнадцатикратное резервирование использует шестнадцать транзисторов вместо одного, парировуются три отказа (сбоя) (рис. 3):

$$\text{КНФ: } (x_{1.1} \vee x_{1.5} \vee x_{1.9} \vee x_{1.13})$$

$$(x_{1.2} \vee x_{1.6} \vee x_{1.10} \vee x_{1.14})$$

$$(x_{1.3} \vee x_{1.17} \vee x_{1.11} \vee x_{1.15})$$

$$(x_{1.4} \vee x_{1.8} \vee x_{1.12} \vee x_{1.16});$$

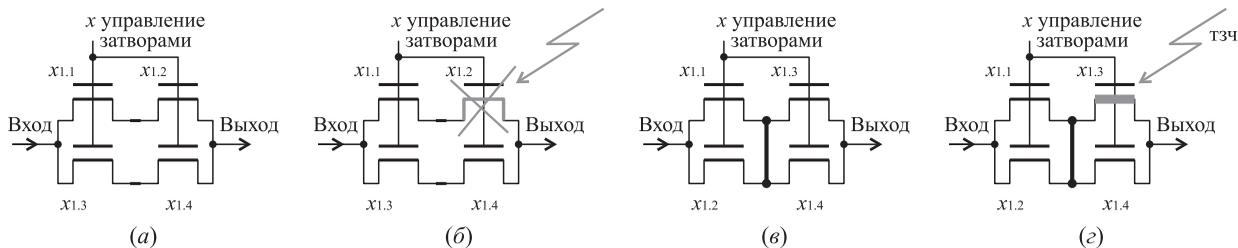


Рис. 1 Примеры схем «расчетверения» одного транзистора по ДНФ (а) и по КНФ (б), постоянно выключенного транзистора — обрыва по $x_{1.2}$ (в) и постоянно включенного транзистора — закорачивания по $x_{1.3}$ (г)

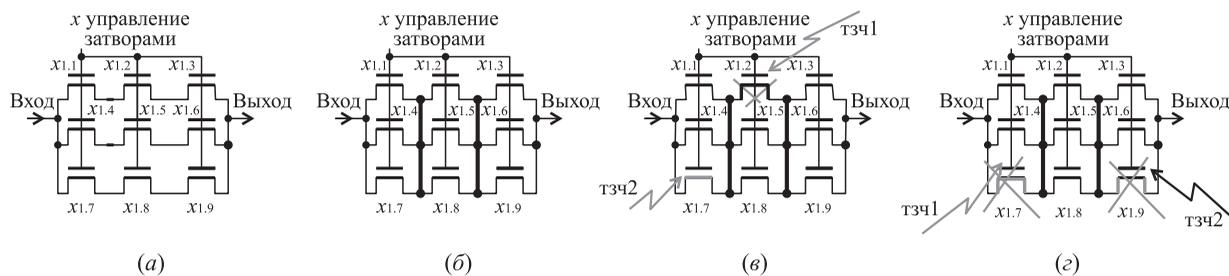


Рис. 2 Примеры схем «удевятерения» по ДНФ (а) и по КНФ (б), постоянно выключенного транзистора — обрыва по $x_{1.2}$ и постоянно включенного транзистора — закорачивания по $x_{1.7}$ (в) и двух постоянно выключенных транзисторов — обрывов по $x_{1.7}, x_{1.9}$ (г)

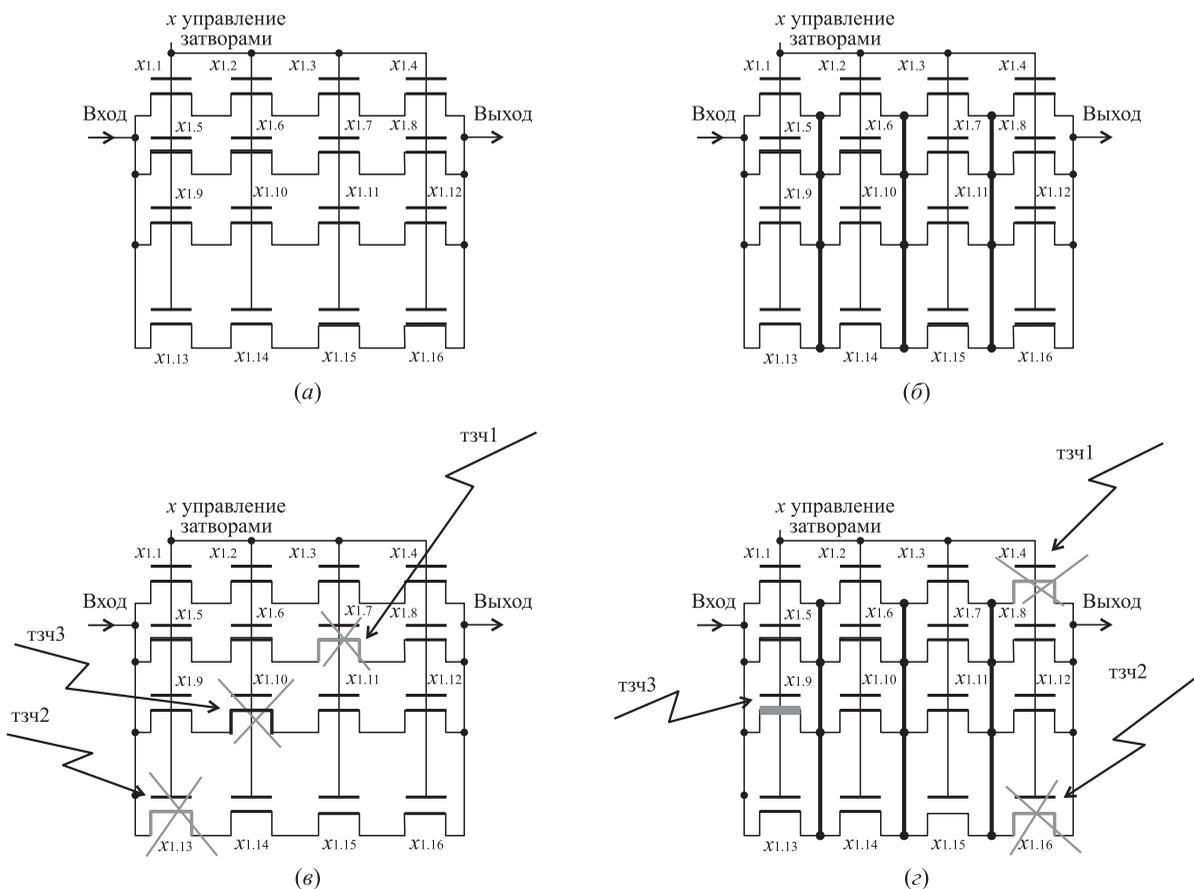


Рис. 3 Примеры схем «расшестнадцатерения» по ДНФ (а) и по КНФ (б), трех постоянно выключенных транзисторов — обрывов по $x_{1.13}, x_{1.10}, x_{1.7}$ (в) и двух постоянно выключенных транзисторов — обрывов по $x_{1.4}, x_{1.16}$, и одного постоянно включенного транзистора — закорачивания по $x_{1.9}$ (г)

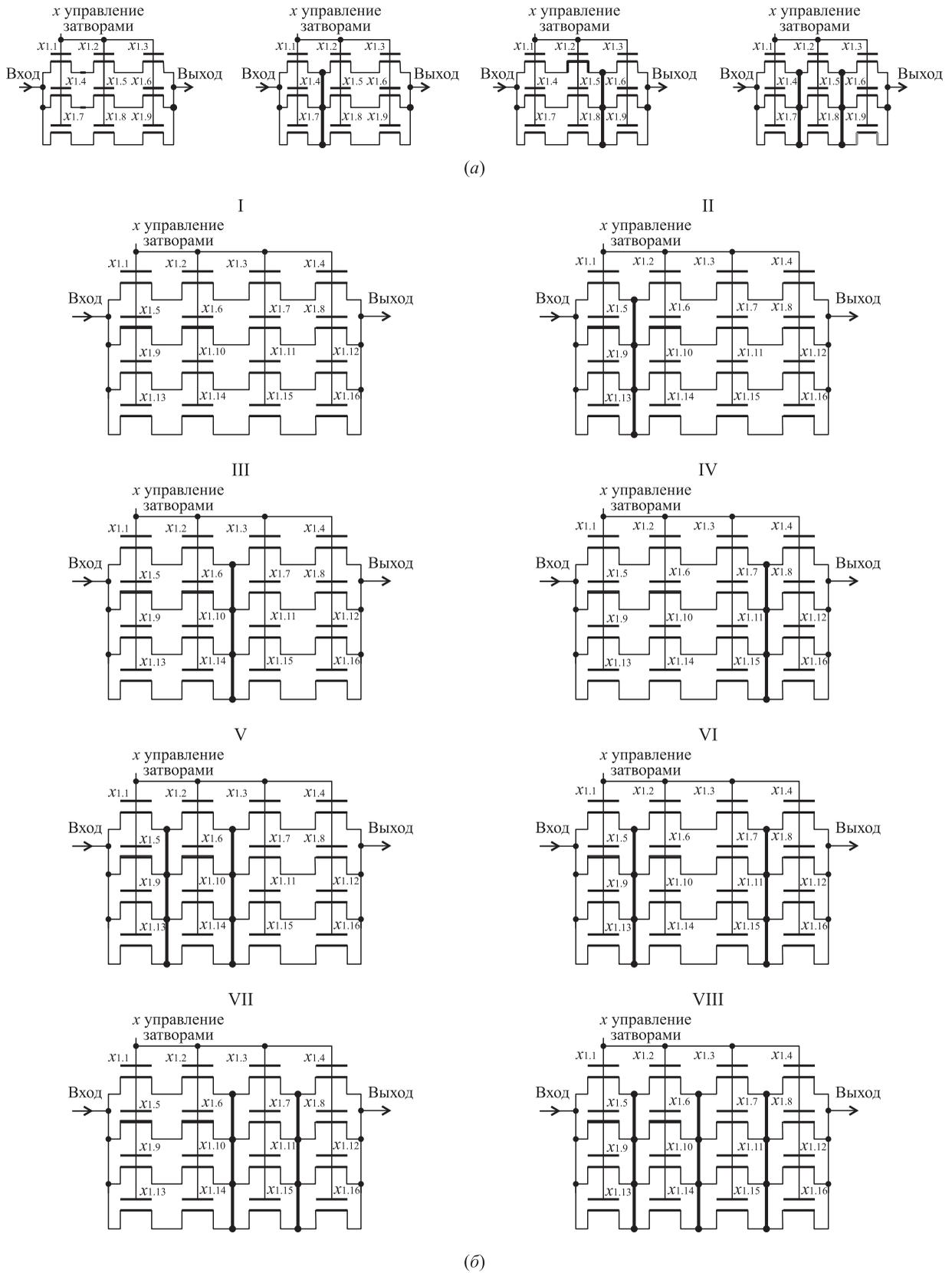


Рис. 4 Варианты резервирования для $r = 2$ — четыре варианта (а) и для $r = 3$ — восемь вариантов (б)

$$\text{ДНФ: } x_{1.1}x_{1.2}x_{1.3}x_{1.4} \vee x_{1.5}x_{1.6}x_{1.7}x_{1.8} \vee \\ \vee x_{1.9}x_{1.10}x_{1.11}x_{1.12} \vee x_{1.13}x_{1.14}x_{1.15}x_{1.16}.$$

При парировании одного отказа (сбоя) ($r = 1$) других вариантов, кроме ДНФ и КНФ, нет. При $r = 2$ получаем третий вариант:

$$(x_{1.1}x_{1.2} \vee x_{1.3}x_{1.4} \vee x_{1.5}x_{1.6})(x_{1.7} \vee x_{1.8} \vee x_{1.9}),$$

вариант

$$(x_{1.1} \vee x_{1.2} \vee x_{1.3})(x_{1.4}x_{1.5} \vee x_{1.6}x_{1.7} \vee x_{1.8}x_{1.9})$$

тоже может быть принят во внимание, поскольку топологические характеристики могут быть разными при учете направления протекания тока (рис. 4, а). При $r = 3$ получаем всего восемь вариантов, включая ДНФ и КНФ (рис. 4, б). При этом варианты резервированных схем транзистора $Q(r + 1) = 2^r$ трактуются как композиции целого числа, например для $r = 3$ их восемь. Варианты могут быть описаны в виде $f_{\text{комб}}^{r=\alpha+\beta}$, α — части ДНФ ($\alpha \geq 2$), β — части КНФ ($\beta = 1$). Например, VI вариант на рис. 4, б выглядит как (1, 2, 1), где «1» — элемент КНФ; «2» — элемент ДНФ. Для нескольких передающих КМОП-транзисторов (КМОП — комплементарный металл–оксид–полупроводник) и элементов варианты могут комбинаторно умножаться.

4 Моделирование и оценка вероятности безотказной работы

Моделирование подтвердило работоспособность и эффективность транзисторного резервирования и позволило получить параметры различных вариантов резервирования. На рис. 5 показаны схемы «расчетверенного» инвертора, полученные в системе топологического моделирования Mircrowind [24]. В табл. 2 сравниваются различные варианты элемента НЕ (инвертора) для 32-нанометровой КМОП-технологии, а на рис. 6 — показатели его надежности, полученные в системе компьютерной алгебры Mathcad.

Оказалось, что при $r = 1$ вариант «везде ДНФ» для КМОП-схем имеет меньшую потребляемую мощность, чем варианты с КНФ, хотя логически все варианты эквивалентны. В то же время при «расчетверении» одного передающего транзистора по потребляемой мощности выигрывает вариант с КНФ (0,566 мкВт против 0,572 мкВт у ДНФ). По показателю энергии переключения — произведению средней задержки переключения на потребляемую мощность — выигрывает также вариант КНФ.

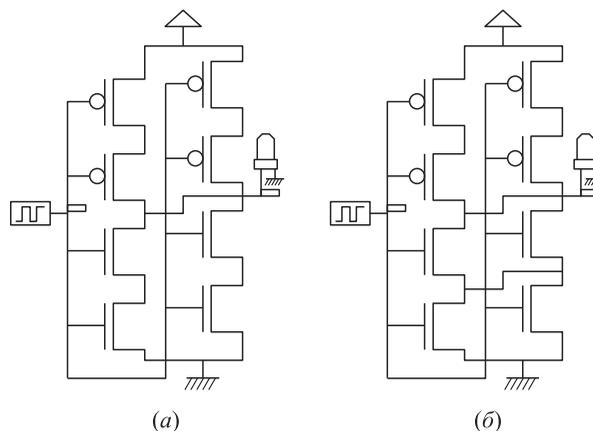


Рис. 5 Варианты резервирования инвертора по ДНФ (а) и по КНФ (б)

Для инвертора: ДНФ = 0,8001 мкВт·пс; КНФ = 0,7616 мкВт·пс. Для одного n-МОП-транзистора: КНФ = 68,6275 мкВт·пс; ДНФ = 73,502 мкВт·пс.

Предложим метод обеспечения сбое- и отказоустойчивости с учетом логико-топологических особенностей.

5 Метод обеспечения сбое- и отказоустойчивости с учетом логико-топологических особенностей

Основа предлагаемого метода — учет характеристик элементов, полученных при топологическом моделировании, и выбор предпочтительного варианта резервирования на транзисторном уровне при $r > 1$.

Дано: число транзисторов, подлежащих резервированию, v , нерезервированная функциональная электрическая схема элемента (устройства), требуемое число парируемых отказов в группе резервированного транзистора r .

Получить: предпочтительный вариант резервирования по (1) и (2).

Решение

1. По заданному r получаем эквивалентные по числу парируемых отказов (сбоев) логические выражения ДНФ: $f_v^r = \bigvee_{i=1}^{r+1} \bigwedge_{j=1}^{r+1} x_{j,i}$; КНФ: $f_v^r = \bigwedge_{i=1}^{r+1} \bigvee_{j=1}^{r+1} x_{j,i}$ и $f_{\text{комб}}^{r=\alpha+\beta}$; α — части ДНФ ($\alpha \geq 2$); β — части КНФ ($\beta = 1$).
2. По заданному v получаем множество комбинаций трех вариантов (третий вариант может содержать подварианты): $F : \{ \{ f_{v, \text{ДНФ}}^r \}, \{ f_{v, \text{КНФ}}^r \}, \{ f_{v, \text{комб}}^{r=\alpha+\beta} \} \}$.

Таблица 2 Параметры автоматически созданных различных вариантов топологий элемента НЕ и одного п-МОП-транзистора

№ п/п	Название варианта	Задержка по фронту, пс	Задержка по срезу, пс	Площадь, мкм ²	Потребляемая мощность, мкВт	Произведение мощности на задержку, мкВт·пс	Число транзисторов, шт.	Вероятность безотказной работы на $t = 1000$ ч, $\lambda = 10^{-6}$, $\alpha = 1,5$	Достоверность функционирования на $t = 1000$ ч	Примечание 1	Примечание 2
НЕ											
1	Нерезервированный	3,9	3,6	2,8	0,05	0,1875	2	0,93	0,93	Неотказоустойчивый	Одна линия передачи переменных
2	Дублирование	4,1	3,5	5,7	0,106	0,4028	4	0,87	0,99 при идеальности элемента XOR	Только повышение достоверности, активная отказоустойчивость	Две линии передачи переменных
3	«Расчетверение» по ДНФ	7,0	5,6	15,7	0,127	0,8001	8	0,98	0,98	Повышение ВБР, пассивная отказоустойчивость	Четыре линии передачи переменных
4	«Расчетверение» по КНФ	6,7	5,2	16,9	0,128	0,7614	8	0,98	0,98	Повышение ВБР, пассивная отказоустойчивость	Четыре линии передачи переменных
5	Тройное резервирование в мажоритаре	8,1	6,6	22,4	0,181	1,33035	10	0,93	0,93	Повышение ВБР, пассивная отказоустойчивость	Три линии передачи переменных
п-МОП-транзистор											
6	Один п-МОП-транзистор по КНФ	50,5	192	12,5	0,566	68,6275	4	0,9945	0,9945	Повышение ВБР, пассивная отказоустойчивость	Затворы объединены, на входе и выходе нерезервированные элементы НЕ
7	Один п-МОП-транзистор по ДНФ	51	206	11,6	0,572	73,502	4	0,9945	0,9945		

Замечания: напряжение источника питания 0,35 В; частота — 0,5 ГГц; модель BSim4; 32-нанометровая КМОП-технология.

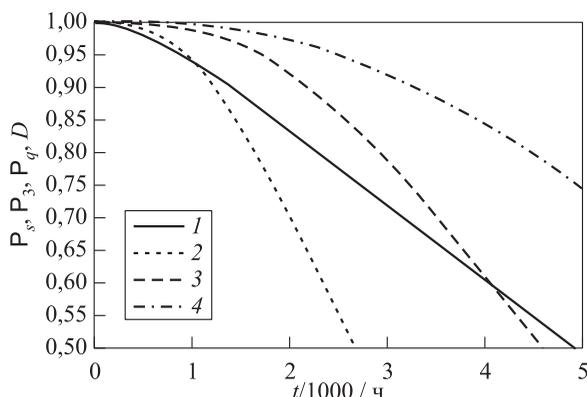


Рис. 6 Сравнение ВБР нерезервированного инвертора P_s (1), «расчетверенного» инвертора $P_q(t)$ (2), мажоритирования $P_3(t)$ (3) и достоверности контроля при дублировании $D(t)$ (4) в соответствии с распределением Вейбулла при $\lambda = 10^{-6}$ и $\alpha = 1,5$

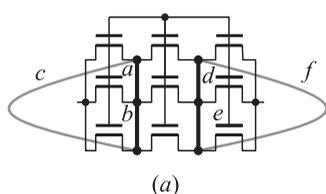


Рис. 7 Комбинирование соединений, исходя из КНФ, для варианта с кодом «11111» (а) и для варианта, описываемого кодом «100010» (б)

3. По полученному множеству F выполняем моделирование и определяем предпочтительный вариант схемы. Возможны разные структуры для одинаковых логических выражений.
4. Опционально оцениваем целесообразность разных r в пределах одного элемента (устройства). Оценки сложности (стоимости) в числе транзисторов и задержке переключения с учетом сложности блоков питания имеют вид:

$$L = [3L_{\geq n} + 3(L_{МЭ}) + 3(L_{БП})] + [3L_{\geq n ГМ} + 3k(L_{МЭ}) + 3k(L_{БП})] + [\chi(r + 1)^2];$$

$$T_{\geq n} = T + T_{МЭ}; \quad T_{\geq n ГМ} = T + kT_{МЭ};$$

$$T_{\chi_r TrR^*} = (r + 1)T,$$

где $L_{\geq n}$ — число транзисторов (сложность) для канального мажоритирования; $L_{БП}$ — сложность блока питания; $L_{МЭ}$ — сложность мажоритарного элемента; $L_{\geq n ГМ}$ — сложность глубокого мажоритирования (k слоев); $L_{\chi_r TrR^*} = \chi(r + 1)^2 + 2(L_{БП})$, $2(L_{БП})$ — сложность транзисторного резервирования, два блока питания $2(L_{БП})$ для КМОП-схем; χ_r — число транзисторов для парирования r отказов

(сбоев). «Чистые» варианты не содержат внутри других: $P_{\geq n} P_{\geq n ГМ} P_{\chi_r TrR^*}$. «Смешанные» варианты, например $P_{\geq n} [P_{\geq n ГМ} (P_{\chi_r TrR^*})]$, используют мажоритар с транзисторным резервированием: TrR^* . Расширенные комбинации $Q^*(r)$ учитывают дополнительные «смешанные» формы представления логических функций резервирования, возникающие, например, при варьировании соединений в КНФ. Рисунок 7 показывает два варианта для $r = 2$, описываемые кодами «11111» (рис. 3, а) и «100010» (рис. 3, б) в формате «abcdef». Единица означает наличие соответствующего соединения, ноль — отсутствие. Таким образом, получается $2^6 = 64$ варианта. Для $r = 3$ получаем 2^{12} вариантов. Легко видеть, что с учетом двух дополнительных вариантов для $r = 1$ (не принимая во внимание направление «петель») $Q^*(r) = 2^{r^2+r}$.

6 Заключение

Предложен новый подход к обеспечению сбое- и отказоустойчивости цифровых элементов и устройств, отличающийся тем, что используется как канальное и глубокое мажоритирование, так и топологические особенности транзисторного резервирования, причем предполагается как полное резервирование по данному варианту, так и комбинирование внутри одного варианта, например транзисторное резервирование внутри мажоритарного, обратное невозможно. Применяется также различная степень транзисторного резервирования в пределах одного устройства. Предложен метод формирования различных вариантов резервирования транзисторов, причем находятся возможные различные логические выражения транзисторного

резервирования (комбинаторные структуры композиций $Q_i(r+1)$), которые оцениваются по показателям потребляемой мощности, задержки переключения и площади схемы на кристалле путем топологического моделирования. Такое резервирование применяется для элементов, содержащих в наиболее длинной цепочке транзисторов не более чем d транзисторов, $d(r+1) \leq e$. Если же это неравенство не выполняется, то необходима декомпозиция исходной схемы для удовлетворения ограничению Мида–Конвей, после чего выполняется резервирование.

Метод может быть применен для обеспечения надежности цифровых устройств критического применения.

Литература

1. Sekar V. How foundries calculate die yield. <https://www.viksnewsletter.com/p/how-foundries-calculate-die-yield>.
2. Надежность и эффективность в технике: Справочник: в 10 т. — М.: Машиностроение, 1986–1990. 3092 с.
3. Hill I., Singh P.C.R., Sheikholeslam S.A., Ivanov A. CMOS reliability from past to future: A survey of requirements, trends, and prediction methods // IEEE T. Device Mat. Rel., 2022. Vol. 22. Iss. 1. P. 1–18. doi: 10.1109/TDMR.2021.3131345.
4. Rajpoot J., Gupta M., Verma S. Enhancing the reliability of hybrid MTJ/CMOS circuits with auto write termination // Symposium (International) on the Physical and Failure Analysis of Integrated Circuits Proceedings. — IEEE, 2023. Art. 10249113. 6 p. doi: 10.1109/IPFA58228.2023.10249113.
5. Ray A., Jaiswal A., Mishra G. P. Total Ionizing Dose (TID) analysis of FinFET-based CMOS NAND and NOR logic gates // Devices for integrated circuit. — IEEE, 2025. P. 723–726. doi: 10.1109/DevIC63749.2025. 11012226.
6. Cannon M.J., Keller A.M., Perez-Celis A., Wirthlin M.J. Modeling common cause failures in systems with triple modular redundancy and repair // Annual Reliability and Maintainability Symposium Proceedings. — IEEE, 2020. Art. 9153662. 6 p. doi: 10.1109/RAMS48030.2020.9153662.
7. Ismail M.M.C., Halim I.S.A., Hassan S.L.M., Ab Rahim A.A., Abdullah N.E. Fault tolerant design comparison study of TMR and 5MR // Symposium on Industrial Electronics & Applications Proceedings. — IEEE, 2021. Art. 9509996. 6 p. doi: 10.1109/ISIEA51897.2021. 9509996.
8. Ahmed S.A., Rani M.S.D. Decision tree based 5modular redundancy algorithm for fault identification in full adder circuit // Conference (International) on Recent Innovation in Smart and Sustainable Technology Proceedings. — IEEE, 2024. Art. 10922023. 5 p. doi: 10.1109/ICRISST59181.2024.10922023.
9. Saravanakumar C., Usha Bhanu N. Fault diagnosis of gate level 2-to-1 multiplexer in FinFET technology // Conference (International) on System, Computation, Automation and Networking Proceedings. — IEEE, 2021. Art. 9526525. 4 p. doi: 10.1109/ICSCAN53069. 2021.9526525.
10. Jamil M., Mukhopadhyay S., Ghoneim M., Shailos A., Prasad C., Meric I., Ramey S. Reliability studies on advanced FinFET transistors of the Intel 4 CMOS technology // Reliability Physics Symposium (International) Proceedings. — IEEE, 2023. Art. 10117992. 5 p. doi: 10.1109/IRPS48203.2023.10117992.
11. Huang Y.-Y., Huang P.-T., Lee P.-Y., Su P. Novel complementary FeFET-based lookup table and routing switch design and their applications in energy/area-efficient FPGA // 7th Electron Devices Technology & Manufacturing Conference Proceedings. — IEEE, 2023. Art. 10103081. 3 p. doi: 10.1109/EDTM55494.2023. 10103081.
12. Соколов И. А., Степченко Ю. А., Дьяченко Ю. Г., Рожественский Ю. В. Повышение сбоеустойчивости самосинхронных схем // Информатика и её применения, 2020. Т. 14. Вып. 4. С. 63–68. doi: 10.14357/19922264200409. EDN: SCPZVU.
13. Соколов И. А., Степченко Ю. А., Дьяченко Ю. Г., Рожественский Ю. В., Каменских А. Н. Базис реализации сбоеустойчивых электронных схем // Информатика и её применения, 2021. Т. 15. Вып. 4. С. 65–71. doi: 10.14357/19922264210409. EDN: PASSDD.
14. Соколов И. А., Степченко Ю. А., Дьяченко Ю. Г., Рожественский Ю. В. Оценка надежности синхронного и самосинхронного конвейеров // Информатика и её применения, 2022. Т. 16. Вып. 4. С. 2–7. doi: 10.14357/19922264220401. EDN: GWXJHM.
15. Ефанов Д. В., Погодина Т. С. Особенности обнаружения ошибок при контроле вычислений в цифровых устройствах по признаку самодвойственности булевых функций // Надежность, 2024. Т. 24. № 2. С. 24–37. doi: 10.21683/1729-2646-2024-24-2-24-37.
16. Kumawat H., Nagar P. On modified-Weibull distribution under left censoring // Conference (International) on Artificial Intelligence and Emerging Technology (Global AI Summit) Proceedings. — IEEE, 2024. P. 1348–1353. doi: 10.1109/GlobalAISummit62156.2024.10947937.
17. Avizienis A., Laprie J.-C. Dependable computing: From concepts to application // Proceedings IEEE, 1986. Vol. 74. Iss. 5. P. 629–638. doi: 10.1109/PROC.1986. 13527.
18. Сапожников В. В., Сапожников В. В., Ефанов Д. В. Основы теории надежности и технической диагностики. — СПб.: Лань, 2019. 588 с. EDN: IZBJPZ.
19. Remala V.N., Reddy A.A., Vidyadhar R.P., Bandi S.N. Circuit reliability through probabilistic transfer matrix //

- 6th Conference (International) on Communication and Electronics Systems Proceedings. — IEEE, 2021. P. 165–172. doi: 10.1109/ICCES51350.2021.9489243.
20. Carmichael C. Triple module redundancy design techniques for virtex FPGAs. Xilinx. 37 p. https://www.xilinx.com/support/documentation/application_notes/xapp197.pdf.
21. El-Maleh A. H., Al-Hashimi B. M., Melouki A., Al-Yamani A. Transistor-level based defect-tolerance for reliable nanoelectronics // Robust computing with nano-scale devices / Ed. C. Huang. — Lecture notes in electrical engineering ser. — Dordrecht: Springer, 2010. Vol. 58. P. 29–49. doi: 10.1007/978-90-481-8540-5_3.
22. Tyurin S. Hyper redundancy for super reliable FPGAs // Радиоелектронні і комп'ютерні системи (Харьков), 2021. № 1. С. 119–132. doi: 10.32620/reks.2021.1.11.
23. Mead C. A., Conway L. Introduction to VLSI systems. — 2nd ed. — Addison-Wesley, 1979. 394 p.
24. Skorniyakova A. Yu., Vikhorev R. V. Self-timed LUT layout simulation // Conference of Russian Young Researchers in Electrical and Electronic Engineering Proceedings. — IEEE, 2020. P. 176–179. doi: 10.1109/EIConRus49466.2020.9039374.

Поступила в редакцию 21.05.2025

Принята к публикации 15.08.2025

MULTIOPTION REDUNDANCY TAKING INTO ACCOUNT LOGICAL AND TOPOLOGICAL FEATURES OF TRANSISTOR CIRCUIT

S. F. Tyurin^{1,2}, M. S. Nikitin¹, Yu. A. Stepchenkov³, and Yu. G. Diachenko³

¹Perm National Research Polytechnic University, 7 Prof. Pozdeev Str., Perm 614013, Russian Federation

²Perm State University, 15 Bukireva Str., Perm 614990, Russian Federation

³Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Passive fault tolerance of digital cells and devices is considered using multioption reliability taking into account features of the transistor redundancy topological simulation. A model is built that includes channel majority redundancy with the majority voters redundancy, allowing for the channel “collapse” during diagnostics, deep redundancy with redundancy at the level of individual channel’s layers with special majority voters that ensure the configuration of layers into channels. The known methods are combined in a relationship that optimizes a given objective function with the required constraints. In addition, redundancy is used at the individual transistor level with varying degrees of failure protection. The topological features of such reservation are investigated by constructing various variants of circuits based on disjunctive normal, conjunctive normal, and intermediate forms. The power of the set of such variants is established. A method for searching for the topologically best variant with a large device dimension is proposed. By means of topological modeling, the preferred backup option is established based on the indicator of the consumed power and the switching delay product. Parameters examples of created topologies are given.

Keywords: fault tolerance; redundancy; majority voter; topological simulation

DOI: 10.14357/19922264250307

EDN: TXXPPO

References

1. Sekar, V. 2004. How foundries calculate die yield. Available at: <https://www.viksnewsletter.com/p/how-foundries-calculate-die-yield> (accessed August 29, 2025).
2. *Nadezhnost' i effektivnost' v tekhnike: Spravochnik* [Reliability and efficiency in engineering: Handbook]. 1986–1990. Moscow: Mashinostroenie. 10 vols. 3092 p.
3. Hill, I., P. C. R. Singh, S. A. Sheikholeslam, and A. Ivanov. 2022. CMOS reliability from past to future: A survey of requirements, trends, and prediction methods. *IEEE T. Device Mat. Re.* 22(1):1–18. doi: 10.1109/TDMR.2021.3131345.
4. Rajpoot, J., M. Gupta, and S. Verma. 2023. Enhancing the reliability of hybrid MTJ/CMOS circuits with auto write termination. *Symposium (International) on the Physical and Failure Analysis of Integrated Circuits Proceedings*. IEEE. Art. 10249113. 6 p. doi: 10.1109/IPFA58228.2023.10249113.
5. Ray, A., A. Jaiswal, and G. P. Mishra. 2025. Total ionizing dose (TID) analysis of FinFET-based CMOS NAND and NOR logic gates. *Devices for integrated circuit*. IEEE. 723–726. doi: 10.1109/DevIC63749.2025.11012226.
6. Cannon, M. J., A. M. Keller, A. Pèrez-Celis, and M. J. Wirthlin. 2020. Modeling common cause failures in systems with triple modular redundancy and repair. *Annu-*

- al Reliability and Maintainability Symposium Proceedings*. IEEE. Art. 9153662. 6 p. doi: 10.1109/RAMS48030.2020.9153662.
7. Ismail, M. M. C., I. S. A. Halim, S. L. M. Hassan, A. A. Ab Rahim, and N. E. Abdullah. 2021. Fault tolerant design comparison study of TMR and 5MR. *Symposium on Industrial Electronics & Applications Proceedings*. IEEE. Art. 9509996. 6 p. doi: 10.1109/ISIEA51897.2021.9509996.
 8. Ahmed, S. A., and M. S. D. Rani. 2024. Decision tree based 5modular redundancy algorithm for fault identification in full adder circuit. *Conference (International) on Recent Innovation in Smart and Sustainable Technology Proceedings*. IEEE. Art. 10922023. 5 p. doi: 10.1109/ICRISST59181.2024.10922023.
 9. Saravanakumar, C., and N. Usha Bhanu. 2021. Fault diagnosis of gate level 2-to-1 multiplexer in FinFET technology. *Conference (International) on System, Computation, Automation and Networking Proceedings*. IEEE. Art. 9526525. 4 p. doi: 10.1109/ICSCAN53069.2021.9526525.
 10. Jamil, M., S. Mukhopadhyay, M. Ghoneim, A. Shailos, C. Prasad, I. Meric, and S. Ramey. 2023. Reliability studies on advanced FinFET transistors of the Intel 4 CMOS technology. *Reliability Physics Symposium (International) Proceedings*. IEEE. Art. 10117992. 5 p. doi: 10.1109/IRPS48203.2023.10117992.
 11. Huang, Y.-Y., P.-T. Huang, P.-Y. Lee, and P. Su. 2023. Novel complementary FeFET-based lookup table and routing switch design and their applications in energy/area-efficient FPGA. *7th Electron Devices Technology & Manufacturing Conference Proceedings*. IEEE. Art. 10103081. 3 p. doi: 10.1109/EDTM55494.2023.10103081.
 12. Sokolov, I. A., Yu. A. Stepchenkov, Yu. G. Diachenko, and Yu. V. Rogdestvenski. 2020. Povyshenie sboustoychivosti samosinkhronnykh skhem [Improvement of self-timed circuit soft error tolerance]. *Informatika i ee Primeneniya — Inform. Appl.* 14(4):63–68. doi: 10.14357/19922264200409. EDN: SCPZVU.
 13. Sokolov, I. A., Yu. A. Stepchenkov, Yu. G. Diachenko, Yu. V. Rogdestvenski, and A. N. Kamenskiy. 2021. Baza realizatsii sboustoychivyykh elektronnykh skhem [The electronic component base of failure resilience digital circuits]. *Informatika i ee Primeneniya — Inform. Appl.* 15(4):65–71. doi: 10.14357/19922264210409. EDN: PASSDD.
 14. Sokolov, I. A., Yu. A. Stepchenkov, Yu. G. Diachenko, and Yu. V. Rogdestvenski. 2022. Otsenka nadezhnosti sinkhronnogo i samosinkhronnogo konveyerov [Synchronous and self-timed pipeline's reliability estimation]. *Informatika i ee Primeneniya — Inform. Appl.* 16(4):2–7. doi: 10.14357/19922264220401. EDN: GWXJHM.
 15. Efanov, D. V., and T. S. Pogodina. 2024. Osobennosti obnaruzheniya oshibok pri kontrole vychisleniy v tsifrovyykh ustroystvakh po priznaku samodvoystvennosti bulevyykh funktsiy [The specificity of error detection as part of computing testing in digital devices based on self-duality of Boolean functions]. *Nadezhnost' [Dependability]* 24(2):24–37. doi: 10.21683/1729-2646-2024-24-2-24-37. EDN: TGVARR.
 16. Kumawat, H., and P. Nagar. 2024. On modified-Weibull distribution under left censoring. *Conference (International) on Artificial Intelligence and Emerging Technology (Global AI Summit) Proceedings*. IEEE. 1348–1353. doi: 10.1109/GlobalAISummit62156.2024.10947937.
 17. Avizienis, A., and J.-C. Laprie. 1986. Dependable computing: From concepts to application. *Proceedings IEEE* 74(5):629–638. doi: 10.1109/PROC.1986.13527.
 18. Sapozhnikov, V. V., V. V. Sapozhnikov, and D. V. Efanov. 2019. *Osnovy teorii nadezhnosti i tekhnicheskoy diagnostiki* [Fundamentals of reliability theory and technical diagnostics]. St. Petersburg: Lan'. 588 p. EDN: IZBJPZ.
 19. Remala, V. N., A. A. Reddy, R. P. Vidyadhar, and S. N. Bandi. 2021. Circuit reliability through probabilistic transfer matrix. *6th Conference (International) on Communication and Electronics Systems Proceedings*. IEEE. 165–172. doi: 10.1109/ICCES51350.2021.9489243.
 20. Carmichael, C. 2006. Triple module redundancy design techniques for virtex FPGAs. Xilinx. 37 p. Available at: https://www.xilinx.com/support/documentation/application_notes/xapp197.pdf (accessed August 29, 2025).
 21. El-Maleh, A. H., B. M. Al-Hashimi, A. Melouki, and A. Al-Yamani. 2010. Transistor-level based defect-tolerance for reliable nanoelectronics. *Robust computing with nano-scale devices*. Ed. C. Huang. Lecture notes in electrical engineering ser. Dordrecht: Springer. 58:29–49. doi: 10.1007/978-90-481-8540-5_3.
 22. Tyurin, S. 2021. Hyper redundancy for super reliable FPGAs. *Radioelektronika i kompyuterni sistemi* [Radioelectronic and Computer Systems] 1:119–132. doi: 10.32620/reks.2021.1.11.
 23. Mead, C. A., and L. Conway. 1979. *Introduction to VLSI systems*. 2nd ed. Addison-Wesley. 396 p.
 24. Skornyakova, A. Yu., and R. V. Vikhorev. 2020. Self-timed LUT layout simulation. *Conference of Russian Young Researchers in Electrical and Electronic Engineering Proceedings*. IEEE. 176–179. doi: 10.1109/EIConRus49466.2020.9039374.

Received May 21, 2025

Accepted August 15, 2025

Contributors

Tyurin Sergey F. (b. 1953) — Doctor of Science in technology, honored inventor of RF, professor, Department of Automation and Telemechanics, Perm National Research Polytechnic University, 7 Prof. Pozdeev Str., Perm

614013, Russian Federation; professor, Department of Software Computing Systems, Perm State University, 15 Bukireva Str., Perm 614990, Russian Federation; tyurinsergfeo@yandex.ru

Nikitin Maxim S. (b. 1998) — PhD student, Department of Automation and Telemecanic, Perm National Research Polytechnic University, 7 Prof. Pozdeev Str., Perm 614013, Russian Federation; mann1k@yandex.ru

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

КЛАССИФИКАЦИЯ МАЛЫХ НАБОРОВ ДАННЫХ БОЛЬШОЙ РАЗМЕРНОСТИ

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

Аннотация: Рассматривается задача классификации данных очень большой размерности, при этом используется только ограниченный набор обучающих образцов таких данных. В этих условиях проверяется возможность использования причинно-следственных связей в решении классификационных задач указанного типа. Решение задач основано на существовании причинно-следственных связей неизвестных причин с наблюдаемыми частично детерминированными следствиями этих причин в поступающих новых данных. Использовано обучение на малых наборах данных. Задачи решаются в условиях, когда размер данных и число возможных свойств данных стремятся к бесконечности. Найдены асимптотические условия однозначной классификации новых данных. В частном случае исследована задача классификации при наличии случайных искажений детерминированных следствий в данных. Сформулированы условия возможности обучения без учителя. Работа показывает принципиальные возможности применения причинно-следственных связей в задачах медицинской диагностики, выявления мошеннических схем в финансовой сфере и оценки ситуационной осведомленности в кибербезопасности.

Ключевые слова: классификация данных большой размерности; искусственный интеллект; причинно-следственные связи

DOI: 10.14357/19922264250308

EDN: SUSRRC

1 Введение

Использование причинно-следственных связей в задачах классификации исследовалось во многих работах [1–4]. В данной статье рассматривается задача классификации данных очень большой размерности, при этом используется только ограниченный набор обучающих образцов таких данных. В этих условиях проверяется возможность использования причинно-следственных связей в решении классификационных задач указанного типа.

В тех случаях, когда причины известны, такой метод называется контрфактуальным [5]. Когда причины неизвестны, используется обучение на прецедентах, имеющих одну и ту же причину.

Приведем примеры таких практических задач с малыми наборами данных большой размерности.

Пример 1. В медицинской диагностике сложных болезней, например диагностике онкологических заболеваний, врачам предоставляют большой объем результатов различных анализов, результаты компьютерной томографии, магнитно-резонансной томографии, рентгеновских снимков, биопсии. Вместе с тем информативные значения параметров собираются по ограниченному множеству прецедентов, которые успешно выявлены ранее для конкретных форм онкологии [6, 7].

Пример 2. Для выявления мошеннических схем в банковской сфере (создание фиктивных документов) для анализа предоставляется большое число документов, а решения принимаются часто методами, такими как модель Бенейша [8] и закон Бенфорда [9] по значениям ограниченного числа параметров.

Эти модели не исчерпывают подходы в такого рода задачах и, например, широко применяются методы искусственного интеллекта [10].

Пример 3. Для формирования решения одобрить или отказать в предоставлении кредита выстраивается процесс анализа значимого объема параметров, характеризующих клиента: внутренние данные банка, данные бюро кредитных историй (БКИ), данные телеком-операторов, соцсетей, госорганов и пр. Число рассматриваемых свойств может достигать до двух–трех сотен, а сам процесс принятия решения представляет собой ориентированный граф. Как частный случай — выявление потенциально мошеннических случаев на «быстрых» данных БКИ, т. е. данных, получаемых на коротком интервале времени до подачи заявки. Для построения набора правил использован случайный лес, который позволил сформировать на множестве из 130+ свойств набор правил (5–13 свойств),

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

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

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

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

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

которые решают задачу min-max: минимизация объема отсекаемых выдач при максимизации доли фрод/просроченных кредитов (NPL — nonperforming loans). Таким образом, проблема кредитования сводится к исследованиям большого объема данных о клиенте, но подозрения на возможность невозврата кредита клиентом возникают при рассмотрении малого числа причин, которые могут привести к неплатежам [11].

Пример 4. Действия инсайдеров по нанесению ущерба предприятию обнаруживаются по типовым сценариям, выявленным на ограниченном множестве проведенных инсайдером операций по нанесению ущерба [12].

Пример 5. Для анализа ситуационной осведомленности в задачах кибербезопасности осуществляется сбор большого количества неоднородных данных. При этом оценка текущей ситуации определяется ограниченным числом причин [13].

2 Математическая модель и постановки задач

Обозначим через R конечное множество свойств, $|R| = M$. Обозначим через D множество данных, которые могут поступать для классификации. Элементами множества D служат векторы x одинаковой длины N , состоящие из свойств множества R .

Структура поступающих данных устроена двумя способами.

Рассмотрим первый способ построения данных. Предположим, что существует первопричина C , детерминированно порождающая следствия, т. е. первопричина C порождает множество свойств, которые являются следствиями C и могут располагаться в различных информационных пространствах [14]. Эти следствия сами могут быть причинами для множества других следствий. Если представить порождения следствий ориентированным графом, то получим DAG (Directed Acyclic Graph). Множество наблюдаемых конечных вершин этого графа имеет размер k , $k < N$, и представляет собой множество свойств, которое образует детерминированную часть наблюдаемых данных x , хотя порядок детерминированно появляющихся следствий в x может быть произвольным. Для того чтобы конечные следствия наблюдались в данных, необходима связь по передаче информации в наблюдаемые данные, т. е. не все конечные точки дерева, порожденного причиной C , появляются в наблюдаемых данных. Остальная часть данных x появляется случайно в соответствии с равномерной полиномиальной схемой I на множестве R .

Второй способ порождения данных состоит в том, что причины C нет и данные представляют собой случайную последовательность по схеме I .

Рассмотрим сначала простую постановку задачи классификации, состоящую в том, что поступившие данные x надо классифицировать на предмет наличия в них свойств, порожденных C . В простейшем случае причина C неизвестна, но предполагается, что она единственная. Сложность задачи в этом случае связана с тем, что N велико. С другой стороны, есть m образцов данных, в которых известно присутствие следствий, порожденных C , и есть несколько образцов, про которые известно, что в них нет следствий, детерминированно порожденных C .

Возможны разные подходы к решению задачи классификации. Существует много вариантов использования методов математической статистики. Так, большая размерность поступивших данных может быть сокращена, например, методом случайных деревьев и методом главных компонент. Имеющиеся образцы можно использовать для обоснования близости редуцированных и исходных данных, а затем применить критерий согласия новых данных после их редукции с полученными ранее данными.

В данной работе предлагается способ, основанный на причинно-следственных связях, который состоит в том, чтобы выбрать в обучающей выборке размера m множество свойств, возникающих с большой вероятностью из-за порождения их причиной C . Затем необходимо отыскать эти свойства в новых данных x . Найдем условия применимости этого метода в рассматриваемой задаче классификации. Пусть выбраны свойства, которые присутствуют одновременно во всех m множествах свойств обучающей выборки. Часть этих свойств представляет собой следствия C или потомков C , а часть свойств возникает случайно в соответствии с полиномиальной схемой.

Пусть детерминированные свойства, порожденные C , это — Y_1, \dots, Y_k . Рассмотрим множество свойств в новых данных x . Если в данных x имеются детерминированные следствия C , а случайные свойства, порождающие такой набор, маловероятны, то в x присутствует набор Y_1, \dots, Y_k . Отсюда следует, что данные x принадлежат к классу, в котором присутствуют следствия, порожденные C .

Оценим вероятность того, что в обучающем наборе образцов кроме детерминированных общих свойств Y_1, \dots, Y_k содержатся случайные свойства, не являющиеся следствием первопричины C , которые попали в множество пересечений обучающей выборки. Предположим, что в обучающей выборке

ке из случайно выбранных $N - k$ свойств, добавленных к Y_1, \dots, Y_k (повторы маловероятны и ими можно пренебречь), есть общие свойства в каждом из образцов обучающей выборки. Возьмем первое случайное свойство из $N - k$ первого образца обучающей выборки. Вероятность отсутствия этого свойства во втором образце обучающей выборки равна $(1 - 1/(M - k))^{N-k}$. Тогда вероятность присутствия первого свойства во втором образце равна $1 - (1 - 1/(M - k))^{N-k}$.

Поскольку все остальные случайные добавления одинаково распределены и независимы друг от друга, то вероятность присутствия первого свойства первого образца во всех образцах обучающей выборки равна $(1 - (1 - 1/(M - k))^{N-k})^{m-1}$. Отсюда следует, что математическое ожидание числа случайных добавок к Y_1, \dots, Y_k равно $(N - k)(1 - (1 - 1/(M - k))^{N-k})^{m-1}$.

Пусть N и M стремятся к бесконечности так, что $N^m/M^{m-1} \rightarrow 0$. Тогда по неравенству Чебышёва вероятность хотя бы одного случайного добавления к Y_1, \dots, Y_k стремится к 0.

Оценим вероятность того, что в целиком случайной последовательности может возникнуть набор свойств Y_1, \dots, Y_k , позволяющий отнести эту последовательность к данным, имеющим следствия, порожденные C .

Рассмотрим событие A_1 , состоящее в том, что Y_1 присутствует в случайном векторе длины N . Очевидно, вероятность этого события равна $1 - (1 - 1/M)^N$. Тогда вероятность присутствия всего набора Y_1, \dots, Y_k равна

$$P\left(\bigcap_{i=1}^k A_i\right) < P(A_1) = 1 - \left(1 - \frac{1}{M}\right)^N.$$

Если при этом N и M стремятся к бесконечности так, что $N/M \rightarrow 0$, то

$$P\left(\bigcap_{i=1}^k A_i\right) \rightarrow 0.$$

Таким образом, доказана следующая теорема.

Теорема. Если $N, M \rightarrow \infty$ таким образом, что выполняется условие $N^m/M^{m-1} \rightarrow 0$, то с вероятностью, стремящейся к единице, для отнесения x к данным, имеющим следствия, порожденные C , необходимо и достаточно появление в новых данных x вектора свойств Y_1, \dots, Y_k .

Доказательство следует из приведенных выше оценок и соотношения

$$\frac{N}{M} < \frac{N^m}{M^{m-1}} \rightarrow 0.$$

Отметим, что если порождающий класс содержит одновременно две причины C_1 и C_2 , то классификация данных x , порожденных сразу двумя причинами C_1 и C_2 при альтернативе полностью случайных последовательностей, решается так же, как в теореме.

3 Классификация в условиях помех

Введем следующую модель помех. Пусть в данных, порожденных C , в векторе свойств Y_1, \dots, Y_k с вероятностью p исчезает ровно одно случайно выбранное свойство, которое заменяется на случайное свойство по полиномиальной схеме I , и с вероятностью $1 - p$ искажений в Y_1, \dots, Y_k нет. Такие замены могут происходить независимо друг от друга на любых данных, порожденных C . Эти ошибки могут происходить также на обучающей части данных. Определим условия, когда задача классификации решается с той же точностью, как и без помех.

На этапе построения детерминированных следствий в обучающей выборке построим следующую матрицу размера $m \times t$. В каждом столбце будем выписывать наиболее повторяющиеся свойства, встретившиеся одновременно в образцах обучающей выборки. Если в образцах обучающей выборки есть свойства, встретившиеся во всех образцах, то их выпишем в первых столбцах. Если нет ни одного такого свойства, то рассмотрим наборы по $m - 1$ образцов обучающей выборки и построим матрицы размеров $(m - 1) \times t$ и т.д. Число шагов следует измерять с вероятностью случайного появления повторений одинаковых свойств в случайных последовательностях обучающей выборки. Если эта вероятность становится значимо большой, то при таких p задача неразрешима. Будем считать, что найдется матрица с числом строк $m - 1$, в которой хоть один столбец заполнен полностью. Тогда при условии $N^{m-1}/M^{m-2} \rightarrow 0$ можно считать, что в первом столбце определены $m - 1$ свойств и в одном известном месте произошел сбой (это место определено отсутствующим образцом обучающей выборки). Тогда вместо отличающегося свойства соответствующего образца обучающей выборки на этом месте можно поставить свойство, совпадающее с остальными элементами столбца, и перейти к восстановлению второго столбца. Заметим, что, когда установлено первое свойство, тогда второе свойство детерминированного набора следствий C устанавливается независимо от первого восстановленного свойства. Поэтому столбцы с разными местами ошибок можно объединять, учитывая исходный порядок строк, соответствующих образцам

обучающей выборки. Самый худший вариант состоит в том, что все искажения сосредоточены на одном свойстве. В этом случае одно свойство будет потеряно в наборе Y_1, \dots, Y_k , что при $k > 2$ по-прежнему позволяет решать задачу классификации.

В остальных случаях решение задачи восстановления детерминированных свойств в обучающей выборке возможно до тех пор, пока случайное появление встречаемости случайных добавлений свойств к детерминированным свойствам обучающей выборки стремится к 0 при N и M , стремящихся к бесконечности. Если p достаточно мало, т. е. искажений мало, то присутствие случайно одинаковых свойств в большом числе образцов обучающей выборки будет маловероятно, что означает получение оценки k . Отсюда следует, что при малых p набор свойств Y_1, \dots, Y_k с вероятностью, стремящейся к 1, определяется однозначно.

Рассмотрим возможность классификации x при возможных искажениях поступивших данных. Как следует из доказательства теоремы, в ее условиях вероятность ложной классификации случайной последовательности стремится к 0, даже когда восстановлено только одно детерминированное свойство.

4 Классификация данных при наличии нескольких порождающих причин

В рамках построенной модели рассмотрим задачу построения обучающей выборки, когда известны несколько причин C_1, \dots, C_s . Сложность задачи заключается в том, что некоторые классы могут появляться только в объединении нескольких причин в данных. Тогда либо классы неоднозначно классифицируются, либо определяются классы с детерминированными свойствами, порожденными несколькими причинами одновременно. Задача выявления невозможности однозначной классификации определяется на ансамбле классификаторов, соответствующих каждой отдельной причине. В самом деле, причина, попавшая в пару и присутствующая в данных отдельно при возможности искажений, будет определять один класс при автономном появлении и другой класс, в котором она порождает следствия одновременно с другой причиной.

Наиболее удобный способ исключения противоречий состоит в том, чтобы нумеровать классы множествами выявленных причин. Если в анализе используется ансамбль классификаторов на каждую причину, то результаты обработки данных x на

всех классификаторах сводят задачу сложной классификации к вычислениям логических выражений над полем из двух элементов.

5 Создание классов в условиях отсутствия учителя

Основная проблема в составлении обучающей выборки, когда есть одна причина C , состоит в доле данных, которые содержат свойства, порожденные C . Если доля данных в D , имеющих следствия, порожденные C , велика, например не меньше некоторого положительного ε , то набрать нужное для определения устойчивого пересечения множества образцов можно различными способами. Например, если имеются всего t данных, то ожидаемое число данных, имеющих детерминированные свойства, порожденные C , равно $t\varepsilon$. Тогда если найдется не менее двух образцов из t , имеющих общее пересечение, то их пересечение уже позволит искать претендентов на принадлежность к классу с порожденными C свойствами. Необходимые для более достоверного решения m обучающих образцов могут быть найдены, если при выполнении условий теоремы имеется достаточно большой размер данных.

6 Заключение

В работе рассмотрены практически важные задачи классификации ансамблями классификаторов в условиях малого набора данных большой размерности.

Решение задач основано на существовании причинно-следственных связей неизвестных причин C_1, \dots, C_s с наблюдаемыми частично-детерминированными следствиями этих причин в поступающих новых данных.

Использовано обучение на малых наборах данных. Задачи решаются в условиях, когда размер данных N и числа возможных свойств данных стремятся к бесконечности.

Найдены асимптотические условия однозначной классификации новых данных. В частном случае исследована задача классификации при наличии случайных искажений детерминированных следствий в данных. Сформулированы условия возможности обучения без учителя.

Работа показывает принципиальные возможности применения причинно-следственных связей в задачах медицинской диагностики, выявления мошеннических схем в финансовой сфере и оценки ситуационной осведомленности в кибербезопасности.

Литература

1. Грушо А. А., Грушо Н. А., Забежайло М. И., Зацаринский А. А., Тимонина Е. Е., Шоргин С. Я. Анализ цепочек причинно-следственных связей // Информатика и её применения, 2022. Т. 16. Вып. 2. С. 68–74. doi: 10.14357/19922264220209. EDN: HPSLTU.
2. Грушо А. А., Грушо Н. А., Забежайло М. И., Кульченков В. В., Тимонина Е. Е., Шоргин С. Я. Причинно-следственные связи в задачах классификации // Информатика и её применения, 2023. Т. 17. Вып. 1. С. 43–49. doi: 10.14357/19922264230106. EDN: DTQZPK.
3. Грушо А. А., Грушо Н. А., Забежайло М. И., Тимонина Е. Е., Шоргин С. Я. Сложные причинно-следственные связи // Информатика и её применения, 2023. Т. 17. Вып. 2. С. 84–89. doi: 10.14357/19922264230212. EDN: TGXQIW.
4. Грушо А. А., Грушо Н. А., Забежайло М. И., Смирнов Д. В., Тимонина Е. Е. Классификация с помощью причинно-следственных связей // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 71–75. doi: 10.14357/19922264230310. EDN: AKWBZD.
5. Höfler M. Causal inference based on counterfactuals // BMC Med. Res. Methodol., 2005. Vol. 5. Art. 28. 12 p. doi: 10.1186/1471-2288-5-28.
6. Забежайло М. И., Трунин Ю. Ю. К проблеме доказательности медицинского диагноза: интеллектуальный анализ данных о пациентах в выборках ограниченного размера // Научно-техническая информация. Сер. 2, 2019. № 12. С. 12–18.
7. Грушо А. А., Забежайло М. И., Тимонина Е. Е. О каузальной репрезентативности обучающих выборок прецедентов в задачах диагностического типа // Информатика и её применения, 2020. Т. 14. Вып. 1. С. 80–86. doi: 10.14357/19922264200111. EDN: LAQCMA.
8. Beneish M. D. The detection of earnings manipulation // *Financ. Anal. J.*, 1999. Vol. 55. No. 5. P. 24–36. doi: 10.2469/faj.v55.n5.2296.
9. Benford F. The law of anomalous numbers // *P. Am. Philos. Soc.*, 1938. P. 551–572.
10. Zhang Ch., Gill A., Liu B., Anwar M. J. AI-based identity fraud detection: A systematic review. — Cornell University, 2025. arXiv:2501.09239v1 [cs.AI]. 31 p.
11. Akinjole A., Shobayo O., Popoola J., Okoyeigbo O., Ogunl-eye B. Ensemble-based machine learning algorithm for loan default risk prediction // *Mathematics*, 2024. Vol. 12. No. 21. Art. 3423. 31 p. doi: 10.3390/math12213423.
12. Грушо А. А., Забежайло М. И., Смирнов Д. В., Тимонина Е. Е. Модель множества информационных пространств в задаче поиска инсайдера // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 65–69. doi: 10.14357/19922264170408. EDN: ZXWUOP.
13. Li Zheng. Overview of cyber security situation awareness // *Applied Computational Engineering*, 2024. Vol. 30. P. 149–154. doi: 10.54254/2755-2721/30/20230089.
14. Grusho A., Grusho N., Timonina E. Method of several information spaces for identification of anomalies // *Intelligent distributed computing XIII / Eds. I. Kotenko, C. Badica, V. Desnitsky, D. El Baz, M. Ivanovic.* — *Studies in computational intelligence ser.* — Cham: Springer, 2020. Vol. 868. P. 515–520. doi: 10.1007/978-3-030-32258-8_60.

Поступила в редакцию 19.05.2025

Принята к публикации 15.08.2025

CLASSIFICATION OF SMALL SETS OF DATA OF LARGE DIMENSION

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

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

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

Abstract: The problem of classifying of data of very large dimension is considered, while only a limited set of training samples of such data is used. Under these conditions, the possibility of using cause-and-effect relationships in solving classification problems of the specified type is checked. Problem solving is based on the existence of cause-and-effect relationships of unknown causes with the observed partially determined effects of these causes in incoming new data. Training on small set of data is used. The problems are solved in conditions when the size of the data and the number of possible data properties tend to infinity. Asymptotic conditions for unambiguous classification of new data were found. In a particular case, the classification problem was investigated in the presence of random distortions of deterministic effects in the data. The conditions for the possibility of training without a teacher are formulated. The work shows the fundamental possibilities of applying cause-and-effect relationships in the tasks of medical diagnostics, identifying fraudulent schemes in the financial sector, and assessing situational awareness in cybersecurity.

Keywords: classification of data of large dimension; artificial intelligence; cause-and-effect relationships

DOI: 10.14357/19922264250308

EDN: SUSRRC

References

1. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, A. A. Zatsarinny, E. E. Timonina, and S. Ya. Shorgin. 2022. Analiz tsepochek prichinno-sledstvennykh svyazey [Cause-and-effect chain analysis]. *Informatika i ee Primeneniya — Inform. Appl.* 16(2):68–74. doi: 10.14357/19922264220209. EDN: HPSLTU.
2. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, V. V. Kulchenkov, E. E. Timonina, and S. Ya. Shorgin. 2023. Prichinno-sledstvennye svyazi v zadachakh klassifikatsii [Causal relationships in classification problems]. *Informatika i ee Primeneniya — Inform. Appl.* 17(1):43–49. doi: 10.14357/19922264230106. EDN: DTQZPK.
3. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, E. E. Timonina, and S. Ya. Shorgin. 2023. Slozhnye prichinno-sledstvennye svyazi [Complex cause-and-effect relationships]. *Informatika i ee Primeneniya — Inform. Appl.* 17(2):84–89. doi: 10.14357/19922264230212. EDN: TGXQIW.
4. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2023. Klassifikatsiya s pomoshch'yu prichinno-sledstvennykh svyazey [Classification by cause-and-effect relationships]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):71–75. doi: 10.14357/19922264230310. EDN: AKWBZD.
5. Höfler, M. 2005. Causal inference based on counterfactuals. *BMC Med. Res. Methodol.* 5:28. 12 p. doi: 10.1186/1471-2288-5-28.
6. Zabezhailo, M. I., and Y. Y. Trunin. 2019. On the problem of medical diagnostic evidence: Intelligent analysis of empirical data on patients in samples of limited size. *Automatic Documentation Mathematical Linguistics* 53:322–328. doi: 10.3103/S0005105519060086.
7. Grusho, A. A., M. I. Zabezhailo, and E. E. Timonina. 2020. O kauzal'noy reprezentativnosti obuchayushchikh vyborok pretsedentov v zadachakh diagnosticheskogo tipa [On causal representativeness of training samples of precedents in diagnostic type tasks]. *Informatika i ee Primeneniya — Inform. Appl.* 14(1):80–86. doi: 10.14357/19922264200111. EDN: LAQCMA.
8. Beneish, M. D. 1999. The detection of earnings manipulation. *Financ. Anal. J.* 55(5):24–36. doi: 10.2469/faj.v55.n5.2296.
9. Benford, F. 1938. The law of anomalous numbers. *P. Am. Philos. Soc.* 78(4):551–572.
10. Zhang, Ch., A. Gill, B. Liu, and M. J. Anwar. 2025. AI-based identity fraud detection: A systematic review. Cornell University. 31 p. Available at: <https://arxiv.org/pdf/2501.09239v1> (accessed August 24, 2025).
11. Akinjole, A., O. Shobayo, J. Popoola, O. Okoyeigbo, and B. Ogunleye. 2024. Ensemble-based machine learning algorithm for loan default risk prediction. *Mathematics* 12(21):3423. 31 p. doi: 10.3390/math12213423.
12. Grusho, A. A., M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2017. Model' mnozhestva informatsionnykh prostranstv v zadache poiska insaydera [The model of the set of information spaces in the problem of insider detection]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):65–69. doi: 10.14357/19922264170408. EDN: ZXWUOP.
13. Li, Zheng. 2024. Overview of cyber security situation awareness. *Applied Computational Engineering* 30:149–154. doi: 10.54254/2755-2721/30/20230089.
14. Grusho, A., N. Grusho, and E. Timonina. 2020. Method of several information spaces for identification of anomalies. *Intelligent distributed computing XIII*. Eds. I. Kotenko, C. Badica, V. Desnitsky, D. El Baz, and M. Ivanovic. Studies in computational intelligence ser. Cham: Springer. 868:515–520. doi: 10.1007/978-3-030-32258-8_60.

Received May 19, 2025

Accepted August 15, 2025

Contributors

Grusho Alexander A. (b. 1946) — 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; grusho@yandex.ru

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

Zabezhailo Michael I. (b. 1956) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; m.zabezhailo@yandex.ru

Kulchenkov Vladislav V. (b. 1989) — deputy head, Portfolio Analysis Department, VTB Bank, 43-1 Vorontsovskaya Str., Moscow 109147, Russian Federation; vlad.kulchenkov@gmail.com

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

АВТОМАТИЗАЦИЯ РАЗМЕТКИ ИМПЛИЦИТНЫХ ЛОГИКО-СЕМАНТИЧЕСКИХ ОТНОШЕНИЙ: ВОЗМОЖНОСТИ И ОГРАНИЧЕНИЯ*

А. А. Гончаров¹, П. В. Ярошенко²

Аннотация: Описаны основные проблемы, возникающие в ходе автоматизации разметки имплицитных логико-семантических отношений (ЛСО), проанализированы причины этих проблем, а также предложены подходы к их решению. Последовательно рассмотрены основные этапы этого процесса: (1) поиск примеров с имплицитными ЛСО; (2) выявление границ аргументов отношения; (3) выбор признаков размечаемых фрагментов. Проанализированы результаты применения метода поиска с исключением в параллельных текстах и отмечены ограничения этого метода. Рассмотрены два фактора, затрудняющих автоматизацию разметки аргументов ЛСО: длина аргументов может сильно варьироваться и аргумент не обязательно включает последовательность соседних токенов, а может прерываться. Детально проанализированы пути автоматизации выбора признаков размечаемых фрагментов. Продемонстрировано, что даже обработка формальных признаков может требовать привлечения экспертов. Что же касается семантических признаков, то среди них есть как признаки, обработка которых может быть частично автоматизирована, так и те, которые на данном этапе могут размечаться только вручную. Выводы иллюстрируются примерами из корпуса.

Ключевые слова: лингвистическое аннотирование; дискурсивные отношения; логико-семантические отношения; имплицитность; параллельные тексты

DOI: 10.14357/19922264250309

EDN: NKEJQY

1 Введение

В настоящее время значительная часть задач, связанных с обработкой естественного языка, с неизбежностью требует привлечения материала лингвистических корпусов. В зависимости от специфики задачи тексты корпуса должны снабжаться необходимой разметкой. Особую актуальность представляет собой создание дискурсивно аннотированных корпусов. Эти корпуса снабжены не только морфологической (формальные характеристики вхождений отдельных слов) и, возможно, синтаксической (формальные характеристики предложений) разметкой, но и разметкой логико-семантической структуры текста (семантика связей между частями текста, в том числе более протяженными, чем предложение).

Определение, с одной стороны, факта наличия логической связи между фрагментами текста, а с другой стороны, семантики этой связи может способствовать улучшению результатов во многих приложениях компьютерной лингвистики: так, при составлении реферата на основе исходного доку-

мента (summarization) фактор связности порожденного текста становится одним из ключевых [1]. Кроме того, в настоящее время приобрело актуальность направление, в рамках которого рассматриваются возможности оценивать текстовую связность с помощью больших языковых моделей (см., например, [2]).

Важную роль играет наличие достаточного количества размеченных данных. Значительная часть имеющихся ресурсов англоязычная (речь прежде всего о Пенсильванском дискурсивно аннотированном корпусе [3, 4]), в то время как для других языков, в частности для русского, размеченных данных не хватает [5].

Логико-семантические отношения — это отношения, которые устанавливаются между отрезками текста³ и описывают, как именно эти отрезки связаны друг с другом по смыслу. В англоязычной литературе для их обозначения по большей части используется термин *discourse relations* — дискурсивные отношения. В данной статье вслед за [6] используется термин *логико-семантические отношения*. Распознавание ЛСО (*discourse relation recog-*

* Исследование выполнено за счет гранта Российского научного фонда № 24-28-00527, <https://rscf.ru/project/24-28-00527/>.

¹ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, a.gonch48@gmail.com

² Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Научно-исследовательский вычислительный центр МГУ им. М. В. Ломоносова, polina.iaroshenko@yandex.ru

³ Как правило, имеющими пропозициональное содержание.

tion) — одна из актуальных задач в области обработки естественного языка¹.

С точки зрения маркированности ЛСО могут быть эксплицитными — выраженными при помощи языкового средства (коннектора или иного показателя, явно указывающего на семантику отношения, см. [10]), и имплицитными — не выраженными ни одним из таких средств². Особенно остро проблема автоматизации обработки стоит для имплицитных ЛСО, тогда как при обработке эксплицитных ЛСО можно ориентироваться на наличие их показателей (например, *потому что* — коннектор, выражающий отношение причины), имплицитные ЛСО требуют более сложного подхода [11–13].

Цель статьи — описать основные проблемы, возникающие в ходе автоматизации разметки имплицитных ЛСО, проанализировать причины этих проблем, а также предложить подходы к их решению. В ходе исследования использовался материал параллельных текстов (для пар языков «русский—английский» и «русский—французский»), причем рассматривались только такие примеры, где ЛСО имплицитно в русскоязычном фрагменте³. Поиск и аннотирование примеров проводились при помощи действующей информационной системы, разрабатываемой для анализа ЛСО. Приводимые далее рисунки представляют собой скриншоты пользовательского интерфейса данной системы.

2 Автоматизация поиска имплицитных логико-семантических отношений

Сложности работы с имплицитными ЛСО начинаются уже на этапе поиска. Как было показано в [14], моноязычные тексты не позволяют целенаправленно отбирать примеры с имплицитными ЛСО. Эта возможность появляется только при работе с параллельными текстами, где можно применить метод поиска с исключением [15, 16]. Результатом применения метода становится поисковая выдача, где пары фрагментов параллельных текстов таковы, что фрагмент текста на одном из языков содержит показатель искомого ЛСО, тогда как фрагмент текста на другом языке, напротив, не содержит заданных показателей этого же ЛСО.

Хотя метод значительно ускоряет формирование выборки примеров с имплицитными ЛСО, он имеет как минимум три ограничения. Во-первых, сами по себе параллельные корпуса намного менее репрезентативны и объемны, чем моноязычные. Во-вторых, метод по определению позволяет обнаруживать только те примеры имплицитных ЛСО, где в другом языке это ЛСО имеет показатель (случаи, где ЛСО имплицитно в оригинале и в переводе, в поисковую выдачу не попадают). В-третьих, даже среди оставшихся примеров обнаруживается немало нерелевантных.

В качестве иллюстрации рассмотрим задачу поиска примеров имплицитного ЛСО причины в русском языке (использовались параллельные тексты для языковой пары «русский—английский»). Для решения этой задачи был разработан, в частности, следующий запрос:

- (1) англоязычный фрагмент должен содержать единицу *because*;
- (2) русскоязычный фрагмент не должен содержать ни одну из следующих единиц: *ибо, потому что / потому как, поскольку, так как, ведь, оттого, из-за, за, от, ввиду, вследствие, затем что, тем более что, в связи с*.

Затем в ходе экспертного анализа пары фрагментов делились на две группы: релевантные и нерелевантные. В данном случае к релевантным относились те пары фрагментов, где в русском языке действительно имеется имплицитное ЛСО причины; все остальные считались нерелевантными. Далее проводился лингвистический анализ релевантных пар, а результат этого анализа записывался в надкорпусную базу данных [17] в форме аннотированных переводных соответствий (АПС) [18, 19]. Оказалось, что для первых 300 пар, попавших в поисковую выдачу для направлений перевода «русский—английский» и «английский—русский», были сформированы 134 и 141 АПС соответственно, т.е. доля релевантных примеров составила около 47%.

Для разных языковых пар и разных запросов доля таких примеров будет различаться, но едва ли можно представить себе ситуацию, где релевантными будут 100% примеров. Во-первых, в запросе могут не быть учтены какие-либо способы эксплицитного выражения ЛСО (например, *по тому одному что* в отличие от *потому что* для ЛСО причины). Во-вторых, при переводе семантика ЛСО

¹ При этом отдельной проблемой оказывается классификация ЛСО [6, 7] и составление списка возможных ЛСО, где каждое ЛСО было бы снабжено структурированным определением [8, 9].

² Так, в примере «*я не был на совещании, потому что заболел*» ЛСО причины является эксплицитным (выражено при помощи *потому что*), а в примере «*я не был на совещании, заболел*» — имплицитным (не выражено).

³ Несмотря на то что все примеры, использованные в статье, взяты из параллельного корпуса, в ряде случаев они даются в сокращенном виде — либо только оригинал, либо только перевод, если этого достаточно для иллюстрации описываемых положений.

может быть изменена (например, одно ЛСО заменяется другим). В-третьих, при переводе могут добавляться или опускаться фрагменты текста. Любое из таких изменений приводит к тому, что вместо имплицитного ЛСО с заданной семантикой фрагмент текста либо содержит другое ЛСО, либо не содержит никакого.

Таким образом, хотя метод поиска с исключением в параллельных текстах позволяет значительно облегчить процесс формирования набора данных, даже этот метод не позволяет полностью автоматизировать поиск целевых примеров так, чтобы не требовалась ручная постобработка.

3 Автоматизация разметки аргументов

Разметка примеров имплицитных ЛСО предполагает, среди прочего, выявление границ отрезков текста, непосредственно связанных размечаемым ЛСО. Такие отрезки будем называть аргументами ЛСО. Их объем должен быть необходимым и достаточным для анализа отношения: сокращать аргументы так, чтобы семантика ЛСО осталась неясной, или включать в состав аргументов то, что не требуется для анализа ЛСО, не следует. На рисунках ниже аргументы отношения выделены цветом.

Автоматизация разметки аргументов ЛСО на данном этапе исследования затруднительна, что обусловлено двумя факторами. Во-первых, длина аргументов может сильно варьироваться; во-вторых, аргумент не обязательно включает в себя

исключительно соседние токены, а может прерываться.

Основную проблему представляет вариативность аргументов с точки зрения их длины. Аргументы отношения могут включать одно слово, словосочетание, предложение и даже несколько предложений. Такой разброс объясняется природой ЛСО, поскольку отношение одного и того же типа теоретически может связывать отдельные слова (рис. 1), части одного предложения, самостоятельные предложения или даже целые абзацы (рис. 2) в зависимости от описываемой ситуации и детальности ее описания.

Пример на рис. 2 демонстрирует еще одну сложность автоматизации разметки параллельного корпуса — деление текста на сегменты может различаться в оригинале и переводе. Это может быть связано как со спецификой синтаксиса языка, так и с частными решениями переводчика. Так, в примере на рис. 2 в русскоязычном оригинале аргумент 1 включает два предложения, в то время как в переводе на английский — одно сложное предложение.

Разрыв аргументов зачастую связан с наличием в размечаемых фрагментах прямой речи. В примере на рис. 3 в русскоязычном блоке текста разорваны оба аргумента, а в переводе — только аргумент 2.

Таким образом, вариативная структура аргументов, их возможная разрывность, а также специфика размечаемого материала — параллельные тексты, где разбиение оригинала и перевода на части может различаться, не позволяют на данном этапе автоматизировать разметку границ аргументов.

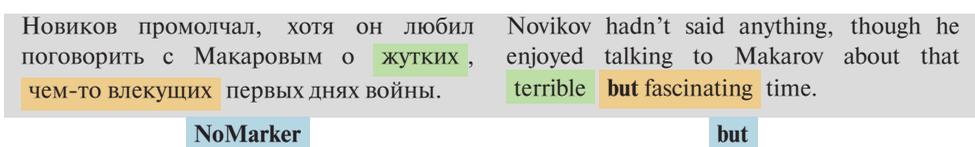


Рис. 1 Пример однословного аргумента ЛСО противопоставления. (Источник примера: [В. С. Гроссман. Жизнь и судьба (ч. 1) (1960) | Life and fate (pt. 1) (пер. R. Chandler; 1985)])

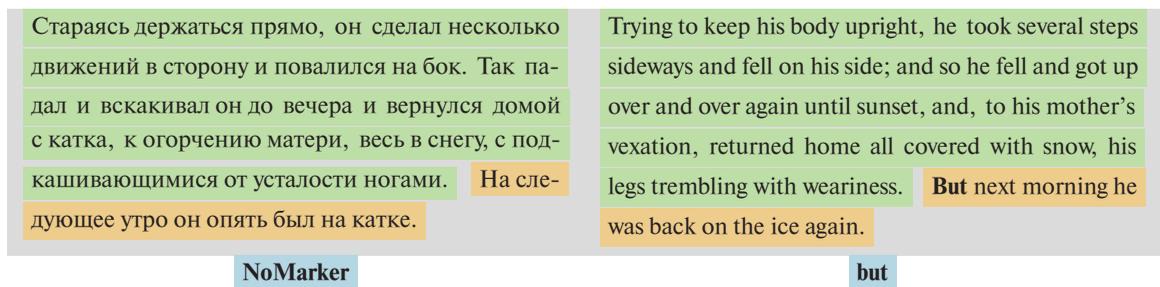


Рис. 2 Пример аргументов ЛСО противопоставления, представляющих собой одно или несколько предложений. (Источник примера: [Б. Н. Полевой. Повесть о настоящем человеке (1946) | A story about a real man (пер. J. Fineberg; 1950)])

<p>— Ну, ну! ну ставь еще четыре тысячи гульденов на красную! Вот бумажник, бери. — Она вынула из кармана и подала мне бумажник. — Ну, бери скорей, тут двадцать тысяч рублей чистыми деньгами. — Бабушка, — прошептал я, — такие куши ... Жива не хочу быть — отыграюсь. Ставь!</p>	<p>“ Very well. Stake another four thousand gulden upon the red. Take this banknote to do it with.” “I have still got twenty thousand roubles in actual cash.” “ But ,” I whispered, “ such a quantity of money — ” “Never mind. I cannot rest until I have won back my losses. Stake!</p>
<p>NoMarker</p>	<p>but</p>

Рис. 3 Пример разрывных аргументов ЛСО противопоставления. (Источник примера: [Ф. М. Достоевский. Игрок (1866) | The gambler (пер. С. J. Hogarth; 1935)])

4 Автоматизация разметки признаков

Признаки, размечаемые в ходе аннотирования примеров с имплицитными ЛСО, весьма разнообразны; они подробно описаны в работах [19, 20]. В контексте же автоматизации разметки важно разделить их на две большие группы: формальные и семантические. Формальные признаки связаны прежде всего с планом выражения (пунктуация, морфология, синтаксис); семантические — с планом содержания (использование слов той или иной семантической категории, наличие заданных семантических структур).

Наиболее простой с точки зрения автоматизации оказывается разметка ряда формальных признаков, таких как наличие некоторых знаков препинания между аргументами, наличие в составе аргумента двух и более предложений. Однако даже среди формальных признаков есть те, которые требуют экспертной постобработки, не говоря уже о семантических. Далее рассмотрим примеры таких признаков.

4.1 Формальные признаки

Признак «Лексический повтор» указывает на то, что первый и второй аргументы ЛСО включают формы одних и тех же слов. На первый взгляд этот признак может быть полностью автоматизирован при наличии морфологической разметки, но это не так. Прежде всего необходимо исключить повтор служебных частей речи — так называемых «стоп-слов»: в рамках данной работы перечни таких слов составлялись для каждого языка (русский, английский и французский) путем расширения списков стоп-слов из библиотеки Natural Language Toolkit [21]. Кроме того, из-за неснятой омонимии к повторам могут быть отнесены те единицы, для которых имеется более одного морфологического разбора. Например, в одном из разборов слова *видимо* в качестве леммы указан глагол

видеть. Следовательно, если в одном аргументе используется слово *видимо* в значении ‘может быть’, а в другом — какая-либо из форм глагола *видеть*, при автоматической обработке такой пример будет снабжен знаком «Лексический повтор», что некорректно.

Признак «Императив» присваивается аргументу, если в нем присутствует глагол в форме повелительного наклонения, и на первый взгляд при наличии морфологической разметки данный признак тоже достаточно легко поддается автоматизации. Однако при анализе результатов автоматической простановки было выявлено, что и в этом случае требуется ручная постобработка. Так, признак «Императив» не нужен, если словоформа в императиве присутствует, но утрачивает свое изначальное значение. К подобным «ложным» императивам можно отнести, например, устойчивое сочетание (*не дай бог* или глаголы в повелительном наклонении, употребляемые в дискурсивном значении: *извольте, разрешите*).

4.2 Семантические признаки

При разметке примеров, содержащих имплицитные ЛСО, особенно значимыми представляются признаки, связанные с семантикой. Эти признаки можно условно разделить на «скорее конкретные» и «скорее абстрактные». В первом случае для простановки признака, как правило, требуется найти лексические единицы или их сочетания, относящиеся к заданному семантическому классу; во втором же смысловое значение может быть настолько абстрактным и выражаться лексически настолько по-разному, что составление списков искомым единиц представляется затруднительным.

Рассмотрим некоторые из семантических признаков, которые относятся к «скорее конкретным».

Признак «Ментальный предикат» указывает на то, что предикат главной клаузы аргумента можно отнести к категории ментальных (т. е. касающихся знаний, рассуждений, мышления; например, гла-

Лили никогда не знала меры в выпивке;
тут не было ничего удивительного : Ли-
ли вообще не знала, что такое мера .

NoMarker

Рис. 4 Пример, где нужен признак «Ментальный предикат». (Источник примера: [Лорен Вайсбергер. Дьявол носит Прада (пер. М. Маяков и др.; 2006)])

— Утро считаешь , надо быть au courant
всего, знать новости .

NoMarker

Рис. 5 Пример, где не нужен признак «Ментальный предикат». (Источник примера: [И. А. Гончаров. Обломов (1848–1859)])

— Почему же не может? — сдерживая слезы, проговорила Анна, очевидно уже не приписывая никакого значения тому, что он скажет . Она чувствовала, что судьба ее была решена [. . .]

NoMarker

Рис. 6 Пример, где имеется ментальный предикат, не входящий в список для поиска. (Источник примера: [Л. Н. Толстой. Анна Каренина (1873–1877)])

голы *знать, мечтать, предполагать* и др.). Чтобы найти в исследуемом материале лексические единицы, входящие в целевую группу, и присвоить соответствующий признак, были использованы данные Национального корпуса русского языка: для семантически размеченной части корпуса доступна функция поиска по лексико-семантическим признакам [22]. Таким образом, можно выгрузить¹ список глаголов, которые имеют семантическую помету «ментальная сфера» (по состоянию на 30.05.2025 он включает 245 единиц).

Далее проводился поиск аргументов, которые включают леммы из списка. Анализ полученных результатов позволил выявить несколько факторов, затрудняющих автоматизацию разметки. Во-первых, из-за неснятой омонимии в выдачу попадали непредикативные единицы, которые в качестве леммы при этом имели глагол (*кажущийся* по лемме *казаться*, *определенный* по лемме *определить*). Во-вторых, признак актуален только в том случае, когда к ментальным относится именно главный предикат аргумента (ср. примеры, приведенные на рис. 4 и 5, содержащие формы глагола *знать*, который включен в список ментальных глаголов из

Национального корпуса русского языка. В первом случае *знала* выступает в качестве сказуемого главной клаузы аргумента (т. е. признак нужен); во втором же эту функцию выполняет *надо* (т. е. признак не нужен).

Наряду с зашумленностью выдачи при автоматическом поиске существует также обратная проблема: ментальный предикат может быть выражен словом, которое не входит в изначальный список для поиска. В примере на рис. 6 в роли главного ментального предиката клаузы выступает глагол *чувствовать*. Сам по себе этот глагол скорее относится к семантическому классу эмоциональной лексики, но здесь он употреблен в качестве ментального предиката, сближаясь по значению с глаголом *понимать*.

Таким образом, даже если признак таков, что его потенциальные показатели теоретически можно задать в виде списка, полностью автоматизировать разметку не представляется возможным. Тем не менее использование списков целевой лексики позволяет значительно облегчить разметку и сделать ее более последовательной, а также проводить дополнительную проверку уже размеченных данных.

Рассмотрим некоторые из признаков, которые относятся к «скорее абстрактным».

Признак «Смена состояния» подразумевает, что один из аргументов включает указание на состояние, а другой — указание на то, что это состояние впоследствии изменилось. В примере на рис. 7 лексическими маркерами смены состояния служат слова *проспал—проснулся*. Однако реализация этого признака может быть крайне разнообразной. Переход от сна к бодрствованию — лишь один из многих вариантов смены состояния, и заранее составить ограниченный набор целевых лексических единиц не представляется возможным.

Еще один признак — «Действие—результат» — присваивается блоку текста в случае, если один из аргументов — указание на действие, а другой — указание на результат действия². В примере на рис. 8 описана ситуация, где действующее лицо ищет некий предмет (действие), но не находит его (результат). Лексическими маркерами признака в русском языке служат слова *обошел — не нашел*. Данный пример хорошо иллюстрирует проблему, обозначенную выше: ввиду разнообразия лексических маркеров автоматизировать разметку затруднительно. Так, более ожидаемой парой глаголов для описания ситуации безрезультатного поиска было бы *искать — не найти*, а не *обойти — не найти*.

¹ <https://ruscorpora.ru/page/instruction-semantic/>.

² В случае с ЛСО противопоставления результат, как правило, не соответствует ожидаемому.

Старик проспал весь процесс. Пиджак его был засыпан пеплом. Он проснулся при виде Поли Белоцерковской.	The old man had slept throughout the trial. His jacket was covered in cigarette ash. But he had woken up when Paulina Byelotserkovskaya had appeared.
NoMarker	but

Рис. 7 Пример, где необходим признак «Смена состояния». (Источник примера: [И. Э. Бабель. Карл-Янкель (1931) | Karl-Yankel (пер. P. Constantine; 2002)])

Just before dinner two telegrams were handed in. The first ran: — “Have just heard that Barrymore is at the Hall. — BASKERVILLE.” The second: — “Visited twenty-three hotels as directed, but sorry to report unable to trace out sheet of Times. — CARTWRIGHT.”	Перед самым обедом нам подали две телеграммы. Первая гласила: «Только что сообщили Бэрримор дома. Баскервиль». Вторая: «Обошел двадцать три гостиницы с сожалением изрезанной страницы «Таймса» не нашел. Картрайт».
but	NoMarker

Рис. 8 Пример, где необходим признак «Действие–результат». (Источник примера: [Arthur Conan Doyle. The hound of the Baskervilles (1902) | Собака Баскервилей (пер. Н. Волжина; 1948)])

Таким образом, разметка семантических признаков, которые относятся к «скорее абстрактным», сложнее всего поддается автоматизации.

5 Заключение

В статье были рассмотрены основные этапы разметки имплицитных ЛСО: поиск целевых фрагментов в корпусном материале, выделение аргументов ЛСО, а также присвоение признаков целевым фрагментам текста.

Показано, что разнообразие семантических критериев, которым должны удовлетворять размеченные данные, приводит к необходимости выходить далеко за пределы собственно анализа имплицитных ЛСО: это и вопросы анализа синтаксической структуры текста, и проблемы снятия омонимии, и задачи составления списков лексики, относящейся к заданному семантическому классу. Полная автоматизация на данном этапе не представляется возможной, сохраняется необходимость привлекать эксперта, по крайней мере для постобработки, однако представленные подходы к автоматизации разметки способны значительно ускорить создание наборов данных с имплицитными ЛСО и сделать разметку более согласованной.

Информационная система для анализа ЛСО разрабатывается авторами с учетом всех вышеописанных сложностей: она позволяет не только создавать АПС в ручном режиме, но и обрабатывать ранее сформированные АПС, при необходимости корректируя разметку в автоматизированном режиме. Таким образом, система представляет собой эффективный инструмент для работы разметчика и эксперта.

Литература

1. *Parmar M., Deilamsalehy H., Dernoncourt F., Yoon S., Rossi R. A., Bui T.* Towards enhancing coherence in extractive summarization: dataset and experiments with LLMs // Conference on Empirical Methods in Natural Language Processing Proceedings. — Miami, FL, USA: Association for Computational Linguistics, 2024. P. 19810–19820. doi: 10.18653/v1/2024.emnlp-main.1106.
2. *Barbosa B., Campelo C.* LLMs as tools for evaluating textual coherence: A comparative analysis // Anais do XV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana. — Porto Alegre: Sociedade Brasileira de Computação, 2024. P. 278–287. doi: 10.5753/stil.2024.245379.
3. *Prasad R., Webber B., Joshi A.* The Penn Discourse Treebank: An annotated corpus of discourse relations // Handbook of linguistic annotation / Eds. N. Ide, J. Pustejovsky. — Dordrecht: Springer Science + Business Media, 2017. P. 1197–1217. doi: 10.1007/978-94-024-0881-2_45.
4. *Webber B., Prasad R., Lee A., Joshi A.* The Penn Discourse Treebank 3.0: Annotation manual. — Philadelphia, PA, USA: Linguistic Data Consortium, University of Pennsylvania, 2019. 81 p. <https://catalog ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>.
5. *Jiang D., He J.* Tree framework with BERT word embedding for the recognition of Chinese implicit discourse relations // IEEE Access, 2020. Vol. 8. P. 162004–162011. doi: 10.1109/ACCESS.2020.3019500.
6. *Инькова О. Ю.* Логико-семантические отношения: проблемы классификации // Связность текста: междисциплинарные логико-семантические отношения / Под ред. О. Ю. Иньковой, Э. Манзотти. — М.: ЯСК, 2019. С. 11–98.
7. *Гончаров А. А.* Классификации внутритекстовых отношений: основания и принципы структурирова-

- ния // Вопросы языкознания, 2021. № 3. С. 97–119. doi: 10.31857/0373-658X.2021.3.97-119. EDN: OKPZEI.
8. *Инькова О. Ю., Кружков М. Г.* Структурированные определения дискурсивных отношений в Надкорпусной базе данных коннекторов // Информатика и её применения, 2021. Т. 15. Вып. 4. С. 27–32. doi: 10.14357/19922264210404. EDN: EZJXVI.
 9. *Инькова О. Ю., Кружков М. Г.* Критерии определения семантической близости дискурсивных отношений // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 100–106. doi: 10.14357/19922264230314. EDN: UJZJZI.
 10. *Гончаров А. А., Инькова О. Ю.* Извлечение знаний о средствах выражения логико-семантических отношений при помощи Надкорпусной базы данных // Информатика и её применения, 2021. Т. 15. Вып. 2. С. 96–103. doi: 10.14357/19922264210214. EDN: GRPWIB.
 11. *Xiang W., Wang B.* A survey of implicit discourse relation recognition // ACM Comput. Surv., 2023. Vol. 55. Iss. 12. Art. 258. 34 p. doi: 10.1145/3574134.
 12. *Xiang W., Liu S., Wang B.* Parsing and encoding interactive phrase structure for implicit discourse relation recognition // Neural Comput. Appl., 2024. Vol. 36. P. 13783–13797. doi: 10.1007/s00521-024-09709-8.
 13. *Cai M., Yang Zh., Jian P.* Improving implicit discourse relation recognition with semantics confrontation // Joint Conference (International) on Computational Linguistics, Language Resources and Evaluation Proceedings. — Torino, Italia: ELRA and ICCL, 2024. P. 8828–8839.
 14. *Гончаров А. А.* Методы поиска имплицитных логико-семантических отношений в моноязычных текстах // Системы и средства информатики, 2022. Т. 32. № 3. С. 92–102. doi: 10.14357/08696527220309. EDN: NUVZGN.
 15. *Гончаров А. А.* Методы поиска имплицитных логико-семантических отношений в параллельных текстах // Системы и средства информатики, 2022. Т. 32. № 4. С. 32–44. doi: 10.14357/08696527220404.
 16. *Гончаров А. А.* Поиск с исключением в параллельных текстах // Системы и средства информатики, 2023. Т. 33. № 4. С. 102–114. doi: 10.14357/08696527230410. EDN: CVPFDV.
 17. *Кружков М. Г.* Концепция построения надкорпусных баз данных // Системы и средства информатики, 2021. Т. 31. № 3. С. 101–112. doi: 10.14357/08696527210309. EDN: UMWNIU.
 18. *Гончаров А. А.* Аннотирование параллельных корпусов: подходы и направления развития // Информатика и её применения, 2023. Т. 17. Вып. 4. С. 81–87. doi: 10.14357/19922264230411. EDN: GDKDOZ.
 19. *Гончаров А. А., Ярошенко П. В.* Принципы аннотирования имплицитных логико-семантических отношений в параллельных текстах // Информатика и её применения, 2024. Т. 18. Вып. 3. С. 106–114. doi: 10.14357/19922264240313. EDN: NPXQNX.
 20. *Goncharov A. A., Iaroshenko P. V.* How to describe implicit discourse relations: The experience of creating a dataset in Russian // Вестник Московского университета. Сер. 9: Филология, 2025 (в печати).
 21. *Bird St., Loper E., Klein E.* Natural language processing with Python. — O'Reilly Media Inc., 2009. 502 p.
 22. *Савчук С. О., Архангельский Т. А., Бонч-Осмоловская А. А., Донина О. В., Кузнецова Ю. Н., Ляшевская О. Н., Орехов Б. В., Подрядчикова М. В.* Национальный корпус русского языка 2.0: новые возможности и перспективы развития // Вопросы языкознания, 2024. № 2. С. 7–34. doi: 10.31857/0373-658X.2024.2.7-34. EDN: AATSXV.

Поступила в редакцию 15.06.2025

Принята к публикации 15.08.2025

AUTOMATION OF ANNOTATING IMPLICIT DISCOURSE RELATIONS: CHALLENGES AND OPPORTUNITIES

A. A. Goncharov¹ and P. V. Iaroshenko^{1,2}

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

²Research Computing Center Lomonosov Moscow State University, 1, bld. 4 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

Abstract: The article outlines the principal challenges encountered in the automation of annotating implicit discourse relations, analyzes the underlying causes of these challenges, and suggests possible solutions. The article examines the main stages of the process: (i) the extraction of examples with implicit discourse relations; (ii) the delimitation of relational argument boundaries; and (iii) the selection of features for annotation of the extracted fragments. The results of applying the method of search with exclusion in parallel texts are presented along with a critical assessment of its limitations. Two factors significantly hindering the automation of argument identification

in text spans with implicit discourse relations are analyzed: the considerable variability in argument length and the noncontiguous nature of arguments, which may be interrupted by intervening tokens. A comprehensive analysis of methods for automating feature selection for the linguistic data is provided. It has been demonstrated that even the processing of formal features may require the involvement of experts. Furthermore, while some semantic features are amenable to partial automation, others currently require manual annotation. The conclusions are illustrated by examples from the corpus.

Keywords: linguistic annotation; discourse relations; logical-semantic relations; implicitness; parallel texts

DOI: 10.14357/19922264250309

EDN: NKEJQY

Acknowledgments

The research was supported by the Russian Science Foundation, project No. 24-28-00527 (<https://rscf.ru/en/project/24-28-00527/>).

References

1. Parmar, M., H. Deilamsalehy, F. Deroncourt, S. Yoon, R. A. Rossi, and T. Bui. 2024. Towards enhancing coherence in extractive summarization: Dataset and experiments with LLMs. *Conference on Empirical Methods in Natural Language Processing Proceedings*. Miami, FL: Association for Computational Linguistics. 19810–19820. doi: 10.18653/v1/2024.emnlp-main.1106.
2. Barbosa, B., and C. Campelo. 2024. LLMs as tools for evaluating textual coherence: A comparative analysis. *Anais do XV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*. Porto Alegre: Sociedade Brasileira de Computação. 278–287. doi: 10.5753/stil.2024.245379.
3. Prasad, R., B. Webber, and A. Joshi. 2017. The Penn Discourse Treebank: An annotated corpus of discourse relations. *Handbook of linguistic annotation*. Eds. N. Ide and J. Pustejovsky. Dordrecht: Springer Science + Business Media. 1197–1217. doi: 10.1007/978-94-024-0881-2_45.
4. Webber, B., R. Prasad, A. Lee, and A. Joshi. 2019. The Penn Discourse Treebank 3.0: Annotation manual. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania. 81 p. Available at: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf> (accessed August 28, 2025).
5. Jiang, D., and J. He. 2020. Tree framework with BERT word embedding for the recognition of Chinese implicit discourse relations. *IEEE Access* 8:162004–162011. doi: 10.1109/ACCESS.2020.3019500.
6. Inkova, O. Yu. 2019. Logiko-semanticheskie otnosheniya: problemy klassifikatsii [Logical-semantic relations: Classification problems]. *Svyaznost' teksta: mereologicheskie logiko-semanticheskie otnosheniya* [Text coherence: Mereological logical semantic relations]. Eds. O. Inkova and E. Manzotti. Moscow: LRC Publishing House. 11–98.
7. Goncharov, A. A. 2021. Klassifikatsii vnutritekstovykh otnosheniy: osnovaniya i printsipy strukturovaniya [Classifications of intratextual relations: Bases and structuring principles]. *Voprosy yazykoznaniya* [Topics in the Study of Language] 3:97–119. doi: 10.31857/0373-658X.2021.3.97-119. EDN: OKPZEI.
8. Inkova, O. Yu., and M. G. Kruzhkov. 2021. Strukturnirovannye opredeleniya diskursivnykh otnosheniy v Nadkorpustnoy baze dannykh konnektorov [Structured definitions of discourse relations in the Supracorpora Database of Connectives]. *Informatika i ee Primeneniya — Inform. Appl.* 15(4):27–32. doi: 10.14357/19922264210404. EDN: EZJXVI.
9. Inkova, O. Yu., and M. G. Kruzhkov. 2023. Kriterii opredeleniya semanticheskoy blizosti diskursivnykh otnosheniy [Evaluation criteria for discourse relations semantic affinity]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):100–106. doi: 10.14357/19922264230314. EDN: UJZJZI.
10. Goncharov, A. A., and O. Yu. Inkova. 2021. Izvlechenie znaniy o sredstvakh vyrazheniya logiko-semanticheskikh otnosheniy pri pomoshchi nadkorpustnoy bazy dannykh [Extracting knowledge about means of expression of logical-semantic relations from the supracorpora database]. *Informatika i ee Primeneniya — Inform. Appl.* 15(2):96–103. doi: 10.14357/19922264210214. EDN: GRPWIB.
11. Xiang, W., and B. Wang. 2023. A survey of implicit discourse relation recognition. *ACM Comput. Surv.* 55(12):258. 34 p. doi: 10.1145/3574134.
12. Xiang, W., S. Liu, and B. Wang. 2024. Parsing and encoding interactive phrase structure for implicit discourse relation recognition. *Neural Comput. Appl.* 36:13783–13797. doi: 10.1007/s00521-024-09709-8.
13. Cai, M., Zh. Yang, and P. Jian. 2024. Improving implicit discourse relation recognition with semantics confrontation. *Joint Conference (International) on Computational Linguistics, Language Resources and Evaluation Proceedings*. Torino, Italia: ELRA and ICCL. 8828–8839.
14. Goncharov, A. A. 2022. Metody poiska implitsitnykh logiko-semanticheskikh otnosheniy v monoyazychnykh tekstakh [Methods for retrieval of implicit logical-semantic relations from monolingual texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(3):92–102. doi: 10.14357/08696527220309. EDN: NUVZGN.
15. Goncharov, A. A. 2022. Metody poiska implitsitnykh logiko-semanticheskikh otnosheniy v paralel'nykh tekstakh [Methods for retrieval of implicit logical-semantic relations from parallel texts]. *Sistemy i Sredstva Informatiki*

- ki* — *Systems and Means of Informatics* 32(4):32–44. doi: 10.14357/08696527220404.
16. Goncharov, A. A. 2023. Poisk s isklyucheniem v paralel'nykh tekstakh [Search with exclusion in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(4):102–114. doi: 10.14357/08696527230410. EDN: CVPFDV.
 17. Kruzhev, M. G. 2021. Kontsepsiya postroeniya nad-korpusnykh baz dannykh [Conceptual framework for supracorpora databases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(3):101–112. doi: 10.14357/08696527210309. EDN: UMWNIU.
 18. Goncharov, A. A. 2023. Annotirovanie paralel'nykh korpusov: podkhody i napravleniya razvitiya [Parallel corpus annotation: Approaches and directions for development]. *Informatika i ee Primeneniya — Inform. Appl.* 17(4):81–87. doi: 10.14357/19922264230411. EDN: GDKDOZ.
 19. Goncharov, A. A., and P. V. Iaroshenko. 2024. Printsipy annotirovaniya implitsitnykh logiko-semanticheskikh ot-nosheniy v paralel'nykh tekstakh [Implicit logical-semantic relations in parallel texts: Annotation principles]. *Informatika i ee Primeneniya — Inform. Appl.* 18(3):106–114. doi: 10.14357/19922264240313. EDN: NPXQNX.
 20. Goncharov, A. A., and P. V. Iaroshenko. 2025 (in press). How to describe implicit discourse relations: The experience of creating a dataset in Russian. *Vestnik Moskovskogo universiteta. Ser. 9: Filologiya* [Moscow University Philology Bulletin].
 21. Bird, S., E. Loper, and E. Klein. 2009. *Natural language processing with Python*. O'Reilly Media Inc. 502 p.
 22. Savchuk, S. O., T. A. Arkhangel'skiy, A. A. Bonch-Osmolovskaya, O. V. Donina, Yu. N. Kuznetsova, O. N. Lyashevskaya, B. V. Orekhov, and M. V. Podryadchikova. 2024. Natsional'nyy korpus russkogo yazyka 2.0: novyye vozmozhnosti i perspektivy razvitiya [Russian National Corpus 2.0: New opportunities and development prospects]. *Voprosy yazykoznaniiya* [Topics in the Study of Language] 2:7–34. doi: 10.31857/0373-658X.2024.2.7-34. EDN: AATSXV.

Received June 15, 2025

Accepted August 15, 2025

Contributors

Goncharov Alexander A. (b. 1994) — scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; a.gonch48@gmail.com

Iaroshenko Polina V. (b. 1994) — Candidate of Science (PhD) in philology, scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; programmer, Research Computing Center Lomonosov Moscow State University, 1, bld. 4 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; polina.iaroshenko@yandex.ru

ИЕРАРХИЯ АКОФФА И ЗАДАЧИ ИЗВЛЕЧЕНИЯ ЗНАНИЯ ИЗ ТЕКСТОВ

И. М. Зацман¹

Аннотация: Цель статьи состоит в описании информационных моделей процессов извлечения из текстов лингвистического знания об исследуемых языковых единицах на основе иерархии Акоффа, которая была предложена им в 1989 г. Компоненты иерархии, которые рассматриваются в данной статье, — это данные, информация и знание. Основным результатом Акоффа состоит в разграничении смыслового содержания слов, которые обозначают компоненты иерархии, и довольно общей формулировке задачи описания трансформаций компонентов. Описание Акоффом этой задачи порождает дискуссии уже не одно десятилетие. При этом задача описания трансформаций компонентов до сих пор остается нерешенной. Не претендуя на решение этой задачи в общем виде, в статье предлагается подойти к решению ее частных случаев для предметной области извлечения знания из текстов на основе детализации смыслового содержания слов «данные», «информация» и «знание». На основе предлагаемого подхода описаны две модели извлечения знания из текстов, каждая из которых специфицирует перечень трансформаций детализированных компонентов иерархии Акоффа. Первая модель стала основой для проектирования технологий без учета неудачных исходов извлечения знания из текстов, а вторая модель — для проектирования технологий с их учетом. Проведенные эксперименты показали, что спроектированные технологии, использующие надкорпусные базы данных, обеспечивают извлечение из текстов уже известного и нового лингвистического знания, а также формирование новых классификаций.

Ключевые слова: иерархия Акоффа; данные; информация; знание; извлечение знания; классификация

DOI: 10.14357/19922264250310

EDN: JAJCQY

1 Введение

Иерархия Акоффа, которая была предложена им в 1989 г., включает четыре компонента (данные, информация, знания, мудрость). Последний компонент рассматривается только в обзорной части данной статьи. Основным результатом Акоффа состоит в разграничении смыслового содержания слов, которые обозначают компоненты иерархии, и довольно общей формулировке задачи описания трансформаций компонентов [1]. Описание Акоффом этой задачи порождает дискуссии уже не одно десятилетие. При этом задача описания трансформаций компонентов до сих пор остается нерешенной.

Другое часто используемое название иерархии Акоффа — иерархия DIKW (Data, Information, Knowledge, Wisdom). Четвертый компонент иногда опускается и тогда используется название *иерархия DIK* (Data, Information, Knowledge). По словам Дж. Роули [2, с. 163], иерархия DIKW «является одной из фундаментальных, широко признанных и «само собой разумеющихся (taken for granted)» моделей в литературе по информации и знаниям».

Согласно Акоффу, «данные... есть результат наблюдения. Информация... извлекается из данных путем анализа, со многими аспектами которого

компьютеры справляются... В любом случае информация выводится из данных» [1, с. 3]. При этом Акофф не уточняет, с какими аспектами анализа данных компьютеры справляются, а с какими — нет.

Главный вывод, к которому пришла Дж. Роули после изучения иерархии Акоффа, заключается в следующем: «... информация определяется на основе данных, знание — на основе информации... но имеется меньше консенсуса в описании процессов, которые преобразуют компоненты, находящиеся ниже в иерархии, в компоненты, находящиеся выше» [2, с. 163].

Согласно Дэвиду Вайнбергеру, описание Акоффом перехода от информации к знаниям в иерархии DIKW оказалось более спорным, чем описание перехода от данных к информации. Недостатки этого описания он сформулировал так: «знание — это не просто результат алгоритмического преобразования [компонентов]... Оно — результат гораздо более сложного процесса, который является социальным, целенаправленным, контекстно и культурно обусловленным... Наиболее важным в этом отношении является то, что там, где решения сложные и знание [необходимое для решения] получить трудно, оно не определяется на основе информации» [3]. В работах [4, 5] сопоставлены существующие

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

ющие точки зрения на иерархию DIKW в части описания перехода от информации к знанию и отмечается, что это один из самых сложных вопросов для ее понимания.

Продолжение довольно общего обсуждения иерархии Акоффа и задачи описания трансформаций ее компонентов (далее — задача Акоффа) не ставится целью данной статьи. Ее цель состоит в описании подхода к решению частных случаев этой задачи для предметной области извлечения знания из текстов на основе детализации смыслового содержания слов «данные», «информация» и «знание».

Кроме иерархии Р. Акоффа в статье используется принцип Н. Белкина, который был им сформулирован в фундаментальном обзоре 1978 г. Суть этого принципа заключается в том, что смысловое содержание многозначных слов (таких как слово «информация») необходимо детализировать исходя из контекста постановки и решения задач в конкретной предметной области [6]. В данной статье — это задачи извлечения знания из текстов предложений.

Согласно типологии проблем информатики Дэвида Арела по основанию видов их сложности (вычислительная, системная и когнитивная) извлечение знания из текстов относится к задачам когнитивной сложности [7], информационные технологии (ИТ) решения которых относятся к классу человеко-машинных. Поэтому созданные две модели включают трансформации компонентов, которые выполняются экспертно.

В заключение этого раздела отметим, что иерархия Акоффа в ее классической форме применяется в информатике при проектировании ИТ визуальной аналитики [8, 9].

2 Предлагаемый подход

Предлагаемый ниже подход к решению частных случаев задачи Акоффа для предметной области извлечения знания из текстов строится на основе следующих принципов:

- разграничение смыслового содержания слов «данные», «информация» и «знание» согласно иерархии Акоффа;
- деление предметной области информатики на среды разной природы [10];

- детализация смыслового содержания многозначных слов исходя из контекста постановки и решения задач в конкретной предметной области (принцип Н. Белкина).

Основная идея построения предлагаемого подхода состоит в принципиально иной последовательности трансформаций (переходов) между данными, информацией и знанием в предметной области извлечения знания из текстов по сравнению с иерархией Акоффа. В иерархии Акоффа постулируется следующая последовательность трансформаций: данные → информация → знание. Однако в нейролингвистике доказано экспериментально, что в предметной области извлечения знания из текстовых данных и его представления в форме знаковой информации реализуется иная последовательность трансформаций: данные → знание → информация [11, 12]. В информатике эта последовательность описана, например, в работе [13].

Таким образом, предлагаемый подход представляет собой сочетание последовательности трансформаций (переходов): данные → знание → информация и трех вышеперечисленных принципов. Отметим, что описание принципа деления предметной области информатики на среды разной природы дано в работе [10] согласно определению информатики по К. Нюгору [14] и его последующему развитию в работах [15, 16]. В результате деления предметной области в работе [10] рассмотрены следующие среды:

- ментальная среда — совокупность когнитивных феноменов, формируемых в процессах познания, происходящих в сознании людей (например, знание и концепты);
- сенсорно воспринимаемая (кратко — информационная) среда — совокупность сенсорно воспринимаемых феноменов, находящихся вне сознания, но которые могут с ним взаимодействовать (например, текстовое описание извлеченного знания¹);
- цифровая среда — совокупность компьютерных кодов;
- нейросреда — электрические потенциалы и магнитные поля, генерируемые мозгом в процессе сенсорного восприятия феноменов, находящихся вне сознания; такие потенциалы

¹ В статье слова «значение» и «концепт» становятся синонимами, если для описания (дефиниций) значений исследуемых языковых единиц используется только один естественный язык. В случае двуязычного описания значений смысловое содержание этих слов отличается. Слово «значение» используется *до описания* значения языковой единицы средствами знаковой системы естественного языка (например, в словосочетании «знание о значении»). Слово «концепт» — это смысловое содержание его описания, сформированного средствами некоторой знаковой системы. Это отличие подчеркивает сложность *идентичного* описания одного и того же значения в разных языках.

и поля используются, например, в ИТ управления роботизированной рукой [17] и в других ИТ, применяющих интерфейсы «мозг—компьютер»;

- ДНК-среда – совокупность синтезированных цепочек РНК и ДНК, которые используются для хранения больших объемов данных.

Согласно принципу Н. Белкина [6] о детализации смыслового содержания многозначных слов (в статье это слова «данные», «информация» и «знание»), для построения двух моделей ИТ, обеспечивающих извлечение лингвистического знания из текстов с учетом и без учета неудачных исходов его извлечения, введем следующие семь определений компонентов создаваемых моделей, которые будем использовать как терминологическую основу проектирования ИТ:

- (1) сенсорно воспринимаемые человеком данные (далее — СВ-данные) — это тексты предложений, содержащих *исследуемые языковые единицы*;
- (2) цифровые данные — это компьютерные коды символов СВ-данных;
- (3) сенсорно воспринимаемая (знаковая) информация — это описание значений исследуемых языковых единиц как концептов используемого естественного языка или *неудачных исходов извлечения*;
- (4) цифровая информация — это компьютерные коды символов знаковой информации;
- (5) знание об исследуемых языковых единицах (известное или новое):
 - (а) известное знание — это совокупность значений исследуемых языковых единиц, описание которых включает словарь, выбранный в качестве эталона новизны для решения задач извлечения знания из текстов;
 - (б) новое знание — это извлеченные значения исследуемых языковых единиц, описание которых отсутствует в словаре;
- (6) концепт — это результат членения знания в той или иной знаковой системе естественного языка;
- (7) нейрообразы символов СВ-данных — это результат их сенсорного восприятия человеком до начала стадии их семантического анализа [11, 12].

3 Модель без учета неудачных исходов извлечения знания

Для построения первой модели ИТ, обеспечивающей решение задач извлечения лингвистического знания без учета неудачных исходов его извлечения из текстов, необходимы первые шесть компонентов и три среды (ментальная, информационная и цифровая): извлеченное известное и новое знание человека, а также его составляющие (концепты) принадлежат к *ментальной среде*; СВ-данные и знаковая (сенсорно воспринимаемая) информация — к *информационной среде*; цифровые данные и цифровая информация — к *цифровой среде* компьютеров.

В соответствии с экспериментальными данными нейролингвистики [11, 12], в результате применения принципа Н. Белкина к иерархии DIK (редуцированный вариант иерархии Акоффа без компонента «мудрость») была построена модель под названием «трехсредовой», которая содержит *первые шесть компонентов* из семи вышеперечисленных и пять их трансформаций, пронумерованных как 1, А, 2–4 на рис. 1. Двойная нумерация трансформации на рис. 1 и 3 (см. ниже) упрощает сопоставление этих двух рисунков. Арабскими цифрами пронумерованы совпадающие по смысловому содержанию трансформации (переходы), а буквами обозначены различающиеся трансформации.

Опишем пять трансформаций трехсредовой модели извлечения из текстовых данных лингвистического знания об исследуемых языковых единицах без учета неудачных исходов его извлечения:

- 1 — декодирование *цифровых данных*, т. е. формирование *СВ-данных* предложения, содержащего исследуемую языковую единицу;
- А — семантический анализ *СВ-данных* предложения и извлечение из него *знания* о том значении (известном или новом) исследуемой языковой единицы, в котором она используется в этом предложении;

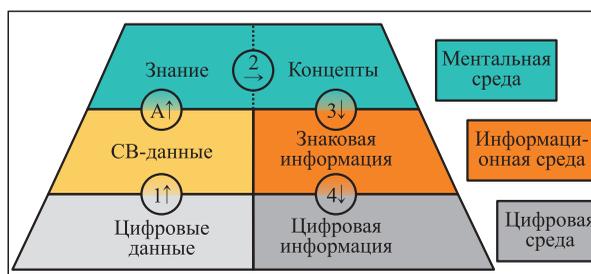


Рис. 1 Трехсредовая модель извлечения знания

- 2 — соотнесение *знания* о значении с *концептами* того естественного языка, на котором в информационной среде готовится дефиниция значения (если необходимо получить двуязычное описание значения, то деление извлеченного знания на концепты выполняется в рамках каждого из двух естественных языков);
- 3 — формирование дефиниции извлеченного значения как *знаковой информации*, знание о котором соотнесено с *концептами* используемого естественного языка; для известного значения исследуемой языковой единицы его дефиниция (описание) копируется из словаря;
- 4 — формирование *цифровой информации* в процессе кодирования символов *знаковой информации*.

Технология, разработанная на основе трехсредовой модели, была применена для обнаружения и извлечения новых значений немецких модальных глаголов, которые не описаны в немецко-русском словаре. В этой технологии первая и четвертая трансформации (декодирование и кодирование) выполняются программно, а трансформация «А», вторая и третья трансформации выполняются экспертами. Они используют надкорпусную базу данных немецко-русских параллельных текстов, разработанную в ФИЦ ИУ РАН, которая содержит тексты предложений (около 18 млн словоупотреблений) и дефиниций значений немецких модальных глаголов как исследуемых языковых единиц. Дефиниции их значений формируются в процессе семантического анализа предложений. Отметим, что кроме дефиниций значений эксперт заполняет таблицу с формализованным описанием контекста глагола в этом значении. Дефиниции с таблицами планируется использовать как обучающее множество для нейронной сети, которой в будущем планируется передать часть работы экспертов.

С помощью этой технологии эксперты выполнили семантический анализ около 4000 предложений с известными и новыми значениями немецких модальных глаголов (*dürfen, können, mögen, müssen, sollen* и *wollen*). Один процент проанализированных предложений (40 из 4000) содержал новое языковое знание о значениях немецких модальных глаголов. Предложения с известными значениями используются для пополнения списка примеров, иллюстрирующих эти значения в словарных статьях исследуемых языковых единиц. Предложения с новыми значениями используются для включения в эти словарные статьи их описаний с иллюстрирующими примерами. Это обеспечивает актуализацию словарей на основе знания, извлеченного из текстов надкорпусной базы данных.

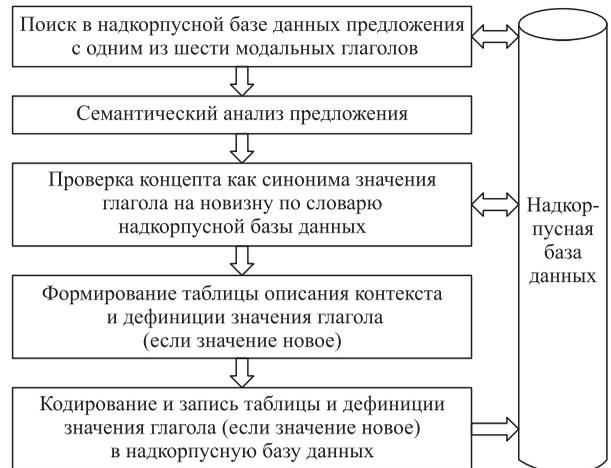


Рис. 2 Технологический цикл обработки предложения

Опишем пример решения задачи семантического анализа предложений с немецкими модальными глаголами на основе трехсредовой модели и моноязычного описания их значений. Как отмечено выше, эта модель содержит шесть компонентов и пять их трансформаций (см. рис. 1). Технологический цикл обработки каждого предложения с глаголом содержит пять этапов (рис. 2), которые соответствуют этим пяти трансформациям.

На первом этапе осуществляется поиск в надкорпусной базе данных предложения с одним из шести перечисленных глаголов в любой его форме и его отображение на экране эксперта. На втором этапе эксперт проводит семантический анализ предложения и в результате интроспекции формируется знание эксперта о значении глагола, в котором он используется в этом предложении. Для моноязычного описания на третьем этапе концепт как синоним значения глагола проверяется на новизну. Ее критерием служит описание значений этого глагола в словаре надкорпусной базы данных. Если в словаре значение глагола, в котором он используется в этом предложении, не описано, то оно считается новым. На четвертом этапе эксперт формирует таблицу описания контекста и дефиницию значения глагола (если значение новое), которые на пятом этапе кодируются и записываются в надкорпусную базу данных.

Проиллюстрируем этот цикл обработки на предложении «Was sollte denn dies wieder bedeuten?» (Что это должно означать?) из книги Ф. Дюрренматта «Подозрение» (перевод Н. Савинкова). После извлечения этого предложения из надкорпусной базы данных эксперт в результате семантического анализа определяет значение глагола *sollen* (в предложении он в форме *sollte*) как *долженствование*. Это значение не рассматривается как новое, так как оно

уже описано в словаре. Эксперт заполняет таблицу с формализованным описанием контекста глагола в этом предложении и записывает ее в надкорпусную базу данных.

4 Модель, учитывающая неудачные исходы извлечения знания

Для построения второй модели ИТ, обеспечивающей решение задач извлечения лингвистического знания с учетом неудачных исходов его извлечения из текстов, необходимы все вышеперечисленные семь компонентов и четыре среды (ментальная, информационная, цифровая и нейросреда). По сравнению с трехсредовой моделью, которая не учитывает случаи неудачного исхода извлечения знания, добавляются нейросреда и нейрообразы символов СВ-данных предложений, содержащих исследуемые языковые единицы. Вторая модель, названная четырехсредовой (рис. 3), дает возможность описать такие случаи.

Отметим, что заранее не известно, из СВ-данных каких предложений возможно извлечение известного или нового лингвистического знания об исследуемых языковых единицах, а из каких предложений экспертам этого сделать не удастся. Четырехсредовая модель на рис. 3 учитывает такие случаи.

Когда экспертам не удастся интерпретировать нейрообразы символов СВ-данных предложения, тогда в надкорпусную базу данных вместо таблицы и описания значения исследуемой языковой единицы заносится тег unknown для этого предложения. К семантическому анализу предложений с такими тегами эксперты возвращаются через некоторый период времени. Между моментами времени повторного семантического анализа такие предложения с тегами unknown хранятся в надкорпусной базе данных.

Четырехсредовая модель содержит все семь вышеопределенных компонентов, а также тег unknown и его коды в цифровой среде. Между ними есть восемь трансформаций, пронумерованных как 1, В–Е, 2–4 на рис. 3. Трансформации 1–4 описаны при рассмотрении трехсредовой модели. Как было отмечено выше, эти трансформации совпадают в обеих моделях. Отличие между ними состоит в том, что четырехсредовая модель вместо одной трансформации А (СВ-данные предложения → извлеченное из него знание о значении) включает четыре следующих трансформации, обозначенных на рис. 3 буквами В–Е:

- В — формирование нейрообразов символов на основе СВ-данных предложения в процессе их сенсорного восприятия человеком [11, 12];
- С — генерация знания о значении исследуемой языковой единицы в процессе семантического анализа нейрообразов символов на основе СВ-данных предложения (для случая удачного извлечения знания);
- D — простановка тега unknown для анализируемого предложения для случая неудачного извлечения знания;
- Е — кодирование тега unknown и запись его кодов в надкорпусную базу данных.

В технологии, разработанной на основе четырехсредовой модели, трансформация «Е», первая и четвертая трансформации (декодирование цифровых данных, кодирование тега и знаковой информации) выполняются программно, а трансформации В–D, вторая и третья трансформации выполняются экспертами. В этой технологии они используют надкорпусную базу данных русско-французских параллельных текстов, разработанную в ФИЦ ИУ РАН, которая содержит тексты предложений (около 2 млн словоупотреблений) и дефиниции значений коннекторов как исследуемых языковых единиц. Дефиниции их значений формируются в процессе семантического анализа предложений для случаев удачных исходов извлечения знания.

С помощью технологии, разработанной на основе четырехсредовой модели, эксперты выполнили семантический анализ около 18 500 предложений с коннекторами русского языка. Основная цель применения этой технологии состоит в построении новой классификации внутритекстовых (дискурсивных) отношений, выражаемых или устанавливаемых коннекторами [18]. В 6,75% проанализированных предложений (1250 из 18 500) эксперты поставили тег unknown.

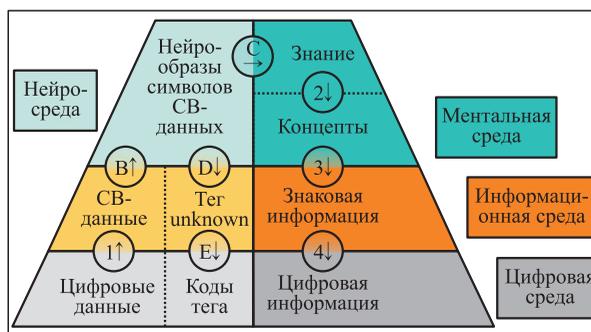


Рис. 3 Четырехсредовая модель извлечения знания

В результате семантического анализа 17 250 предложений был сформирован список известных и новых внутритекстовых отношений как значений коннекторов русского языка. На основе этого списка в ФИЦ ИУ РАН была создана принципиально новая классификация внутритекстовых отношений. Ее сопоставительный анализ с ранее созданными классификациями продемонстрировал, что она обладает существенно большей полнотой и детальностью описания внутритекстовых отношений. Уникальное ее свойство состоит в возможности определения меры близости внутритекстовых отношений, определяемой по степени общности их различительных признаков [19, 20].

5 Заключение

Для создания двух моделей извлечения из текстов лингвистического знания (см. рис. 1 и 3) использовались компоненты четырех из пяти сред, так как для решения рассмотренных частных случаев задачи Акоффа ДНК-среда не используется. Трехсредовая модель состоит из шести компонентов и пяти их трансформаций. Четырехсредовая модель состоит из девяти компонентов и восьми их трансформаций.

Для решения частных случаев задачи Акоффа каждый из трех компонентов классической иерархии Акоффа (данные, информация, знание) был заменен на два компонента. Также были добавлены седьмой компонент (нейрообразы), тег `unknown` и его коды для решения задач извлечения знания из текстов с учетом случаев неудачных исходов извлечения знания. Методическая особенность предлагаемого в статье подхода состоит в том, что он позволяет преодолевать огромный разрыв между абстрактным уровнем иерархии Акоффа и реальными ИТ, обеспечивающими решение практических задач извлечения знания из текстов.

В заключение отметим, что теоретической основой решения рассмотренных частных случаев задачи Акоффа служат принцип Н. Белкина (о детализации смыслового содержания многозначных слов исходя из контекста постановки и решения задач в конкретной предметной области) и принцип деления предметной области информатики на среды разной природы [14–16]. Для решения задач извлечения из текстов лингвистического знания с применением систем искусственного интеллекта планируется использовать эти два принципа, а также информационный ресурс, сформированный по технологиям, разработанным согласно созданным двум моделям, который может служить обучающим множеством для искусственных нейронных сетей.

Автор признателен анонимному рецензенту и редактору за предложения и замечания, учет которых позволил существенно улучшить статью.

Литература

1. *Ackoff R.* From data to wisdom // *J. Applied Systems Analysis*, 1989. Vol. 16. No. 1. P. 3–9.
2. *Rowley J.* The wisdom hierarchy: representations of the DIKW hierarchy // *J. Inf. Sci.*, 2007. Vol. 33. No. 2. P. 163–180. doi: 10.1177/0165551506070706.
3. *Weinberger D.* The problem with the data–information–knowledge–wisdom hierarchy // *Harvard Bus. Rev.*, 2010. <https://hbr.org/2010/02/data-is-to-info-as-info-is-not>.
4. *Bratianu C.* Knowledge creation // *Organizational knowledge dynamics: Managing knowledge creation, acquisition, sharing, and transformation.* — Hershey, PA, USA: IGI Global, 2015. P. 180–206. doi: 10.4018/978-1-4666-8318-1.ch008.
5. *Bratianu C., Bejinaru R.* From knowledge to wisdom: Looking beyond the knowledge hierarchy // *Knowledge*, 2023. Vol. 3. No. 2. P. 196–214. doi: 10.3390/knowledge3020014.
6. *Belkin N.J.* Information concepts for information science // *J. Doc.*, 1978. Vol. 34. No. 1. P. 55–85. doi: 10.1108/eb026653.
7. *Harel D.* *Algorithmics — the spirit of computing.* — Reading, MA, USA: Addison-Wesley, 1987. 514 p.
8. *Chen M., Ebert D., Hagen H., Laramée R., Van Liere R., Ma K.-L., Ribarsky W., Scheuermann G., Silver D.* Data, information, and knowledge in visualization // *IEEE Comput. Graph.*, 2009. Vol. 29. No. 1. P. 12–19. doi: 10.1109/MCG.2009.6.
9. *Federico P., Wagner M., Rind A., Amor-Amoros A., Miksch S., Aigner W.* The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics // *IEEE Conference on Visual Analytics Science and Technology Proceedings.* — New York, NY, USA: IEEE, 2017. P. 92–103. doi: 10.1109/VAST.2017.8585498.
10. *Зацман И. М.* Теоретические основания компьютерного образования: среды предметной области информатики как основание классификации ее объектов // *Системы и средства информатики*, 2022. Т. 32. № 4. С. 77–89. doi: 10.14357/08696527220408. EDN: SAUWDF.
11. *Pylkkänen L., Bemis D. K., Elorrieta E. B.* Building phrases in language production: An MEG study of simple composition // *Cognition*, 2014. Vol. 133. No. 2. P. 371–384. doi: 10.1016/j.cognition.2014.07.001.
12. *Walenski M., Europa E., Caplan D., Thompson C. K.* Neural networks for sentence comprehension and production: An ALE-based meta-analysis of neuroimaging studies // *Hum. Brain Mapp.*, 2019. Vol. 40. No. 8. P. 2275–2304. doi: 10.1002/hbm.24523.
13. *Информатика как наука об информации / Под ред. Р. С. Гиляревского.* — М.: Фаир-Пресс, 2006. 592 с.

14. Nygaard K. Program development as a social activity // 10th World Computer Congress Proceedings / Ed. H.-J. Kugler. — North Holland: Elsevier Science Publs., 1986. P. 189–198.
15. Зацман И. М. Научная парадигма информатики как третьей культуры // Научно-техническая информация. Сер. 1: Организация и методика информационной работы, 2023. № 11. С. 1–14.
16. Зацман И. М. Парадигма информатики: актуальность и постулаты построения // Научно-техническая информация. Сер. 1: Организация и методика информационной работы, 2025. № 4. С. 1–12.
17. Зацман И. М. Интерфейсы третьего порядка в информатике // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 82–89. doi: 10.14357/19922264190312. EDN: EHRQLF.
18. Зацман И. М., Инькова О. Ю., Нуриев В. А. Построение классификационных схем: методы и технологии экспертного формирования // Научно-техническая информация. Сер. 2: Информационные процессы и системы, 2017. № 1. С. 8–22.
19. Инькова О. Ю., Кружков М. Г. Критерии определения семантической близости дискурсивных отношений // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 100–106. doi: 10.14357/19922264230314. EDN: UJZJZI.
20. Инькова О. Ю., Кружков М. Г. Степень семантической близости дискурсивных отношений: методы и инструменты расчета // Информатика и её применения, 2023. Т. 17. Вып. 4. С. 88–95. doi: 10.14357/19922264230412. EDN: FXTSPZ.

Поступила в редакцию 11.07.2025

Принята к публикации 15.08.2025

ACKOFF'S HIERARCHY AND TASKS OF KNOWLEDGE DISCOVERY IN TEXTS

I. M. Zatsman

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

Abstract: The aim of the paper is to describe information models of the processes of discovering in texts linguistic knowledge about the studied language units based on the Ackoff's hierarchy proposed by him in 1989. The components of the hierarchy considered in the paper are data, information, and knowledge. Ackoff's principal outcome consists in dividing the semantic content of words denoting the components of the hierarchy and a general formulation of the problem of describing component transformations. Ackoff's description of this problem has generated discussions for decades. At the same time, the problem of describing the component transformations still remains unsolved. Without claiming to solve this problem in general, the paper proposes an approach to solving its particular cases for the subject domain of knowledge discovery in texts based on detailing the semantic content of the words “data,” “information,” and “knowledge.” Based on the proposed approach, the paper describes two models of knowledge discovery in texts, each of which specifies a list of transformations of the detailed components of Ackoff's hierarchy. The first model became the basis for designing technologies without taking into account unsuccessful outcomes of knowledge discovery in texts and the second model — for designing technologies taking them into account. The experiments conducted showed that the designed technologies using supracorpora databases provide for discovering in texts already known and new linguistic knowledge from texts as well as formation of new classifications.

Keywords: Ackoff's hierarchy; data; information; knowledge; knowledge discovery; classification

DOI: 10.14357/19922264250310

EDN: JAJCQY

References

1. Ackoff, R. 1989. From data to wisdom. *J. Applied Systems Analysis* 16(1):3–9.
2. Rowley, J. 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. *J. Inf. Sci.* 33(2):163–180. doi: 10.1177/0165551506070706.
3. Weinberger, D. 2010. The problem with the data–information–knowledge–wisdom hierarchy. *Harvard Bus. Rev.* Available at: <https://hbr.org/2010/02/data-is-to-info-as-info-is-not> (accessed August 26, 2025).
4. Bratianu, C. 2015. Knowledge creation. *Organizational knowledge dynamics: Managing knowledge creation, acquisition, sharing, and transformation*. Hershey, PA: IGI Global. 180–206. doi: 10.4018/978-1-4666-8318-1.ch008.
5. Bratianu, C., and R. Bejinaru. 2023. From knowledge to wisdom: Looking beyond the knowledge hierarchy. *Knowledge* 3(2):196–214. doi: 10.3390/knowledge3020014.
6. Belkin, N. J. 1978. Information concepts for information science. *J. Doc.* 34(1):55–85. doi: 10.1108/eb026653.
7. Harel, D. 1987. *Algorithmics: The spirit of computing*. Reading, MA: Addison-Wesley. 514 p.

8. Chen, M., D. Ebert, H. Hagen, R. Laramée, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. 2009. Data, information, and knowledge in visualization. *IEEE Comput. Graph.* 29(1):12–19. doi: 10.1109/MCG.2009.6.
9. Federico, P., M. Wagner, A. Rind, A. Amor-Amoros, S. Miksch, and W. Aigner. 2017. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. *IEEE Conference on Visual Analytics Science and Technology Proceedings*. New York, NY: IEEE. 92–103. doi: 10.1109/VAST.2017.8585498.
10. Zatsman, I. M. 2022. Teoreticheskie osnovaniya komp'yuternogo obrazovaniya: sredi predmetnoy oblasti informatiki kak osnovanie klassifikatsii ee ob'ektov [Theoretical foundations of digital education: Subject domain media of informatics as the base of its objects' classification]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(4):77–89. doi: 10.14357/08696527220408. EDN: SAUWDF.
11. Pylkkänen, L., D. K. Bemis, and E. B. Elorrieta. 2014. Building phrases in language production: An MEG study of simple composition. *Cognition* 133(2):371–384. doi: 10.1016/j.cognition.2014.07.001.
12. Walenski, M., E. Europa, D. Caplan, and C. K. Thompson. 2019. Neural networks for sentence comprehension and production: An ALE-based meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 40(8): 2275–2304. doi: 10.1002/hbm.24523.
13. Gilyarevskiy, R. S., ed. 2006. *Informatika kak nauka ob informatsii* [Informatics as information science]. Moscow: FAIR-PRESS. 592 p.
14. Nygaard, K. 1986. Program development as a social activity. *10th World Computer Congress Proceedings*. Ed. H.-J. Kugler. North Holland: Elsevier Science Pubs. 189–198.
15. Zatsman, I. M. 2023. Scientific paradigm of informatics as a third culture. *Scientific Technical Information Processing* 50(4):246–258. doi: 10.3103/S0147688223040111.
16. Zatsman, I. M. 2025. Informatics paradigm: Relevance and postulates of its construction. *Scientific Technical Information Processing* 52(2):92–103.
17. Zatsman, I. M. 2019. Interfeysy tret'ego poryadka v informatike [Third-order interfaces in informatics]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):82–89. doi: 10.14357/19922264190312. EDN: EHRQLF.
18. Zatsman, I., O. Inkova, and V. Nuriev. 2017. The construction of classification schemes: Methods and technologies of expert formation. *Automatic Documentation Mathematical Linguistics* 51(1):27–41. EDN: ZADENJ.
19. Inkova, O. Yu., and M. G. Kruzhkov. 2023. Kriterii opredeleniya semanticheskoy blizosti diskursivnykh otnosheniy [Evaluation criteria for discourse relations semantic affinity]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):100–106. doi: 10.14357/19922264230314. EDN: UJZJZI.
20. Inkova, O. Yu., and M. G. Kruzhkov. 2023. Stepen' semanticheskoy blizosti diskursivnykh otnosheniy: metody i instrumenty rascheta [Evaluating the degree of discourse relations semantic affinity: Methods and instruments]. *Informatika i ee Primeneniya — Inform. Appl.* 17(4):88–95. doi: 10.14357/19922264230412. EDN: FXTSPZ.

Received July 11, 2025

Accepted August 15, 2025

Contributor

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, head of department, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; izatsman@yandex.ru

Борисов Андрей Владимирович (р. 1965) — доктор физико-математических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Борисов Василий Андреевич (р. 2002) — студент магистратуры кафедры «Математическая кибернетика» факультета «Компьютерные науки и прикладная математика» Московского авиационного института (национального исследовательского университета)

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гончаров Александр Анатольевич (р. 1994) — научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Николай Александрович (р. 1982) — кандидат физико-математических наук, старший научный сотрудник 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

Забейхайло Михаил Иванович (р. 1956) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Федерального исследовательского центра «Информатика и управление» Российской академии наук

Игнатов Алексей Николаевич (р. 1991) — кандидат физико-математических наук, доцент кафедры «Теория вероятностей и компьютерное моделирование» Московского авиационного института (национального исследовательского университета)

Козлов Михаил Васильевич (р. 1966) — ведущий математик Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кульченков Владислав Владимирович (р. 1989) — заместитель начальника управления портфельного анализа Банка ВТБ

Малашенко Юрий Евгеньевич (р. 1946) — доктор физико-математических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Назарова Ирина Александровна (р. 1966) — кандидат физико-математических наук, научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Назарьин Артем Игоревич (р. 2000) — магистрант кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов

* **Никитин Максим Сергеевич** (р. 1998) — аспирант кафедры «Автоматика и телемеханика» Пермского национального исследовательского политехнического университета

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

Сопин Эдуард Сергеевич (р. 1987) — кандидат физико-математических наук, доцент кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов; старший научный сотрудник Федерального исследовательского цент-

ра «Информатика и управление» Российской академии наук

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

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тюрин Сергей Феофанович (р. 1953) — доктор технических наук, профессор, заслуженный изобретатель РФ, профессор кафедры «Автоматика и телемеханика» Пермского национального исследовательского политехнического университета; профессор кафедры «Математическое обеспечение вычислительных систем» Пермского государственного национального исследовательского университета имени А. М. Горького

Шестаков Олег Владимирович (р. 1976) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; ведущий научный сотрудник Московского центра фундаментальной и прикладной математики

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

Ярошенко Полина Владимировна (р. 1994) — кандидат филологических наук, научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; программист Научно-исследовательского вычислительного центра МГУ имени М. В. Ломоносова

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

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

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Если при подготовке статьи авторы использовали инструменты на основе искусственного интеллекта, они обязаны включить эту информацию в текст статьи.

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

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

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или \LaTeX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 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 почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp и
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний. Если источник имеет DOI и/или EDN, то их необходимо указывать.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

Описание статьи из журнала:

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

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

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — New York, NY, USA: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. New York: Wiley. 521 p.

Описание неопубликованного документа:

Laturov, A. R., M. M. Khasanov, and V. A. Baikov. 2004 (unpubl.). *Geologiya i dobycha (NGT GiD)* [Geology and production (NGT GiD)]. Certificate on official registration of the computer program No. 2004611198.

Описание интернет-ресурса:

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

Описание диссертации или автореферата диссертации:

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

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

Описание ГОСТа:

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

Описание патента:

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

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- 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 «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:
Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН
Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05
e-mail: iiep@frccsc.ru (Стригина Светлана Николаевна)
<http://www.ipiran.ru/journal/issues/>

Requirements for manuscripts submitted to Journal “Informatics and Applications”

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

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

The topics covered include the following areas:

- theoretical fundamentals of informatics;
- mathematical methods for studying complex systems and processes;
- information systems and networks;
- information technologies; and
- architecture and software of computational complexes and networks.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations. Author(s) signature(s): (name(s), address(es), date).

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

If authors used artificial intelligence (AI)-based tools in preparing their manuscript, they must include this information in the text of the article.

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or L^AT_EX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 10 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site:

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

8. The articles should enclose data both in **Russian and English**:

- title;
- author’s name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format (see site):

http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp and

http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;

- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;
 - indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.
Important! Keywords must not be sentences; and
 - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.net/ru/bgn/>).
- Please take into account the following examples of Russian references appearance:
- Article in journal:**
Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.
- Journal article in electronic format:**
Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).
- Article from the continuing publication (collection of works, proceedings):**
Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy "stal'-kompozit" [Experimental study of the strength of joints "steel-composite"]. *Trudy MGTU "Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem" [Bauman MSTU "Mathematical Modeling of Complex Technical Systems" Proceedings]*. 593:125–130.
- Conference proceedings:**
Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya 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 slovarya informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.
- State standards and patents:**
GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform 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 email;
 - an article and additional materials have to be attached using the "attach" function; and
 - an electronic version of the article should contain the file with the text and a separate file with figures.
12. "Informatics and Applications" journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia
Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05
e-mail: iiep@frccsc.ru (to Svetlana Strigina)
<http://www.ipiran.ru/english/journal.asp>