ISSN 0002-3388

Номер 4

Июль - Август 2024

ИЗВЕСТИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК

ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ

НАУКА

1727

СОДЕРЖАНИЕ

Номер 4, 2024

КОМПЬЮТЕРНЫЕ МЕТОДЫ

ДИСКРЕТНЫЕ СИСТЕМЫ

СИСТЕМНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

НАВИГАЦИОННЫЕ СИСТЕМЫ

CONTENTS

Number 4, 2024

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, с. 3–32

КОМПЬЮТЕРНЫЕ МЕТОДЫ

УДК 004.021, 519.677

ПРЕДОБУСЛОВЛИВАТЕЛЬ БЫСТРОГО НОРМАЛЬНОГО РАЗЛОЖЕНИЯ ДЛЯ ПРИТЯГИВАЮЩИХ СВЯЗАННЫХ НЕЛИНЕЙНЫХ УРАВНЕНИЙ ШРЕДИНГЕРА С ДРОБНЫМ ЛАПЛАСИАНОМ[1]

© 2024 г. Я. Ченг[a], *, Ш. Янг[a], И. А. Матвеев[b], **

[a]*Колледж математики Нанкинского ун-та аэронавтики и космонавтики, Нанкин, КНР*

[b]*ФИЦ ИУ РАН, Москва, Россия*

**e-mail: lyandcxh@nuaa.edu.cn*

***e-mail: matveev@frccsc.ru*

Поступила в редакцию 18.03.2024 г.

После доработки 20.03.2024 г.

Принята к публикации 13.05.2024 г.

Линейная консервативная разностная схема применяется для дискретизации притягивающих связанных нелинейных уравнений Шредингера с дробным лапласианом. В этом случае возникают сложные симметричные линейные системы, матрицы которых неопределенны и теплиц-плюс-диагональны. Стандартные быстрые методы прямого решения или итераций с использованием предобусловливателя не применимы для таких систем. Предлагается новый итерационный метод, основанный на нормальном разложении эквивалентной вещественной блочной формы линейных систем. Доказывается безусловная сходимость, определяется квазиоптимальный параметр итерации. Предобусловливатель для данного метода получается естественным путем, он строится и эффективно реализуется с помощью быстрого преобразования Фурье. Теоретический анализ показывает, что собственные значения предобусловленной матрицы системы тесно кластеризованы. Численные эксперименты показывают, что новый предобусловливатель значительно ускоряет скорость сходимости итерационных методов подпространства Крылова. В частности, поведение сходимости соответствующего предобусловленного итерационного метода минимальной невязки не зависит от размера пространственной сетки и почти нечувствительно к дробному порядку. Более того, линейно неявная консервативная разностная схема в этом случае сохраняет массу и энергию с заданной точностью.

Ключевые слова: циркулянтная матрица, нелинейные уравнения Шредингера, дробный лапласиан, предобусловливание, матрица Теплица.

DOI: 10.31857/S0002338824040014 **EDN:** UERATM

A FAST NORMAL SPLITTING PRECONDITIONER FOR ATTRACTIVE COUPLED NONLINEAR SCHROEDINGER EQUATIONS WITH FRACTIONAL LAPLACIAN

Y. Cheng[a], *, X. Yang[a], I. A. Matveev[b], **

[a]*College of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing, China*

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russia

**e-mail: lyandcxh@nuaa.edu.cn*

***e-mail: matveev@frccsc.ru*

A linearly implicit conservative difference scheme is applied to discretize the attractive coupled nonlinear Schroedinger equations with fractional Laplacian. In this case complex symmetric linear systems appear, with indefinite and Toeplitz-plus-diagonal system matrices. Standard fast methods of direct solution or iteration using a preconditioner are not applicable for such systems. A novel iterative method is proposed, based

[1] Работа выполнена при частичной финансовой поддержке Государственного фонда естественных наук Китая (гранты № 11101213; 12071215).

on the normal splitting of the equivalent real block form of linear systems. Unconditional convergence is proved and the quasi-optimal iteration parameter is deducted. The preconditioner for this method is obtained naturally; it is constructed and efficiently implemented using the fast Fourier transform. Theoretical analysis shows that the eigenvalues of the preconditioned system matrix are closely clustered. Numerical experiments demonstrate new preconditioner significantly speeds up the convergence rate of iterative Krylov subspace methods. In particular, the convergence behavior of the corresponding preconditioned generalized minimum residual method is independent of the mesh size and almost insensitive to the fractional order. Moreover, the linearly implicit conservative difference scheme in this case preserves mass and energy with a given accuracy.

Keywords: circulant matrix, nonlinear Schroedinger equations, fractional Laplacian, preconditioning, Toeplitz matrix.

Введение. Одним из самых популярных методов расчета в квантовой механике [1,2] является интеграл Фейнмана по броуновским путям, который приводит к стандартным уравнениям Шредингера. Если вместо броуновского движения рассматривать полет Леви, то возникает система дробных уравнений Шредингера (fractional Schroedinger equations, FSE) [3]. Вместо лапласиана в стандартных уравнениях Шредингера [4] FSE включает дробную производную порядка α ($1 < \alpha < 2$). При $\alpha = 2$ FSE сводится к стандартному случаю. Кроме того, FSE имеет важные приложения в квантовой механике, полупроводниках и других областях [5]. Из-за нелокальной природы оператора дробного дифференциала точное решение FSE получить трудно. Для изучения природы FSE было разработано большое количество численных методов, таких, как методы конечных элементов [6,7], спектральные методы [8,9], методы коллокации [10,11], методы конечных разностей [12–16] и пр.

В статье рассматриваются следующие одномерные (1D) пространственные дробно-связанные нелинейные уравнения Шредингера (coupled nonlinear Schroedinger, CNLS):

$$\begin{cases} ru_t - \gamma\left(-\Delta\right)^{\frac{\alpha}{2}} u + \rho\left(|u|^2 + \beta|v|^2\right)u = 0, \\ ru_t - \gamma\left(-\Delta\right)^{\frac{\alpha}{2}} v + \rho\left(|v|^2 + \beta|u|^2\right)u = 0, \end{cases} \quad x \in \mathbb{R}, 0 < t \le T, \tag{0.1}$$

с начальными условиями

$$u(x,0) = u_0(x), \quad u(x,0) = v_0(x), \quad x \in \mathbb{R}, \tag{0.2}$$

где $1 = \sqrt{-1}$, параметры $\gamma > 0$, $\beta \geqslant 0$, ρ – вещественные константы, дробный лапласиан

$$\left(-\Delta\right)^{\frac{\alpha}{2}} = -\frac{\partial^\alpha}{\partial|x|^\alpha} \tag{0.3}$$

определяется дробной производной одномерного пространства Рисса [17,18] при $1 < \alpha < 2$. При параметре $\beta = 0$ можно получить несвязанные нелинейные уравнения Шредингера (decoupled nonlinear Schroedinger, DNLS) [19]. Что касается параметра ρ, то он определяет три различных варианта нелинейного члена FSE. При $\rho = 0$ нелинейные члены исчезают и система (0.1) описывает свободные частицы [17,20]. При $\rho < 0$ нелинейные члены в системе (0.1) представляют собой отталкивающее взаимодействие частиц [21–23]. При $\rho > 0$ нелинейные члены в системе (0.1) представляют собой притягивающее взаимодействие частиц [21,24,25]. В данной статье рассматривается только $\rho > 0$, в этом случае матрица коэффициентов дискретизированной линейной системы является комплексной симметричной неопределенной. Случай $\rho < 0$, приводящий к сложным симметричным отрицательно определенным линейным системам, обсуждается в [26–30].

Теоретически решение FSE сохраняет массу и энергию, например система (0.1). В частности, решения u(x,t),v(x,t) задачи (0.1) удовлетворяют сохранению массы [13]:

$$\begin{aligned} \left\|u(\cdot,t)\right\|_{L_2}^2 &= \int_{\mathbb{R}} |u(x,t)|^2 dx = \left\|u(\cdot,0)\right\|_{L_2}^2, \\ \left\|v(\cdot,t)\right\|_{L_2}^2 &= \int_{\mathbb{R}} |v(x,t)|^2 dx = \left\|u(\cdot,0)\right\|_{L_2}^2, \end{aligned} \tag{0.4}$$

и сохранению энергии:

$$E(t) = \frac{\gamma}{2}\int_{\mathbb{R}}\left(\bar{u}(-\Delta)^{\frac{\alpha}{2}}u + \bar{v}(-\Delta)^{\frac{\alpha}{2}}v\right)dx - \frac{\rho}{4}\int_{\mathbb{R}}\left(|u|^4 + |v|^4 + 2\beta|u|^2|v|^2\right)dx, \tag{0.5}$$

где функционал энергии удовлетворяет $E(\mathrm{t}) = E(0)$. Вышеуказанные законы сохранения остаются в силе и при дискретизации системы (0.1).

В этой статье вместо всей вещественной оси $\mathbb{R}$ рассмотрение ведется на ограниченном интервале $a \leqslant x \leqslant b$ и приняты граничные условия Дирихле:

$$u(a,t) = u(b,t) = 0, \quad v(a,t) = v(b,t) = 0, \quad 0 \le t \le T.$$

Для дискретизации дробных пространственных уравнений CNLS (0.1) используется схема линейно-неявной консервативной разности (linearly implicit conservative difference, LICD), предложенная в [13]. На каждом временном уровне необходимо решать сложные симметричные линейные системы с теплиц-плюс-диагональной структурой $(D - T + \imath I)\,\mathbf{u} = \mathbf{b}$, где D – диагональная, а T – теплицева симметричная положительно определенная матрица. При $\rho > 0$ матрица D положительно полуопределена, а $D - T$ неопределена. Тогда $D - T + \imath I$ является комплексно-симметричной неопределенной.

Существует множество численных методов для работы со сложными симметричными линейными системами. Первый класс методов – это методы неявной итерации с чередующимся направлением [31–35]. В [31,32] были предложены итерационные методы эрмитова и косоэрмитова разложения (skew-hermitian splitting, HSS) для работы с общими неэрмитовыми положительно определенными линейными системами. Кроме того в этих публикациях были разработаны модифицированные итерационные методы HSS (modified HSS, MHSS) [33] и предварительно обусловленные итерационные методы MHSS (preconditioned MHSS, PMHSS) [34,35] с учетом симметричной структуры матрицы системы. Второй класс методов – итерационные C-to-R [36], которые требуют высококачественных предобусловливателей для дополнения Шура для обеспечения устойчивости. Построение дополнения Шура затруднено, поскольку матрица дополнения Шура плотна и имеет вид $S + \widetilde{S}^{-1}$, где $S,\ \widetilde{S} \in \mathbb{R}^{M\times M}$ – плотные матрицы с теплиц-плюс-диагональной структурой. К сожалению, методы типа HSS и итерации C-to-R не могут быть применены для прямой обработки комплексной симметричной неопределенной матрицы $D - T + \imath I$, поскольку они предназначены для неэрмитовой положительно определенной матрицы.

Для линейных систем с теплиц-плюс-диагональной структурой быстрые прямые решатели не разработаны. Следовательно, лучшим вариантом является использование предварительно обусловленных методов итерации подпространства Крылова [37,38]. Реализация умножения матрицы на вектор и решение линейной системы с обобщенной невязкой (generalized residual, GR) всегда необходимы на каждой итерации предобусловливаемых методов итерации подпространства Крылова. Если существуют быстрые реализации умножения матрицы на вектор и можно эффективно сконструировать и реализовать высококачественные предобусловливатели, итерации подпространства Крылова будут работать очень хорошо. Многие эффективные предобусловливатели работают с эрмитовыми положительно определенными теплиц-плюс-диагональными системами. Это, например, класс полосовых (banded) предобусловливателей для эрмитовых систем «теплиц-плюс-полоса» [39]; класс приближенных предобусловливателей обратного циркулянта-плюс-диагонали (approximate inverse circulant-plus-diagonal, AICD) [40]; предобусловливатели диагонального и теплицевого разложения (diagonal and toeplitz splitting, DTS), а также предобусловливатели диагонального и циркулянтного разложения (diagonal and circulant splitting, DCS), предложенные в [41] (1D-случай) и [42] (многомерный случай). Тем не менее, никакие предобусловливатели, описанные выше, не могут быть напрямую применены для решения сложных линейных систем $(D - T + \imath I)\,\mathbf{u} = \mathbf{b}$, поскольку они предназначены для эрмитовых положительно определенных систем, но не для неопределенных симметричных теплиц-плюс-диагональных.

Для разработки эффективных итерационных методов рассматривается вещественная блочная (два на два) форма линейной системы $(D - T + \imath I)\,\mathbf{u} = \mathbf{b}$. Матрица блочной системы раскладывается на нормальную блочную матрицу и антисимметричную блочную матрицу и строится класс итерационных методов нормального и антисимметричного разложения (normal and anti-symmetric splitting, NASS), основанный на подходе неявных итерационных методов чередующегося направления (alternating direction implicit, ADI) [43,44]. На каждом шаге итерационного метода NASS используются две линейные подсистемы с матрицами коэффициентов:

$$\begin{bmatrix} (\omega+1)I & T \\ -T & (\omega+1)I \end{bmatrix}, \quad \begin{bmatrix} \omega I & -D \\ D & \omega I \end{bmatrix}, \tag{0.6}$$

которые необходимо решить. Первое можно решить итеративно с помощью метода предобусловленного обобщенного метода минимальной невязки (generalized minimal residual method, GMRES) с блочным предобусловливателем:

$$\begin{bmatrix} (\omega+1)I & C \\ -C & (\omega+1)I \end{bmatrix} \tag{0.7}$$

с циркулянтной аппроксимацией C. Каждая итерация предобусловленного GMRES может быть выполнена за $\mathcal{O}(M \log M)$ операций с плавающей точкой (флоп, flops), если применяется быстрый алгоритм, например быстрое преобразование Фурье (БПФ) [37]. Второе можно решить непосредственно с помощью разреженного исключения Гаусса (Gaussian elimination, GE) за $\mathcal{O}(M)$ флоп. Теоретически итерационный метод NASS безусловно сходится к единственному решению. Получена верхняя оценка коэффициента сжатия итерационного метода NASS, зависящая только от спектров симметричной положительно определенной матрицы Теплица $T \in \mathbb{R}^{M\times M}$. Оптимальное значение параметра итерации, минимизирующее верхнюю границу коэффициента сжатия итерационного метода NASS, определяется нижней и верхней границами собственных значений T.

Класс предобусловливателей разложения матрицы, называемый предобусловливателями NASS, может быть получен естественным путем из итерационного метода NASS. Чтобы уменьшить вычислительную сложность предобусловливателей NASS, для замены блоков Теплица используются циркулянтные аппроксимации, что приводит к созданию класса более экономичных циркулянтных улучшенных нормальных и антисимметричных предобусловливателей (circulant improved normal and anti-symmetric, CNAS). Реализация предобусловливателя CNAS требует разрешения двух линейных подсистем. Матрицы подсистем имеют вид

$$\begin{bmatrix} (\omega+1)I & C \\ -C & (\omega+1)I \end{bmatrix}, \quad \begin{bmatrix} \omega I & -D \\ D & \omega I \end{bmatrix}. \tag{0.8}$$

Первое можно решить за $\mathcal{O}(M \log M)$ флоп на основе БПФ, а второе можно решить непосредственно с помощью разреженного GE за $\mathcal{O}(M)$ флоп. Теоретический анализ показывает, что собственные значения матрицы предварительно обусловленной системы NASS группируются около 1. Кроме того, можно доказать, что предварительное обусловливание системы CNAS представляет собой небольшое возмущение нормы и коррекцию низкого ранга матрицы предварительно обусловленной системы NASS. Это значит, что большинство собственных значений матрицы предобусловленной системы CNAS сгруппированы вокруг аналогов для NASS, т.е. могут быть расположены около 1. Численные эксперименты показывают, что предобусловливатель CNAS значительно повышает вычислительную эффективность методов итерации подпространства Крылова (таких, как GMRES), а соответствующий предобусловленный метод (PGMRES) ведет себя независимо от размера пространственной сетки и нечувствителен к дробному порядку. Кроме того, замечено, что дискретная масса и энергия сохраняются по схеме LICD в сочетании с предварительно обусловленным методом CNAS GMRES с точки зрения заданной точности.

Изложение построено следующим образом. В разд. 1 схема LICD применяется для дискретизации дробно-пространственных уравнений CNLS, что приводит к неопределенной комплексной симметричной линейной системе $(D - T + \varepsilon I)\,\mathbf{u} = \mathbf{b}$. В разд. 2 представлен итерационный метод NASS и установлена соответствующая асимптотическая теория сходимости. В разд. 3 создаются и анализируются новые предобусловливатели. В разд. 4 рассматриваются и обсуждаются реализации и вычислительные сложности новых предобусловливателей. В разд. 5 приведены числовые результаты и даются заключительные замечания.

1. Дискретизация и линейная система. Дробные по пространству уравнения CNLS (0.1) дискретизируются по схеме LICD [13]. Пусть N и M – целые положительные числа, через $\tau = T/N$ и $h = (\mathrm{b} - \mathrm{a})/(M + 1)$ заданы размеры временных шагов и пространственных сеток соответственно. Тогда $t_n = n\tau$ для $n = 0, N$ и $x_j = \mathrm{a} + jh$ для $j = 0, M + 1$ представляют временные уровни и узлы пространственных ячеек.

Пусть $u_j^n \approx u(x_j,t_n)$ и $v_j^n \approx v(x_j,t^n)$ – аппроксимации сеточной функции. Одномерный дробный лапласиан $(-\Delta)^{\frac{\alpha}{2}}$ можно дискретизировать дробной центрированной разностью [45,46] в ограниченном пространственном интервале [a, b] как

$$(-\Delta)^{\frac{\alpha}{2}} u\left(x_j,t\right) = \frac{1}{h^\alpha}\Delta_h^\alpha u_j + \mathcal{O}\left(h^2\right).$$

Здесь

$$\Delta_h^\alpha u_j = \sum_{k=1}^{M} c_{j-k}u_k \tag{1.1}$$

с коэффициентами

$$c_k = (-1)^k\,\Gamma(\alpha+1)\Big/\left[\Gamma(\alpha/2-k+1)\Gamma(\alpha/2+k+1)\right],\ \forall\, k\in\mathbb{Z},$$

где $\Gamma(\cdot)$ – гамма-функция. Кроме того, коэффициенты c_k удовлетворяют следующим свойствам [45]:

$$c_0 \ge 0,\quad c_k = c_{-k} \le 0\,(\forall\, k \ge 1),\quad \sum_{k=-\infty,k\ne 0}^{+\infty} |c_k| = c_0.$$

Схема LICD [13] для системы (0.1) выглядит следующим образом:

$$\begin{cases} l\dfrac{u_j^{n+1}-u_j^{n-1}}{2\tau} - \dfrac{\gamma}{h^\alpha}\Delta_h^\alpha \hat{u}_j^n + \rho\left(\left|u_j^n\right|^2 + \beta\left|v_j^n\right|^2\right)\hat{u}_j^n = 0,\\ l\dfrac{v_j^{n+1}-v_j^{n-1}}{2\tau} - \dfrac{\gamma}{h^\alpha}\Delta_h^\alpha \hat{v}_j^n + \rho\left(\left|v_j^n\right|^2 + \beta\left|u_j^n\right|^2\right)\hat{v}_j^n = 0, \end{cases} \tag{1.2}$$

где $\hat{u}_j^n = (u_j^{n+1} + u_j^{n-1})/2$, $\hat{u}_j^n = (v_j^{n+1} + v_j^{n-1})/2$, для $j = 1, 2,...,M, n = 1, 2,...,N-1$. Начальные и граничные условия: $u_j^0 = u_0(x_j)$, $v_j^0 = v_0(x_j)$, $u_0^n = u_{M+1}^n = 0$, $v_0^n = v_{M+1}^n = 0$. Эквивалентно первая строка (1.2) приводит к сложной линейной системе:

$$A^{n+1}\mathbf{u}^{n+1} = \mathbf{b}^{n+1},\ \forall\, n \ge 1, \tag{1.3}$$

где $A^{n+1} = D^{n+1} - T + \imath I \in \mathbb{C}^{M\times M}$, $D^{n+1} = \mathrm{diag}\{d_1^{n+1}, d_2^{n+1},..., d_M^{n+1}\} \in \mathbb{R}^{M\times M}$ диагонально с $d_j^{n+1} = \rho\tau(|u_j^n|^2 + \beta|v_j^n|^2)$, $I \in \mathrm{R}^{M\times M}$ – единичная матрица, а $T \in \mathbb{R}^{M\times M}$ является теплицевой:

$$T = \mu\begin{bmatrix} c_0 & c_{-1} & \cdots & c_{2-M} & c_{1-M} \\ c_1 & c_0 & \ddots & \ddots & c_{2-M} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c_{M-2} & \ddots & \ddots & c_0 & c_{-1} \\ c_{M-1} & c_{M-2} & \cdots & c_1 & c_0 \end{bmatrix} \tag{1.4}$$

с $\mu = \gamma\tau/h^\alpha$. Очевидно, что вторая строка (1.2) также допускает связанную линейную систему, такую же, как (1.3).

Ввиду того, что $\rho > 0$, $\gamma > 0$, $\beta \geqslant 0$ и свойств коэффициентов c_k, D^{n+1} – неотрицательная диагональная матрица, а T строго диагонально доминантна и симметрична. Следовательно, $D^{n+1} - T$ симметричная неопределенная, а A^{n+1} комплексно симметричная и неопределенная.

В [13] доказано, что дискретизированные связанные линейные системы (1.2) сохраняют дискретную массу и энергию, т.е.

$$Q_1^n = Q_1^0,\quad Q_2^n = Q_2^0,\quad E^n = E^0,\quad 1 \le n \le N, \tag{1.5}$$

где $Q_1^n = (\|u^{n+1}\|^2 + \|u^n\|^2)/2$ и $Q_2^n = (\|v^{n+1}\|^2 + \|v^n\|^2)/2$ с $\|u\|^2 = \langle u,u\rangle$ и

$$\langle u,v\rangle = h\sum_{k=1}^{M-1} u_j\,\bar{v}_j,$$

$$E^n = \frac{\gamma n}{4h^\alpha}\sum_{j=1}^{M-1}\left(\bar{u}_j^{n+1}\Delta_h^\alpha u_j^{n+1} + \bar{u}_j^n\Delta_h^\alpha u_j^n + \bar{v}_j^{n+1}\Delta_h^\alpha v_j^{n+1} + \bar{v}_j^n\Delta_h^\alpha v_j^n\right) -$$
$$-\frac{\rho h}{4}\sum_{j=1}^{M-1}\left[\left(\left|u_j^n\right|^2\left|u_j^{n+1}\right|^2 + \left|v_j^n\right|^2\left|v_j^{n+1}\right|^2\right) + \beta\left|u_j^n\right|^2\left|v_j^{n+1}\right|^2 + \left|v_j^n\right|^2\left|u_j^{n+1}\right|^2\right]. \quad (1.6)$$

2. Итеративный метод NASS. Рассматривается построение итерационного метода для сложных линейных систем вида (1.3). Если игнорировать верхние индексы, он записывается как

$$A\mathbf{u} = \mathbf{b}, \quad (2.1)$$

где $A = D - T + \imath I \in \mathbb{C}^{M\times M}$ комплексно-симметричная и неопределенная, причем $D \in \mathbb{R}^{M\times M}$ – положительная полуопределенная диагональная матрица и $T \in \mathbb{R}^{M\times M}$ – симметричная положительно определенная теплицевая матрица. Комплексный неизвестный вектор имеет вид $\mathbf{u} = y + \imath z \in \mathbb{C}^M$, а комплексный вектор $\mathbf{b} = p + \imath q \in \mathbb{C}^M$ – правая часть, где $y, z, p, q \in \mathbb{R}^M$ – вещественные векторы. Затем сложную линейную систему (2.1) можно преобразовать в следующую вещественную блочную линейную систему:

$$\widehat{\mathcal{R}}\hat{x} \equiv \begin{bmatrix} D-T & -I \\ I & D-T \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} p \\ q \end{bmatrix} \equiv \widehat{f} \quad (2.2)$$

где $\widehat{\mathcal{R}} \in \mathbb{R}^{2M\times 2M}$ несимметрично и неопределенно.

Определим перестановки блоков

$$\mathcal{P} = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix}, \quad \mathcal{Q} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \quad (2.3)$$

и пусть

$$\mathcal{R} = \mathcal{P}\widehat{\mathcal{R}}\mathcal{Q}, \quad x = \mathcal{Q}\hat{x}, \quad f = \mathcal{P}\widehat{f}.$$

Тогда вещественная блочная линейная система (2.2) эквивалентна вещественной несимметричной положительно определенной блочной линейной системе следующим образом:

$$\mathcal{R}x \equiv \begin{bmatrix} I & T-D \\ D-T & I \end{bmatrix}\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} -p \\ q \end{bmatrix} \equiv f. \quad (2.4)$$

Матрица $\mathcal{R} \in \mathbb{R}^{2M\times 2M}$ допускает следующее нормальное и антисимметричное (NAS) разложение:

$$\mathcal{R} = \mathcal{T} + \mathcal{D}, \quad (2.5)$$

$$\mathcal{T} = \begin{bmatrix} I & T \\ -T & I \end{bmatrix}, \quad \mathcal{D} = \begin{bmatrix} 0 & -D \\ D & 0 \end{bmatrix}. \quad (2.6)$$

Основываясь на разложении NAS (2.5) и духе неявной итерации с чередующимся направлением (ADI) [43,44], можно построить метод итерации NASS следующим образом.

Итерационный метод NASS. Пусть $x^{(0)} \in \mathbb{R}^{2M}$ – произвольное начальное предположение. Для $k = 0, 1, 2,...$, пока последовательность итераций $\{x^{(k)}\}$ $k\geqslant 0$ не сойдется, вычислить следующую итерацию $x^{(k+1)}$, согласно следующей процедуре:

$$\begin{cases} (\omega I + \mathcal{T})x^{\left(k+\frac{1}{2}\right)} = (\omega I - \mathcal{D})x^{(k)} + f, \\ (\omega I + \mathcal{D})x^{(k+1)} = (\omega I - \mathcal{T})x^{\left(k+\frac{1}{2}\right)} + f, \end{cases} \quad (2.7)$$

где $\omega > 0$ – произвольный параметр итерации.

Итерационная схема (2.7) может быть переформулирована для итерации с фиксированной точкой как $x^{(k+1)} = \mathcal{L}_\omega x^{(k)} + \mathcal{F}_\omega^{-1} f$, где матрица итераций это

$$L_\omega = \mathcal{F}_\omega^{-1}\mathcal{G}_\omega. \tag{2.8}$$

Здесь

$$\mathcal{F}_\omega = \frac{1}{2\omega}(\omega I + \mathcal{T})(\omega I + \mathcal{D}) \tag{2.9}$$

и

$$\mathcal{G}_\omega = \frac{1}{2\omega}(\omega I - \mathcal{T})(\omega I - \mathcal{D}).$$

Действительно, матрицы $\mathcal{F}_\omega$ и $\mathcal{G}_\omega$. образуют разложение $\mathcal{R}$, т.е.

$$\mathcal{R}_\omega = \mathcal{F}_\omega - \mathcal{G}_\omega. \tag{2.10}$$

В следующей теореме обеспечивается свойство сходимости итерационного метода NASS.

Т е о р е м а 1. Пусть $\mathcal{R} \in \mathbb{R}^{2M\times 2M}$ – матрица системы (2.4), $\mathcal{T} \in \mathbb{R}^{2M\times 2M}$ и $\mathcal{D}\in \mathbb{R}^{2M\times 2M}$ – блочные матрицы, образующие разбиение $\mathcal{R}$ в (2.5). Если ω – положительная константа, то итерационный метод NASS (2.7) корректно определен, а матрица итераций $\mathcal{L}_\omega$ задается (2.8). Спектральный радиус $\rho(\mathcal{L}_\omega)$ ограничен:

$$\sigma(\omega) = \max_{\lambda_i \in \lambda(T)} \sqrt{\frac{(\omega-1)^2 + \lambda_i^2}{(\omega+1)^2 + \lambda_i^2}}, \tag{2.11}$$

где λ(T) – спектр T. Таким образом, можно записать

$$\rho(\mathcal{L}_\omega) \le \sigma(\omega) < 1, \forall\, \omega > 0, \tag{2.12}$$

т. е. итерация NASS сходится к единственному решению блочной линейной системы (2.4).

Д о к а з а т е л ь с т в о. Поскольку $\mathcal{T}$ положительно определен, $\mathcal{D}$ антисимметричен и $\omega > 0$, то $\omega I + \mathcal{T}$ и $\omega I + \mathcal{D}$ обратимы и положительно определены. Таким образом, итерация NASS (2.7) четко определена. Заметим, что итерационная матрица $\mathcal{L}_\omega$ удовлетворяет условию $\mathcal{L}_\omega = (\omega I + \mathcal{D})^{-1}\mathcal{U}_\omega\mathcal{V}_\omega(\omega I + \mathcal{D})$, где $\mathcal{U}_\omega = (\omega I + \mathcal{T})^{-1}(\omega I - \mathcal{T})$ и $\mathcal{V}_{\text{omega}} = (\omega I - \mathcal{D})(\omega I + \mathcal{D})^{-1}$, можно установить следующее соотношение:

$$\rho(\mathcal{L}_\omega) = \rho(\mathcal{U}_\omega\mathcal{V}_\omega) \le \|\mathcal{U}_\omega\|_2 \|\mathcal{V}_\omega\|_2. \tag{2.13}$$

Теперь сосредоточимся на оценке $\|\mathcal{U}_\omega\|_2$ и $\|\mathcal{V}_\omega\|_2$. Во-первых, легко проверить, что $\mathcal{T}^\top\mathcal{T} = \mathcal{T}\mathcal{T}^\top$, тогда имеем

$$\begin{aligned}
\mathcal{U}_\omega{}^\top\mathcal{V}_\omega &= (\omega I - \mathcal{T})^\top(\omega I + \mathcal{T})^{-\top}(\omega I + \mathcal{T})^{-1}(\omega I - \mathcal{T}) = \\
&= (\omega I + \mathcal{T})^{-\top}(\omega I + \mathcal{T})^{-1}(\omega I - \mathcal{T})^\top(\omega I - \mathcal{T}) = \\
&= \left[(\omega I + \mathcal{T})(\omega I + \mathcal{T})^\top\right]^{-1}(\omega I - \mathcal{T})^\top(\omega I - \mathcal{T}) = \\
&= \begin{bmatrix} (\omega+1)^2 I + T^2 & 0 \\ 0 & (\omega+1)^2 I + T^2 \end{bmatrix}_{-1} \begin{bmatrix} (\omega-1)^2 I + T^2 & 0 \\ 0 & (\omega-1)^2 I + T^2 \end{bmatrix}.
\end{aligned} \tag{2.14}$$

Ввиду того, что T симметрична положительно определена, справедливо равенство

$$\|\mathcal{U}_\omega\|_2 = \max_{\lambda_i \in \lambda(T)} \sqrt{\frac{(\omega-1)^2 + \lambda_i^2}{(\omega+1)^2 + \lambda_i^2}} < 1. \tag{2.15}$$

Во-вторых, рассмотрим оценку $\|\mathcal{V}_\omega\|_2$. Поскольку $\mathcal{D}^\top\mathcal{D} = \mathcal{D}\mathcal{D}^\top$, то следует, что

$$\mathcal{V}_\omega^\top \mathcal{V}_\omega = \left(\omega I + \mathcal{D}^\top\right)^{-1}\left(\omega I - \mathcal{D}^\top\right)(\omega I - \mathcal{D})(\omega I + \mathcal{D})^{-1} =$$

$$= \left(\omega I - \mathcal{D}^\top\right)(\omega I - \mathcal{D})\left[(\omega I + \mathcal{D})\left(\omega I + \mathcal{D}^\top\right)\right]^{-1} =$$

$$= \begin{bmatrix} \omega^2 I + D^2 & 0 \\ 0 & \omega^2 I + D^2 \end{bmatrix}\begin{bmatrix} \omega^2 I + D^2 & 0 \\ 0 & \omega^2 I + D^2 \end{bmatrix}^{-1} = I.$$

Тогда

$$\left\|\mathcal{V}_\omega\right\|_2 = 1. \tag{2.16}$$

На основании (2.13)–(2.16) можно получить оценку (2.12).

З а м е ч а н и е 1. Для дальнейшего понимания теоретического результата, изложенного в приведенной выше теореме, сделаем следующие замечания.

1. Скорость сходимости итерации NASS ограничена величиной $\sigma(\omega)$ и зависит только от спектра теплицевой матрицы T.

2. Если ввести взвешенную векторную норму $\|x\|_{\mathcal{D}} = \|(\omega I + \mathcal{D})x\|_2$ для $x \in \mathbb{R}^{2M}$ и соответствующую норму индуцированной взвешенной матрицы $\|\mathcal{X}\|_{\mathcal{D}} = \|(\omega I + \mathcal{D})\mathcal{X}(\omega I + \mathcal{D})^{-1}\|_2$ для $\mathcal{X} \in \mathbb{R}^{2M\times 2M}$ имеет место равенство

$$\left\|\mathcal{L}_\omega\right\|_{\mathcal{D}} = \left\|\mathcal{U}_\omega\mathcal{V}_\omega\right\|_2 < \sigma(\omega), \quad \forall\, \omega > 0.$$

Более того, на основе формы итерации с фиксированной точкой итерации NASS можно проверить, что последовательность итераций $\{x^{(k)}\}$ $k \geqslant 0$ удовлетворяет условию

$$\left\|x^{(k+1)} - x^*\right\|_{\mathcal{D}} \leq \sigma(\omega)\left\|x^{(k)} - x^*\right\|_{\mathcal{D}}, \quad \forall\, k \geq 0.$$

Следовательно, $\sigma(\omega)$ также служит верхней границей коэффициента сжатия итерации NASS в смысле $\|\cdot\|_{\mathcal{D}}$-нормы. Замечено, что при $TD = DT$ выполняется $\mathcal{U}_\omega\mathcal{V}_\omega = \mathcal{V}_\omega\mathcal{U}_\omega$, что приводит к факту $\rho(\mathcal{L}_\omega) = \|\mathcal{L}_\omega\|_{\mathcal{D}} = \sigma(\omega)$. Тогда эти величины можно минимизировать при том же оптимальном значении ω.

3. Верхняя граница $\sigma(\omega)$ может быть дополнительно ограничена величиной $\hat{\sigma}(\omega)$, т.е.

$$\sigma(\omega) \leq \max_{\lambda_{\min}\leq\lambda\leq\lambda_{\max}} \sqrt{\frac{(\omega-1)^2 + \lambda_i^2}{(\omega+1)^2 + \lambda_i^2}} \equiv \hat{\sigma}(\omega),$$

где $\lambda_{\min} > 0$ и $\lambda_{\max} > 0$ – минимальное и максимальное собственные значения T. Поскольку функция $g(\omega;\lambda) = \sqrt{\left[(\omega-1)^2 + \lambda^2\right]\Big/\left[(\omega+1)^2 + \lambda^2\right]}$ монотонно возрастает при $\lambda > 0$, она имеет вид

$$\hat{\sigma}(\omega) = \sqrt{\frac{(\omega-1)^2 + \lambda_{\max}^2}{(\omega+1)^2 + \lambda_{\max}^2}}.$$

Пусть

$$\frac{(\hat{\sigma}(\omega))}{\omega} = \frac{2\left(\omega^2 - 1 - \lambda_{\max}^2\right)}{\left(\omega-1\right)^2 + \lambda_{\max}^2\ {}^{\frac{1}{2}}\quad \left(\omega+1\right)^2 + \lambda_{\max}^2\ {}^{\frac{3}{2}}} = 0. \tag{2.17}$$

тогда $\hat{\sigma}(\omega)$ минимизируется в $\omega^\star = \sqrt{\lambda_{\max}^2 + 1}$. Если применяется параметр $\omega = \omega^\star$, скорость сходимости итерации NASS для решения блочной линейной системы (2.4) ограничивается следующим образом:

$$\sigma(\omega^*) \leq \hat{\sigma}(\omega^*) = \frac{\lambda_{\max}}{1 + \sqrt{\lambda_{\max}^2 + 1}}.$$

3. Предварительное обусловливание. Итерационный метод NASS естественным образом вызывает матрицу предварительной обработки $\mathcal{F}_\omega$, определенную в (2.9) для матрицы $\mathcal{R}$ в (2.4). Соответствующая предобусловленная линейная система имеет вид

$$\mathcal{F}_\omega^{-1}\mathcal{R}x = \mathcal{F}\omega^{-1}f. \tag{3.1}$$

предобусловливатель $\mathcal{F}_\omega$ называется предобусловливателем NASS. Очевидно, $\mathcal{F}_\omega$ есть произведение блочных нормальных матриц $\omega I + \mathcal{T}$, $\omega I + \mathcal{D}$ и скалярного коэффициента $1/(2\omega)$.

В реализации основной задачей предварительной подготовки NASS является решение последовательности линейной системы с обобщенной невязкой:

$$\mathcal{F}_\omega z^{(k)} = r^{(k)}, \forall\, k \geq 0, \tag{3.2}$$

где $r^{(k)}$ – текущий вектор невязки, а $z^{(k)}$ – вектор обобщенной невязки. Разрешение (3.2) состоит из решения линейных подсистем с матрицами коэффициентов $\omega I + \mathcal{T}$ и $\omega I + \mathcal{D}$. Чтобы снизить стоимость процесса предварительной обработки, рассматриваются циркулянтные аппроксимации для замены матрицы Теплица T в $\mathcal{F}_\omega$ [47–49]. Теоретическое рассмотрение дается только для циркулянта Стрэнга C для T [47]. Для простоты анализа пусть M четное, в дальнейшем будем записывать как $C = \mu[\varsigma_{ij}] \in \mathbb{R}^{M\times M}$ с $\varsigma_{ij} = \varsigma_{i-j}$, $\varsigma_k = \varsigma_{-k}$ для $0 \leqslant k < M$, $\varsigma_{\frac{M}{2}} = 0$ и

$$\varsigma_k = \begin{cases} c_k, & \text{if } 0 \leq k < \dfrac{M}{2}, \\ c_{M-k}, & \text{if } \dfrac{M}{2} < k < M, \end{cases}$$

т.е.

$$C = \mu\begin{bmatrix} c_0 & c_{-1} & \cdots & c_{-\left(\frac{M}{2}-1\right)} & 0 & c_{-\left(\frac{M}{2}-1\right)} & \cdots & c_{-1} \\ c_{-1} & c_0 & \ddots & \ddots & c_{-\left(\frac{M}{2}-1\right)} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & c_{-\left(\frac{M}{2}-1\right)} \\ c_{-\left(\frac{M}{2}-1\right)} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & c_{-\left(\frac{M}{2}-1\right)} \\ c_{-\left(\frac{M}{2}-1\right)} & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & c_{-1} \\ c_{-1} & c_{-2} & \cdots & 0 & c_{-\left(\frac{M}{2}-1\right)} & \ddots & \cdots & c_0 \end{bmatrix}, \tag{3.3}$$

которая является вещественной симметричной. Затем строится улучшенный эффективный предобусловливатель на основе циркулянта $\widetilde{\mathcal{F}}_\omega$:

$$\widetilde{\mathcal{F}}_\omega = \frac{1}{2\omega}(\omega I + \mathcal{C})(\omega I + \mathcal{D}), \quad \mathcal{C} = \begin{bmatrix} I & C \\ -C & I \end{bmatrix}.$$

Очевидно, $\mathcal{C}$ нормален, а $\mathcal{D}$ антисимметричен. Таким образом, предобусловливатель $\widetilde{\mathcal{F}}_\omega$ называется циркулянтным улучшенным нормальным и антисимметричным предобусловливателем (CNAS).

Теперь изучим свойство кластеризации собственных значений предварительно обусловленной матрицы системы $\widetilde{\mathcal{F}}_\omega{}^{-1}\mathcal{R}$. Благодаря тому факту, что

$$\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{R} = \underbrace{\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{F}_\omega}\,\underbrace{\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{R}} \tag{3.4}$$

это рассмотрение сводится к анализу свойств $\mathcal{F}_\omega{}^{-1}\mathcal{R}$ и $\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{F}_\omega$ соответственно.

3.1. Свойство кластеризации $\mathcal{F}_\omega{}^{-1}\mathcal{R}$. Рассмотрим свойство кластеризации $\mathcal{F}_\omega{}^{-1}\mathcal{R}$.

Т е о р е м а 2. Пусть $\mathcal{R} \in \mathbb{R}^{2M\times 2M}$ – матрица системы в (2.4), $\mathcal{T} \in \mathbb{R}^{22M\times 2M}$ и $\mathcal{D} \in \mathbb{R}^{22M\times 2M}$ – блочные матрицы, образующие разбиение $\mathcal{R}$ в (2.5), ω – положительная константа и σ(ω) определяется (2.11).

Когда предобусловливатель NASS $\mathcal{F}_\omega$ применяется к блочной линейной системе (2.4), собственные значения предобусловленной матрицы $\mathcal{F}_\omega{}^{-1}\mathcal{R}$ в (3.1) расположены в круге радиуса $\sigma(\omega) < 1$ с центром в точке 1.

Д о к а з а т е л ь с т в о. Согласно (2.8) и (2.10), оно записывается как $\mathcal{L}_\omega = \mathrm{I} - \mathcal{F}_\omega{}^{-1}\mathcal{R}$. Таким образом, пусть η – собственное значение $\mathcal{F}_\omega{}^{-1}\mathcal{R}$, тогда $1 - \eta$ – собственное значение $\mathcal{L}_\omega$. Кроме того, теорема 1 показывает, что $\rho(\mathcal{L}_\omega) \leqslant \sigma(\omega)$, т. е. $|1 - \eta| \leqslant \sigma(\omega)$ для всех $\eta \in \lambda(\mathcal{F}_\omega{}^{-1}\mathcal{R})$, что и является результатом этой теоремы.

Т е о р е м а 3. Пусть $\mathcal{R} \in \mathbb{R}^{22M\times 2M}$ – матрица системы в (2.4), $\mathcal{T} \in \mathbb{R}^{22M\times 2M}$ и $\mathcal{D} \in \mathbb{R}^{22M\times 2M}$ – блочные матрицы, образующие разбиение $\mathcal{R}$ в (2.5), ω – положительная константа. Тогда собственные значения предвобусловленной матрицы $\mathcal{F}_\omega{}^{-1}\mathcal{R}$ группируются в точке 0_+ на правой полукомплексной плоскости, если ω стремится к положительной бесконечности.

Д о к а з а т е л ь с т в о. Пусть λ – собственное значение $\mathcal{L}_\omega$, а x – соответствующий собственный единичный вектор, тогда оно записывается как $\mathcal{L}_\omega x = \lambda x$, т.е.

$$(\omega I - \mathcal{T})(\omega I - \mathcal{D})x = \lambda(\omega I + \mathcal{T})(\omega I + \mathcal{D})x$$

Обозначим

$$\widehat{\mathcal{T}} = \begin{bmatrix} I & T \\ -T & I \end{bmatrix}. \tag{3.5}$$

Если верно, что $\mathcal{T} = I + \widehat{\mathcal{T}}$, то имеем

$$\left[(\omega - 1)I - \widehat{\mathcal{T}}\right](\omega I - \mathcal{D})x = \lambda\left[(\omega + 1)I + \widehat{\mathcal{T}}\right](\omega I + \mathcal{D})x.$$

Умножая x^* слева на обе части приведенного выше уравнения, получим, что

$$\lambda = \frac{\omega(\omega - 1) - \omega x^*\widehat{\mathcal{T}}x - (\omega - 1)x^*\mathcal{D}x + x^*\widehat{\mathcal{T}}\mathcal{D}x}{\omega(\omega + 1) + \omega x^*\widehat{\mathcal{T}}x + (\omega + 1)x^*\mathcal{D}x + x^*\widehat{\mathcal{T}}\mathcal{D}x}.$$

Поскольку $x^*\widehat{\mathcal{T}}x$ и $x^*\mathcal{D}x$ являются чисто мнимыми числами в силу того, что $\widehat{\mathcal{T}}$ и $\mathcal{D}$ вещественны и антисимметричны, запишем $x^*\widehat{\mathcal{T}}x = \gamma_T \imath$ и $x^*\mathcal{D}x = \gamma_D \imath$ с $\gamma_T, \gamma_D \in \mathbb{R}$. Кроме того, обозначим $x^*\widehat{\mathcal{T}}\mathcal{D}x = \gamma_{TD}^{(R)} + \imath\gamma_{TD}^{(I)}$ с $\gamma_{TD}^{(R)}, \gamma_{TD}^{(I)} \in \mathbb{R}$. Из чего следует

$$\lambda = \frac{\omega(\omega - 1) + \gamma_{TD}^{(R)} + I\left[-\omega\gamma_T - (\omega - 1)\gamma_D + \gamma_{TD}^{(I)}\right]}{\omega(\omega - 1) + \gamma_{TD}^{(R)} + I\left[\omega\gamma_T + (\omega - 1)\gamma_D + \gamma_{TD}^{(I)}\right]}.$$

Поскольку $\mathcal{F}_\omega{}^{-1}\mathcal{R} = I - \mathcal{L}_\omega$, то 1-λ является собственным значением $\mathcal{F}_\omega{}^{-1}\mathcal{R}$ и получается

$$1 - \lambda = \frac{2}{\omega}\frac{1 + I(\gamma_T + \gamma_D) + \mathcal{O}(1/\omega)}{\left[1 + 1/\omega + \gamma_{TD}^{(R)}/\omega^2\right]^2 + \left[\gamma_T/\omega + (\omega + 1)\gamma_D/\omega^2 + \gamma_{TD}^{(I)}/\omega^2\right]^2}$$

Очевидно, что вещественная часть 1-λ стремится к 0_+, а мнимая часть 1-λ – к 0, когда параметр ω стремится к положительной бесконечности.

3.2. Свойство $\widetilde{\mathcal{F}}_\omega{}^{-1}\mathcal{F}_\omega$. Теперь обсудим свойство $\widetilde{\mathcal{F}}_\omega{}^{-1}\mathcal{F}_\omega$. Для удобства определим следующие константы:

$$\theta = \frac{\left(1 - \frac{1+\alpha}{5+\alpha/2}\right)^{5+\frac{\alpha}{2}} e^{1+\alpha}\Gamma(\alpha+1)\sin\left(\frac{\pi\alpha}{2}\right)}{\pi\alpha}$$
$$\theta_0 = \frac{\sqrt{2e}^{13/12}\Gamma(\alpha+1)\sin\left(\frac{\pi\alpha}{2}\right)}{\pi\alpha}. \tag{3.6}$$

Далее вводятся две леммы следующего содержания.

Л е м м а 1. Эта и следующая леммы приведены в [30]. Пусть $c_j = (-1)^j\Gamma(\alpha+1)/[\Gamma(\alpha/2 - j + 1)\Gamma(\alpha/2 + j + 1)]$, $k_0 \geqslant 3$, и $1 < \alpha < 2$, тогда

$$\frac{\theta}{(k_0 + 1/2)^\alpha} < \sum_{j=k_0+1} |c_j| < \frac{\theta_0}{(k_0 - 1)^\alpha}. \tag{3.7}$$

На основе леммы 1 оценки собственных значений матрицы Теплица T и ее циркулянтной аппроксимации C представлены в следующей лемме.

Л е м м а 2. Пусть T – матрица Теплица в (1.4), C – циркулянт Стрэнга в (3.3) и M четно, тогда справедливо равенство

$$\frac{2\gamma\tau\theta}{(b-a)^\alpha} < \lambda_T < \frac{2\gamma\tau}{h^\alpha}\left[\frac{\Gamma(\alpha+1)}{\Gamma(\alpha/2+1)} - \frac{\theta h^\alpha}{(b-a)^\alpha}\right], \quad M \geq 4, \tag{3.8}$$

и

$$\frac{2^{\alpha+1}\gamma\tau\theta}{(b-a)^\alpha} < \lambda_C < \frac{2\gamma\tau}{h^\alpha}\left[\frac{\Gamma(\alpha+1)}{\Gamma(\alpha/2+1)^2} - \frac{2^\alpha\theta h^\alpha}{(b-a)^\alpha}\right], \quad M \geq 8, \tag{3.9}$$

где λ_T и λ_C – собственные значения T и C.

На основе лемм 1 и 2 свойство $\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{F}_\omega$ равно суммированы в следующей теореме.

Т е о р е м а 4. Пусть $1 < \alpha < 2$ и $M \geqslant 8$ четно, $\epsilon > 0$ – небольшая константа и

$$k_0 = \left\lceil \left(\frac{\mu\theta_0}{\epsilon}\right)^{\frac{1}{\alpha}} \right\rceil + 1, \tag{3.10}$$

где $\lceil . \rceil$ представляет собой округление вещественного числа до положительной бесконечности. Тогда существуют две матрицы $\tilde{\mathcal{E}} \in \mathbb{R}^{2M\times 2M}$ и $\widehat{\mathcal{F}} \in \mathbb{R}^{2M\times 2M}$, удовлетворяющее $\operatorname{rank}\left(\tilde{\mathcal{E}}\right) = 4k_0$:

$$\left\|\tilde{\mathcal{E}}\right\|_2 < \frac{\left(\omega^2+\nu^2\right)^{\frac{1}{2}} M^{\frac{1}{2}}\mu}{\omega\sqrt{(\omega+1)^2 + \left[\frac{2^{\alpha+1}\gamma\tau\theta}{(b-a)^\alpha}\right]^2}}\left[\frac{c_0}{2} - \frac{\theta}{\left(M - \frac{1}{2}\right)^\alpha}\right], \tag{3.11}$$

$$\left\|\widetilde{\mathcal{F}}\right\|_2 < \frac{\left(\omega^2+\nu^2\right)^{\frac{1}{2}} M^{\frac{1}{2}}\epsilon}{\omega\sqrt{(\omega+1)^2 + \left[\frac{2^{\alpha+1}\gamma\tau\theta}{(b-a)^\alpha}\right]^2}}, \tag{3.12}$$

где $\nu = \max_{\mu i\in\lambda(D)}|\mu_i|$, такой, что

$$\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{F}_\omega = I + \tilde{\mathcal{E}} + \widetilde{\mathcal{F}}. \tag{3.13}$$

Доказательство. Поскольку M четно, оно записывается как

$$T - C = \mu \begin{bmatrix} 0 & \widehat{F}_{12} & \widehat{E}_{13} \\ \widehat{\mathcal{F}}_{12}^{\top} & 0 & 0 \\ \widehat{\mathcal{F}}_{13}^{\top} & 0 & 0 \end{bmatrix},$$

где $\widehat{F}_{12} \in \mathbb{R}^{\frac{M}{2} \times (\frac{M}{2} - k_0)}$ и $\widehat{E}_{13} \in \mathbb{R}^{\frac{M}{2} \times k_0}$ имеют вид

$$\widehat{F}_{12} = \begin{bmatrix} c_{\frac{M}{2}} & c_{\frac{M}{2}+1} - c_{\frac{M}{2}-1} & \cdots & c_{M-(k_0+1)} - c_{k_0+1} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_{\frac{M}{2}+1} - c_{\frac{M}{2}-1} \\ \vdots & \ddots & \ddots & c_{\frac{M}{2}} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix},$$

$$\widehat{E}_{13} = \begin{bmatrix} c_{M-k_0} - c_{k_0} & \cdots & \cdots & c_{M-1} - c_1 \\ \vdots & \ddots & \ddots & \vdots \\ c_{\frac{M}{2}+1} - c_{\frac{M}{2}-1} & \ddots & \ddots & \vdots \\ c_{\frac{M}{2}} & \ddots & \ddots & c_{M-k_0} - c_{k_0} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_{\frac{M}{2}+1} - c_{\frac{M}{2}-1} \\ 0 & \cdots & 0 & c_{\frac{M}{2}} \end{bmatrix}.$$

Следовательно, выполняется

$$T - C = \widehat{E} + \widehat{F}, \tag{3.14}$$

где

$$\widehat{E} = \mu \begin{pmatrix} 0 & 0 & \widehat{E}_{13} \\ 0 & 0 & 0 \\ \widehat{E}_{13}^{\top} & 0 & 0 \end{pmatrix}, \quad \widehat{F} = \mu \begin{pmatrix} 0 & \widehat{F}_{12} & 0 \\ \widehat{F}_{12}^{\top} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Очевидно, оно записывается как $\operatorname{rank} \widehat{E} = 2k_0$. Более того, ввиду лемм 1 и 2, а также структуры $\widehat{E}$ и $\widehat{F}$ можно получить следующие оценки:

$$\begin{gathered} \left\| \widehat{E} \right\|_{\infty} = \mu \max \left\{ \left\| \widehat{E}_{13} \right\|_{\infty}, \left\| \widehat{E}_{13}^{\top} \right\|_{\infty} \right\} = \mu \left\| \widehat{E}_{13}^{\top} \right\|_{\infty} \leq \\ \leq \mu \sum_{j=1}^{M-1} \left| c_j \right| = \mu \left(\frac{c_0}{2} - \sum_{j=M}^{\infty} \left| c_j \right| \right) < \\ < \mu \left[\frac{c_0}{2} - \frac{\theta}{\left(M - \frac{1}{2} \right)^{\alpha}} \right], \end{gathered} \tag{3.15}$$

$$\begin{aligned}\left\|\widehat{F}\right\|_\infty &= \mu \max\left\{\left\|\widehat{F}_{12}\right\|_\infty, \left\|\widehat{F}_{12}^\top\right\|_\infty\right\} = \mu\left\|\widehat{F}_{12}\right\|_\infty \le \\ &\le \mu \sum_{j=k_0+1}^{M-k_0-1} c_j < \mu \sum_{j=k_0+1}^{\infty} \left|c_j\right| < \\ &< \frac{\mu\theta_0}{\left(k_0-1\right)^\alpha} < \epsilon.\end{aligned} \tag{3.16}$$

Можно проверить, что $\widehat{\mathcal{F}}_{\omega}^{-1}\mathcal{F}_{\omega} - I$ и $(\omega I + \mathcal{C})^{-1}(\mathcal{T} - \mathcal{C})$ подобны, т.е.

$$\widetilde{\mathcal{F}}_{\omega}^{-1}\mathcal{F}_{\omega} - I = \left(\omega I + \mathcal{D}\right)^{-1}\left[\left(\omega I + \mathcal{C}\right)^{-1}\left(\omega I + \mathcal{T}\right) - I\right]\left(\omega I + \mathcal{D}\right) =$$
$$\left(\omega I + \mathcal{D}\right)^{-1}\left(\omega I + \mathcal{C}\right)^{-1}\left(\mathcal{T} - \mathcal{C}\right)\left(\omega I + \mathcal{D}\right),$$

причем

$$\left(\omega I + \mathcal{C}\right)^{-1}\left(\mathcal{T} - \mathcal{C}\right) = \begin{bmatrix}\left(\omega+1\right)I & C \\ -C & \left(\omega+1\right)I\end{bmatrix}^{-1}\begin{bmatrix}0 & T-C \\ C-T & 0\end{bmatrix}. \tag{3.17}$$

Учитывая (3.14) запишем

$$\left(\omega I + \mathcal{C}\right)^{-1}\left(\mathcal{T} - \mathcal{C}\right) = \hat{\mathcal{E}} + \widehat{\mathcal{F}},$$

где

$$\hat{\mathcal{E}} = \begin{bmatrix}\left(\omega+1\right)I & C \\ -C & \left(\omega+1\right)I\end{bmatrix}^{-1}\begin{bmatrix}0 & \widehat{E} \\ -\widehat{E} & 0\end{bmatrix},$$
$$\widehat{\mathcal{F}} = \begin{bmatrix}\left(\omega+1\right)I & C \\ -C & \left(\omega+1\right)I\end{bmatrix}^{-1}\begin{bmatrix}0 & \widehat{F} \\ -\widehat{F} & 0\end{bmatrix}.$$

Таким образом, выполняется

$$\mathcal{F}_{\omega}^{-1}\mathcal{F}_{\omega} - I = \tilde{\mathcal{E}} + \widetilde{\mathcal{F}},$$

где

$$\tilde{\mathcal{E}} = \left(\omega I + \mathcal{D}\right)^{-1}\tilde{\mathcal{E}}\left(\omega I + \mathcal{D}\right) \tag{3.18}$$

и

$$\widetilde{\mathcal{F}} = \left(\omega I + \mathcal{D}\right)^{-1}\widetilde{\mathcal{F}}\left(\omega I + \mathcal{D}\right). \tag{3.19}$$

Дополнительно имеем

$$\begin{aligned}\left\|\tilde{\mathcal{E}}\right\|_2 &\le \left\|\begin{bmatrix}\omega I & -D \\ D & \omega I\end{bmatrix}^{-1}\right\|_2\left\|\begin{bmatrix}\omega I & -D \\ D & \omega I\end{bmatrix}\right\|_2\left\|\begin{bmatrix}\left(\omega+1\right)I & C \\ -C & \left(\omega+1\right)I\end{bmatrix}^{-1}\right\|_2\left\|\widehat{E}\right\|_2 = \\ &= \frac{\max_{\mu_i\in\lambda(D)}\sqrt{\omega^2+\mu_i^2}}{\min_{\mu_i\in\lambda(D)}\sqrt{\omega^2+\mu_i^2}}\frac{1}{\min_{\lambda_i^{(c)}\in\lambda(C)}\sqrt{\left(\omega+1\right)^2+\left(\lambda_i^{(c)}\right)^2}}\left\|\widehat{E}\right\|_2 \le \\ &\le \frac{\max_{\mu_i\in\lambda(D)}\sqrt{\omega^2+\mu_i^2}}{\min_{\mu_i\in\lambda(D)}\sqrt{\omega^2+\mu_i^2}}\frac{1}{\min_{\lambda_i^{(c)}\in\lambda(C)}\sqrt{\left(\omega+1\right)^2+\left(\lambda_i^{(c)}\right)^2}}M^{\frac{1}{2}}\left\|\widehat{E}\right\|_\infty <\end{aligned}$$

$$< \frac{\left(\omega^2 + v^2\right)^{\frac{1}{2}} M^{\frac{1}{2}}\mu}{\omega\sqrt{(\omega+1)^2 + \left[\dfrac{2^{\alpha+1}\gamma\tau\theta}{(b-a)^{\alpha}}\right]^2}} \left[\frac{c_0}{2} - \frac{\theta}{\left(M - \frac{1}{2}\right)^{\alpha}}\right].$$

Первое "=" связано с тем, что собственные значения

$$\begin{bmatrix} \omega I & -D \\ D & \omega I \end{bmatrix} \tag{3.20}$$

равны $\omega \pm \mu_i$ı с $\mu_i \in \lambda(D)$ и $\imath = \sqrt{-1}$, а собственные значения

$$\begin{bmatrix} (\omega+1) I & C \\ -C & (\omega+1) I \end{bmatrix} \tag{3.21}$$

представляют собой $\omega + 1 \pm \lambda_i^{(c)}$ı с $\lambda_i^{(c)} \in \lambda(C)$. Второй знак "$\leqslant$" обусловлен связью между $\|\cdot\|_2$ и $\|\cdot\|_\infty$, а последнее строгое неравенство возникает вследствие леммы 2, оценки $\left\|\widehat{E}\right\|_\infty$ и учета того, что $\omega > 0$ и $0 \leqslant \mu_i \leqslant \nu$. Аналогично имеем

$$\left\|\widehat{\mathcal{F}}\right\|_2 < \frac{\left(\omega^2 + v^2\right)^{\frac{1}{2}} M^{\frac{1}{2}}\epsilon}{\omega\sqrt{(\omega+1)^2 + \left[\dfrac{2^{\alpha+1}\gamma\tau\theta}{(b-a)^{\alpha}}\right]^2}}.$$

З а м е ч а н и е 2. Рассмотрим следующее соотношение, сформулированное в теореме 4:

$$\widehat{\mathcal{F}}_\omega^{-1}\mathcal{F}_\omega = \underbrace{I + \widetilde{\mathcal{F}}} + \tilde{\mathcal{E}}.$$

С одной стороны, матрица $I + \widehat{\mathcal{F}}$ является малым возмущением единичной матрицы I. В частности, поскольку $\|\widehat{\mathcal{F}}\|_2 \leq \epsilon$, теорема Бауэра-Фике [50] приводит к тому, что

$$|\xi - 1| \leq \epsilon, \forall\, \xi \in \lambda\left(I + \widetilde{\mathcal{F}}\right),$$

т.е. собственные значения $I + \widetilde{\mathcal{F}}$ сгруппированы в небольшом диске с центром в точке 1 и радиусом ϵ. С другой стороны, поскольку матрица $\tilde{\mathcal{E}}$ имеет ограниченную ℓ_2-норму и низкий ранг, матрицу $\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{F}_\omega$ можно рассматривать как низкоранговую модификацию $I + \widetilde{\mathcal{F}}$. Тогда ожидаем, что собственные значения $\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{F}_\omega$ также расположены в том же диске с центром в точке 1 и радиусом ϵ, за исключением небольшого количества выбросов.

Наконец, следующая теорема устанавливает связь между $\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{R}$ и $\mathcal{F}_\omega^{-1}\mathcal{R}$.

Т е о р е м а 5. Пусть $1 < \alpha < 2$, $M \geqslant 8$ четно, ϵ – небольшая положительная константа, такая, что

$$\frac{2^{\alpha}\mu\theta_0}{(M-2)^{\alpha}} < \epsilon \leq \mu\theta_0, \quad k_0 = \left\lceil \left(\frac{\mu\theta_0}{\epsilon}\right)^{\frac{1}{\alpha}} \right\rceil + 1, \quad v = \max_{\mu_i \in \lambda(D)} |\mu_i|. \tag{3.22}$$

Тогда существуют две матрицы $\mathcal{P}_\omega \in \mathbb{R}^{2M \times 2M}$ и $\mathcal{Q}_\omega \in \mathbb{R}^{2M \times 2M}$, удовлетворяющие

$$\operatorname{rank}\left(\mathcal{P}_\omega\right) = 4k_0, \tag{3.23}$$

$$\left\|\mathcal{P}_\omega\right\|_2 < \frac{2\left(\omega^2 + v^2\right) M^{\frac{1}{2}}\mu}{\omega\sqrt{(\omega+1)^2 + \left[\dfrac{2^{\alpha+1}\gamma\tau\theta}{(b-a)^{\alpha}}\right]^2}} \left[\frac{c_0}{2} - \frac{\theta}{\left(M - \frac{1}{2}\right)^{\alpha}}\right], \tag{3.24}$$

$$\left\|\mathcal{Q}_\omega\right\|_2 < \frac{2\left(\omega^2+v^2\right)M^{\frac{1}{2}}\varepsilon}{\omega\sqrt{(\omega+1)^2+\left[\dfrac{2^{\alpha+1}\gamma\tau\theta}{(b-a)^\alpha}\right]^2}} \tag{3.25}$$

при этом

$$\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{R} = \mathcal{F}_\omega^{-1}\mathcal{R} + \mathcal{P}_\omega + \mathcal{Q}_\omega \tag{3.26}$$

Доказательство. Отметим, что

$$\mathcal{F}_\omega^{-1}\mathcal{R} = I - \widetilde{\mathcal{F}}_\omega^{-1}\mathcal{Q}_\omega =$$
$$= \begin{bmatrix}\omega I & -D\\ D & \omega I\end{bmatrix}^{-1}\left(I - \mathcal{U}_\omega\mathcal{V}_\omega\right)\begin{bmatrix}\omega I & -D\\ D & \omega I\end{bmatrix},$$

где $\mathcal{U}_\omega$ и $\mathcal{V}_\omega$ определены в доказательстве теоремы 1. Согласно (2.15) и (2.16), известно, что $\left\|\mathcal{U}_\omega\mathcal{V}_\omega\right\|_2 < 1$, и в этом случае $\left\|I - \mathcal{U}_\omega\mathcal{V}_\omega\right\|_2 \leqslant \|I\|2 + \left\|\mathcal{U}_\omega\mathcal{V}_\omega\right\|_2 < 2$.
Далее, согласно доказательству теоремы 4, имеем

$$\left\|\begin{bmatrix}\omega I & -D\\ D & \omega I\end{bmatrix}^{-1}\right\|_2 \left\|\begin{bmatrix}\omega I & -D\\ D & \omega I\end{bmatrix}\right\|_2 < \frac{\sqrt{\omega^2+v^2}}{\omega}.$$

Таким образом $\left\|\mathcal{F}_\omega^{-1}\mathcal{R}\right\|_2 < 2\sqrt{\omega^2+v^2}\,/\omega$.
Ввиду (3.4) и (3.13) выполняется

$$\widetilde{\mathcal{F}}_\omega^{-1}\mathcal{R} = \left(I + \tilde{\mathcal{E}} + \widetilde{\mathcal{F}}\right)\mathcal{F}_\omega^{-1}\mathcal{R} = \mathcal{F}_\omega^{-1}\mathcal{R} + \mathcal{P}_\omega + \mathcal{Q}_\omega, \tag{3.27}$$

где $\mathcal{P}_\omega = \tilde{\mathcal{E}}\,\mathcal{F}_\omega{}^{-1}\mathcal{R}$ и $\mathcal{Q}_\omega = \widetilde{\mathcal{F}}\,\mathcal{F}_\omega{}^{-1}\mathcal{R}$. Из теоремы 4 следует $\operatorname{rank}\left(\mathcal{P}_\omega\right) = 4k_0$,

$$\left\|\mathcal{P}_\omega\right\|_2 \leq \left\|\tilde{\mathcal{E}}\right\|_2 \left\|\mathcal{F}_\omega^{-1}\mathcal{R}\right\| <$$
$$< \frac{2\left(\omega^2+v^2\right)M^{\frac{1}{2}}\mu}{\omega^2\sqrt{(\omega+1)^2+\left[\dfrac{2^{\alpha+1}\gamma\tau\theta}{(b-a)^\alpha}\right]^2}}\left[\frac{c_0}{2} - \frac{\theta}{\left(M-\frac{1}{2}\right)^\alpha}\right],$$
$$\left\|\mathcal{Q}_\omega\right\|_2 \leq \left\|\widetilde{\mathcal{F}}\right\|_2 \left\|\mathcal{F}_\omega^{-1}\mathcal{R}\right\| <$$
$$< \frac{2\left(\omega^2+v^2\right)M^{\frac{1}{2}}\varepsilon}{\omega^2\sqrt{(\omega+1)^2+\left[\dfrac{2^{\alpha+1}\gamma\tau\theta}{(b-a)^\alpha}\right]^2}}$$

Замечание 3. Теорема 5 показывает, что $\mathcal{Q}_\omega$ – матрица малой нормы, поэтому собственные значения $\mathcal{F}_\omega{}^{-1}\mathcal{R} + \mathcal{Q}_\omega$ можно рассматривать как малые возмущения собственных значений $\mathcal{F}_\omega{}^{-1}\mathcal{R}$. Кроме того, $\mathcal{P}_\omega$ имеет ограниченную ℓ_2-норму и низкий ранг, поэтому можно ожидать, что большинство собственных значений $\widetilde{\mathcal{F}}_\omega{}^{-1}\mathcal{R}$ распределены вблизи собственных значений $\mathcal{F}_\omega{}^{-1}\mathcal{R} + \mathcal{Q}_\omega$. Таким образом, собственные значения $\widetilde{\mathcal{F}}_\omega{}^{-1}\mathcal{R}$ группируются вокруг собственных значений $\mathcal{F}_\omega{}^{-1}\mathcal{R}$, за исключением нескольких выбросов.

4. Реализация и сложность. В предыдущих разделах были предложены и проанализированы новые предобусловливатели для блочной линейной системы (2.4), а именно NASS $\mathcal{F}_\omega$ и CNAS $\widetilde{\mathcal{F}}_\omega$, реализация которых подробно обсуждается в этом разделе. Фактически можно игнорировать скалярный множитель $1/(2\omega)$ для $\mathcal{F}_\omega$ и $\widetilde{\mathcal{F}}_\omega$ по той причине, что он не влияет на свойства предобусловленных матриц системы. Тогда упрощенные предобусловливатели записываются как

$$\mathcal{F}_{\text{NASS}} = \begin{bmatrix} \widehat{\omega} I & T \\ -T & \widehat{\omega} I \end{bmatrix} \begin{bmatrix} \omega I & -D \\ D & \omega I \end{bmatrix} =$$

$$= \begin{bmatrix} I & 0 \\ \widehat{L}_{21} & I \end{bmatrix} \begin{bmatrix} \widehat{\omega} I & T \\ 0 & \widehat{U}_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ L_{21} & I \end{bmatrix} \begin{bmatrix} \omega I & -D \\ 0 & U_{22} \end{bmatrix},$$

$$\mathcal{F}_{\text{CNAS}} = \begin{bmatrix} \widehat{\omega} I & C \\ -C & \widehat{\omega} I \end{bmatrix} \begin{bmatrix} \omega I & -D \\ D & \omega I \end{bmatrix} =$$

$$= \begin{bmatrix} F & 0 \\ 0 & F \end{bmatrix}^{-1} \begin{bmatrix} I & 0 \\ \widetilde{L}_{21} & I \end{bmatrix} \begin{bmatrix} \widehat{\omega} I & \Lambda \\ 0 & \widetilde{U}_{22} \end{bmatrix} \begin{bmatrix} F & 0 \\ 0 & F \end{bmatrix} \begin{bmatrix} I & 0 \\ L_{21} & I \end{bmatrix} \begin{bmatrix} \omega I & -D \\ 0 & U_{22} \end{bmatrix},$$

где $\omega > 0$ служит параметром $\mathcal{F}_{\text{NASS}}$ и $\mathcal{F}_{\text{CNAS}}$, $\mathcal{F} \in \mathbb{C}^{M \times M}$ – дискретное преобразование Фурье (ДПФ), $\Lambda = \text{diag}(Fc) \in \mathbb{C}^{M \times M}$, где $c \in \mathbb{C}^M$ является первым столбцом C, $\widehat{\omega} = \omega + 1$, $\widehat{L}_{21} = -T/\widehat{\omega}$, $\widehat{U}_{22} = \widehat{\omega} I + T^2/\widehat{\omega}$, $L_{21} = D/omega$, $U_{22} = \omega I + D^2/\omega$, $\widetilde{L}_{21} = -\Lambda/\widehat{\omega}$, $\widetilde{U}_{22} = \widehat{\omega} I + \Lambda^2/\widehat{\omega}$. Диагональные матрицы $L_{21}, U_{22}, \widetilde{L}_{21}, \widetilde{U}_{22}$ можно вычислить заранее.

В методах предварительно обусловленных подпространств Крылова вектор обобщенной невязки вычисляется путем решения соответствующего уравнения, связанного с новыми предобусловливателями, на каждой итерации. Пусть $x = (x_1^\top, x_2^\top)^\top \in \mathbb{R}^{2M}$ с $x_1, x_2 \in \mathbb{R}^M$ и $r = (r_1^\top, r_2^\top)^\top \in \mathbb{R}^{2M}$ с $r_1, r_2 \in \mathbb{R}^M$. Подробности для реализация предварительной подготовки NASS и CNAS приведены в описании алгоритмов 1 и 2.

Стоимость алгоритма 2 складывается из операций быстрого преобразования Фурье (БПФ) и его обратного (ОБПФ) размерности M и векторных операций размерности M (включая сложение векторов, диагональное умножение матрицы на вектор и решение диагональной линейной системы). Согласно [51], БПФ/ОБПФ могут быть выполнены за $\mathcal{O}(M\log M)$ флоп. Все операции с векторами размерности M могут быть произведены за $\mathcal{O}(M)$ флоп. Таким образом, в вычислениях метода $\mathcal{F}_{\text{CNAS}}$ преобладают операции БПФ/ОБПФ. Это означает, что вектор обобщенной невязки может быть вычислен за $\mathcal{O}(\text{M}\log M)$ флоп на каждой итерации предобусловленных методов подпространств Крылова.

Стоимость алгоритма 1 складывается из векторных операций размерности M (около $\mathcal{O}(M)$ флоп), умножений теплицевой матрицы $M \times M$ на вектор (около $\mathcal{O}(M\log M)$ флопс, поскольку его можно реализовать на основе БПФ/ОБПФ) и однократного решения плотной линейной системы размерности $M \times M$ с матрицей коэффициентов $(\omega + 1)I + T^2/(\omega + 1)$. Решение линейной системы может быть реализовано на основе прямого метода (около $\mathcal{O}(\text{M}^3)$) или метода предварительно обусловленного сопряженного градиента (PCG) с предобусловливателем $(\omega + 1)I + C^2/(\omega + 1)$ (около $\mathcal{O}(kM\log M)$ флоп, где k – количество итераций PCG). Очевидно, что вычислительная нагрузка по реализации $\mathcal{F}_{\text{NASS}}$ во многом зависит от стоимости решения плотной линейной системы, что намного дороже, чем реализация $\mathcal{F}_{\text{CNAS}}$. Поэтому настоятельно рекомендуется использовать новый предобусловливатель $\mathcal{F}_{\text{CNAS}}$.

А л г о р и т м 1. Решение уравнения с обобщенной невязкой $\mathcal{F}_{\text{NASS}}\, x = r$.

1: $x_1 = r_1$, $x_2 = r_2 - \widehat{L}_{21} r_1$;
2: $\widehat{U}_{22} x_2 = x_2$ и $x_1 = (r_1 - Tx_2)/\widehat{\omega}$;
3: $x_2 = x_2 - L_{21} x_1$;
4: $U_{22} x_2 = x_2$ и $x_1 = (x_1 + Dx_2)/\omega$.

На первом шаге этого алгоритма используется одно умножение теплицевой матрицы $M \times M$ на вектор и две векторные операциии размерности M; на втором – одно решение плотной линейной системы $M \times M$, одно умножение теплицевой матрицы на вектор и две векторные операции; на третьем – две векторные операции; на четвертом – четыре векторные операции.

А л г о р и т м 2. Решение уравнение обобщенной невязки $\mathcal{F}_{\text{CNAS}}\, x = r$.

1: $x_j = \text{БПФ}(r_j), j = 1, 2;$
2: $x_2 = x_2 - \widetilde{L}_{21} x_1;$
3: $\widetilde{U}_{22} x_2 = x_2$ и $x_1 = (x_1 - \Lambda x_2)/\widehat{\omega}$;
4: $x_j = \text{ОБПФ}(x_j), j = 1, 2;$
5: $x_2 = x_2 - L_{21} x_1;$
6: $U_{22} x_2 = x_2$ и $x_1 = (x_1 + Dx_2)/\omega$.

На первом шаге этого алгоритма осуществляются два БПФ размерности M; на втором и пятом шагах – по две векторные операции; на третьем и шестом шагах – по четыре векторные операции; на четвертом – два ОБПФ. размерности M.

5. Численные эксперименты. В этом разделе представлены доказательства свойств новых предобусловливателей $\mathcal{F}_{\text{NASS}}$ и $\mathcal{F}_{\text{CNAS}}$. Метод GMRES применяется с предобусловливателем CNAS $\mathcal{F}_{\text{CNAS}}$ для решения дробных уравнений CNLS в дискретизированном пространстве в случае притягивающего взаимодействия частиц. Дробные дискретизированные уравнения CNLS получены на основе схемы LICD. Для запуска схемы LICD требуется начальное значение на начальном временном уровне и приблизительное значение второго или более высокого порядка на первом временном уровне. Например, приближенное значение второго порядка можно получить с помощью неявной консервативной схемы второго порядка [12].

Во всех численных экспериментах блочная линейная система (2.4) на втором временном уровне дискретизированных дробных уравнений CNLS выбирается в качестве тестируемой линейной системы, а начальное предположение предварительно обусловленного метода GMRES устанавливается равным нулевому вектору. Кроме того, метод GMRES (предварительно обусловленный) выполняется без перезапуска и завершает либо относительную невязку ℓ_2-нормы тестируемой линейной системы, упавшую ниже 10^{-6}, либо количество итераций, превышающее 3000.

5.1. Случай DNLS. Пусть $\beta = 0$, тогда система (0.1) расцеплена. Рассмотрим следующую усеченную систему:

$$lu_t - \gamma\left(-\Delta\right)^{\frac{\alpha}{2}} u + \rho|u|^2 u = 0, \quad -20 \le x \le 20, \quad 0 < t \le T, \tag{5.1}$$

при соблюдении начальных и граничных условий

$$u(x,0) = \operatorname{sech}(x)\, e^{2lx}, \quad u(-20,t) = u(20,t) = 0. \tag{5.2}$$

Здесь взяты параметры $\gamma = 1$, $\rho = 2$, $1 < \alpha \leqslant 2$. Дискретизированные дробные уравнения DNLS получаются путем применения схемы LICD к (5.1) и (5.2). Требуется решить сложную симметричную линейную систему вида (2.1) на каждом временном уровне t_n для $1 < n \leqslant N$, что эквивалентно решению блочной линейной системы (2.4). В частности, блочная линейная система (2.4) решается методом предварительно обусловленной GMRES CNAS (CNAS-GMRES). В этом подразделе «IT» представляет количество итераций тестируемого метода.

На рис. 1 показаны кривые IT в зависимости от параметра $\omega \in (0, 8]$ CNAS-GMRES при $\alpha = 1.1 : 0.2 : 1.9$, $M = 6400$, $N = 200$. Видно, что IT быстро увеличивается, когда ω стремится к нулю. Однако когда ω растет, IT быстро достигает минимума, а затем растет очень медленно, а оптимальное значение ω составляет примерно [2.2, 2.8] для всех случаев α. Таким образом, сходимость CNAS-GMRES менее чувствительна к параметру ω, пока ω не слишком близок к нулю, что делает предобусловливатель CNAS $\mathcal{F}_{\text{CNAS}}$ прост в использовании. Кроме того, больший α приводит к большему IT, а это означает, что с увеличением α систему становится сложнее решить.

На рис. 2 показаны кривые IT CNAS-GMRES в зависимости от размера пространственной сетки M схемы LICD, примененной к дробным пространственным уравнениям DNLS, когда $\alpha = 1.1 : 0.2 : 1.9$, $N = 200$. Выбрано эмпирическое оптимальное значение ω. Показано, что IT CNAS-GMRES остается практически неизменным с увеличением количества пространственных дискретных точек, что указывает на независимость свойства сходимости CNAS-GMRES от размера пространственной сетки. Кроме того, больший дробный порядок α приводит к более высокому положению кривой на графике.

В табл. 1 перечислены IT CNAS-GMRES в сочетании с различными циркулянтными матрицами, а также эмпирическое оптимальное значение параметра ω CNAS-GMRES при $\alpha = 1.9$, $M = 6400$, $N = 200$. Интервал поиска параметра ω равен (0, 4]. Перечислены только

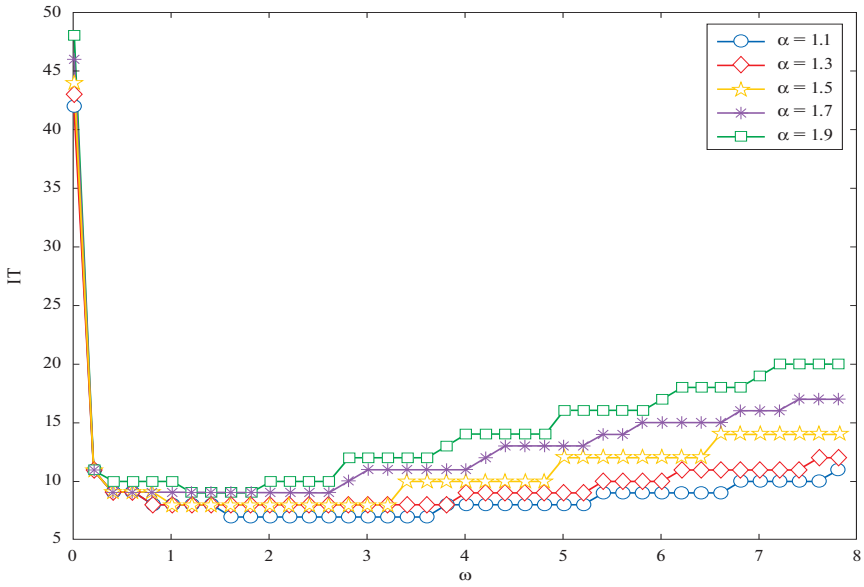

Рис. 1. Кривые IT от параметра $\omega \in (0, 8]$ CNAS-GMRES при $\alpha = 1.1 : 0.2 : 1.9$, $M = 6400$, $N = 200$

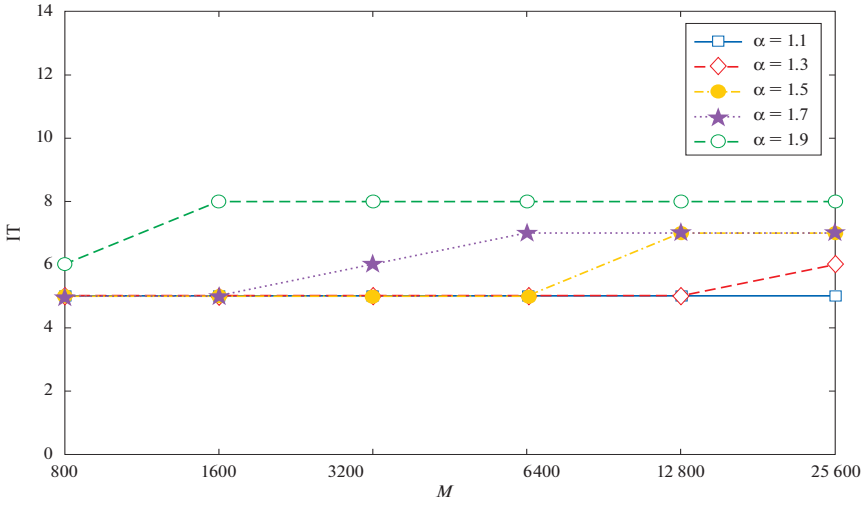

Рис. 2. Кривые IT CNAS-GMRES от размера пространственной сетки M при $\alpha = 1.1 : 0.2 : 1.9$, $N = 200$

Таблица 1. IT CNAS-GMRES с различными циркулянтными матрицами и эмпирическое оптимальное значение параметра ω CNAS-GMRES при $\alpha = 1.9$, $M = 6400, N = 200$

Circulant Matrix	IT	ω
T. Chan	7	[0.41,0.81]
Strang	8	[0.21,1.01]
R. Chan	8	[0.21,1.01]
Modified Dirichlet kernel	8	[0.21,1.01]
Hann kernel	8	[0.21,1.01]
Hamming kernel	8	[0.21,1.01]
Superoptimal	220	[0.52,0.55]

результаты нескольких репрезентативных циркулянтных матриц [51], включая циркулянтную матрицу Т.Чана, циркулянтную матрицу Стрэнга, циркулянтную матрицу Р.Чана, циркулянтные матрицы, построенные на основе некоторых известных ядер (например, модифицированного ядра Дирихле, ядра фон Ханна, ядра Хэмминга), и супероптимальная циркулянтная матрица. За исключением супероптимальной циркулянтной матрицы, IT CNAS-GMRES со всеми остальными циркулянтными матрицами меньше 10, а оптимальное эмпирическое значение ω практически одинаково, т.е. [0.21, 1.01] ([0.41, 0.81] для циркулянтной матрицы Т.Чана). Эти эксперименты показывают, что большинство циркулянтных матриц в литературе эффективны, когда они применяются к CNAS-GMRES, что позволяет легко выбрать циркулянтное приближение в CNAS-GMRES. Супероптимальная циркулянтная матрица плохо работает в наших экспериментах, и возможная причина заключается в том, что оптимальное значение ω CNAS-GMRES в этом случае остается за пределами интервала поиска (0, 4].

На рис. 3–5 показано распределение собственных значений матрицы $\mathcal{R}$, предобусловленной матрицы NASS $\mathcal{F}_{\text{NASS}}{}^{-1}\mathcal{R}$ и предобусловленной матрицы CNAS $\mathcal{F}_{\text{CNAS}}{}^{-1}\mathcal{R}$, когда $\alpha = 1.1 : 0.4 : 1.9$, $M = 1600, 3200$. Левые графики относятся к случаю $M = 1600$, а правые – к случаю $M = 3200$. На рис. 3 вещественные части собственных значений $\mathcal{R}$ равны 1, а мни-

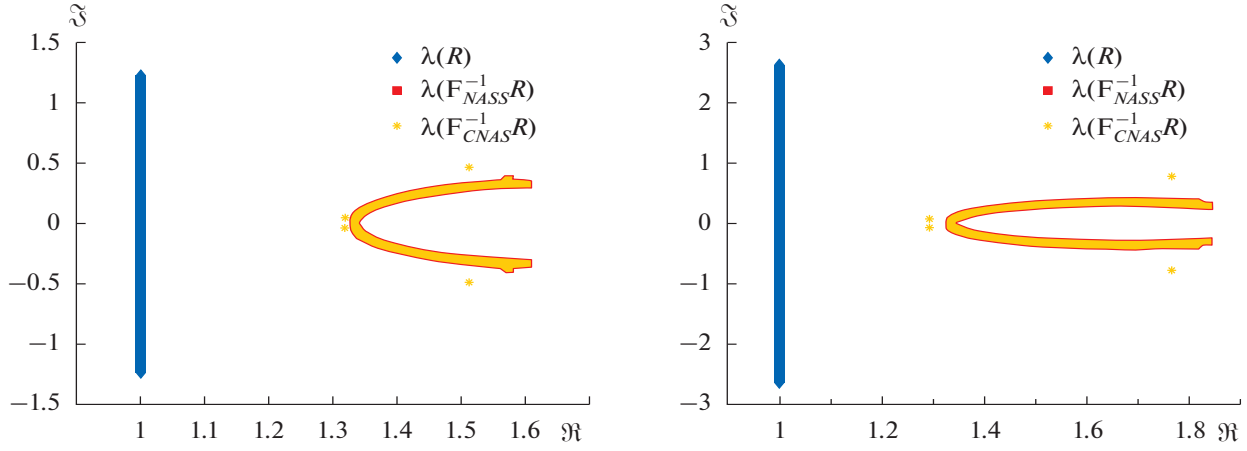

Рис. 3. Распределение собственных значений $\mathcal{R}$, $\mathcal{F}_{\text{NASS}}{}^{-1}\mathcal{R}$ и $\mathcal{F}_{\text{CNAS}}{}^{-1}\mathcal{R}$, при $\alpha = 1.1$, $M = 1600$, $N = 200$ (слева) и $M = 3200$, $N = 200$ (справа)

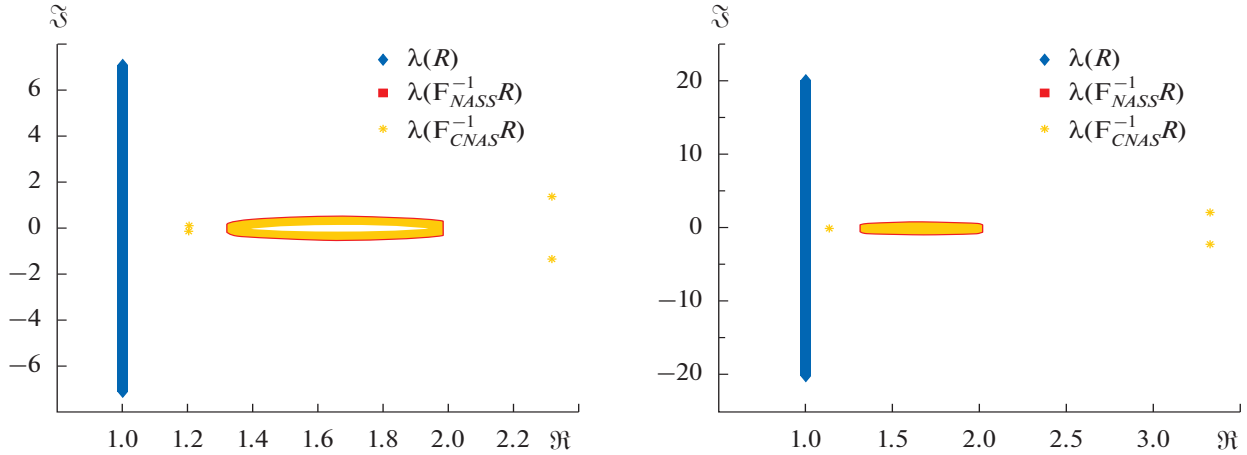

Рис. 4. Распределение собственных значений $\mathcal{R}$, $\mathcal{F}_{\text{NASS}}{}^{-1}\mathcal{R}$ и $\mathcal{F}_{\text{CNAS}}{}^{-1}\mathcal{R}$, при $\alpha = 1.5$, $M = 1600, N = 200$ (слева) и $M = 3200$, $N = 200$ (справа)

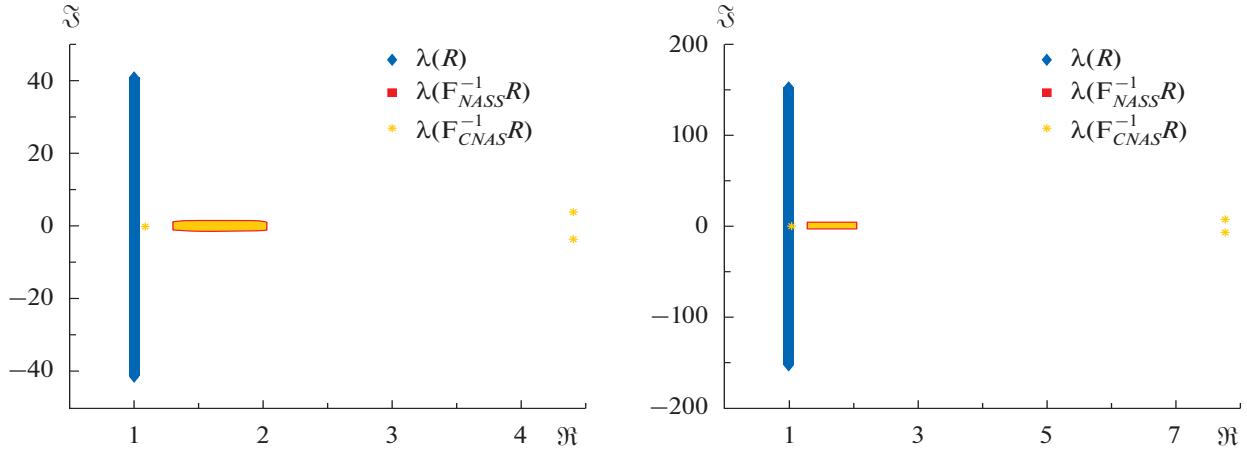

Рис. 5. Распределение собственных значений $\mathcal{R}$, $\mathcal{F}_{\text{NASS}}{}^{-1}\mathcal{R}$ и $\mathcal{F}_{\text{CNAS}}{}^{-1}\mathcal{R}$, при $\alpha = 1.9$, $M = 1600$, $N = 200$ (слева) и $M = 3200$, $N = 200$ (справа)

мые части распределены от −1.3 до 1.3 при $M = 1600$ и от −2.7 до 2.7 при $M = 3200$. На рис. 4 вещественные части собственных значений $\mathcal{R}$ равны 1, а мнимые части распределены от −8 до 8 при $M = 1600$ и от −20 до 20 при $M = 3200$. На рис. 5 вещественные части собственных начений $\mathcal{R}$ равны 1, а мнимые части распределены от −42 до 42 при $M = 1600$ и от −150 до 150 при $M = 3200$. При этом вещественные части собственных значений $\mathcal{F}_{\text{NASS}}{}^{-1}\mathcal{R}$ и $\mathcal{F}_{\text{CNAS}}{}^{-1}\mathcal{R}$ распределены от 1.3 до 2, а мнимые – от −0.5 до 0.5 на всех графиках. Очевидно, собственные значения $\mathcal{F}_{\text{NASS}}{}^{-1}\mathcal{R}$ и $\mathcal{F}_{\text{CNAS}}{}^{-1}\mathcal{R}$ более кластеризованы, чем $\mathcal{R}$ (особенно для больших α). Кроме того, большинство собственных значений $\mathcal{F}_{\text{CNAS}}{}^{-1}\mathcal{R}$ группируются вокруг значений $\mathcal{F}_{\text{NASS}}{}^{-1}\mathcal{R}$и только несколько собственных значений $\mathcal{F}_{\text{CNAS}}{}^{-1}\mathcal{R}$ отклоняются сильнее, что соответствует предсказанию замечаний 2, 3. Распределение собственных значений предварительно обусловленных случаев остается почти таким же, когда M увеличивается с 1600 до 3200, что указывает на свойство сходимости NASS-GMRES и CNAS-GMRES, не зависящее от размера пространственной сетки.

На рис. 6–9 слева показано численное решение u_{CNAS}, полученное с помощью CNAS-GMRES, а справа – его ошибка $\text{err}_u = |u_{\text{CNAS}} - u_{\text{GE}}|$, т.е. отклонение от решения, полученного в схеме LICD с использованием решения плотной линейной системы методом Гаусса, для дробно-пространственных уравнений DNLS (5.1) при $M = 800$, $N = 200$, $\alpha = 1.1 : 0.4 : 1.9$ и $\alpha = 2$. Замечено, что дробный порядок α будет влиять на форму волнового фронта как по высоте, так и по ширине. Чем меньше α, тем быстрее будет меняться форма волнового фронта. При

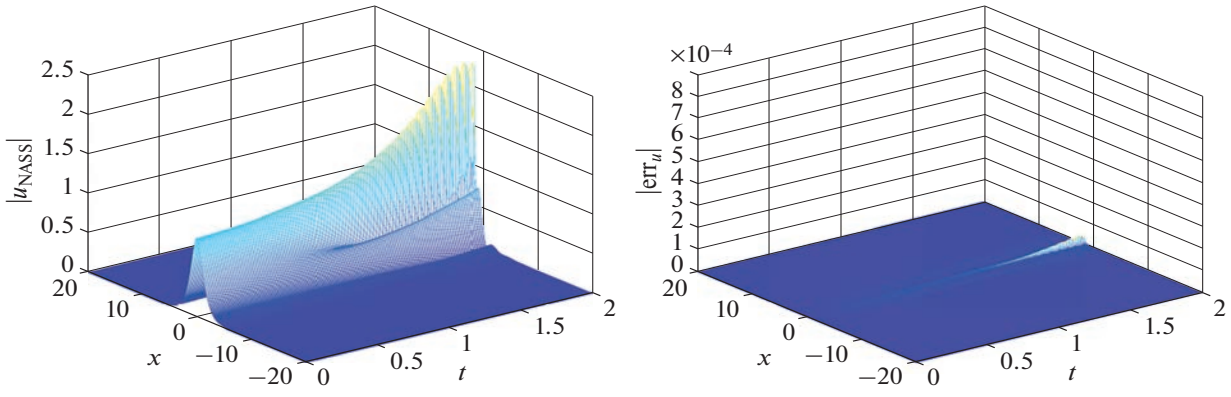

Рис. 6. Численное решение (слева) и его ошибка (справа) дробного пространственного уравнения DNLS (5.1) при $\alpha = 1.1$, $M = 800$, $N = 200$

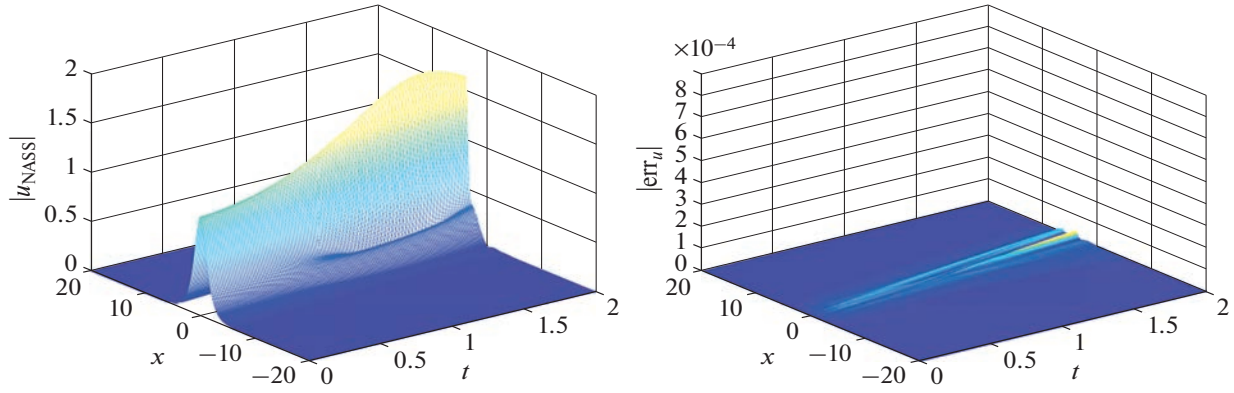

Рис. 7. Численное решение (слева) и его ошибка (справа) дробного пространственного уравнения DNLS (5.1) при $\alpha = 1.5$, $M = 800$, $N = 200$

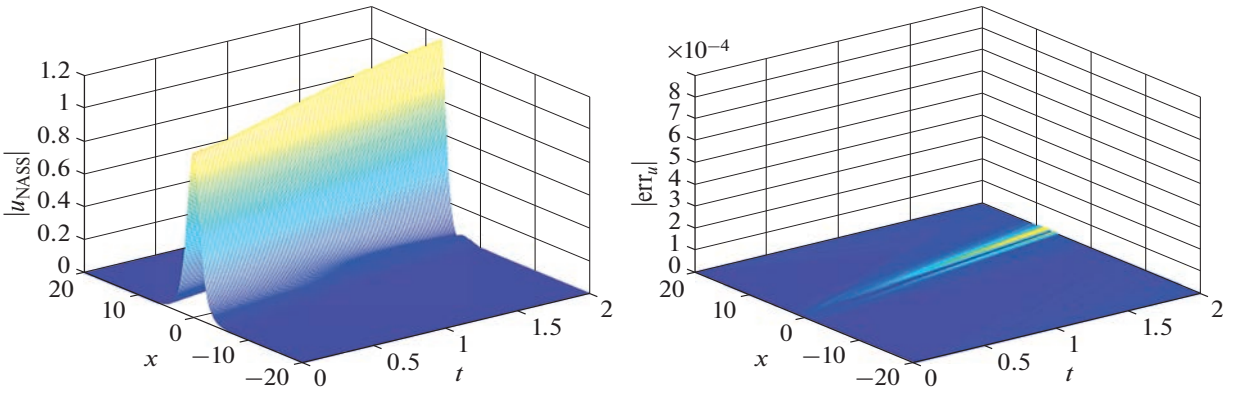

Рис. 8. Численное решение (слева) и его ошибка (справа) дробного пространственного уравнения DNLS (5.1) при $\alpha = 1.9$, $M = 800$, $N = 200$

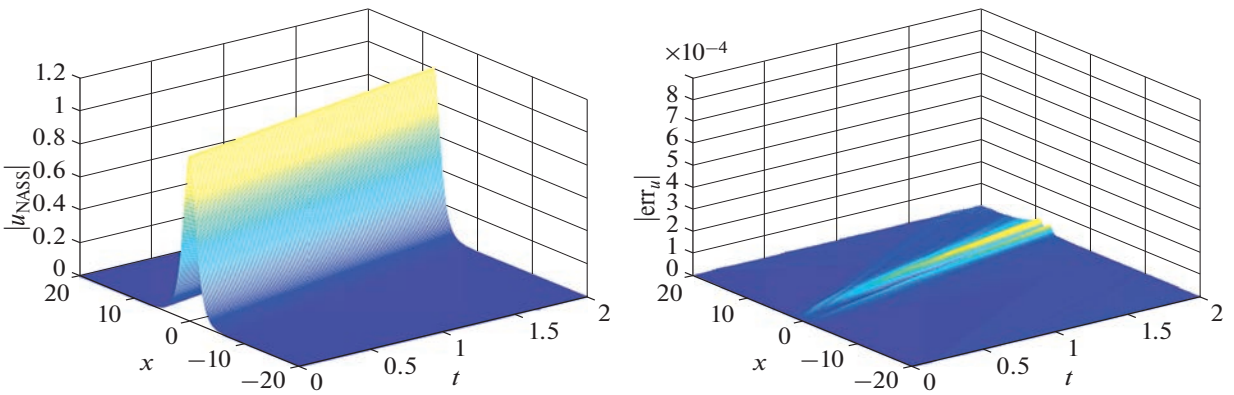

Рис. 9. Численное решение (слева) и его ошибка (справа) дробного пространственного уравнения DNLS (5.1) при $\alpha = 2$, $M = 800$, $N = 200$

стремлении α к 2 волновой фронт пространственного дробного уравнения DNLS сходится к волновому фронту недробного уравнения, и высота волнового фронта стабильна. Кроме того, ошибка между численным решением и точным решением схемы LICD остается всего лишь около 10^{-4} во всей задействованной пространственно-временной области, следовательно, численное решение, полученное с помощью CNAS-GMRES, является надежным.

В табл. 2 и на рис. 10 показаны относительные погрешности дискретной массы и энергии для различных значений α соответственно. Размер пространственной сетки и размер временного шага составляют $h = 0.2$ и $\tau = 0.05$ для табл. 2 и $h = 0.2$ и $\tau = 0.001$ для рис. 10. На каждом временном уровне блочная линейная система решается с помощью CNAS-GMRES и завершается, когда относительная невязка по норме ℓ_2 блочной линейной системы снижается ниже 10^{-15}. Указанные ошибки остаются очень небольшими, а это означает, что линейный решатель CNAS-GMRES сохраняет устойчивость схемы LICD.

Таблица 2. Относительные ошибки дискретной массы, т. е. $|(Q^n - Q^0)/Q^0|$ при $h = 0.2$, $\tau = 0.05$

α	t = 1	t = 2	t = 3	t = 4
1.4	4.8850E-015	7.5495E-015	6.4393E-015	9.1038E-015
1.7	5.9952E-015	4.2188E-015	3.3307E-015	3.7748E-015
1.9	1.1102E-015	2.2209E-015	3.1086E-015	1.5543E-015
2	6.6613E-016	5.7732E-015	5.3291E-015	6.6615E-015

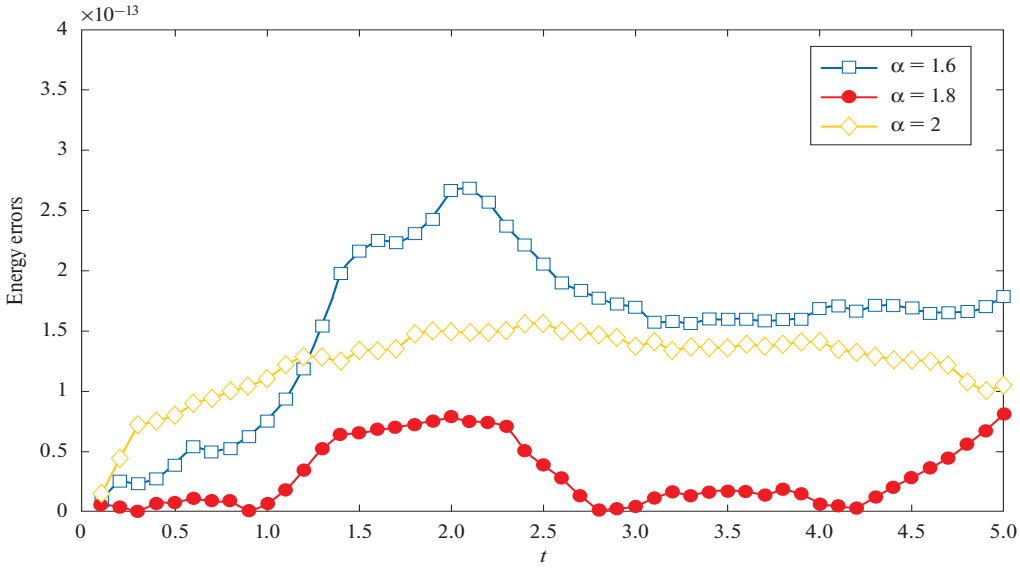

Рис. 10. Относительные погрешности дискретной энергии, т.е. $|(E^n - E^0)/E^0|$, при $h = 0.2$, $\tau = 0.001$

5.2. Случай CNLS. Эксперименты здесь проводятся в связанном случае, т. е. усеченной системе уравнений CNLS:

$$\begin{cases} lu_t - \gamma(-\Delta)^{\frac{\alpha}{2}} u + \rho\left(|u|^2 + \beta|v|^2\right)u = 0, \\ lu_t - \gamma(-\Delta)^{\frac{\alpha}{2}} v + \rho\left(|v|^2 + \beta|u|^2\right)v = 0, \end{cases} \quad -20 \le x \le 20, \quad 0 < t \le T, \tag{5.3}$$

при соблюдении начальных и граничных условий:

$$\begin{cases} u(x,0) = \operatorname{sech}(x+5)\, e^{3lx}, \quad u(x,0) = \operatorname{sech}(x-5)\, e^{-3lx}, \\ u(-20,t) = u(20,t) = 0, \qquad v(-20,t) = v(20,t) = 0, \end{cases} \tag{5.4}$$

где $\gamma = 1$, $\rho = 1$, $\beta = 1$, $1 < \alpha \leqslant 2$. Схема LICD, примененная к (5.3) и (5.4), приводит к дробным уравнениям CNLS в дискретизированном пространстве. Требуется последовательно решить две сложные симметричные линейные системы вида (2.1) на каждом временном уровне t_n для $1 < n \leqslant N$, что эквивалентно решению двух блочных линейных систем вида (2.4). В дальнейшем «CPU» и «IT» – это общее время вычислений в секундах и общее количество итераций для решения двух связанных линейных систем в точке t_n.

На рис. 11 показаны кривые IT CNAS-GMRES в зависимости от размера пространственной сетки M схемы LICD, применяемой к дробным пространственным уравнениям CNLS, когда $\alpha = 1.1 : 0.2 : 1.9$, $M = 800, 1600, 3200, 6400, 12800, 25600$ и $N = 200$. Выбрано эмпирическое оптимальное значение ω. Результат аналогичен случаю дробных уравнений DNLS с дискретизированным пространством, что подтверждает, что CNAS-GMRES также не зависит от размера пространственной сетки в связанном случае.

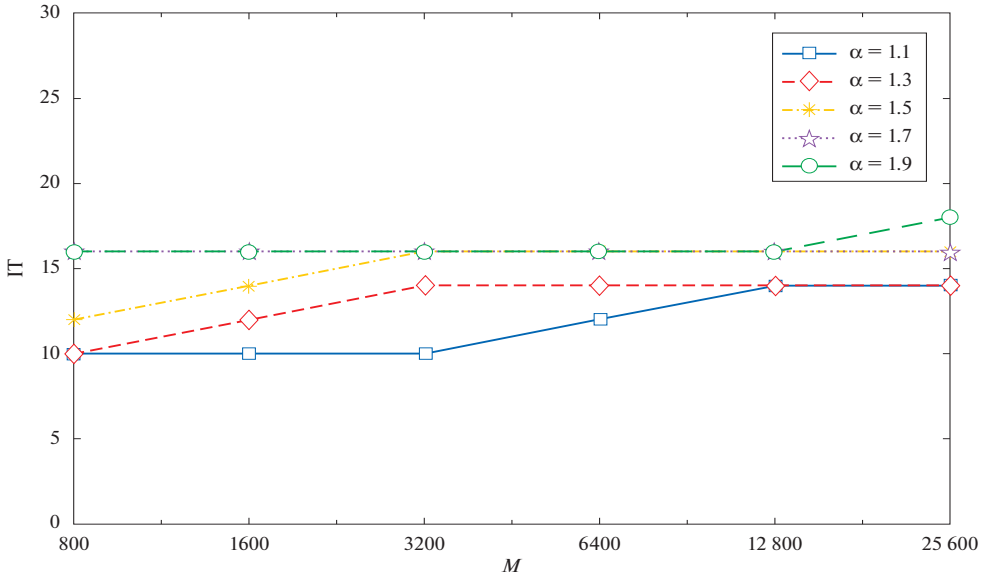

Рис. 11. Кривые IT CNAS-GMRES от размера пространственной сетки M при $\alpha = 1.1 : 0.2 : 1.9$, $N = 200$

На рис. 12 показано влияние силы нелинейного члена (управляемого параметром ρ) на количество итераций GMRES, NASS-GMRES и CNAS-GMRES при $\alpha = 1.1 : 0.4 : 1.9$, $M = 1600$. Параметр ρ увеличивается с 1 до 64. На этих графиках IT схем NASS-GMRES и CNAS-GMRES очень мало и медленно увеличивается по мере роста ρ. При этом IT схемы GMRES очень велико и быстро растет. Видно, что чем сильнее нелинейный член, тем труднее решать связанные линейные системы. Новые предобусловленные методы GMRES могут эффективно справляться с сильно нелинейным случаем.

В табл. 3–7 указаны CPU и IT GMRES, CNAS-GMRES и GE при $\alpha = 1.1 : 0.2 : 1.9$, $M = 3200, 6400, 12800, 25600$, $N = 200$. В этих таблицах '–' означает, что GMRES не сходится за заданное количество итераций, N/A означает, что данные IT для GE недоступны. Эмпирические оптимальные параметры CNAS-GMRES, соответствующие результатам табл. 3–7, приведены в табл. 8. В частности, ω_u и ω_v обозначаются эмпирическими оптимальными параметрами CNAS-GMRES для блочных линейных систем (2.4), связанных с u и v соответственно. Согласно таб. 3–7, GE требует самого большого CPU во всех тестах, а CPU CNAS-GMRES всегда меньше CPU GMRES. Кроме

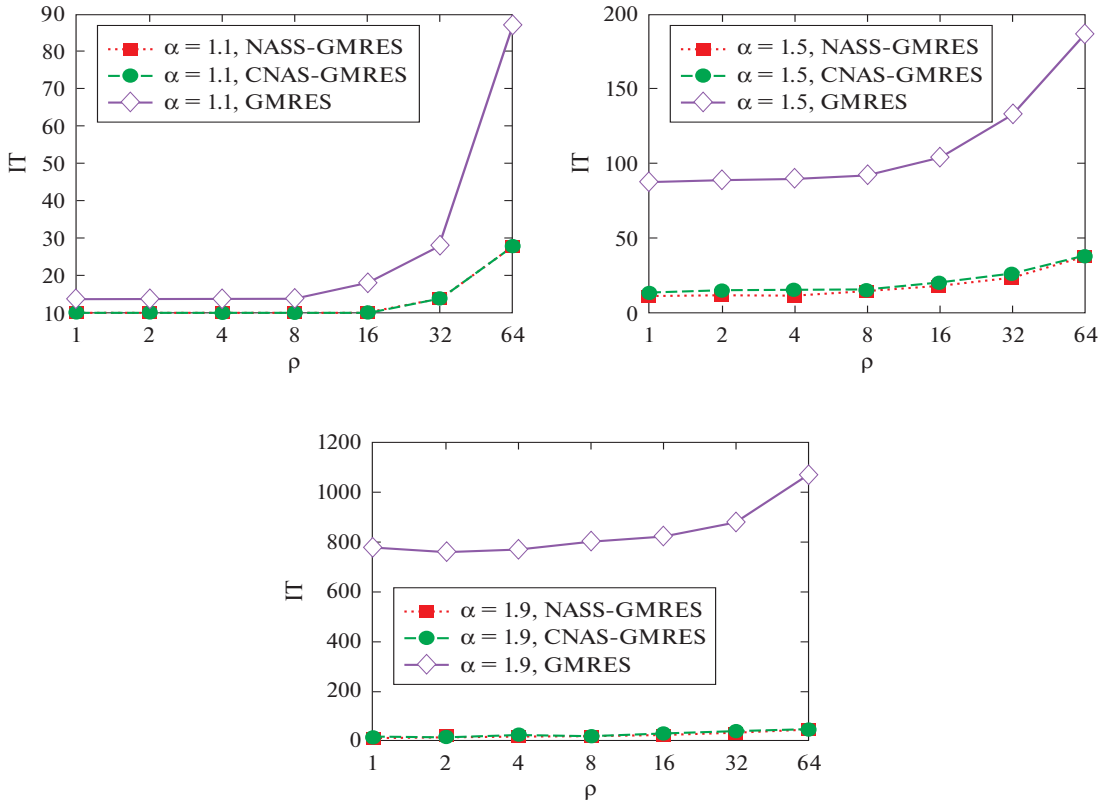

Рис. 12. Кривые IT NASS-GMRES, CNAS-GMRES и GMRES от нелинейного термопараметра ρ при α = 1.1 : 0.4 : 1.9, $M = 1600$, $N = 200$

Таблица 3. CPU и IT GMRES для CNAS-GMRES и GE при α = 1.1, $N = 200$

M	3200		6400		12800		25600	
	CPU	IT	CPU	IT	CPU	IT	CPU	IT
GMRES	8.33E-02	22	5.69E-01	50	2.67E-00	110	1.10E+01	236
CNAS-GMRES	6.10E-02	10	2.61E-01	12	6.98E-01	14	1.21E-00	14
GE	2.10E+01	N/A	1.67E+02	N/A	1.36E+03	N/A	1.88E+04	N/A

Таблица 4. CPU и IT для GMRES, CNAS-GMRES и GE при α = 1.3, N = 200

M	3200		6400		12800		25600	
	CPU	IT	CPU	IT	CPU	IT	CPU	IT
GMRES	3.84E-01	66	2.82E-00	182	2.47E+01	514	1.61E+02	1087
CNAS-GMRES	6.52E-02	14	2.53E-01	14	7.12E-01	14	1.23E-00	14
GE	2.10E+01	N/A	1.64E+02	N/A	1.36E+03	N/A	1.94E+04	N/A

Таблица 5. CPU и IT для GMRES, CNAS-GMRES и GE при $\alpha = 1.5$, $N = 200$

M	3200		6400		12800		25600	
	CPU	IT	CPU	IT	CPU	IT	CPU	IT
GMRES	1.86E-00	256	2.89E+01	832	3.26E+02	2700	–	–
CNAS-GMRES	7.24E-02	16	2.60E-01	16	7.56E-01	16	1.26E-00	16
GE	2.10E+01	N/A	1.66E+02	N/A	1.34e+03	N/A	1.92E+04	N/A

Таблица 6. CPU и IT для GMRES, CNAS-GMRES и GE при $\alpha = 1.7$, $N = 200$

M	3200		6400		12800		25600	
	CPU	IT	CPU	IT	CPU	IT	CPU	IT
GMRES	1.31E+01	1086	4.46E+02	4056	–	–	–	–
CNAS-GMRES	9.40E-02	16	2.95E-01	16	8.09E-01	16	1.53E-00	16
GE	2.11E+01	N/A	1.68E+02	N/A	1.39E+03	N/A	1.96E+04	N/A

Таблица 7. CPU и IT для GMRES, CNAS-GMRES и GE при $\alpha = 1.9$, $N = 200$

M	3200		6400		12800		25600	
	CPU	IT	CPU	IT	CPU	IT	CPU	IT
GMRES	2.51E+02	4696	–	–	–	–	–	–
CNAS-GMRES	1.27E-01	16	2.66E-01	16	7.88E-01	16	1.45E-00	18
GE	2.12E+01	N/A	1.70E+02	N/A	1.47E+03	N/A	1.98E+04	N/A

Таблица 8. Эмпирические оптимальные параметры CNAS-GMRES при $\alpha = 1.1 : 0.2 : 1.9$, $N = 200$

α	M	3200	6400	12800	25600
1.1	ω_u ω_v	[0.16,0.24] [0.18,0.25]	[0.16,0.24] [0.18,0.25]	[0.15,0.24] [0.16,0.22]	[0.17,0.24] [0.20,0.25]
1.3	ω_u ω_v	[0.19,0.24] [0.20,0.24]	[0.19,0.24] [0.20,0.25]	[0.18,0.24] [0.21,0.23]	[0.17,0.24] [0.19,0.23]
1.5	ω_u ω_v	[0.09,0.24] [0.10,0.25]	[0.20,0.24] [0.18,0.25]	[0.17,0.24] [0.17,0.24]	[0.17,0.24] [0.18,0.24]
1.7	ω_u ω_v	[0.26,0.34] [0.30,0.43]	[0.26,0.34] [0.28,0.34]	[0.18,0.24] [0.20,0.25]	[0.14,0.24] [0.16,0.25]
1.9	ω_u ω_v	[0.09,0.34] [0.10,0.35]	[0.08,0.34] [0.09,0.34]	[0.06,0.24] [0.09,0.25]	[0.21,0.24] [0.22,0.25]

того, IT GMRES быстро увеличивается с увеличением дробного порядка α и размера пространственной сетки M, что указывает на то, что систему (5.3) становится сложнее решить по мере роста α. Между тем IT CNAS-GMRES меньше, чем IT GMRES, особенно для больших α и M. Таким образом, CNAS-GMRES значительно повышает эффективность вычислений, ведет себя независимо от размера пространственной сетки M и почти нечувствителен к дробному порядку α.

Таблицы 9, 10 и рисунок 13 сообщают об относительных ошибках дискретной массы и энергии. Соответствующий размер пространственной сетки и размер временного шага составляют $h = 0.1$, $\tau = 0.01$. На каждом временном уровне блочные линейные системы, связанные с u и v, решаются с помощью CNAS-GMRES и прекращаются, когда относительные невязки

Таблица 9. Относительные ошибки дискретной массы в u, т. е. $|(Q_1^n - Q_1^0)/Q_1^0|$, при $h = 0.1$, $\tau = 0.01$

α	β	t = 2	t = 4	t = 6	t = 8	t = 10
2	1	4.4409E-015	3.8857E-015	6.2172E-015	7.9936E-015	1.0749E-014
1.6	1	8.8818E-016	3.5527E-015	3.7748E-015	4.2188E-015	3.7748E-015
1.5	2	1.1102E-015	6.6613E-016	2.8866E-015	5.5511E-015	5.3291E-015

Таблица 10. Относительные ошибки дискретной массы в v, т.е. $|(Q_2^n - Q_2^0)/Q_2^0|$, при $h = 0.1$, $\tau = 0.01$

α	β	t = 2	t = 4	t = 6	t = 8	t = 10
2	1	9.9920E-016	4.4409E-016	2.6645E-015	3.9968E-015	9.6589E-015
1.6	1	1.5543E-015	4.4409E-016	3.2204E-015	1.3323E-015	1.3323E-015
1.5	2	4.4409E-016	2.2204E-016	1.1102E-015	4.2188E-015	3.5527E-015

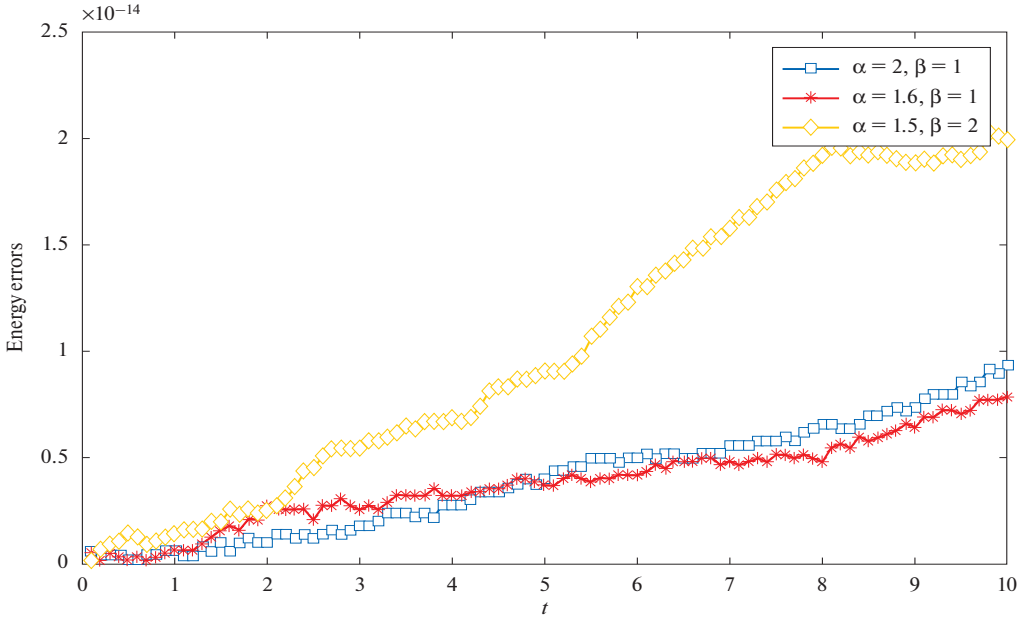

Рис. 13. Относительные погрешности дискретной энергии, т.е. $|(E^n - E^0)/E^0|$, при $h = 0.1$, $\tau = 0.01$

блочных линейных систем по ℓ_2-норме уменьшаются ниже 10^{-15}. Небольшие ошибки, показанные в табл. 9, 10 и на рис. 13, показывают, что как дискретная масса, так и энергия схемы LICD сохраняются линейным решателем CNAS-GMRES при $< \alpha = 2, \beta = 1 >$, $< \alpha = 1.6, \beta = 1 >$ и $< \alpha = 1.5, \beta = 2 >$.

На рис. 14–17 слева изображены численные решения u_{CNAS} и v_{CNAS}, полученные с помощью CNAS-GMRES, а справа – соответствующие ошибки $\mathrm{err}_u = |u_{\mathrm{CNAS}} - u_{\mathrm{GE}}|$ и $\mathrm{err}_v = |v_{\mathrm{CNAS}} - v_{\mathrm{GE}}|$ относительно точных решений, полученных схемой LICD с решением линейной системы методом Гаусса, для дробно-пространственных уравнений CNLS (5.3) при $\alpha = 1.1 : 0.4 : 1.9$ и $\alpha = 2$, $M = 800$, $N = 600$. Форма волновых фронтов меняется в зависимости от дробного порядка α и сходится к волновым фронтам стандартных уравнений CNLS при приближении α к 2. Дробный порядок α влияет на время столкновения волновых фронтов. Чем больше значение α, тем быстрее движутся волновые фронты и тем раньше происходит столкновение. Из рис. 16–17 видно, что отражения происходят после того, как волновые фронты достигают границы пространственно-временной области. Очевидно, что никакого отражения волнового фронта не произойдет, если пространственный интервал не усечен. Кроме того, ошибки численного решения остаются очень небольшими, что означает надежность линейного решателя CNAS-GMRES.

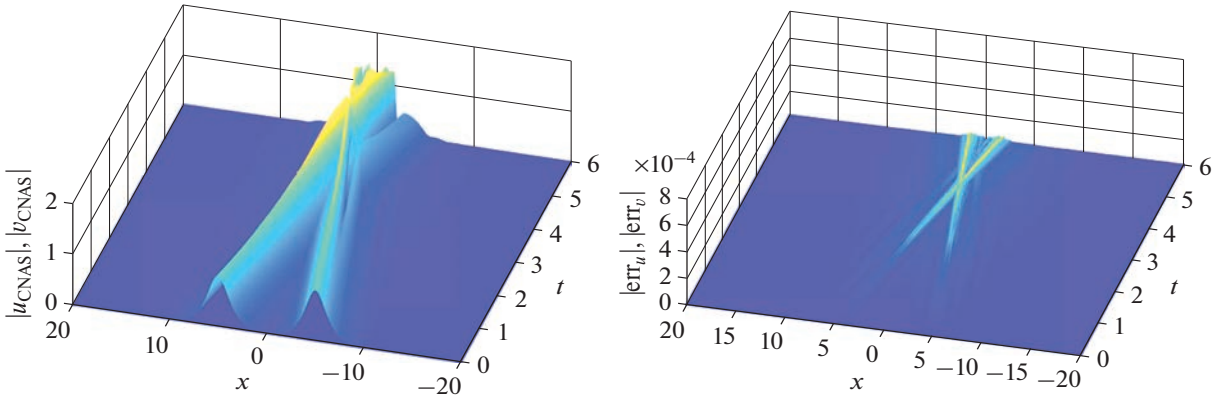

Рис. 14. Численные решения (слева) и их ошибки (справа) дробно-пространственных уравнений CNLS (5.3) при $\alpha = 1.1$, $M = 800$, $N = 600$

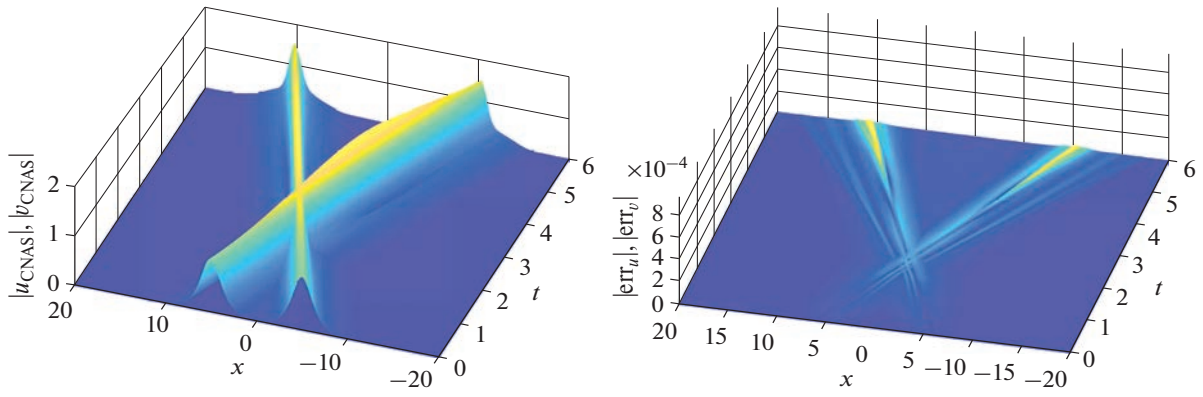

Рис. 15. Численные решения (слева) и их ошибки (справа) дробно-пространственных уравнений CNLS (5.3) при $\alpha = 1.5$, $M = 800$, $N = 600$

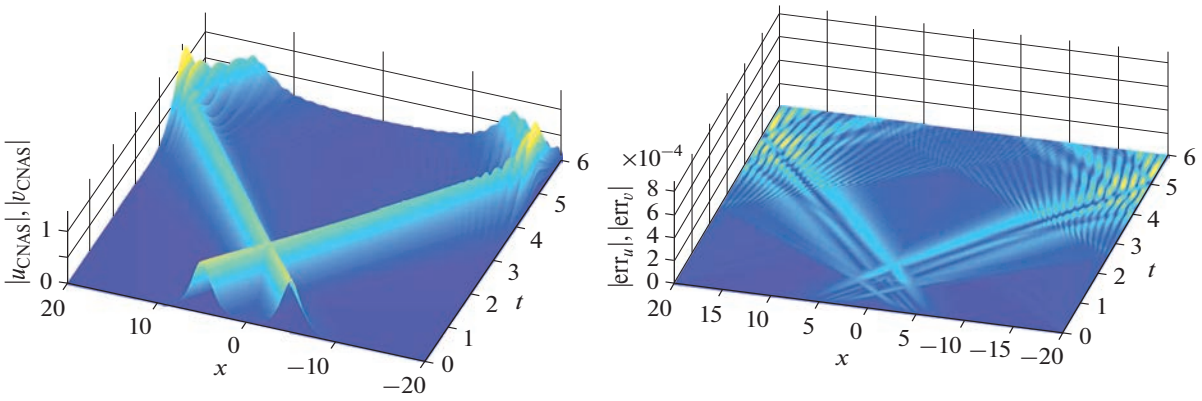

Рис. 16. Численные решения (слева) и их ошибки (справа) дробно-пространственных уравнений CNLS (5.3) при $\alpha = 1.9$, $M = 800$, $N = 600$

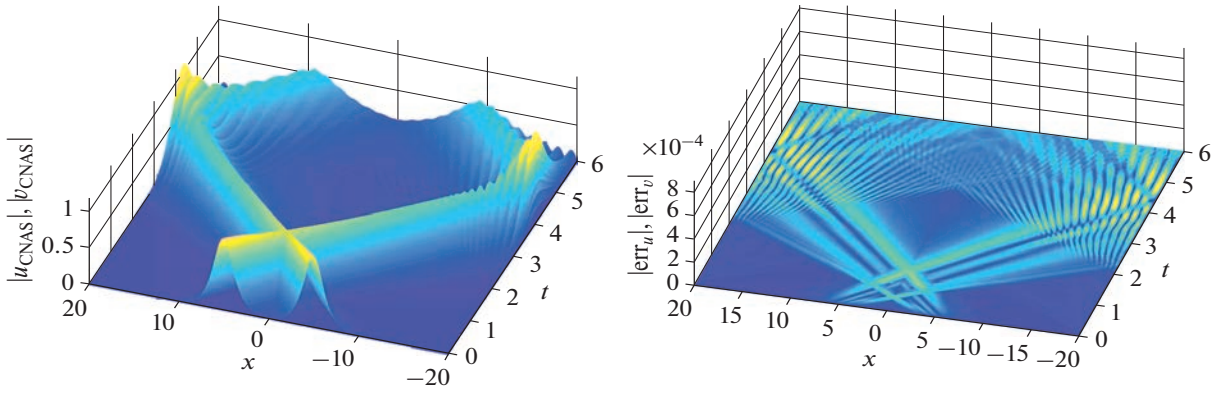

Рис. 17. Численные решения (слева) и их ошибки (справа) дробно-пространственных уравнений CNLS (5.3) при $\alpha = 2$, $M = 800$, $N = 600$

Заключение. Проведенная работа сосредоточена на создании эффективных линейных решателей для сложных симметричных и неопределенных линейных систем вида $(D - T + \imath I)\,\mathbf{u} = \mathbf{b}$, которые имеют теплиц-плюс-диагональ и комплексную симметричную структуры. Эти линейные системы возникают из одномерных пространственных дробных уравнений Шредингера с притягивающим взаимодействием частиц, дискретизированных по схеме LICD. В частности, предлагаются итерационный метод NASS и, естественно, порождаемый предобусловливатель NASS. Простой в реализации и чрезвычайно эффективный предобусловливатель создается путем точной аппроксимации теплицевых блоков циркулянтными матрицами, которая называется предобусловливателем CNAS. Теоретически исследованы получаемые итерационный метод и предобусловливатели. Метод GMRES с предварительной обработкой CNAS подтвержден как эффективный и действенный линейный решатель для $(D - T + \imath I)u = b$ посредством численных экспериментов, основанных на одномерных дробных уравнениях CNLS. Однако эти результаты получены при условии постоянных коэффициентов одномерной задачи и для равномерных сеток. В будущем можно расширить итерационный метод NASS и связанный с ним предобусловливатель для задач более высокой размерности. Также, когда в задаче появляются переменные коэффициенты, в полученных линейных системах не может быть найдена явная структура теплиц-плюс-диагональ, поэтому нахождение возможной неявной структуры и расширение описанных методов для построения эффективных решателей может стать большой проблемой. Наконец, когда задача дискретизируется на неравномерных сетках, теплиц-плюс-диагональная структура полученной линейной системы может быть полностью утрачена. Тогда возможным способом построения быстрых решателей является объединение иерархически-матричного подхода [52,53] и структуры, предложенной в этой статье.

СПИСОК ЛИТЕРАТУРЫ

1. *Feynman R.P.* Statistical Mechanics: A Set of Lectures. 1st edn. CRC Press, 1998.
2. *Feynman R.P., Hibbs A.R., Styer D.F.* Quantum Mechanics and Path Integrals. Dover Publications, 2010.
3. *Laskin N.* Fractional Quantum Mechanics and Levy Path Integrals // Phys. Lett. A. 2000. V. 268. P. 298–305.
4. *Laskin N.* Fractional Quantum Mechanics // Phys. Rev. E. 2000. V. 62. P. 3135–3145.
5. *Guo X.Y., Xu M.Y.* Some Physical Applications of Fractional Schroedinger Equation // J. Math. Phys. 2006. V. 47. P. 082104.
6. *Li M., Gu X.M., Huang C.M. et al.* A Fast Linearized Conservative Finite Element Method for the Strongly Coupled Nonlinear Fractional Schroedinger Equations // J. Comput. Phys. 2018. V. 358. P. 256–282.
7. *Li M., Huang C.M., Wang P.D.* Galerkin Finite Element Method for Nonlinear Fractional Schroedinger Equations // Numer. Algorithms. 2017. V. 74. P. 499–525.
8. *Duo S.W., Zhang Y.Z.* Mass-conservative Fourier Spectral Methods for Solving the Fractional Nonlinear Schroedinger Equation // Comput. Math. Appl. 2016. V. 71. P. 2257–2271.
9. *Wang Y., Mei L.Q., Li Q. et al.* Split-step Spectral Galerkin Method for the Two-dimensional Nonlinear Space-fractional Schroedinger Equation // Appl. Numer. Math. 2019. V. 136. P. 257–278.
10. *Amore P., Fernandez F.M., Hofmann C.P. et al.* Collocation Method for Fractional Quantum Mechanics // J. Math. Phys. 2010. V. 51. P. 122101.
11. *Bhrawy A.H., Zaky M.A.* An Improved Collocation Method for Multi-dimensional Space-time Variable-order Fractional Schroedinger Equations // Appl. Numer. Math. 2017. V. 111. P. 197–218.
12. *Wang D.L., Xiao A.G., Yang W.* Crank-Nicolson Difference Scheme for the Coupled Nonlinear Schroedinger Equations with the Riesz Space Fractional Derivative // J. Comput. Phys. 2013. V. 242. P. 670–681.
13. *Wang D.L., Xiao A.G., Yang W.* A Linearly Implicit Conservative Difference Scheme for the Space Fractional Coupled Nonlinear Schroedinger Equations // J. Comput. Phys. 2014. V. 272. P. 644–655.
14. *Wang P.D., Huang C.M.* An Energy Conservative Difference Scheme for the Nonlinear Fractional Schroedinger Equations // J. Comput. Phys. 2015. V. 293. P. 238–251.
15. *Zhang R.P., Zhang Y.T., Wang Z. et al.* A Conservative Numerical Method for the Fractional Nonlinear Schroedinger Equation in Two Dimensions // Sci. China Math. 2019. V. 62. P. 1997–2014.
16. *Zhao X., Sun Z.Z., Hao Z.P.* A Fourth-order Compact ADI Scheme for Two-dimensional Nonlinear Space Fractional Schroedinger Equation // SIAM J. Sci. Comput. 2014. V. 36. P. A2865–A2886.
17. *Laskin N.* Fractional Schroedinger Equation // Phys. Rev. E. 2002. V. 66. P. 056108.
18. *Riesz M.* Lintegrale de Riemann-Liouville et le Probleme de Cauchy // Acta Math. 1949. V. 81. P. 1–222.
19. *Guo B.L., Han Y.Q., Xin J.* Existence of the Global Smooth Solution to the Period Boundary Value Problem of Fractional Nonlinear Schroedinger Equation // Appl. Math. Comput. 2008. V. 204. P. 468–477.
20. *Luchko Y.* Fractional Schroedinger Equation for a Particle Moving in a Potential Well // J. Math. Phys. 2013. V. 54. P. 012111.
21. *Bao W.Z., Cai Y.Y.* Mathematical Theory and Numerical Methods for Bose-Einstein Condensation // arXiv preprint. 2012. arXiv:1212.5341
22. *Carr L.D., Clark C.W., Reinhardt W.P.* Stationary Solutions of the One Dimensional Nonlinear Schroedinger Equation I. Case of Repulsive Nonlinearity // Phys. Rev. A. 2000. V. 62. P. 063610.
23. *Jin S., Levermore C.D., McLaughlin D.W.* The Semiclassical Limit of the Defocusing NLS Hierarchy // Comm. Pure Appl. Math. 1999. V. 52. P. 613–654.
24. *Bao W.Z., Jaksch D.* An Explicit Unconditionally Stable Numerical Method for Solving Damped Nonlinear Schroedinger Equations with a Focusing Nonlinearity // SIAM J. Numer. Anal. 2003. V. 41. P. 1406–1426.
25. *Saito H., Ueda M.* Intermittent Implosion and Pattern Formation of Trapped Bose-Einstein Condensates with an Attractive Interaction // Phys. Rev. Lett. 2001. V. 86. P. 1406–1409.
26. *Ran Y.H., Wang J.G., Wang D.L.* On HSS-like Iteration Method for the Space Fractional Coupled Nonlinear Schroedinger Equations // Appl. Math. Comput. 2015. V. 271. P. 482–488.
27. *Ran Y.H., Wang J.G., Wang D.L.* On Partially Inexact HSS Iteration Methods for the Complex Symmetric Linear Systems in Space Fractional CNLS Equations // J. Comput. Appl. Math. 2017. V. 317. P. 128–136.
28. *Ran Y.H., Wang J.G., Wang D.L.* On Preconditioners Based on HSS for the Space Fractional CNLS Equations // East Asian J. Appl. Math. 2017. V. 7. P. 70–81.
29. *Wang Z.Q., Yin J.F., Dou Q.Y.* Preconditioned Modified Hermitian and Skew-Hermitian Splitting Iteration Methods for Fractional Nonlinear Schroedinger Equations // J. Comput. Appl. Math. 2020. V. 367. P. 112420.
30. *Zhang F.Y., Yang X.* Diagonal and Normal with Toeplitz-block Splitting Iteration Method for Space Fractional Coupled Nonlinear Schroedinger Equations with Repulsive Nonlinearities // arXiv preprint. 2023. arXiv: 2039.11106
31. *Bai Z.Z., Golub G.H., Ng M.K.* Hermitian and Skew-Hermitian Splitting Methods for Non-Hermitian Positive Definite Linear Systems // SIAM J. Matrix Anal. Appl. 2003. V. 24. P. 603–626.
32. *Bai Z.Z., Golub G.H., Pan J.Y.* Preconditioned Hermitian and Skew-Hermitian Splitting Methods for Non-Hermitian Positive Semidefinite Linear Systems // Numer. Math. 2004. V. 98. P. 1–32.
33. *Bai Z.Z., Benzi M., Chen F.* Modified HSS Iteration Methods for a Class of Complex Symmetric Linear Systems // Computing. 2010. V. 87. P. 93–111.
34. *Bai Z.Z., Benzi M., Chen F.* On Preconditioned MHSS Iteration Methods for Complex Symmetric Linear Systems // Numer. Algorithms. 2011. V. 56. P. 297–317.
35. *Bai Z.Z., Benzi M., Chen F. et al.* Preconditioned MHSS Iteration Methods for a Class of Block Two-by-two Linear Systems with Applications to Distributed Control Problems // IMA J. Numer. Anal. 2013. V. 33. P. 343–369.

36. *Axelsson O., Kucherov A.* Real Valued Iterative Methods for Solving Complex Symmetric Linear Systems // Numer. Linear Algebra Appl. 2000. V. 7. P. 197–218.
37. *Golub G.H., van Loan C.F.* Matrix Computations // 4th Edn. Baltimore: Johns Hopkins University Press, 2013.
38. *Saad Y.* Iterative Methods for Sparse Linear Systems // 2nd Edn. Philadelphia: Society for Industrial and Applied Mathematics, 2003.
39. *Chan R.H., Ng K.P.* Fast Iterative Solvers for Toeplitz-plus-band Systems // SIAM J. Sci. Comput. 1993. V. 14. P. 1013–1019.
40. *Ng M.K., Pan J.Y.* Approximate Inverse Circulant-plus-diagonal Preconditioners for Toeplitz-plus-diagonal Matrices // SIAM J. Sci. Comput. 2010. V. 32. P. 1442–1464.
41. *Bai Z.Z., Lu K.L., Pan J.Y.* Diagonal and Toeplitz Splitting Iteration Methods for Diagonal-plus-Toeplitz Linear Systems from Spatial Fractional Diffusion Equations // Numer. Linear Algebra Appl. 2017. V. 24. P. e2093.
42. *Bai Z.Z., Lu K.Y.* Fast Matrix Splitting Preconditioners for Higher Dimensional Spatial Fractional Diffusion Equations // J. Comput. Phys. 2020. V. 404. P. 109117.
43. *Peaceman D.W., Rachford H.H., Jr. T*he Numerical Solution of Parabolic and Elliptic Differential Equations // J. Soc. Ind Appl. Math. 1955. V. 3. P. 28–41.
44. *Douglas J.* Alternating Direction Methods for Three Space Variables // Numer. Math. 1962. V. 4. P. 41–63.
45. *Celik C., Duman M. C*rank-Nicolson Method for the Fractional Diffusion Equation with the Riesz Fractional Derivative // J. Comput. Phys. 2012. V. 231. P. 1743–1750.
46. *Ortigueira M.D.* Riesz Potential Operators and Inverses via Fractional Centred Derivatives // Int. J. Math. Math. Sci. 2006. P. 1–12. (Aticle ID 48391).
47. *Chan R.H., Strang G.* Toeplitz Equations by Conjugate Gradients with Circulant Preconditioner // SIAM J. Sci. Stat. Comput. 1989. V. 10. P. 104–119.
48. *Chan T.* An Optimal Circulant Preconditioner for Toeplitz Systems // SIAM J. Sci. Stat. Comput. 1988. V. 9. P. 766–771.
49. *Chan R.H., Ng M.K.* Conjugate Gradient Methods for Toeplitz Systems // SIAM Rev. 1996. V. 38. P. 427–482.
50. *Bauer F.L., Fike C.T.* Norms and Exclusion Theorems // Numer. Math. 1960. V. 2. P. 137–141.
51. *Chan R.H., Jin X.Q.* An Introduction to Iterative Toeplitz Solvers. Philadelphia: Society for Industrial and Applied Mathematics, 2007.
52. *Bebendorf M.* Hierarchical Matrices. Heidelberg: Springer-Verlag, 2008.
53. *Ho K.L., Ying L.* Hierarchical Interpolative Factorization for Elliptic Operators: Differential Equations // Commun. Pur. Appl. Math. 2016. V. 69. P. 1415–1451.

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, с. 33–39

КОМПЬЮТЕРНЫЕ МЕТОДЫ

УДК 519.7

ЛОГИЧЕСКАЯ КЛАССИФИКАЦИЯ НА ОСНОВЕ ПОИСКА ПРАВИЛЬНЫХ ПРЕДСТАВИТЕЛЬНЫХ ЭЛЕМЕНТАРНЫХ КЛАССИФИКАТОРОВ

© 2024 г. Н. А. Драгунов[a], *, Е. В. Дюкова[a], А. П. Дюкова[a]

[a]ФИЦ ИУ РАН, Москва, Россия

**e-mail: nikitadragunovjob@gmail.com*

Поступила в редакцию 29.01.2024 г.

После доработки 19.03.2024 г.

Принята к публикации 13.05.2024 г.

Рассмотрен подход к задаче классификации по прецедентам, базирующийся на применении аппарата дискретной математики (логических методов анализа данных). Исследована возможность сокращения временных затрат на стадии обучения корректного логического классификатора. Предложены новые модели классификаторов, основанные на поиске в описаниях прецедентов часто встречающихся фрагментов специального вида, названных правильными элементарными классификаторами. Описания моделей классификаторов даны с использованием понятий теории логических функций. Для построения искомых фрагментов авторами разработан и реализован оригинальный алгоритм. Эффективность предлагаемых моделей классификаторов обоснована экспериментально и подтверждена теоретическими оценками сложности их обучения. Получена верхняя асимптотическая оценка типичного числа правильных элементарных классификаторов.

Ключевые слова: задача классификации по прецедентам, корректная классификация, представительный элементарный классификатор, правильный элементарный классификатор, тупиковое покрытие целочисленной матрицы.

DOI: 10.31857/S0002338824040027 **EDN:** UENRUE

LOGICAL CLASSIFICATION BASED ON FINDING REGULAR REPRESENTATIVE ELEMENTARY CLASSIFIERS

N. Dragunov[a], *, E. Djukova[a], A. Djukova[a]

[a]Federal Research Center «Computer Science and Control» of the Russian Academy of Sciences, Moscow, Russia

**e-mail: nikitadragunovjob@gmail.com*

An approach to the supervised classification problem based on the apparatus of discrete mathematics (logical methods of data analysis) is considered. The possibility of time costs reducing at the stage of correct logical classifier training is investigated. New models of classifiers are proposed. These models are based on finding frequently occurring fragments of a special type in the descriptions of precedents — regular elementary classifiers. Descriptions of classifier models are given using the concepts of logical functions theory. To construct sought fragments, the authors have developed and implemented an original algorithm. The effectiveness of proposed classifier models has been experimentally substantiated and confirmed by theoretical estimates of their training complexity. An upper asymptotic estimate of the typical number of regular elementary classifiers is obtained.

Keywords: supervised classification problem, correct classification, representative elementary classifier, regular elementary classifier, irredundant covering of an integer matrix.

Введение. Задача классификации по прецедентам является одной из основных задач интеллектуального анализа данных и формулируется следующим образом. Исследуется некоторое множество объектов M, описываемых в системе числовых признаков $x_1,\ldots,x_n$. Известно, что M редставимо в виде объединения l подмножеств $K_1,\ldots,K_l$, называемых классами. Дан набор объектов из M, о которых известно, каким классам они принадлежат. Это прецеденты или

обучающие объекты. Требуется на базе анализа множества прецедентов построить алгоритм, определяющий класс любого объекта из M.

Дискретный или логический подход к задаче классификации предполагает, что каждый признак имеет ограниченное число допустимых значений, каждое из которых кодируется целым числом. Рассматриваемый подход имеет целью построение корректных моделей классификаторов, обеспечивающих безошибочное распознавание прецедентов.

Одними из известных направлений логической классификации являются LAD (logical analysis of data) и CVP (correct voting procedures). Каждое из направлений базируется на поиске таких фрагментов описаний прецедентов, которые позволяют отличать прецеденты из разных классов. В LAD искомые фрагменты называют логическими закономерностями, а в CVP — представительными элементарными классификаторами. Различным образом определяется понятие информативности фрагмента. В первом случае ищутся «максимальные» логические закономерности и решается сложная в вычислительном плане оптимизационная задача линейного программирования. Во втором случае ищутся «тупиковые» (в некотором смысле минимальные) представительные элементарные классификаторы, при этом возникают труднорешаемые дискретные перечислительные задачи. Направление LAD предложено в [1] и в основном развивается за рубежом. В России это направление представлено работами [2, 3]. Для направления CVP основополагающими являются публикации отечественных ученых [4–10].

Логические классификаторы наиболее эффективны в случае целочисленной информации низкой значности. Их описание может быть дано с использованием аппарата функций k-значной логики ($k \geq 2$). Тогда представительный элементарный классификатор (логическая закономерность) класса K является элементарной конъюнкцией над переменными $x_1,\ldots,x_n$, принимающей значение 1 на описании хотя бы одного прецедента из класса K и значения 0 на описаниях всех прецедентов из других классов [6].

Поиск тупиковых представительных элементарных классификаторов класса K основан, как правило, на первоначальном анализе множества прецедентов из других классов и сводится к решению сложной перечислительной задачи, называемой монотонной дуализацией, или к обобщениям этой задачи [6, 8]. Фактически сначала строятся элементарные конъюнкции над переменными $x_1,\ldots,x_n$, принимающие значение 0 на описаниях тех прецедентов, которые не принадлежат классу K, и теряющие это свойство при удалении хотя бы одного сомножителя. Затем из найденных конъюнкций отбираются те, которые не менее p ($p \geq 1$) раз принимают значение 1 на описаниях прецедентов класса K, т.е. отбираются тупиковые p-представительные элементарные классификаторы класса K (здесь p-настраиваемый параметр). В данной модели классификатора, обозначаемой далее A_0, вычисление оценки принадлежности распознаваемого объекта классу K существляется на основе проведения классической процедуры «голосования» [2], в которой участвуют все отобранные элементарные классификаторы.

В настоящей работе предлагаются и исследуются модели A_1, A_2, A_3 корректных логических классификаторов, обучение которых осуществляется путем поиска для каждого класса K так называемых правильных p-представительных элементарных классификаторов, т.е. таких представительных элементарных классификаторов этого класса, которые имеют ранг p ($p \geq 1$) и не менее p раз принимают значение 1 на описаниях прецедентов класса K. При этом классификатор A_1 действует по схеме классификатора A_0, но в голосовании участвуют только те тупиковые p-представительные элементарные классификаторы класса K, которые имеют ранг p. Классификаторы A_2 и A_3 действуют по иной схеме. Первоначально анализируются описания прецедентов класса K и строятся элементарные конъюнкции, которые не менее p раз принимают значение 1 на описаниях прецедентов этого класса и имеют ранг p. Такие конъюнкции называются правильными элементарными классификаторами. Затем просматриваются прецеденты из других классов и в A_2 из найденных конъюнкций отбираются представительные элементарные классификаторы класса K, а в A_3 отбираются тупиковые представительные элементарные классификаторы класса K. Процедура вычисления оценки принадлежности распознаваемого объекта классу K такая же, как и в алгоритме A_0.

Таким образом, на этапе обучения модель A_1 решает задачу монотонной дуализации, а модели A_2 и A_3 осуществляют поиск правильных элементарных классификаторов, базирующийся на предложенном в работе оригинальном алгоритме. Идея применения методов поиска часто встречающихся фрагментов в данных на этапе обучения логического классификатора была анонсирована авторами в [11].

Экспериментальное сравнение рассматриваемых алгоритмов на реальных и случайных модельных данных свидетельствует о целесообразности (в плане сокращения временных затрат) предлагаемого подхода к построению логических классификаторов. Получены теоретические результаты, характеризующие сложность обучения классификаторов A_2 и A_3 для случая, когда число

прецедентов класса K существенно больше числа признаков n. В экспериментах значение параметра p выбиралось согласно оценке типичного ранга правильного элементарного классификатора.

1. Основные понятия. Описание классификаторов A_1, A_2 и A_3. Пусть E_k^n, $k \geq 2$ – множество наборов вида $(\alpha_1,\ldots,\alpha_n)$, где $\alpha_i \in \{0,1,\ldots,k-1\}$.

Элементарной конъюнкцией над переменными $x_1,\ldots,x_n$ называется функция вида $x_{j_1}^{\sigma_1} \& \ldots \& x_{j_r}^{\sigma_r}$, где $\sigma_i \in \{0,1,\ldots k-1\}$, $x_{j_i} \in \{x_1,\ldots,x_n\}$ при $i = \overline{1,r}$, и при $r \geq 2$ выполнено $x_{j_q} \neq x_{j_t}$, $t = \overline{1,r}$, $q = \overline{1,r}$, $t \neq q$. Для краткости знак & опускается. Конъюнкция $B = x_{j_1}^{\sigma_1} \ldots x_{j_r}^{\sigma_r}$ обращается в 1 на тех наборах $(\alpha_1,\ldots,\alpha_n)$ из E_k^n, в которых $\alpha_{j_i} = \sigma_i$, $i = \overline{1,r}$. Множество наборов из E_k^n, на которых B принимает значение 1, обозначается через N_B, а через $\mathcal{B}(n,k)$ – множество всех элементарных конъюнкций рассматриваемого вида. Не ограничивая общности, можно считать, что объекты из исследуемого множества M описаны признаками, каждый из которых принимает значения из множества $\{0,1,\ldots,k-1\}$.

Пусть $K \in \{K_1,\ldots K_l\}$. Зададим на множестве прецедентов двузначную частичную (не всюду определенную) функцию $f_K(x_1,\ldots,x_n)$, которая принимает значение 1 на наборах, являющихся описаниями прецедентов класса K, и значение 0 на наборах, описывающих остальные обучающие объекты. Функция $f_K(x_1,\ldots,x_n)$ называется характеристической функцией класса K. Решение задачи классификации заключается в доопределении f_K на наборах, не входящих в обучающую выборку.

Далее U_K и Z_K обозначают соответственно множества прецедентов, на которых функция f_K равна 1 и 0. Положим $|U_K| = m_1$, $|Z_K| = m_2$, $1 \leq p \leq m_1$ (здесь и далее $|W|$ – мощность множества W).

Элементарным классификатором (*ЭК*) *ранга r* называется элементарная конъюнкция из $\mathcal{B}(n,k)$, зависящая от r переменных. ЭК B называется покрытием для Z_K, если $N_B \cap Z_K = \varnothing$. ЭК B, являющийся покрытием для Z_K, называется тупиковым покрытием для Z_K, если не существует покрытия B' для Z_K, такого, что $N_B \subset N_{B'}$.

Пусть $p \in \{1,2\ldots,m_1\}$. ЭК B называется p-частым в U_K, если $|N_B \cap U_K| \geq p$. ЭК B называется *p-представительным* для класса K, если B – p-частый в U_K и B – покрытие для Z_K. ЭК B называется *тупиковым p-представительным* для класса K, если B – p-частый в U_K и B – тупиковое покрытие для Z_K.

ЭК B ранга p называется *правильным* для U_K, если B – p-частый в U_K. ЭК B ранга p называется *правильным p-представительным* для класса K, если B – p-частый в U_K и B – покрытие для Z_K.

Приведем подробное описание моделей корректных классификаторов A_1, A_2 и A_3, о которых говорилось во Введении. Пусть $T_1(p, K)$ – множество всех тупиковых правильных p-представительных ЭК для класса K; $T_2(p, K)$ – множество всех правильных p-представительных ЭК для класса K; $T_3(p, K) = T_1(p, K)$; P_B^i, $B \in \mathrm{T}_i(p,K)$, $i \in \{1,2,3\}$, – число объектов S в U_K, таких, что $S \in N_B$.

На стадии обучения классификатор A_i, $i \in \{1,2,3\}$, строит некоторое множество ЭК из $\mathrm{T}_i(p,K)$. На следующей стадии (стадии распознавания) каждый найденный ЭК B участвует в процедуре голосования, заключающейся в вычислении величин P_B^i и $\Omega(B,S)$, где S – распознаваемый объект и $\Omega(B, S) = 1$, если $S \in N_B$, иначе $\Omega(B, S) = 0$. В результате получается оценка $\Gamma_i(S,K)$ принадлежности объекта S классу K, имеющая вид

$$\Gamma_i(S,K) = \frac{1}{|T_i(p,K)|} \sum_{B \in T_i(p,K)} P_B^i \ \Omega(B,S).$$

Объект S относится к классу с наибольшей оценкой. Если таких классов несколько, то объект относится к классу с наибольшим числом прецедентов.

В модели A_1 множество $T_1(p,K)$ строится в два этапа. Сначала анализируется множество Z_K и строятся тупиковые покрытия для Z_K ранга p. При этом решается задача монотонной дуализациии, которая относится к труднорешаемым дискретным перечислительным задачам. Затем из найденных ЭК отбираются те, которые являются p-частыми в U_K. Основная вычислительная сложность в этой модели заключается в необходимости решать задачу монотонной дуализации. Эффективность перечислительных задач принято оценивать сложностью нахождения нового решения (сложностью одного шага). В настоящее время для монотонной дуализации не построен алгоритм с полиномиальным шагом (алгоритм с полиномиальной задержкой [12]). Наиболее

эффективными в практическом отношении для этой задачи являются асимптотически оптимальные алгоритмы [10].

В моделях A_2 и A_3 множества $\mathrm{T}_2(p,K)$ и $\mathrm{T}_3(p,K)$ строятся также в два этапа. Однако, в отличие от модели A_1, сначала вместо анализа множества Z_K проводится анализ множества U_K, которое обычно меньше по мощности, чем Z_K, в случае, если число классов больше двух. В результате такого анализа строится множество правильных ЭК для U_K ранга p. На втором этапе в моделях A_2 и A_3 из найденных ЭК отбираются соответственно покрытия для Z_K и тупиковые покрытия для Z_K.

В настоящей работе при реализации классификаторов A_1, A_2 и A_3 к исходным данным применялась известная процедура one-hot кодирования [9]. В результате классификаторы работали с бинарными описаниями объектов. Для поиска правильных ЭК в бинарных данных разработан алгоритм ADR, описание которого приведено в разд. 2.

2. Алгоритм ADR поиска правильных ЭК. Типичное число правильных ЭК. Обозначим через L матрицу, строками которой являются бинарные описания объектов класса K, полученные с помощью one-hot кодирования.

Пусть Q – набор различных столбцов матрицы L, L^Q – подматрица матрицы L образованная набором Q. Набор столбцов Q называется p-частым, если L^Q содержит не менее p строк, все элементы которых равны 1. Набор столбцов Q называется p-правильным, если он p-частый и его мощность равна p. Несложно видеть, что поиск всех правильных ЭК ранга p эквивалентен поиску всех p-правильных наборов столбцов матрицы L.

Обозначим через $R(L,p)$ множество всех столбцов матрицы L, имеющих не менее p элементов, равных 1. Пронумеруем столбцы матрицы L слева направо, начиная с 1. Пусть $e_1(R)$ и $e_2(R)$ – столбцы соответственно с наименьшим номером и наибольшим номером из R, $R \subseteq R(L,p)$. Через $U_p(L)$ обозначим множество всех p-частых наборов столбцов матрицы L, мощность которых не превосходит p. Алгоритм ADR строит множество всех p-правильных наборов столбцов матрицы L, перечисляя с полиномиальной задержкой наборы из $U_p(L)$.

Определим порядок, в котором происходит перечисление наборов из $U_p(L)$. На первом шаге рассматривается набор $Q = \{e_1(R(L,p))\}$.

Пусть на шаге $i\,(i \geq 1)$ построен набор $Q \in U_p(L)$, состоящий из столбцов с номерами $j_1,\ldots,j_r$, $j_1 < \ldots < j_r$, $r \leq p$. Если $Q = \{e_2(R(L,p))\}$, то алгоритм заканчивает работу. Если же $Q \neq \{e_2(R(L,p))\}$, то на шаге $i+1$ алгоритм ADR строит новый набор ΔQ из $U_p(L)$. При этом возможны два случая: $r < p$ и $r = p$. В первом случае алгоритм строит ΔQ согласно приведенным ниже правилам $1-4$. Во втором случае алгоритм строит ΔQ по правилам $2-4$.

Для описания правил построения ΔQ введем обозначения: Q_t, $t = \overline{1,r}$, – набор столбцов матрицы L с номерами $j_1,\ldots,j_t$; R_t, $t = \overline{1,r}$, – множество столбцов в $R(L,p)$, номера которых больше j_t; G_t, $t = \overline{1,r}$, $r < p$, – множество столбцов из R_t, каждый из которых в объединении со столбцами из Q_t образует набор из $U_p(L)$. Положим $G_r = \varnothing$ в случае $r = p$.

Заметим, что в случае $r < p$ для построения G_t в L нужно оставить только те столбцы, номера которых больше j_t и которые имеют не менее p элементов, равных 1 в подматрице, полученной после удаления из L строк, дающих 0 в пересечении со столбцами с номерами $j_1,\ldots,j_t$.

Положим $Q_0 = \varnothing$ и $G_0 = \varnothing$. Перечислим возможные случаи и в каждом из них укажем правила построения ΔQ:

1) $G_r \neq \varnothing : \Delta Q = Q_r \cup \{e_1(G_r)\}$;

2) $G_r = \varnothing,\ G_{r-1} \cap R_r \neq \varnothing : \Delta Q = Q_{r-1} \cup \{e_1(G_{r-1} \cap R_r)\}$;

3) $G_r = \varnothing,\ G_{r-1} \cap R_r = \varnothing,\ r = 1 : \Delta Q = \{e_1(R_r)\}$;

4) $G_r = \varnothing,\ G_{r-1} \cap R_r = \varnothing,\ r > 1 : \Delta Q = Q_{r-2} \cup \{e_1(G_{r-2} \cap R_{r-1})\}$.

Заметим, что $R_r \neq \varnothing$ при $r = 1$, так как $Q \neq \{e_2(R(L,p))\}$, и $G_{r-2} \cap R_{r-1} \neq \varnothing$ при $\mathrm{T}_2(p,K)$, так как столбец с номером j_r принадлежит этому множеству.

Из описания работы алгоритма ADR видно, что в его основе лежит процесс ветвления, который удобно представить в виде обхода дерева решений в глубину. Вершинами этого дерева являются наборы из $U_p(L)$, причем p-правильные наборы столбцов находятся среди висячих вершин. Через L_K обозначим матрицу, строками которой являются описания прецедентов класса K. Правильные ЭК порождаются квадратными подматрицами матрицы L_K, состоящими из одинаковых строк. Такие подматрицы назовем правильными.

Ниже приведены асимптотические оценки типичных значений числа правильных подматриц целочисленной матрицы L_K и порядка такой подматрицы в случае большого числа строк матрицы L_K. Пусть M_{mn}^k – множество всех целочисленных матриц размера $m \times n$ с элементами из $\{0,1,\ldots,k-1\}$; $S(L)$, $L \in M_{mn}^k$, – множество правильных подматриц в матрице L; $\phi_k(m,n)$ –

интервал $(0,\ r(k,m,n))$, где $r(k,m,n) = 0.5\log_k mn - 0.5\log_k\log_k mn + \log_k\log_k\log_k n$; $b_n \sim c_n$, $n\to\infty$ означает, что $lim_{n\to\infty} b_n / c_n = 1$.

Т е о р е м а. Если $n^\alpha \le m \le k^n$, $\alpha > 1$, то при $n\to\infty$ для почти всех матриц L из M^k_{mn} справедливо

$$|S(L)| \sim \sum_{r\in\phi_k(m,n)} C^r_n C^r_m k^{r-r^2}.$$

и порядки почти всех подматриц из $S(L)$ принадлежат интервалу $\phi_k(m,n)$.

Доказательство теоремы аналогично доказательству теоремы 3 из [13], в которой при тех же ограничениях на m и n получена асимптотическая оценка типичного числа так называемых σ -подматриц матрицы L, служащая верхней оценкой числа тупиковых покрытий для Z_K при условии, что $|Z_K| = m$.

Приведенная в теореме оценка типичного порядка подматрицы из $S(L)$ косвенно свидетельствуют о том, что в случае, когда число прецедентов m_1 класса K существенно больше числа признаков n, типичный ранг правильного ЭК в U_K не превосходит $r(k,m_1,n)$.

З а м е ч а н и е 1. В работе [14] получены асимптотические оценки типичного числа правильных ЭК в U_K для двух случаев: 1) $m_1{}^a \le n \le k^{m_1\beta}$, $a>1$, $\beta < 1$; 2) $n \le m_1 \le k^{n^\beta}$, $\beta < 1/2$. Авторами показано, что типичный ранг правильного ЭК в U_K в случаях 1) и 2) соответственно принадлежит интервалу $\phi_k(m_1,n)$ и не превосходит $\log_k m_1 + \log_k\log_k m_1$.

3. Результаты экспериментов. Результаты счета на реальных целочисленных задачах приведены в таблице. Задачи взяты из репозитория UCI [archive.ics.uci.edu] и репозитория ВЦ ФИЦ ИУ РАН. Описанные выше алгоритмы A_1, A_2, A_3 оценивались по качеству классификации и по времени обучения. Алгоритмы реализованы на языке программирования C++. В тестировании на качество классификации также участвовали такие известные алгоритмы, как случайный лес (RF) и логистическая регрессия (LR). Дополнительная настройка алгоритмов RF и LR не производилась.

Результаты счета усреднялись по 10 случайным независимым разбиениям прецедентов, 80% которых использовалось для обучения моделей, а 20% — для оценки качества классификации. В каждом из разбиений распределение прецедентов по классам сохранялось неизменным.

Таблица 1.

m, n_1, l	Время, мс			Качество			
$(p_1, \dots, p_1)$	A_1	A_2	A_3	A_1, A_3	A_2	RF	LR
144, 379, 2 (3, 3)	512.1	**47.0**	48.6	0.691	0.735	0.742	**0.774**
267, 566, 2 (3, 4)	289.2	**18.3**	18.4	0.560	0.570	0.545	**0.578**
957, 27, 2 (3, 3)	71.7	**1.0**	**1.0**	**0.976**	**0.976**	0.939	0.639
79, 160, 2 (2, 3)	238.4	**140.0**	150.0	0.614	**0.623**	0.542	0.553
3195, 73, 2 (4, 4)	5294.0	**903.7**	1061.9	0.903	0.974	**0.988**	0.956
1532, 284, 2 (5, 5)	2763106	**59265**	69387	0.960	**0.971**	0.960	0.922
2056, 83, 3 (4, 4, 4)	35471	**8.3**	9.4	0.641	0.770	**0.905**	0.790
3190, 287, 3 (5, 5, 5)	10487213	**235045**	315275	0.793	0.794	**0.946**	0.831

В таблице последовательно указаны результаты счета для следующих задач: Манелис, Остеосаркома, Крестики-нолики (UCI), Инсульт, Шахматы (UCI), Молекулярная Биология 1 (UCI), Задача 5, Молекулярная Биология 2 (UCI). Для каждой задачи указаны число прецедентов m, число признаков n_1 полученное после one-hot перекодировки, число классов l и ранг p_i,

$i \in \{1,2,\ldots,l\}$, голосующих ЭК класса K_i. Время работы алгоритмов указано в миллисекундах. Функционалом качества выбрана сбалансированная точность классификации, вычисляемая по формуле

$$\psi = \sum_{i=1}^{l} q_i \,/\, l,$$

где q_i — доля верно классифицированных объектов класса K_i. Данный функционал хорошо себя зарекомендовал при несбалансированных классах. В случае равномощных классов сбалансированная точность совпадает с долей верно классифицированных объектов.

Как видно из таблицы, модель A_2 превосходит по качеству и времени работы модели A_1 и A_3 на всех рассмотренных данных, кроме задачи Крестики-нолики, и в среднем не уступает по качеству ни случайному лесу, ни логистической регрессии. На трех задачах (Крестики-нолики, Инсульт, Молекулярная Биология 1) модель A_2 превосходит все модели.

Модель A_1 работает существенно медленнее модели A_3 при том, что оба алгоритма строят множество всех тупиковых p-представительных ЭК ранга p. Однако модель A_1 на первом этапе обучения ищет тупиковые покрытия для Z_K ранга p, а модель A_3 перечисляет правильные ЭК ранга p для U_K. Стоит отметить, что на шести задачах (Манелис, Остеосаркома, Крестики-нолики, Инсульт, Шахматы, Задача 5) модели A_2 и A_3 обучались менее чем за 1 с, что свидетельствует об их высокой вычислительной эффективности.

З а м е ч а н и е 2. В экспериментах ранг p_i, $i \in \{1,2,\ldots,l\}$, голосующих ЭК класса K_i брался равным числу $0.5\log_2 m_i n_1 - 0.5\log_2\log_2 m_i n_1 - \log_2\log_2\log_2 n_1$, где m_i — число прецедентов класса K_i. Обучение с таким рангом в среднем показывало лучшее качество по сравнению с обучением с другими значениями ранга p_i, также принадлежащими интервалу $\phi_2(m_i, n_1)$.

На рис. 1, 2 приведено время обучения моделей A_1 и A_2 на случайных модельных данных из равномерного распределения при $l = 2$, $k = 2$, m_1 — число прецедентов в каждом классе, n_1 — число признаков. Результаты счета усреднены по 20 независимым запускам. Время работы алгоритмов указано в секундах. Время счета модели A_3 не приводится на графиках, так как в рассматриваемых примерах оно практически совпадает с временем работы A_2.

На рис. 1 показан экспоненциальный рост временных затрат на этапе обучения классификаторов A_1 и A_2 при $m_1 = 250$ в зависимости от числа признаков n_1. Видно, что при относительно небольшом n_1 разрыв во времени счета для A_1 и A_2 незначителен. При $n_1 \geq 150$ алгоритм A_1 работает значительно медленнее алгоритма A_2. Например, A_1 обучается примерно в 1.3 раза медленнее A_2 при $n_1 = 150$, а при $n_1 = 250$ — в 1.7 раз медленнее.

На рис. 2 продемонстрирован линейный рост временных затрат на этапе обучения классификаторов A_1 и A_2 при $n_1 = 100$ в зависимости от числа прецедентов m_1. Видно, что время работы A_1 растет быстрее по сравнению с временем работы A_2. Например, A_1 обучается примерно в 1.2 раза медленнее A_2 при $m_1 = 100$ и почти в 2 раз медленнее при $m_1 = 700$.

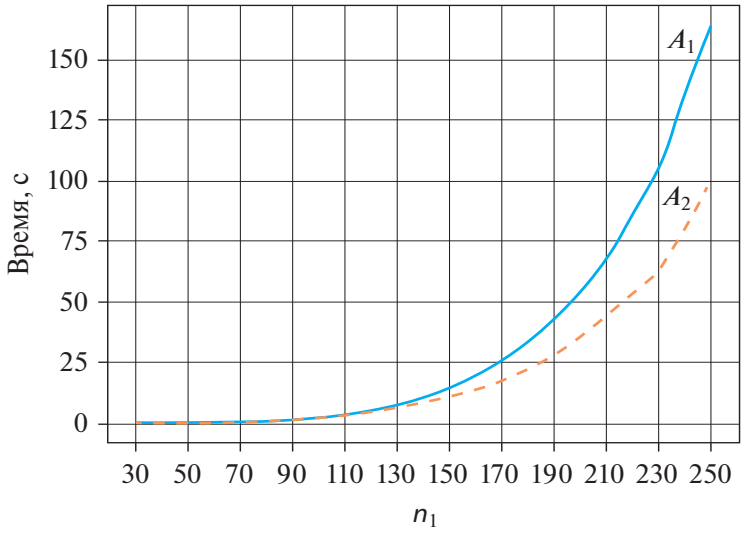

Рис. 1. Зависимость времени обучения моделей A_1 и A_2 от числа признаков при $m_1 = 250$

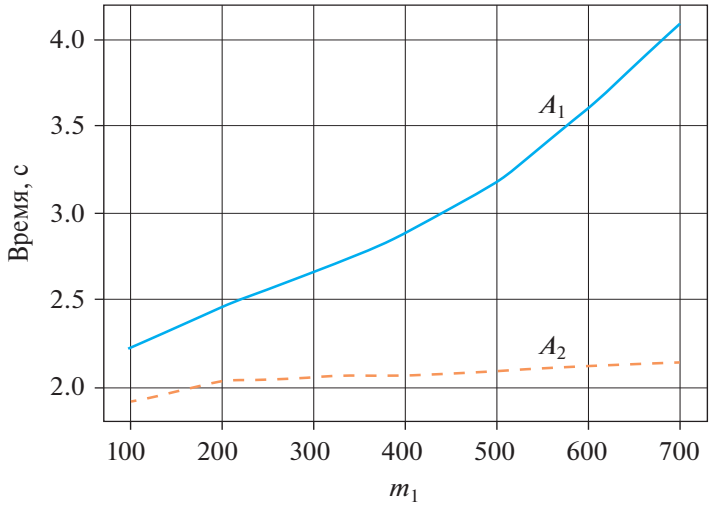

Рис. 2. Зависимость времени обучения моделей A_1 и A_2 от числа прецедентов при $n_1 = 100$

Заключение. Исследованы актуальные вопросы снижения временных затрат, возникающие при логическом анализе данных в задачах классификации на основе прецедентов. Предложены новые модели корректного голосования, базирующиеся на поиске в описаниях прецедентов каждого класса правильных ЭК ранга p (p – настраиваемый параметр модели). Разработан эффективный алгоритм для перечисления искомых правильных ЭК. Получена верхняя асимптотическая оценка типичного числа правильных ЭК для случая, когда число прецедентов существенно больше числа признаков. При этом указан типичный ранг правильного ЭК, который использован в экспериментах для выбора параметра p. Теоретические выводы подтверждены результатами экспериментального исследования на реальных и случайных модельных данных. А именно показано, что время обучения модели A_1, базирующейся на решении задачи монотонной дуализации, растет быстрее времени обучения модели A_2, основанной поиске правильных ЭК.

СПИСОК ЛИТЕРАТУРЫ

1. *Crama Y., Hammer P.L., Ibaraki T.* Cause-effect Relationships and Partially Defined Boolean Functions // Ann. Oper. Res. 1988. V. 16. Iss. 1. P. 299–325.
2. *Журавлёв Ю.И., Рязанов В.В., Сенько О.В.* Распознавание. Математические методы. Программная система. Практические применения. М.: ФАЗИС, 2006. 159 с.
3. *Масич И.С.* Метод оптимальных логических решающих правил для задач распознавания и прогнозирования // Системы управления и информационные технологии. 2019. Т. 75. № 1. С. 31–37.
4. *Бонгард М.М., Вайнцвайг М.Н., Губерман Ш.А., Извекова М.Л., Смирнов М.С.* Использование обучающейся программы для выявления нефтеносных пластов // Геология и геофизика. 1966. № 6.
5. *Баскакова Л.В., Журавлёв Ю.И.* Модель распознающих алгоритмов с представительными наборами и системами опорных множеств // ЖВМ и МФ. 1981. Т. 21. № 5. С. 1264–1275.
6. *Дюкова Е.В., Журавлёв Ю.И.* Дискретный анализ признаковых описаний в задачах распознавания большой размерности // ЖВМ и МФ. 2000. Т. 40. №8. С. 1264–1278.
7. *Яблонский С.В., Чегис И.А.* О тестах для электрических схем // УМН. 1955. Т. 10. Вып. 4(66). С. 182–184.
8. *Дюкова Е.В., Журавлёв Ю.И.* Задача монотонной дуализации и ее обобщения: асимптотические оценки числа решений // ЖВМ и МФ. 2018. Т. 58. № 12. С. 2153–2168.
9. *Дюкова Е.В., Инякин С.А.* Об асимптотически оптимальном построении тупиковых покрытий целочисленной матрицы // Математические вопросы кибернетики. 2008. № 17. С. 247–262.
10. *Дюкова Е.В., Прокофьев П.А.* Об асимптотически оптимальных алгоритмах дуализации // ЖВМ и МФ. 2015. Т. 55. № 5. С. 895–910.
11. *Dragunov N., Djukova E., Djukova. A.* Supervised Classification and Finding Frequent Elements in Data // 8th Intern. Conf. on Information Technology and Nanotechnology Proceedings. N.J.: IEEE, 2022. P. 5.
12. *Johnson D.S., Yannakakis M., Papadimitriou C.H.* On Generating All Maximal Independent Sets // Information Processing Letters. 1988. V. 27. Iss. 3.
13. *Дюкова Е.В., Песков Н.В.* Поиск информативных фрагментов описаний объектов в дискретных процедурах распознавания // ЖВМ и МФ. 2002. Т. 42. № 5. С. 741–753.
14. *Дюкова Е. В., Дюкова А. П.* О числе решений некоторых специальных задач логического анализа целочисленных данных // Изв. РАН. ТиСУ. 2023. № 5. С. 57–66.

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, с. 40–64

КОМПЬЮТЕРНЫЕ МЕТОДЫ

УДК 519.769

МЕТОДЫ РЕШЕНИЯ ЗАДАЧИ ТЕМАТИЧЕСКОЙ СЕГМЕНТАЦИИ ТЕКСТОВ НА ОСНОВЕ ГРАФОВ ЗНАНИЙ

© 2024 г. З. К. Авдеева[a, *], М. С. Гаврилов[a, b, **], Д. В. Лемтюжникова[a, *], А. Ф. Шарафиев[a, ****]**

[a] *Институт проблем управления им. В. А. Трапезникова РАН, Москва, Россия*

[b] *МАИ (национальный исследовательский ун-т), Москва, Россия*

**e-mail: avdeeva@ipu.ru*

***e-mail: cobraj@yandex.ru*

****e-mail: darabbt@gmail.com*

*****e-mail: whiskeydudev@gmail.com*

Поступила в редакцию 14.04.2024 г.

После доработки 12.05.2024 г.

Принята к публикации 15.07.2024 г.

Тематическая сегментация – это задача разделения неструктурированного текста на тематически связные сегменты (такие, в которых речь идет об одном и том же). Граф знаний – графовая структура, вершинами которой являются различные объекты, а ребрами – отношения между ними. Как задача тематической сегментации, так и задача автоматического построения графа знаний не будут новыми, поэтому существует множество алгоритмов для их решения. Однако методы решения задачи тематической сегментации с помощью графов знаний до сих пор исследованы мало. Более того, пока еще нельзя сказать, что задача тематической сегментации решена в общем виде, т.е.существуют алгоритмы, способные при должной настройке решить задачу с требуемым качеством на конкретном наборе данных. Предлагается новый метод решения задачи тематической сегментации на основе графов знаний. Применение графов знаний при сегментации позволяет использовать больше информации о словах в тексте: помимо того чтобы основываться на co-occurrance и семантических расстояниях (как классические алгоритмы), методы на базе графов знаний могут применять расстояние между словами на графе, инкорпорируя тем самым фактологическую информацию из графа знаний в процесс принятия решений о биении текста на сегменты.

Ключевые слова: тематическая сегментация, граф знаний, обработка естественного языка.

DOI: 10.31857/S0002338824040031 **EDN:** UENRQR

METHODS FOR SOLVING THE PROBLEM OF TOPIC SEGMENTATION OF TEXTS BASED ON KNOWLEDGE GRAPHS

Z. K. Avdeeva[a, *], M. S. Gavrilov[a, b, **], D. V. Lemtyuzhnikova[a, *], A. F. Sharafiev[a, ****]**

[a] *V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia*

[b] *Moscow Aviation Institute (National Research University), Moscow, Russia*

**e-mail: avdeeva@ipu.ru*

***e-mail: cobraj@yandex.ru*

****e-mail: darabbt@gmail.com*

*****e-mail: whiskeydudev@gmail.com*

Topic segmentation is the task of dividing unstructured text into thematically connected segments (such as those dealing with the same matter). The knowledge graph is a graph structure, the vertices of which are various objects, and the edges are the relationships between them. Both the task of topic segmentation and the task of automatically constructing a knowledge graph are not new, therefore there are many algorithms for solving them. However, methods for solving the problem of topic segmentation using knowledge graphs have so far been little studied. Moreover, yet it cannot be said that the problem of topic segmentation has been solved in a general way, that is, there are algorithms that, if properly configured, can solve the problem

with the required quality on a specific data set. In this paper, a new method for solving the problem of topic segmentation based on knowledge graphs is proposed. The use of knowledge graphs in segmentation allows us to use more information about words in the text: in addition to being based on co-occurrence and semantic distances (like classical algorithms), knowledge graph-based methods can apply the distance between words on the graph, thereby incorporating factual information from the knowledge graph into the decision-making process of partitioning the text into segments. In this paper, we propose a method for solving the problem of topic segmentation based on knowledge graphs.

Keywords: topic segmentation, knowledge graph, natural language processing.

Введение. В современном мире человеку приходится воспринимать и обрабатывать огромные потоки информации разной природы, основными из которых являются визуальная, слуховая и текстовая. Несмотря на то, что наиболее естественной является именно визуальная информация, людям нужно взаимодействовать с текстовыми данными: рабочими отчетами, литературными произведениями, постами в социальных сетях, субтитрами в фильмах и т.д.

В общем случае, поступающие человеку текстовые данные являются слабоструктурированными. В некоторых случаях, как, например, при чтении на электронном носителе скачанной из интернета книги, это не причиняет особого дискомфорта – достаточно того, что книга разделена на главы. Однако во многих ситуациях отсутствие видимой структуры в тексте затрудняет восприятие информации. Когда необходимо отыскать какую-то конкретную публикацию среди множества источников, поиск сильно усложняется, если это множество плохо структурировано.

Задача структурирования текстов возникает в большом количестве приложений, и для ее решения применяются различные подходы. Одной из ее устойчивых вариаций является задача тематической сегментации. Неформально она формулируется как отыскание разбиения документа на непересекающиеся последовательные подмножества таким образом, чтобы каждое подмножество характеризовалось высокой тематической связностью, другими словами, чтобы внутри каждого сегмента текст был об одном и том же.

Задача тематической сегментации имеет большое значение в области обработки естественного языка. Результаты решения данной задачи представляют интерес для использования в таких задачах, как суммаризация, моделирование диалога, в вопросно-ответных системах и др.

В статье предлагается формальная постановка задачи тематической сегментации в терминах графов, рассматриваются несколько алгоритмов для решения задачи тематической сегментации, исследуется качество их работы на научных статьях.

1. Задачи и алгоритмы тематической сегментации текстов и автоматического построения графов.

1.1. Обзор моделей, задач и алгоритмов. В общем виде предлагаемые методы решения задачи тематической сегментации текста представляют собой комбинацию из графа знаний, в которой отображаются предложения документа, и алгоритма поиска оптимальной группировки подграфов, соответствующих предложениям. Существует большой пласт методов и алгоритмов, которые решают задачу автоматического построения графа знаний. Приведем примеры некоторых работ, в которых они описываются.

В [1] граф знаний строится на основе синтаксического дерева. Из данного дерева извлекаются именные группы, соединенные заданными типами связи. Связи играют роль ребер. Помимо извлечения таких наборов, для каждого ребра рассчитывается вес на базе относительного расстояния между двумя вершинами в тексте, семантической схожести и энтропии. В [2] рассматривается метод построения онтологий на основе поиска семантических шаблонов. Семантический шаблон – это выражение на естественном языке, у которого есть смысл. Утверждается, что существующие методы построения онтологий ограничены поиском таксономических отношений, а нетаксономические отношения не рассматриваются. Предлагаемый в работе метод на базе семантических шаблонов призван помочь в решении этой проблемы. В [3] предлагаемый алгоритм осуществляет поиск и извлечение отношений с помощью структуры английских предложений. Для пар именованных фраз из текстов алгоритм извлекает фразы-отношения, удовлетворяющие определенным условиям: фраза не должна быть слишком длинной и должна удовлетворять заданным шаблонам (например, глагол-предлог). В [4] граф знаний строится следующим образом. Для заданного набора сущностей из сети Интернета набирается объем документов, которые содержат определенное число этих сущностей. Затем текстовые вхождения в документы, содержащие сущности из набора, подаются на вход

заранее обученной нейронной сети, которая, решая задачу классификации, предсказывает отношение между сущностями.

Существуют разные подходы сегментации, приведем четыре характерных на основе: лексического анализа, тематического моделирования, нейронных сетей и графа знаний. Сегментация на основе лексического состава заключается в сравнении употребляемого словаря в соседних блоках текста. К этому типу подходов относятся работы [5, 6]. В [5] документ токенизируется и разбивается на блоки с заданным числом токенов. Затем в соответствии с выбранной стратегией сравнения соседних блоков принимается решение о наличии или отсутствии тематической границы между ними. В статье рассматриваются три стратегии сравнения соседних блоков на основе: наличия одинаковых слов в обоих блоках, числа новых слов в последующем блоке относительно предыдущего и так называемых лексических цепочек, т.е. последовательностей предложений, в которых появляется то или иное слово. В [6] развивается подход публикации [5] с лексическими цепочками. В этой работе к ним добавляются определенным образом подсчитанные веса, в то время как в [5] просто рассматривалось, проходит или нет та или иная лексическая цепочка через блок. Суть второго направления заключается в интеграции задачи сегментации в задачу тематического моделирования текста. Здесь можно выделить два направления работ. В первом используется уже обученная тематическая модель, с помощью которой рассчитываются тематические векторные представления единиц документа. Во втором сегментирование документа встраивается в генеративный процесс документа. К этому типу подходов относятся работы [7, 8]. Так в [7] применяют обученную модель латентное размещение Дирихле (latent dirichlet allocation (LDA)), алгоритм для получения тематического вектора ранее неизвестного модели документа и динамическое программирование. Задача решается следующим образом. Принимается, что если сегмент является тематически более связным в сравнении с другим сегментом, то его функция правдоподобия должна быть больше. Таким образом, документ разделяется на всевозможные сегменты, для каждого из которых с помощью модели LDA выводится свой собственный тематический вектор. Затем с помощью динамического программирования находится та сегментация, которая максимизирует правдоподобие всего документа. В [8] рассматривается фиктивный генеративный процесс создания документов, как это делается, например, в модели LDA. Однако явным образом добавляется эвристика о том, что документ состоит из нескольких сегментов, каждый со своим тематическим распределением. Тематическое распределение для сегментов порождается с помощью процесса Дирихле-Пуассона и общим тематическим вектором для всего документа. Третий подход основывается на применении графа знаний. К этому типу подходов относится работа [9]. Представленный в [9] метод использует для решения задачи сегментации граф знаний. Непосредственно задача сегментации решается для длинных видеолекций, состоящих из слайдов, однако при этом применяется только текстовая информация, так что обобщить предложенный метод на только текстовые данные не составляет большого труда. Идея метода заключается в том, чтобы отобразить информацию, соответствующую каждому слайду, в подграф графа знаний. Данный граф знаний определенным образом строится на самом корпусе текстовой информации. Затем сравнивая, как подграфы соседних слайдов располагаются друг относительно друга, принимается решение о наличии или отсутствии тематической границы. Четвертый подход заключается в различных вариантах применения глубоких нейронных сетей. К этому типу подходов относятся работы [10–12]. Например, в [10] используется последовательно соединенные двунаправленные нейронные сети-кодировщики (bidirectional encoder representations from transformers (BERT)) и модель двунаправленной долгой краткосрочной памяти (bidirectional long short-term memory (Bi-LSTM)). Выходы последней пропускаются через блок пропускного подключения (skip-connection) и специальный блок многоцелевого внимания (multihop-attention), который рекуррентно реализует механизм внимания. Задача решается как классификация наличия в предложении тематического сдвига. В [11] применяются два последовательных трансформера. Первый трансформер используется для кодирования каждого предложения. Вектор предложения конкатенируется из двух частей – вектор прямого кодирования предложения и вектор, полученный кодированием конкатенации данного предложения с соседним. Совокупный вектор далее применяется в качестве входа следующего трансформера, чьи выходы уже используются для предсказания темы предложения и наличия в нем тематического перехода. В [12] для решения задачи сегментации был собран специальный датасет, состоящий из текстов Википедии, разбитых на секции. Каждая секция характеризовалась своим заголовком, а также меткой класса, проставленной в соответствии с заголовком. Архитектура модели представляла собой Bi-LSTM-кодировщик, который отображал каждое предложение текста в латентное представление. На полученных

векторах затем обучалось два классификатора: первый обучался предсказывать тему для предложения, второй – предсказывать слова, входящие в заголовок.

Приведенный обзор работ показывает, что направление решения задачи тематической сегментации с применением графов знаний исследовано на недостаточном уровне.

1.2. Постановка задачи. Неформально задача тематической сегментации текста звучит так: пусть есть некоторый документ, который необходимо разбить на части таким образом, чтобы каждая часть была тематически связной, т.е. в ней говорилось примерно об одном и том же. С формальной точки зрения на эту задачу можно смотреть с двух сторон. С одной стороны, искомой сегментацией можно считать разбиение документа, оптимальное по какому-либо критерию. В таком случае, "оптимальную" сегментацию необходимо искать среди множества всех возможных разбиений документа. Такую задачу сегментации можно назвать тематической сегментацией документа с глобальной оптимизацией. С другой стороны, практикуется иной подход, который заключается в сравнении примыкающих предложений (или блоков предложений), и на его основе принимается решение о наличии/отсутствии тематической границы между этими предложениями (блоками предложений). Другими словами, группировка предложений осуществляется с помощью локального сравнения. Такую задачу сегментации можно назвать тематической сегментацией документа с локальной оптимизацией. Мы предлагаем следующую формальную постановку для задачи глобальной тематической сегментации.

2. Типы задачи тематической сегментации.

2.1. Задача глобальной тематической сегментации. Формально эту задачу в терминах графов можно поставить следующим образом. Пусть есть некоторый документ D, состоящий из предложений d_i, длина документа в предложениях равна T. Считая, что существует заранее построенный граф знаний, поставим в соответствие каждому предложению подграф g_i из этого графа знаний следующим образом. Найдем в предложении термины, содержащиеся в графе знаний, и в качестве подграфа g_i возьмем подграф, порожденный множеством соответствующих этим терминам вершин.

Получим кортеж, состоящий из подграфов, обозначим его $S_1 = (g_1, g_2, ..., g_T)$. Применяя к этому кортежу операцию объединения подряд идущих подграфов в различных комбинациях, получим множество кортежей, элементами которых являются или сами подграфы – gi, или объединение подряд идущих подграфов, или и то и другое. Примеры таких кортежей – $S_2 = ((g_1, g_2), g_3, g_4, ..., g_T)$; $S_3 = (g_1, (g_2, g_3), g_4, ..., g_T)$; $S_i = ((g_1, ..., g_r), (g_{r+1}, ..., g_l), ..., (g_m, ..., g_T))$. Обозначим это множество $P = \{S_1, S_2, ..., S_n\}$. Полагая, что на множестве P задан некоторый функционал качества $L = L(S_i)$, будем называть оптимальной сегментацией кортеж S_i, на котором функция L принимает оптимальное значение (в частности, минимум).

Поскольку в данной постановке задачи рассматриваются всевозможные разбиения документа D, обозначим такую задачу сегментации текста как тематическая сегментация с глобальной оптимизацией.

2.2. Задача локальной тематической сегментации. Помимо этого, в литературе развивается другое направление методов, которые не подпадают под указанное определение. В этих методах решение о наличии или отсутствии тематической границы между предложениями принимается на основе локального сравнения двух соседних предложений (или в некоторых методах блоков предложений).

Формально такую задачу сегментации можно представить следующим образом. Как и в случае глобальной сегментации, документ, состоящий из предложений, отображается в кортеж подграфов общего графа знаний. Таким образом, пусть есть некоторый кортеж $S = (g_1, g_2, ..., g_T)$, соответствующий некоторому документу D. Считая, что существует некоторая функция $s(g_l, g_r)$, вычисляющая схожесть между подграфами g_l и g_r, будем относить к одному сегменту те подграфы, для которых $s(g_l, g_r) > TH$, где TH – некоторое задаваемое значение.

3. Графы знаний. Граф знаний (knowledge graph) – это графовая структура, в которой хранится информация о разных сущностях и взаимосвязях между ними, вершины в графе знаний суть некоторые концепции или понятия, а ребра – связи между концепциями. В работе используются три вида графов знаний: на основе тезаурусов по теории управления, классификатора и синтаксического графа. В контексте исследуемых методов графы знаний рассматриваются в качестве инструмента получения представлений для предложений. В табл. 1–3 представлен внешний вид тезаурусов и классификатора.

3.1. Описание используемых графов знаний. Граф знаний на основе тезаурусов по теории управления строится по табл. 1, 2 определений, которые представляют

Таблица 1. Представление тезауруса 1988 г.

Наименование термина/уровня	Английский термин	Определение	Примечание
Основные понятия			
Объект управления	Controlledobject; controllable object	Объект, для достижения желаемых результатов функционирования которого необходимы и допустимы специально организованные воздействия	1. Объект управления, подвергаемый управляющим воздействиям, можно называть управляемым объектом. 2. Объектами управления могут быть как отдельные объекты, выделенные по определенным признакам (например, конструктивным, функциональным), так и совокупности объектов-комплексов. 3. В зависимости от свойств или назначения объектов управления могут быть выделены технические, технологические, экономические, организационные, социальные и другие объекты управления и комплексы
Цель управления	Control aim	Значения, соотношения значений координат процессов в объекте управления или их изменения во времени, при которых обеспечивается достижение желаемых результатов функционирования объекта	

собой пары – термин и определение этого термина через другие термины. Вершинами в данном графе являются термины из табл. 1, 2. Вершины соединяются ребрами, если один термин определяется через другой. Отношение «определен через» в общем смысле не является симметричным (т.е. связь будет направленной), однако на первых этапах от направленности было решено отказаться.

Для построения таких графов использовалось два предметных тезауруса по теории управления, составленных в 1988 г. [13] и 2024 г. [14], которые отличаются набором и количеством вершин. Помимо этого, рассматривался совокупный граф, представляющий собой объединение графов, описанных выше. Графы объединялись таким образом, что непересекающиеся вершины, т.е. вершины, которые принадлежат только одному графу, добавлялись вместе со всеми связями между данными непересекающимися вершинами. После этого добавлялись пересекающиеся вершины и соответствующие связи с непересекающимися вершинами.

Граф классификатора строится на основе онтологии по теории управления, собранной экспертами. Классификатор представляет собой лес, содержащий три дерева. Каждое дерево имеет четыре уровня, отражающих общность находящихся на этом уровне терминов (т.е. чем ниже находится термин, тем он более конкретный). Пример последовательности терминов в порядке увеличения глубины: математический аппарат, алгебра и теория чисел, теория чисел, простое число. Граф знаний на базе классификатора строился следующим образом. В качестве вершин в графе брались вершины онтологии. Две вершины соединяются ребром, если они принадлежали одному дереву.

Синтаксический граф строится с помощью синтаксического дерева разбора предложений в обучающей выборке. Для каждого предложения из корпуса текстов выполняется синтаксический разбор. Из полученного синтаксического дерева извлекаются имена, глаголы и синтаксические связи между ними. Извлеченные связи и слова записываются в граф. Также

Таблица 2. Представление тезауруса 2024 г.

Термин	Английский термин	Определение
Абдукция	Abduction	Вид рассуждения, использующий абдуктивный вывод, т.е. вывод от следствия к причине. Правила абдуктивного вывода имеют следующий вид: из *A* следует *B*; *B* имеет место; следовательно, причиной *B* будет *A*. Поскольку причин явления *B* может быть много, заключение абдуктивного вывода представляется всего лишь гипотезой, а сам вывод – правдоподобным выводом. Поэтому абдуктивные выводы называют порождением гипотез
Абсолютная устойчивость	Absolute stability	Свойство нелинейного объекта сохранять асимптотическую устойчивость в целом для любых значений параметров нелинейной характеристики объекта из заданного класса нелинейных характеристик
Абстрагирование	Abstracting, abstraction	Процесс формирования образов реальности (представлений, понятий, суждений) посредством отвлечения и пополнения, т.е. путем использования (или усвоения),– лишь части из множества соответствующих данных и прибавления к этой части новой информации, не вытекающей из этих данных

в граф добавляется информация о местонахождении слов в корпусе (номер текста и номер слова) и число раз, которые эти слова и связи встретились.

Расстояние между двумя словами a и b на графе рассчитывается как сумма расстояния по синтаксическим ребрам и ссылочного расстояния. Для определения расстояний по ребрам за вес берется число, обратное количеству раз, которое ребро встретилось в корпусе, а ссылочное расстояние вычисляется по формуле; $1 / D(a,b) + S(a,b) + e(b)$, где $D(a,b)$ – кратчайшее расстояние между словами a и b в тексте, $S(a,b)$ – семантическая схожесть слов a и b, рассчитываемая как разность единицы и косинусного расстояния между векторными представлениями a и b, а $e(b)$ – энтропия Шеннона от слова b. Такой метод расчета расстояния по графу знаний учитывает частоты соупотребления слов, семантическую схожесть и количество информации, содержащейся в целевом слове. Графовое и ссылочное расстояния затем нормируются функцией сигмоиды, и их сумма используется как итоговое расстояние между словами. Итоговое расстояние нормировано в пределах от 0 до 1, и чем оно меньше, тем больше слова похожи друг на друга.

3.2. Характеристики используемых графов знаний. Для исследования взаимосвязей между характеристиками графов и качеством работы предлагаемых методов были подсчитаны следующие характеристики графов: средняя степень вершины; число связных компонент; количество изолированных вершин; коэффициент кластерности; кликовое, цикломатическое и хроматическое числа; степень ассортотивности; ранг матрицы смежности; остовное число. В табл. 4 приводится более подробное описание характеристик графов.

В теории графов коэффициент кластеризации – это мера степени, в которой узлы графа склонны объединяться в группы. Кликовое число графа – число вершин в наибольшей клике графа. Цикломатическое число графа указывает то наименьшее число ребер, которое нужно удалить из данного графа, чтобы получить дерево (для связного графа) или лес (для несвязного графа), т.е. добиться отсутствия у графа циклов. Хроматическое число графа – это минимальное число цветов, которыми можно правильно раскрасить вершины графа. Степень ассортативность – тенденция узлов графа соединяться с другими узлами того же типа.

Используемые в работе графы являются не эйлеровыми, не планарными, нерегулярными и не являются деревьями. Из всех графов только граф на основе классификатора будет хордальным, что следует из способа его построения.

Среди графов на базе тезауруса наибольшую среднюю степень вершины имеет граф на тезаурусе 2024 г., равную 8.852, наименьшую — граф на тезаурусе 1988 г., равную 7.007. Значение средней степени вершины для графа на классификаторе сильно отличается от графов на тезаурусе и равно 862.231. Граф на тезаурусе 1988 г. имеет одну связную компоненту, а граф на тезаурусе 2024 г. – четыре, однако этот граф также имеет три изолированные вершины, у совмещенного графа есть характеристики, как у графа на тезаурусе 2024 г. Исходя из построения графа на классификаторе, он имеет три компоненты связности. Количество вершин

Таблица 3. Представление классификатора

Факторы классификации	Уровень		Термины
	1	2	
Математический аппарат	Алгебра и теория чисел	Лычагин В.В.	–
		Теория чисел	Простое число, совершенное число, взаимно простые числа, алгебраические числа, p-адические числа, группа Галуа, дзета функция, конечные поля
		Линейная алгебра, теория матриц	Линейное пространство, линейная комбинация, векторное пространство, базис, линейный оператор, ранг матрицы, детерминант, обратная матрица
		Алгебраическая геометрия	Базис Гребнера, аффинные многообразия, проективные многообразия, алгебраические множества, бирациональная эквивалентность, спектр кольца, простой идеал
	Геометрия и топология	Лычагин В.В.	–
		Дифференциальная геометрия	Гладкое многообразие, диффеоморфизм, векторное поле, дифференциальная форма, внешний дифференциал, линейная связность, ковариантный дифференциал, кривизна связности, кручение связности
		Топология	Топологическое пространство, компакт, локально компактное пространство, фактор-пространство, нормальное пространство, хаусдорфово пространство, компактизация, связное пространство, непрерывное отображение, гомеоморфизм, гомотопия

в каждой компоненте равно 863, 1042, 417. Граф на классификаторе имеет самое большое значение коэффициента кластерности, равное один. Самое большое значение коэффициента среди графов на тезаурусах имеет граф на тезаурусе 1988 г., равное 0.495, самое маленькое значение у графа на тезаурусе 2024 г., равно 0.213.

Графы на тезаурусах имеют одинаковые значения для кликовых и хроматических чисел, равные пять и семь соответственно. Для графа на классификаторе обе метрики равны 1042. Среди графов на тезаурусах самое большое значение для цикломатического числа у совмещенного графа, будет 4233, а самое маленько – у графа на тезаурусе 1988 г., равное 687. Для графа на классификаторе эта метрика равна 998731. Среди всех графов самое большое значение для степени ассортотивности будет у графа на классификаторе, равное 1. Самое маленькое значение у графа на тезаурусе 1988 г. составляет –0.347. Матрица смежности графа на классификаторе имеет самый высокий ранг, равный 2322, в то время как самый низкий ранг, равный 173, имеет матрица смежности графа на тезаурусе 1988 г. Среди графов на тезаурусах самая большая длина пути (при расчете исключались изолированные вершины) у совмещенного графа (3.029), самая маленькая у графа на основе тезауруса 1988 г. (2.509). Исходя из построения, в графе на базе классификатора значение этой характеристики для каждой компоненты связности равно 1.

Средняя степень вершины в синтаксическом графе равняется 24.771, что больше, чем у всех графов на тезаурусах, но меньше, чем у графа классификатора. Синтаксический граф имеет шесть компонент связности и пять изолированных вершин. Коэффициент кластерности равен 0.209. Кликовое, хроматическое и цикломатическое числа равны 20, 30 и 88870 соответственно. Степень

Таблица 4. Характеристики графов

Характеристика	Год			Классификатор			Синтаксический граф
	2024	1988	1988+2024				
Средняя степень вершины	8.852	7.007	8.618	862.231			24.771
Число связных компонент	4	1	4	3			6
Количество изолированных вершин	3	0	3	0			5
Кластерность	0.213	0.495	0.262	1.0			0.209
Кликовое число	5	5	5	1042			20
Хроматическое число	7	7	7	1042			30
Цикломатическое число	3567	687	4233	998731			88870
Степень ассортотивности	–0.216	–0.347	–0.214	1.0			–0.127
Ранг матрицы смежности	742	173	878	2322			7004
Компонента	1	1	1	1	2	3	1
Размер компоненты	1037	274	1275	863	1042	417	7800
Средняя длина пути	2.991	2.509	3.029	1	1	1	2.864

ассортотивности равна –0.127 и является самой близкой к нулю. Ранг матрицы смежности синтаксического графа равен 7004. По своим характеристикам, если не учитывать степень ассортотивности и ранг матрицы смежности, данный граф находится где-то между графом классификатора и графами на тезаурусах. Средняя длина пути в синтаксическом графе равна 2.864.

4. Методы тематической сегментации текстов. Как уже было сказано, задача тематической сегментации заключается в разбиении документа на тематически связные сегменты. Для решения задачи тематической сегментации текста предлагаются следующие методы: на основе оценки всевозможных сегментаций, на базе функции тематической связности, на основе синтаксического графа и на базе накопительного графа. Рассмотрим каждый из них.

4.1. Методы на основе оценки всевозможных сегментаций. Задача глобальной тематической сегментации решается следующим образом.

Пусть $S_i = ((g_1,...,g_r),(g_{r+1},...,g_l),...,(g_m,...,g_T)) = (G_1,G_2,...,G_h)$ – некоторый кортеж из множества P. Приведем общий вид функции:

$$L(S_i) = \sum_{i=1}^{h-1} L'(G_i, G_{i+1}),$$

где $L'(G_i,G_{i+1})$ – некоторая функция, определенная на парах подряд идущих элементов кортежа S_i. Для функции L' был выбран следующий функциональный вид:

$$L'(G_i, G_{i+1}) = \sum_{n_p \in G_i} \sum_{n_q \in G_{i+1}} F(n_p, n_q),$$

где n_p, n_q – вершины, принадлежащие $G_i, G_{(i+1)}$ соответственно, F – функция взаимодействия между вершинами n_p, n_q. Другими словами, значение функционала L на S_i равно сумме значений некоторой функции L' на парах подряд идущих элементов кортежа. В свою очередь L' равна сумме попарных взаимодействий между вершинами, принадлежащими разным элементам пары. Функция F задает взаимодействие между двумя вершинами и является объектом исследования данной статьи.

Для графов на тезаурусах используется следующая функция взаимодействия:

$$F(n_p, n_q) = \begin{cases} G\dfrac{w_{n_p} w_{n_q}}{r^2}, r < TH, \\ -G\dfrac{w_{n_p} w_{n_q}}{r^\gamma}, r \geq TH, \end{cases}$$

где w_{n_p}, w_{n_q} – частоты терминов, соответствующих вершинам n_p, n_q; r – расстояние между вершинами n_p, n_q; G, γ, TH – гиперпараметры.

Для графов на онтологии применяется следующая функция взаимодействия:

$$F(n_p, n_q) = \begin{cases} G\ w_{n_p} w_{n_q}, Comp(n_p) = Comp(n_q), \\ -G\ w_{n_p} w_{n_q}, Comp(n_p) \neq Comp(n_q), \end{cases}$$

где $Comp(n)$ – функция, возвращающая индекс компоненты связности, в которой лежит вершина n.

Эвристика здесь следующая. Если вершины находятся достаточно близко друг к другу, то они должны стараться объединить два подграфа в один, в противном случае они должны сопротивляться этому объединению. Поскольку для графа на онтологии понятие расстояния теряет какой-либо смысл, в соответствующих формулах этот множитель просто убирается.

4.2. Методы на основе расчета весов промежутков. Рассматриваются несколько типов алгоритмов решения задачи локальной тематической сегментации: на базе функции тематической связности, на основе синтаксического графа и накопительного графа. За основу всех предлагаемых методов берется классический алгоритм TextTiling [5]. Напомним, что идея этого алгоритма заключается в том, что текст разбивается на блоки фиксированной длины, после чего с помощью сравнения тем или иным образом соседних блоков (или групп блоков) принимается решение о наличии/отсутствии тематической границы между ними. При этом в статье предлагаются различные способы того, как можно сравнивать соседние блоки. Представленные алгоритмы берут за основу идею сравнения соседних блоков. Непосредственно в публикации [5] блоки состоят из фиксированного количества слов, что может идти вразрез с синтаксической структурой текста (границы блоков проходят внутри предложений). Данные методы опираются на сравнение предложений как минимальных единиц документа. Для локальной оптимизации справедлива та же эвристика, что и для глобальной оптимизации с той лишь разницей, что в этом случае рассматривается только кортеж, получаемый непосредственно после отображения предложений в подграфы.

4.2.1. Метод на основе функции тематической связности. В качестве метрики схожести соседних графов была выбрана следующая функция, которую будем называть функцией тематической связности (topic coherence):

$$\Delta L^i = \frac{1}{2}(\Delta L(g_i, g_{i+1}) + \Delta L(g_{i+1}, g_i)),$$

$$\Delta L(g_m, g_c) = \sum_{v \in g_c} \Delta L(g_m, g_c, v),$$

$$\Delta L(g_m, g_c, v) = \begin{cases} w_v, v \in N(g_m), \\ -w_v, v \notin N(g_m), \end{cases}$$

где ΔL^i – оценка прироста тематической связности при объединении подграфов g_i, g_{i+1}; g_i, g_{i+1} – подграфы предложений d_i, d_{i+1} соответственно; g_m, g_c – основной и дополняющий подграфы; v – вершина, принадлежащая g_c; w_v – частота термина, соответствующего вершине v в соответствующем предложении.

Рассматривая два соседних графа, сначала один из них выбирается в качестве основного, а второй – в качестве дополняющего. Если очередная вершина лежит в единичной окрестности основного графа, то к метрике схожести добавляется частота этой вершины, так как считается, что в дополняющем графе говорится про то же, что и в основном. Если очередная вершина не лежит в единичной окрестности основного графа, то ее частота вычитается из итогового значения. После прохода по всем вершинам дополняющего графа, графы меняются ролями и аналогичный подсчет повторяется для другого графа. В качестве итогового значения метрики используется среднее значений для двух случаев.

4.2.2. Метод на основе синтаксического графа. Для расчета графовых представлений из предложения извлекаются слова, соответствующие вершинам заранее построенного синтаксического графа, затем по графу находятся расстояния между словами, входящими в графы соседних предложений. Расстояние между словом и подграфом A рассчитывается как минимальное из всех расстояний между этим словом и вершинами из подграфа A. Расстояние от подграфа A до подграфа B вычисляется как среднее всех расстояний между вершинами подграфа A и графом B. Вес промежутка между двумя предложениями рассчитывается как среднее значение расстояний от A до B и от B до A.

Когда расчет промежутков завершен, выполняется сегментация, пороговое значение вычисляется как сумма среднего значения всех промежутков и их среднеквадратичного отклонения. Все промежутки, значение которых больше, чем пороговое значение, считаются разрывами сегментов.

4.2.3. Метод на основе накопительного графа. В данном методе вместо расчета схожести между соседними предложениями, вычисляется схожесть предложения-кандидата с накопленным сегментом. Это позволяет лучше учитывать контекст и аккумулировать информацию о рассматриваемом сегменте.

В этом методе графовое представление предложения считается как структура связей между именованными группами в предложении. Графовые представления предложений рассчитываются с помощью алгоритма 1.

Алгоритм 1. Расчет графовых представлений предложения.

Вход: предложение.

Выход: группа графовое представление предложения.

Шаг 1. Выполним синтаксический разбор предложения с помощью синтаксического парсера.

Шаг 2. Извлечем из синтаксического дерева имена существительные.

Шаг 3. Между соответствующими вершинами, имеющими прямую связь в дереве разбора, добавим ребра.

Шаг 4. Если в предложении два имени связаны через глагол, также запишем эту связь в итоговый граф.

Шаг 5. Дополним граф словами, имеющими в данном предложении наибольший вес, согласно мере tf-idf.

Для решения задачи сегментации используется алгоритм 2.

Алгоритм 2. Алгоритм решения задачи тематической сегментации текста с помощью накопительного графа.

Вход: текст, разбитый на предложения.

Выход: группы из предложений.

Шаг 1. Рассчитаем для каждого предложения графовое представление с помощью алгоритма 1.

Шаг 2. Создадим граф сегмента, представляющий собой графовое представление первого предложения. Установим пороговое значение, равным 0.

Шаг 3. Берем первое доступное графовое представление предложения из списка предложений, пока они есть.

Шаг 3.1. При рассмотрении данного предложения находим схожесть между сегментом и предложением по формуле N_i/N_a, где N_i – количество вершин графового представления приведенного предложения, входящих в граф сегмента, а N_a – это количество всех вершин графового представления этого предложения.

Шаг 3.2. Если величина схожести больше порогового значения.

Шаг 3.2.1. Производим слияние сегмента и предложения.

Шаг 3.2.2. Записываем величину схожести в список истории.

Шаг 3.2.3. Записываем индекс предложения в список индексов предложений сегмента.

Шаг 3.2.4. Помечаем это предложение как недоступное.

Шаг 3.2.5. Рассчитываем новое пороговое значение как сумму среднего и среднеквадратичного отклонения истории.

Шаг 3.3. Если величина схожести меньше либо равна пороговому значению.

Шаг 3.3.1. Считаем величину схожести сегмента с последующим предложением.

Шаг 3.3.2. Если величина схожести больше порогового значения.

Шаг 3.3.2.1. Добавляем оба предложения в сегмент.

Шаг 3.3.2.2. Добавляем оба значения схожести в историю.

Шаг 3.3.2.3. Записываем индексы предложений в список индексов предложений сегмента.

Ш а г 3.3.2.4. Помечаем эти предложения как недоступные.

Ш а г 3.3.3. Если величина схожести меньше либо равна пороговому значению.

Ш а г 3.3.3.1. Добавляем накопленные индексы предложений сегмента в список сегментов.

Ш а г 3.3.3.2. Очищаем историю величин схожести и очищаем сегмент.

Ш а г 4. Переходим к шагу 3.

Ш а г 5. Возвращаем группы предложений в соответствии со списком сегментов.

Результатом работы этого алгоритма является сегментация s, представляющая собой последовательность 0 и 1 длиной $n - 1$, где n – это количество предложений в тексте. Единице соответствует разрыв сегмента, нулю — его отсутствие.

Также рассматривалось предположение, что учет длины приведенного сегмента может положительно сказаться на качестве сегментации. Для его проверки была проведена следующая модификация данного метода.

Вместо статистической меры по истории промежутков пороговое значение вычисляется по формуле $2 / (1 + e^{-l/c}) - 1$, где l – длина сегмента, а c – гиперпараметр, определяющий скорость роста порогового значения с длиной. Эвристика данной формулы заключается в том, что в естественном тексте тематические сегменты не могут совпадать со всем текстом (если он имеет большой объем предложений), а значит, существует некоторое граничное (неизвестное нам) значение для распределения истинных размеров сегментов. Таким образом, чем больше сегмент имеет длину, тем более вероятно, что он будет иметь длину выше пороговой, и тем менее вероятно, что к данному сегменту можно добавить еще предложения.

5. Эксперименты.

5.1. О п и с а н и е д а н н ы х. Данные представляют собой статьи сотрудников научной организации в формате pdf, из которых выбрали 6000 русскоязычных статей. Из них было отобрано 65 статей на 27 тем. Каждую статью вручную разделили на предложения. Затем из этих статей были собраны новые статьи следующим образом. Сначала все предложения, относящиеся к одной теме, объединили в группу. При генерации нового документа сначала случайным образом выбиралась тема, в которой есть свободные предложения, затем из этой группы извлекалось случайное количество предложений, которые добавлялись в генерируемый документ. Этот процесс продолжался до тех пор, пока генерируемый документ не стал содержать заданного объема предложений. Документы генерировались, пока были не пустые темы. Таким образом создали 63 статьи, состоящие минимум из 70 предложений. В табл. 5, 6 представлена статистика синтезированного датасета по темам и описание тем и их кодировок.

Следовательно, для создания тестового датасета было отобрано 65 статей. Оставшиеся 5935 статей из начального набора использовали для построения синтаксического графа.

В качестве базовых методов, с которыми проводилось сравнение предлагаемых алгоритмов, рассматривались алгоритмы локальной оптимизации, применяющие модели векторного представления текста. В качестве таких алгоритмов выбрали широко известный алгоритм частота терминов по обратной частоте документов (term frequency inverse document frequency (TF-IDF)) [15] и языковую модель BERTSciRus (science russian). С их помощью для каждого предложения рассчитывались векторные представления, которые затем использовались для вычисления косинусной близости соседних предложений.

5.2. П о к а з а т е л и к а ч е с т в а. Для оценки качества работы алгоритмов сегментации применяются два основных показателя – вероятностная ошибка (Pk) [16] и разность сегментов (window difference (WD)) [17].

Pk вычисляется с использованием скользящего окна. При движении окна по документу алгоритм определяет, принадлежат ли предложения на концах этого окна одному или разным сегментам в эталонной сегментации. Если эта принадлежность не соответствует предсказанной сегментации, тогда счетчик неправильных окон увеличивается на один. Итоговое количество неправильных окон нормируется на число всевозможных окон в документе. Размер скользящего окна обычно устанавливается равным половине средней длины эталонного сегмента.

Показатель WD работает схожим образом, однако он оценивает несоответствие в эталонной и предсказанной сегментациях в количестве границ в пределах окна. Другими словами, если в эталонной сегментации некоторое окно содержит N границ, а предсказанная сегментация – M и $M \neq N$, то счетчик числа неправильных окон увеличивается на один.

Хронологически показатель WD появился позже показателя Pk и был задуман как более лучшая характеристика, исправляющая ряд проблем второй. По этой причине в качестве ос-

Таблица 5. Статистика датасета по темам

Обозначение темы	**Общее число предложений**	**Общее число блоков**
упр_груп_бпла	375	56
управл_соц_обр	30	4
углеводор	71	13
нейросемант_ис	501	71
упр_соцсеть	338	44
учеба_vr	22	5
множдос_нелин	266	38
имит_мод	129	19
инфобез	57	8
интел_анализ	442	62
онколог_вакцин	490	71
адапт_резв_комп	98	16
времряд_активиы	155	21
эконом_росс	368	51
гис_соц_обр	394	55
упр_косм_роб	335	49
графика_01_т	48	7
интер_эргон	130	20
CRM	91	16
устойч_стабил_колеб	114	17
тепл_ламп	111	16
упр_спин_глобул	82	12
модел_газ	88	14
быстрореаг_произв	196	29
решен	24	4
геосинх_косм_апп	59	9
сист_полет	257	38

новного показателя используется WD, в то время как Pk – как вспомогательный инструмент, например для отладки алгоритмов.

5.3. А н а л и з р е з у л ь т а т о в. Полная таблица с результатами работы алгоритмов представлена в табл. 7. Исходя из анализа этой таблицы можно заключить, что в среднем методы для решения задачи глобальной тематической сегментации показывают одни из самых плохих результатов и в целом работают хуже, чем алгоритмы для решения задачи локальной сегментации. Лучше работает алгоритм, использующий функцию тематической связности. При этом такие алгоритмы работают лучше базовых моделей, если применяются графы знаний на тезаурусах, в противном случае алгоритм проигрывает базовой модели BERT. Из “глобальных” алгоритмов сильно выбивается алгоритм на основе классификатора, который превосходит аналогичные алгоритмы на тезаурусах, а также базовые модели.

Алгоритм, основанный на синтаксическом графе, показывает результаты, сравнимые с результатами работы алгоритмов на функции тематической связности и лучше, чем у базовой модели. И наконец, алгоритм, имеющий самые хорошие метрики качества, – алгоритм на основе накопительного графа.

Таблица 6. Расшифровка тем

Наименование темы	Тема
упр_груп_бпла	Управление группой БПЛА
управл_соц_обр	Управление в сфере социального образования
углеводор	Разведка и добыча углеводородов
нейросемант_ис	Нейросемантические системы
упр_соцсеть	Информационное управление в социальных сетях
учеба_vr	Учебные пособия на основе виртуальной реальности
множдос_нелин	Множества достижимости нелинейных систем
имит_мод	Имитационноемоделирование
инфобез	Информационнаябезопасность
интел_анализ	Интеллектуальныйанализ
онколог_вакцин	Вакцинотерапияонкологии
адапт_резв_комп	Адаптивное резервирование комплексов взаимосвязанных программных модулей
времряд_активы	Прогнозирование временных рядов (активов)
эконом_росс	Экономика России
гис_соц_обр	Геоинформационные системы для социально-образовательной сферы
упр_косм_роб	Управление космическим роботом
графика_01_т	Программный комплекс «графика-01-т»
интер_эргон	Эргономика интерфейсов
CRM	CRM-системы
устойч_стабил_колеб	Устойчивость и стабилизация колебаний
тепл_ламп	Тепловые лампы
упр_спин_глобул	Управление системой спинов
модел_газ	Моделирование сильно нестационарных потоков газа
быстрореаг_произв	Концепция «быстро реагирующего производства»
решен	Системы для оптимизации процесса принятия решений
геосинх_косм_апп	Система геосинхронных низкоорбитальных спутников
сист_полет	Системы управления полетом

Такую ситуацию можно объяснить следующими соображениями. Во-первых, все методы для решения задачи глобальной сегментации опираются только на тезаурусы и классификатор. То же можно сказать и про “локальные” алгоритмы, которые используют функцию тематической связности. Поскольку графы на тезаурусах и классификаторе были собраны вручную, они не отражают всевозможного лексического содержания тем. Другими словами, некоторые предложения отображаются либо в небольшие подграфы (малое число вершин), либо совсем не отображаются в подграфы (число вершин равно нулю). Более того, случается, что предложения, принадлежащие разным темам, могут отображаться или в близкие (минимум попарного расстояния между вершинами подграфов равен единице), или в пересекающиеся подграфы. Видно, что при увеличении числа вершин качество работы алгоритмов, использующих функцию тематической связности, улучшается.

Аномально хорошие результаты “глобального” алгоритма на классификаторе в сравнении с аналогичными алгоритмами на тезаурусах могут свидетельствовать о том, что механизм решения глобальной задачи сегментации хорошо сочленяется с графом классификатора.

Таблица 7. Результаты тестирования алгоритмов

Наименование метода	WD	Pk
Методы на основе перебора всевозможных разбиений		
Тезаурус-1988	0.790	0.432
Тезаурус-2024	0.738	0.612
Тезаурус-1988+2024	0.744	0.617
Классификатор	0.629	0.606
Методы на основе оценки весов промежутков		
Синтаксический граф	0.644	0.383
Базовая модель (BERT)	0.664	0.343
Базовая модель (tfidf)	0.998	0.318
Topic Coherence (Тезаурус-1988)	0.652	0.510
Topic Coherence (Тезаурус-2024)	0.626	0.486
Topic Coherence (Тезаурус-1988+2024)	0.623	0.476
Topic Coherence (классификатор)	0.660	0.528
Topic Coherence (синтаксический граф)	0.668	0.529
Методы на основе накопительного графа		
Без штрафа за длину	0.485	0.404
Со штрафом за длину	0.457	0.376

Методы, применяющие накопительный граф по своей задумке, не используют граф, построенный заранее, вместо этого собирая новые графы во время своего выполнения. Таким образом, концепции (соответствующие вершинам), которые алгоритм извлекает из текста, являются более релевантными самому тексту в сравнении с графами на тезаурусах и классификаторе. Если сравнивать результаты этого подхода с базовыми моделями, можно заметить, что подход значительно выигрывает в качестве. При этом рассматривалось два вида подобного подхода: без штрафования за длину сегмента и с штрафами. Алгоритм, использующий штрафы за длину сегмента, показывает результаты лучше, чем без штрафов, однако для его настройки применяется ряд гиперпараметров, что требует разделять выборку на несколько частей – валидационную и тестовую.

Наконец, алгоритм на основе синтаксического графа, использует заранее построенный в автоматическом режиме граф на базе большого числа релевантных тестовому датасету документов, закладывая через различные расстояния в этот граф информацию о синтаксических и семантических связях. Согласно результатам, такой подход работает чуть лучше по сравнению с моделью BERT. Алгоритм на основе функции тематической связности также показывает не самые хорошие результаты при использовании синтаксического графа. Это может свидетельствовать о том, что алгоритм построения синтаксического графа захватывает слишком много шумовой информации.

Анализируя табл. 5 с метриками графов и результатами, можно сделать следующие выводы. Сравнивая алгоритм локальной сегментации с функцией оценки Topic Coherence для графов на тезаурусах, видно, что увеличение вершин оказывает положительную динамику на результаты сегментации. Однако это справедливо только для графов на тезаурусах, поскольку граф классификатора и синтаксический граф имеют гораздо большее совокупное число вершин и при этом показывают худшие результаты. Помимо этого можно заметить, что данный метод демонстрирует самые лучшие результаты на совмещенном графе тезаурусов 1988 и 2024 гг. Видно при этом, что средняя степень вершины и кластерность не имеют явного влияния на качество сегментации, так как совокупный граф находится по этим метрикам между графами тезаурусов 1988 и 2024 гг. Алгоритм работает тем лучше, чем больше средняя длина пути. Помимо этого видно, что лучшая сегментация имеет самую близкую к нулю степень ассортотивности (среди графов на тезаурусах), что может говорить о том, что для лучшей сегментации не

должно быть тенденции смежности вершин с одинаковыми степенями. Что касается графа на классификаторе, то можно сделать вывод, что его структура в целом не подходит для алгоритма локальной сегментации с функцией тематической сязности. По построению синтаксический граф ближе по структуре к графам на тезаурусах, но в сравнении с ними алгоритм на основе синтаксического графа дает самые плохие результаты. Однако из сравнения характеристик этих графов трудно сделать какой-то вывод. Кластерность синтаксического графа меньше, чем у графов на тезаурусе, но совмещенный граф, дающий лучший результат, имеет промежуточное значение кластерности. У синтаксического графа существует самая большая степень ассортотивности в сравнении с графами на тезаурусах, но степень ассортотивности совмещенного графа ниже, чем у синтаксического графа. На данном этапе подобное поведение списывается на сильную зашумленность синтаксического графа (большое число неинформативных вершин и ребер).

Анализируя связь характеристик с результатами алгоритмов глобальной сегментации, можно увидеть различия в поведении алгоритмов на тезаурусах. Если в случае локальной сегментации лучше всего себя показал совмещенный граф, то в данном случае самые высокие метрики принадлежат алгоритму на основе тезауруса 2024 г. Можно заметить некоторую корреляцию между кластерностью графа и результатами работы, а именно чем ниже кластерность графа такой структуры, тем лучше работает алгоритм. Похожее замечание можно сделать относительно средней степени вершины,т.е. чем она больше, тем лучше результат.

Таблица 8. Статистика по темам метода локальной сегментации на основе функции тематической связности с графом на тезаурусе 1988 г.

topic	inner_mistakes_rate	outer_mistakes_rate
im_top_best		
учеба_vr	0	0
тепл_ламп	0	13.33
быстрореаг_произв	1.23	13.21
углеводор	1.59	12
времряд_активиы	3.1	17.07
im_top_worst		
имит_мод	21.1	13.89
управл_соц_обр	19.23	12.5
нейросемант_ис	18.84	12.98
адапт_резв_комп	18.29	7.14
графика_01_т	14.63	33.33
pr_top_best		
учеба_vr	0	0
геосинх_косм_апп	7.41	0
адапт_резв_комп	18.29	7.14
упр_соцсеть	3.74	7.32
сист_полет	11.42	10.29
pr_top_worst		
решен	10	42.86
графика_01_т	14.63	33.33
упр_косм_роб	13.43	27.47
интел_анализ	9.26	20.69
упр_спин_глобул	5.71	20

Таблица 9. Статистика по темам метода локальной сегментации на основе функции тематической связности с графом на тезаурусе 2024 г.

topic	inner_mistakes_rate	outer_mistakes_rate
im_top_best		
решен	0	57.14
углеводор	0	28
тепл_ламп	0	20
устойч_стабил_колеб	2.06	15.15
онколог_вакцин	2.63	20.63
im_top_worst		
адапт_резв_комп	23.17	28.57
интер_эргон	21.05	18.92
гис_соц_обр	17.4	28.16
имит_мод	16.51	36.11
времряд_активиы	13.95	26.83
pr_top_best		
геосинх_косм_апп	9.26	0
учеба_vr	5.88	0
сист_полет	4.57	10.29
модел_газ	6.76	11.54
управл_соц_обр	3.85	12.5
pr_top_worst		
решен	0	57.14
имит_мод	16.51	36.11
быстрореаг_произв	7.41	32.08
адапт_резв_комп	23.17	28.57
гис_соц_обр	17.4	28.16

В табл. 8–17 приводятся статистики работы алгоритмов по темам. Алгоритмы сегментации могут совершать два вида ошибок. Первый – когда алгоритм ставит границу между предложениями, относящимися к одной теме. Второй – когда алгоритм не разбивает предложения, относящиеся к разным темам. В табл. 8–17 предлагаемые алгоритмы ранжируются по этим ошибкам. В таблицах столбец inner mistakes rate означает количество ошибок первого типа, которые алгоритм допустил для той или иной темы, деленное на общее количество предложений данной темы; столбец outer mistakes rate означает долю тех случаев, когда алгоритм не отделил данную тему от какой-либо другой. Строка im_top_best/im_top_worst показывает начало блоков, содержащих топ пяти тем, на которых алгоритм отработал лучше всего/хуже всего в терминах ошибок первого вида. Строки pr_top_best/pr_top_worst имеют аналогичное значение, но для ошибок второго типа.

Из табл. 8–17 можно увидеть, что алгоритмы, использующие графы тезаурусов 2024 г. и совмещенный, имеют схожий профиль тем по качеству работы. В целом, такое поведение ожидаемо, поскольку граф тезауруса 2024 г. содержит в 10 раз больше вершин, чем граф 1988 г., т.е. при их совмещении основная информация берется из графа тезауруса 2024 г. При этом профили лучших тем по внутренним ошибкам немного отличаются для алгоритма решения задачи глобальной сегментации.

Можно заметить, что для “локального” и “глобального” алгоритмов на классификаторе схожи профили лучших тем по внутренним ошибкам, но отличаются профили лучших тем по внешним ошибкам. Часто можно заметить, что темы “решен” и “учеба_vr” появляются в профилях лучших

Таблица 10. Статистика по темам метода локальной сегментации на основе функции тематической связности с композитным графом на тезаурусах 1988 и 2024 гг.

topic	inner_mistakes_rate	outer_mistakes_rate
im_top_best		
решен	0	57.14
углеводор	0	28
тепл_ламп	0	20
устойч_стабил_колеб	2.06	15.15
управл_соц_обр	3.85	12.5
im_top_worst		
адапт_резв_комп	25.61	25
интер_эргон	21.05	21.62
имит_мод	16.51	36.11
гис_соц_обр	15.93	28.16
времряд_активиы	14.73	31.71
pr_top_best		
геосинх_косм_апп	9.26	0
учеба_vr	5.88	0
модел_газ	6.76	11.54
управл_соц_обр	3.85	12.5
упр_груп_бпла	7.84	12.5
pr_top_worst		
решен	0	57.14
имит_мод	16.51	36.11
быстрореаг_произв	7.41	33.96
времряд_активиы	14.73	31.71
упр_соцсеть	11.22	29.27

тем алгоритмов для обоих типов ошибок. Скорее всего, такое очень частое появление связано с тем, что популяции этих тем (количество предложений) в датасете самые малочисленные, а значит, на этих темах как минимум не получится в принципе сделать больше ошибок, чем для более многочисленных тем.

Из этих же таблиц видно, что “глобальные” алгоритмы хорошо работают на теме «графика_01_т». “Локальные” алгоритмы сегментации на основе тезаурусов в совокупности (объединение по множеству «хороших» темминус множество «плохих» тем обоих типов) хорошо работают на темах 'онколог_вакцин', 'тепл_ламп', 'углеводор', 'устойч_стабил_колеб', 'учеба_vr'по ошибкам первого типа и на темах 'геосинх_косм_апп', 'модел_газ', 'сист_полет', 'упр_груп_бпла', 'учеба_vr' по ошибкам второго типа. «Глобальные» алгоритмы сегментации на тезаурусах хорошо работают с темами 'быстрореаг_произв', 'управл_соц_обр', 'учеба_vr'по ошибкам первого типа и с темами 'быстрореаг_произв', 'интел_анализ', 'управл_соц_обр', 'эконом_росс' по ошибкам второго типа. Методы на базе классификатора хорошо работают на темах 'CRM', 'инфобез', 'устойч_стабил_колеб', 'учеба_vr', 'эконом_росс' по ошибкам первого типа и 'CRM', 'времряд_активиы', 'интел_анализ', 'инфобез', 'множдос_нелин', 'учеба_vr' по ошибкам второго типа.

Хуже всего алгоритмы на тезаурусах работают с темами 'адапт_резв_комп', 'времряд_активиы', 'геосинх_косм_апп', 'гис_соц_обр', 'графика_01_т', 'имит_мод', 'интер_эргон', 'множдос_нелин', 'нейросемант_ис', 'решен', 'тепл_ламп', 'углеводор', 'управл_соц_обр', 'устойч_стабил_колеб' по ошибкам первого типа и 'CRM', 'адапт_резв_комп', 'быстрореаг_произв',

Таблица 11. Статистика по темам метода локальной сегментации на основе функции тематической связности с графом на классификаторе

topic	inner_mistakes_rate	outer_mistakes_rate
im_top_best		
инфобез	0	20
учеба_vr	0	11.11
CRM	1.33	10
упр_спин_глобул	1.43	5
эконом_росс	3.44	10.64
im_top_worst		
упр_косм_роб	19.43	20.88
управл_соц_обр	19.23	12.5
геосинх_косм_апп	16.67	17.65
имит_мод	16.51	16.67
решен	15	14.29
pr_top_best		
времряд_активиы	6.2	4.88
упр_спин_глобул	1.43	5
интел_анализ	13.76	9.48
CRM	1.33	10
множдос_нелин	8.33	10
pr_top_worst		
тепл_ламп	7.37	30
модел_газ	4.05	26.92
быстрореаг_произв	8.02	26.42
графика_01_т	12.2	25
упр_соцсеть	11.22	21.95

'времряд_активиы', 'гис_соц_обр', 'графика_01_т', 'имит_мод', 'интел_анализ', 'интер_эргон', 'инфобез', 'множдос_нелин', 'модел_газ', 'нейросемант_ис', 'решен', 'упр_косм_роб', 'упр_соцсеть', 'упр_спин_глобул', 'устойч_стабил_колеб' по ошибкам второго типа. Алгоритмы на основе графа классификатора хуже всего работают на темах 'геосинх_косм_апп', 'имит_мод', 'решен', 'упр_косм_роб', 'упр_соцсеть', 'управл_соц_обр' по ошибкам первого типа и 'адапт_резв_комп', 'быстрореаг_произв', 'гис_соц_обр', 'графика_01_т', 'модел_газ', 'тепл_ламп', 'упр_соцсеть', 'упр_спин_глобул' по ошибкам второго типа. Эти результаты напрямую свидетельствуют о соответствии используемых графов темам из собранного датасета и будут применены в дальнейшем при рассмотрении/построении новых графов знаний.

Заключение. Исследована модель задачи тематической сегментации текстов в терминах графов знаний. Предложена формализация этой задачи и методы ее решения. Изучена связь характеристик графов знаний с результатами работы алгоритмов. Проведен анализ тем, на которых хорошо работают предлагаемые методы тематической сегментации. Методы представляют собой связку из графа знаний и алгоритма поиска оптимальной сегментации. Для сравнения методов использовались две базовые модели – на основе алгоритма tf-idf и BERT. Алгоритмы на основе накопительных графов, решающие задачу сегментации с локальной оптимизацией, показывают наилучшие результаты, как в сравнении с базовыми моделями, так и с остальными предлагаемыми алгоритмами.

Таблица 12. Статистика по темам метода глобальной сегментации на основе тезауруса 1988 г.

topic	inner_mistakes_rate	outer_mistakes_rate
im_top_best		
быстрореаг_произв	11.73	20.75
времряд_активиы	13.18	31.71
инфобез	14.29	20
тепл_ламп	16.84	26.67
учеба_vr	17.65	22.22
im_top_worst		
углеводор	38.1	32
графика_01_т	31.71	16.67
решен	30	14.29
нейросемант_ис	29.07	34.35
имит_мод	27.52	30.56
pr_top_best		
управл_соц_обр	23.08	12.5
решен	30	14.29
графика_01_т	31.71	16.67
эконом_росс	17.81	19.15
инфобез	14.29	20
pr_top_worst		
устойч_стабил_колеб	24.74	42.42
CRM	18.67	36.67
нейросемант_ис	29.07	34.35
упр_соцсеть	22.11	32.93
интер_эргон	20.18	32.43

Алгоритмы для решения “глобальной” сегментации на базе тезаурусов показали наоборот самые плохие результаты, за исключением алгоритма на основе классификатора. Такое поведение представляет определенный интерес для дальнейшего исследования, поскольку может быть ключом к отысканию как подходящего вида критерия оптимизации, так и структуры графа знаний, для решения “глобальной” задачи сегментации.

Несмотря на то, что алгоритмы, использующие синтаксический граф, показали не самые хорошие результаты, это направление также представляет огромный интерес. Если решить задачу с очисткой графа от шумовой информации, потенциально можно получить хорошие результаты, так как данный граф может содержать в себе гораздо больше информации, чем во всех остальных графах вместе взятых.

Алгоритмы на основе функции тематической связности показали средние результаты. Однако лучше всего они работают с графами на тезаурусах. При этом можно видеть, что при увеличении числа вершин в этих графах и средней длины пути, а также при приближении к нулю степени ассортотивности качество работы алгоритмов растет. Граф на основе классификатора лучше всего работает при решении задачи глобальной сегментации. Графы на тезаурусах лучше работают с функцией тематической связности для решения задачи локальной сегментации. Метод, строящий граф “на лету” работает лучше, чем алгоритмы, использующие вручную построенные графы. Главным итогом данной статьи является вывод о том, что возможно применять графы знаний для решения задачи тематической сегментации.

Таблица 13. Статистика по темам метода глобальной сегментации на основе тезауруса 2024 г.

topic	inner_mistakes_rate	outer_mistakes_rate
im_top_best		
решен	0	28.57
интер_эргон	1.75	5.41
имит_мод	1.83	8.33
быстрореаг_произв	1.85	1.89
упр_соцсеть	2.72	7.32
im_top_worst		
тепл_ламп	12.63	0
времряд_активиы	12.4	14.63
устойч_стабил_колеб	9.28	3.03
геосинх_косм_апп	9.26	5.88
множдос_нелин	9.21	14.29
pr_top_best		
графика_01_т	4.88	0
тепл_ламп	12.63	0
быстрореаг_произв	1.85	1.89
устойч_стабил_колеб	9.28	3.03
интел_анализ	4.23	4.31
pr_top_worst		
решен	0	28.57
модел_газ	6.76	15.38
времряд_активиы	12.4	14.63
множдос_нелин	9.21	14.29
инфобез	4.08	13.33

Дальнейшее направление исследований видится в следующем. Во-первых, необходимо понять, почему «глобальный» метод сегментации с помощью графа классификатора показывает настолько хорошие результаты в сравнении с аналогичными методами на других графах. Во-вторых, как уже было сказано, большой интерес представляет синтаксический граф. Планируется доработать алгоритм его генерации. Например, можно, выделив заранее с помощью другого алгоритма термины из данных, ограничить множество вершин множеством выделенных терминов. В-третьих, необходимо исследовать другие возможные функции взаимодействия между вершинами.

Таблица 14. Статистика по темам метода глобальной сегментации на основе тезаурусов 1988 и 2024 гг.

topic	inner_mistakes_rate	outer_mistakes_rate
im_top_best		
решен	0	28.57
управл_соц_обр	0	12.5
углеводор	1.59	8
имит_мод	1.83	8.33
инфобез	2.04	6.67
im_top_worst		
тепл_ламп	12.63	0
адапт_резв_комп	10.98	17.86
времряд_активиы	10.08	14.63
графика_01_т	9.76	0
геосинх_косм_апп	9.26	5.88
pr_top_best		
графика_01_т	9.76	0
тепл_ламп	12.63	0
быстрореаг_произв	6.79	1.89
интел_анализ	3.44	4.31
геосинх_косм_апп	9.26	5.88
pr_top_worst		
решен	0	28.57
адапт_резв_комп	10.98	17.86
модел_газ	4.05	15.38
упр_спин_глобул	4.29	15
времряд_активиы	10.08	14.63

Таблица 15. Статистика по темам метода глобальной сегментации на основе классификатора

topic	inner_mistakes_rate	outer_mistakes_rate
im_top_best		
инфобез	0	0
учеба_vr	0	0
упр_спин_глобул	0	5
устойч_стабил_колеб	1.03	0
тепл_ламп	1.05	0
im_top_worst		
геосинх_косм_апп	10	0
имит_мод	9.52	0
упр_косм_роб	8.04	0
управл_соц_обр	7.69	0
упр_соцсеть	6.8	0
pr_top_best		
модел_газ	5.41	0
инфобез	0	0
управл_соц_обр	7.69	0
учеба_vr	0	0
геосинх_косм_апп	10	0
pr_top_worst		
упр_спин_глобул	0	5
адапт_резв_комп	1.22	3.57
быстрореаг_произв	6.59	1.89
гис_соц_обр	6.19	0.97
модел_газ	5.41	0

Таблица 16. Статистика по темам метода локальной сегментации на основе синтаксического графа

topic	inner_mistakes_rate	outer_mistakes_rate
im_top_best		
решен	0	0
модел_газ	2.7	15.38
упр_спин_глобул	2.86	20
адапт_резв_комп	4.88	10.71
учеба_vr	5.88	22.22
im_top_worst		
тепл_ламп	29.47	33.33
углеводор	29.31	24
инфобез	24.49	0
геосинх_косм_апп	20	17.65
эконом_росс	18.61	20.21
pr_top_best		
решен	0	0
управл_соц_обр	15.38	0
графика_01_т	12.2	0
инфобез	24.49	0
имит_мод	6.36	8.33
pr_top_worst		
CRM	12	36.67
тепл_ламп	29.47	33.33
углеводор	29.31	24
учеба_vr	5.88	22.22
онколог_вакцин	17.18	20.63

Таблица 17. Статистика по темам метода локальной сегментации на основе накопительного графа

topic	inner_mistakes_rate	outer_mistakes_rate
im_top_best		
учеба_vr	5.88	55.56
решен	10	42.86
имит_мод	11.82	52.78
упр_косм_роб	14.69	46.15
управл_соц_обр	15.38	62.5
im_top_worst		
углеводор	31.03	48
эконом_росс	30.28	48.94
времряд_активиы	27.61	43.9
нейросемант_ис	27.44	48.85
упр_спин_глобул	27.14	45
pr_top_best		
устойч_стабил_колеб	22.68	24.24
инфобез	22.45	40
интер_эргон	20.91	40.54
графика_01_т	17.07	41.67
решен	10	42.86
pr_top_worst		
геосинх_косм_апп	24	64.71
управл_соц_обр	15.38	62.5
быстрореаг_произв	19.76	58.49
учеба_vr	5.88	55.56
онколог_вакцин	16.23	55.56

СПИСОК ЛИТЕРАТУРЫ

1. *Chen H., Luo X.* An Automatic Literature Knowledge Graph and Reasoning Network Modeling Framework Based on Ontology and Natural Language Processing // Advanced Engineering Informatics. 2019. V. 42. https://doi.org/: 10.1016/j.aei.2019.100959
2. *Dahab M., Hassan H.* TextOntoEx: Automatic Ontology Construction from Natural English Text // Expert Systems with Applications. 2008. V. 34(2). P. 1474–1480. https://doi.org/10.1016/j.eswa.2007.01.043
3. *Oren E., Anthony F., Christensen J., Soderland S.* Mausam. Open Information Extraction: The Second Generation // Intern. Joint Conf. on Artificial Intelligence.Barcelona, 2011. https://doi.org/:10.5591/978-1-57735-516-8/IJCAI11-012
4. *Ristoski P., Gentile A.L., Alba A., Gruhl D., Welch S.* Large-scale Relation Extraction from Web Documents and Knowledge Graphs with Human-in-the-loop // J. Web Semantics. 2019. V. 60. https://doi.org/: 100546. doi:10.1016/j.websem.2019.100546
5. *Hearst A.M.* TextTiling: Segmenting Text IntoMulti-paragraph Subtopic Passages // Computational Linguistics. 1997. V. 23(1). P. 33–64.
6. *Galley M., McKeown K., Fosler-Lussier E.* Discourse Segmentation of Multi-Party Conversation // Proc. 41st Annual Meeting on Association for Computational Linguistics (ACL '03). 2003. V. 3. P. 562–569. https://doi.org/:10.3115/1075096.1075167
7. *Misra H., Yvon F., Jose J.M.* Text Segmentation via Topic Modeling: An Analytical Study //Proc. 18th ACM Conf. on Information and Knowledge Management (CIKM '09). Hong Kong, 2009. V. 1. P. 1553–1556. https://doi.org/:10.1145/1645953.1646170
8. *Du L., Buntine W., Jin H.* A Segmented Topic Model Based on the Two-parameter Poisson-Dirichlet Process // Machine Language. 2010. V. 81(2). P. 5–19. https://doi.org/:10.1007/s10994-010-5197-4
9. *Das A., Das P.* Incorporating Domain Knowledge To Improve Topic Segmentation Of Long MOOC Lecture Videos // arXiv:2012.07589 [cs.CL]. https://doi.org/10.48550/arXiv.2012.07589
10. *Nouar F., Belhadef H.* A Deep Neural Network Model with Multihop Self-attention Mechanism for Topic Segmentation of Texts // Innovative Systems for Intelligent Health Informatics. 2021. V. 72. P. 407–417. https://doi.org/:10.1007/978-3-030-70713-2_38
11. *Lo K., Jin Y., Tan W., Liu M., Du L., Buntine W.L.* Transformer over Pre-trained Transformer for Neural Text Segmentation with Enhanced Topic Coherence // Findings of the Association for Computational Linguistics: EMNLP 2021. 2021. V. 1. P. 3334–3340. https://doi.org/:10.18653/v1/2021.findings-emnlp.283
12. *Arnold S., Schneider R., Cudr'e-Mauroux P., Gers F.A.* SECTOR: A Neural Model for Coherent Topic Segmentation and Classification // Transactions of the Association for Computational Linguistics. 2019. V. 7. P. 169–184. https://doi.org/:10.1162/tacl_a_00261
13. Теория управления. Терминология / Под ред. М. М. Гальперина. М.: Наука, 1988. 56 с.
14. Теория управления: словарь системы основных понятий / Под общ. ред. *Д. А. Новикова.* М.: ЛЕНАНД, 2024. 128 с.
15. *Jones K.S.* A Statistical Interpretation of Term Specificity and Its Application in Retrieval // Journal of Documentation. 2004. V. 60(5). P. 493—502. https://doi.org/:10.1108/EB026526
16. *Beeferman D., Berger A. L., Lafferty J.D.* Statistical Models for Text Segmentation // Machine Learning. 1998. V. 34. P. 177–210. https://doi.org/:10.1108/EB026526
17. *Pevzner L., Hearst M.A.* A Critique and Improvement of an Evaluation Metric for Text Segmentation // Computational Linguistics. 2002. V. 28. P. 19–36. https://doi.org/:10.1162/089120102317341756

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, С. 65–76

ДИСКРЕТНЫЕ СИСТЕМЫ

УДК 519.711

РЕАЛИЗАЦИЯ СИСТЕМЫ НЕ ПОЛНОСТЬЮ ОПРЕДЕЛЕННЫХ БУЛЕВЫХ ФУНКЦИЙ СХЕМОЙ ИЗ ДВУХВХОДОВЫХ ЭЛЕМЕНТОВ С ПОМОЩЬЮ АЛГЕБРАИЧЕСКОЙ ДЕКОМПОЗИЦИИ

© 2024 г. Ю.В. Поттосин[a], *

[a]*Объединенный ин-т проблем информатики НАН Беларуси, Минск, Беларусь*

**e-mail: pott@newman.bas-net.by*

Поступила в редакцию 28.09.2023 г.
После доработки 19.12.2023 г.
Принята к публикации 04.02.2024 г.

Задача алгебраической декомпозиции булевой функции (в англоязычной литературе – bi-decomposition) заключается в представлении заданной булевой функции с помощью логической операции над двумя булевыми функциями. Предлагается для реализации систем не полностью определенных (частичных) булевых функций в базисе двухвходовых логических элементов использовать метод, основанный на алгебраической декомпозиции булевых функций. В качестве базиса могут быть следующие базисы элементов: ИЛИ-НЕ, И-НЕ или И, ИЛИ при доступной инверсии входных сигналов. Применяемый метод алгебраической декомпозиции сводится к поиску двухблочного взвешенного покрытия полными двудольными подграфами (бикликами) взвешенного двудольного графа, представляющего собой различия между строками булевых матриц, которые задают рассматриваемую систему функций. Исходная система частичных булевых функций задается двумя булевыми матрицами, одна из которых служит областью булева пространства аргументов, где значения функций определены, а другая – значениями функций на элементах указанной области. Каждой биклике из получаемого покрытия приписывается в качестве веса некоторое множество переменных, являющихся аргументами функций заданной системы. Каждая из этих биклик определяет булеву функцию, аргументы которой – приписанные к биклике переменные. Полученные таким образом функции составляют разложение исходной функции. Процесс синтеза комбинационной схемы заключается в последовательном применении алгебраической декомпозиции к этим функциям. Описан способ получения двухблочного взвешенного покрытия бикликами упомянутого двудольного графа.

Ключевые слова: синтез комбинационных схем, булевы функции, разложение булевых функций, Булевы матрицы, полный двудольный граф, биклика, двухблочное покрытие

DOI: 10.31857/S0002338824040044 **EDN:** UEIHST

IMPLEMENTATION OF A SYSTEM OF INCOMPLETELY SPECIFIED BOOLEAN FUNCTIONS IN A CIRCUIT OF TWO-INPUT GATES BY MEANS OF BI-DECOMPOSITION

Yu. V. Pottosin[a], *

[a]*United Institute of Informatics Problems, National Academy of Sciences of Belarus, Minsk, Republic of Belarus*

**e-mail: pott@newman.bas-net.by*

The problem of bi-decomposition of a Boolean function is to represent a given Boolean function as a logic algebra operation over two Boolean functions. A meth-od based on bi-decomposition of Boolean functions is suggested to implement sys-tems of incompletely specified (partial) Boolean functions in the basis of two-input gates. This basis can be the basis of NOR gates, NAND gates or the basis of AND and OR gates with accessible input complements. The used method for bi-decomposition is reduced to the search for two-block weighted cover of a complete bipartite weighted graph with complete bipartite subgraphs (bi-cliques). The graph represents differences between the rows of Boolean matrices that specify the given system of partial

Boolean functions. The system is given by two Boolean matrices, one of which represents the domain of Boolean space where the values of the given functions are specified, and the other the values of the functions on the elements of the domain. Every bi-clique in the obtained cover is assigned in a certain way with a set of variables that are the arguments of the function. This set is the weight of the bi-clique. Every of those bi-cliques defines a Boolean function whose arguments are the variables assigned to it. The functions obtained in such a way constitute the re-quired decomposition. The process of synthesis of a combinational circuit consists in successive application of bi-decomposition to the obtained functions. The meth-od for two-block covering the orthogonality graph of rows of ternary matrices is de-scribed.

Keywords: synthesis of combinational circuits, Boolean function, decomposition of Boolean functions, Boolean matrix, complete bipartite graph, bi-clique, two-block cover

Введение. Под декомпозицией системы булевых функций понимается ее представление в виде суперпозиции двух или более систем функций, каждая из которых в некотором смысле проще исходной системы. Задача декомпозиции булевых функций является одной из важных и сложных задач из области логического проектирования, успешное решение которой непосредственно влияет на качество и стоимость проектируемых цифровых устройств. Решение этой задачи дает возможность заменить сложную задачу аппаратной реализации булевой функции от большого числа переменных на более простую задачу реализации нескольких функций с гораздо меньшим числом аргументов.

Существует довольно много различных видов декомпозиции булевой функции [1]. Одним из таких видов **в**ыступает алгебраическая декомпозиция. Задача алгебраической декомпозиции (в англоязычной литературе – bi-decomposition) ставится следующим образом. Для заданной булевой функции $y = f(\mathbf{x})$, где компонентами вектора $\mathbf{x} = (x1, x2, ..., x_n)$ являются булевы переменные, составляющие множество X, требуется найти суперпозицию $f(\mathbf{x}) = \varphi(g_1(\mathbf{z}_1), g_2(\mathbf{z}_2))$, где компоненты векторов $\mathbf{z}_1$ и $\mathbf{z}_2$ – переменные из множеств $Z_1 \subseteq X$ и $Z_2 \subseteq X$ соответственно. Вид функции φ от двух переменных также задан. Это может быть любая из 10 булевых функций, существенно зависящих от обеих переменных и представляемых операциями алгебры логики. Обычно множества Z_1 и Z_2 заданы и $Z_1 \cap Z_2 = \varnothing$. Такая декомпозиция называется *разделительной* в отличие от *неразделительной* декомпозиции, где условие $Z_1 \cap Z_2 = \varnothing$ необязательно, но при этом на мощности множеств Z_1 и Z_2 могут быть наложены ограничения.

Существуют разнообразные методы решения как разделительной, так и неразделительной алгебраической декомпозиции при заданных множествах Z_1 и Z_2 [2–7]. Вопрос определения множеств Z_1 и Z_2, для которых существует нетривиальная алгебраическая декомпозиция булевой функции, имеет особый интерес. Нетривиальной считается декомпозиция, если числа аргументов функций g_1 и g_2 меньше числа аргументов функции f. Среди публикаций, где рассматривается задача подходящей пары множеств Z_1, Z_2 для получения нетривиальной декомпозиции, можно назвать работы [1, 5, 8–11]. В [12] решается задача алгебраической декомпозиции, где множества Z_1 и Z_2 определяются в процессе решения задачи. Метод, описанный в [12], использует подход к решению задачи параллельной декомпозиции системы частичных булевых функций, предложенный в [13]. Метод, усовершенствованный для случая, когда функция φ представляется операцией сложения по модулю 2, приведен в [14]. Методы из [12, 14] минимизируют мощности множеств Z_1 и Z_2, применяя полный перебор возможных ситуаций в процессе решения, что значительно ограничивает их практическое использование. В [15] описан эвристический метод алгебраической декомпозиции частичных булевых функций, который не гарантирует абсолютного минимума этих мощностей, но позволяет решить задачу за более короткое время.

Вероятность существования какой-либо нетривиальной декомпозиции для полностью определенных булевых функций весьма низка, но по-другому дело обстоит, когда рассматриваемые функции являются не полностью определенными (частичными), особенно когда они заданы только на небольшой части булева пространства аргументов. Поэтому в литературе основное внимание уделялось декомпозиции (в том числе алгебраической) частичных булевых функций. Если функция φ относится к классу нелинейных функций, то функции g_1 и g_2 оказываются проще функции f в том смысле, что степень зависимости их от некоторых переменных может быть меньше, чем у функции f. Данный параметр рассматривался в [16]. Под степенью зависимости функции f от переменной x_i здесь понимается число пар значений $(\mathbf{x}', \mathbf{x}'')$ вектора $\mathbf{x}$ с различными значениями i-й компоненты, для которых $f(\mathbf{x}') \neq f(\mathbf{x}'')$. Кроме того, если какая-то из функций g_i, $i = 1, 2$, оказалась с тем же числом аргументов, что и полностью определенная функция f, то эта функция g_i в любом случае будет не полностью определенной функцией, что увеличивает вероятность ее разложимости.

Известны примеры применения методов алгебраической декомпозиции для повышения быстродействия схем [17, 18] и при синтезе схем на базе программируемой вентильной матрицы (FPGA) [19]. Далее предлагается метод синтеза комбинационных схем в базисе двухвходовых элементов, реализующих нелинейные функции. Имеются в виду базисы И-НЕ, ИЛИ-НЕ, а также базис элементов И, ИЛИ при доступных инверсиях переменных. Метод основан на последовательном применении алгебраической декомпозиции к получаемым функциям, в которой используется подход, описанный в [15]. Построение функций φ, g_1 и g_2, представляющих искомую декомпозицию, сводится в данном подходе к поиску в некотором двудольном графе взвешенного покрытия его двудольными полными подграфами (бикликами). Применение в задачах логического проектирования аппарата, связанного с (бикликами), описано в [20].

1. Предлагаемый подход. Предполагается, что система не полностью определенных (частичных) булевых функций $\mathbf{f}(\mathbf{x})$, где $\mathbf{f}$ и $\mathbf{x}$ – векторы $(f_1, f_2, ..., f_m)$ и $(x_1, x_2, ..., x_n)$, задана двумя матрицами $\mathbf{X}$ и $\mathbf{Y}$. Строками матрицы $\mathbf{X}$ служат элементы булева пространства аргументов x_1, x_2, ..., x_n, а соответствующими строками матрицы $\mathbf{Y}$ – векторы, представляющие наборы значений функций на соответствующих элементах булева пространства. Для каждой функции f_i заданной системы можно из матрицы $\mathbf{X}$ выделить булевы матрицы $\mathbf{M}^1(f_i)$ и $\mathbf{M}^0(f_i)$, представляющие области булева пространства, где функция f_i имеет соответственно значения 1 и 0. Далее эту же символику будем использовать для задания любых других функций.

Рассмотрим полный двудольный граф $G(f_i) = (V^1, V^0, E)$ с взвешенными ребрами, где вершины из множества V^1 соответствуют строкам матрицы $\mathbf{M}^1(f_i)$, вершины из множества V^0 – строкам матрицы $\mathbf{M}^0(f_i)$. Каждому ребру v^1v^0 ($v^1 \in V^1$, $v^0 \in V^0$) графа $G(f_i)$ в качестве веса приписана элементарная дизъюнкция $x_i \vee x_j \vee ... \vee x_k$ аргументов заданной системы функций, если компоненты строк матриц $\mathbf{M}^1(f_i)$ и $\mathbf{M}^0(f_i)$, связанных с ребром v^1v, в столбцах $x_i, x_j, ..., x_k$ имеют различные значения (0 и 1).

Полному двудольному подграфу, или *биклике*, графа $G(f_i)$ припишем конъюнктивную нормальную форму (КНФ) с элементарными дизъюнкциями, приписанными ребрам, которые принадлежат данной биклике. После удаления возможных поглощаемых элементарных дизъюнкций преобразуем полученную КНФ, раскрыв скобки, в дизъюнктивную нормальную форму (ДНФ). Переменные, составляющие элементарную конъюнкцию минимального ранга в полученной ДНФ, припишем соответствующей биклике.

Пусть требуется выразить некоторую заданную частичную функцию $f(\mathbf{x})$ как $f(\mathbf{x}) = \varphi(g_1(\mathbf{z}_1), g_2(\mathbf{z}_2))$, где φ – булева функция от двух переменных g_1 и g_2, которые являются функциями соответственно от векторных переменных $\mathbf{z}_1$ и $\mathbf{z}_2$, представляющих собой части вектора x. Символ «=» обозначает как отношение реализации, так и равенство, которое можно рассматривать в качестве частного случая отношения реализации. Функция φ, частичная или полностью определенная, *реализует* частичную функцию f, если значения функции φ совпадают со значениями функции f везде, где они определены [21].

Функции g_1 и g_2 построим следующим образом. В графе $G(f)$ выделим две биклики $B_1 = (V_1^1, V_1^0, E_1)$ и $B_2 = (V_2^1, V_2^0, E_2)$ так, чтобы любое ребро графа G присутствовало хотя бы в одном из множеств E_1 или E_2, те биклики B_1 и B_2 должны покрывать своими ребрами все множество E ребер графа $G(f)$. Биклики B_1 и B_2 достаточно задать парами множеств (V_1^1, V_1^0) и (V_2^1, V_2^0), так как в биклике каждая вершина из одной доли связана ребрами со всеми вершинами другой доли.

Аргументами функции g_i, $i = 1, 2$, являются переменные, приписанные биклике B_i. Строки матрицы $\mathbf{M}^1(g_i)$, представляющие собой значения векторной переменной $\mathbf{z}_i$, где функция g_i имеет значение 1, составляют части строк из матрицы $\mathbf{M}^1(f)$ или из $\mathbf{M}^0(f)$ (в зависимости от заданного вида функции φ), соответствующих вершинам из множества V_i^1. Части этих векторов определяются переменными, приписанными биклике B_i, т.е. эти переменные являются компонентами вектора $\mathbf{z}_i$. Аналогично формируется матрица $\mathbf{M}^0(g_i)$ из частей векторов, связанных с вершинами из множества V_i^0. Таким образом, каждой строке из $\mathbf{M}^1(f)$ или из $\mathbf{M}^0(f)$ соответствует пара значений функций g_1 и g_2. Если эта пара связана со строкой из матрицы $\mathbf{M}^1(f)$, то она составляет строку матрицы $\mathbf{M}^1(\varphi)$. Если она соответствует строке из матрицы $\mathbf{M}^0(f)$, то она составляет строку матрицы $\mathbf{M}^0(\varphi)$. Так будет задана функция φ. Заметим, что пары (V_1^1, V_1^0) и (V_2^1, V_2^0) следует считать упорядоченными, поскольку они связаны со значениями функцийg_1 и g_2.

Описываемый метод предполагает дальнейшее подобное разложение функций g_1 и g_2 и последующих получаемых функций до функций от двух переменных из множества $X = \{x_1, x_2, ..., x_n\}$.

2. Получение покрытия графа *G(f)* двумя бикликами. В таблице показано, какие значения должны иметь функции g_1 и g_2 при определенных значениях функции φ и при разных видах этой функции. Из нее видно, что $V_1^1 = V_2^1 = V^1$ должно быть для операции И, $V_1^0 = V_2^0 = V^0$ – для операции ИЛИ, $V_1^1 = V_2^1 = V^0$ – для операции И-НЕ и $V_1^0 = V_2^0 = V^1$ – для операции ИЛИ-НЕ.

Таблица 1. Соотношения значений функций g_1, g_2 и φ

И	ИЛИ	И-НЕ	ИЛИ-НЕ
φ g_1 g_2	φ g_1 g_2	φ g_1 g_2	φ g_1 g_2
1 1 1	0 0 0	0 1 1	1 0 0
0 – 0	1 – 1	1 – 0	0 – 1
0 0 –	1 1 –	1 0 –	0 1 –

Таким образом, одна из долей биклики всегда определена видом функции φ, как одна из долей полного двудольного графа G, и она присутствует в обеих бикликах. Другие доли биклик B_1 и B_2 образуются как блоки разбиения другой доли графа G. Например, если $V_1^0 = V_2^0 = V^1$, то $B_1 = (V_1^1, V^1)$ и $B_2 = (V_2^1, V^1)$, где $V_1^1 \cup V_2^1 = V^0$, и $V_1^1 \cap V_2^1 = \varnothing$.

Исходной информацией для получения искомого покрытия выступает множество *звездных графов*, которые являются подграфами графа $G(f)$. Звездным графом, или *звездой* называется полный двудольный граф $K_{1, n}$ [22]. Одноэлементная доля его представляет собой *центр* звезды. В нашем случае упомянутое множество – это множество всех биклик, у которых одной долей является одноэлементное множество с вершиной $v \in V^0$, а другой – множество V^1 или у которых одна доля – одноэлементное множество с вершиной $v \in V^1$, а другая – множество V^0 двудольного графа G. Назовем их *звездными бикликами*.

Как было сказано выше, каждой биклике приписывается КНФ, которая преобразуется в ДНФ. Из ДНФ выберем элементарную конъюнкцию K минимального ранга и вместо ДНФ и КНФ припишем соответствующей звездной биклике B_i множество переменных X_i из конъюнкции K. Выберем две звездных биклики B_i и B_j, у которых пересечение $X_i \cap X_j$ имеет минимальную мощность среди всех пар рассматриваемых звездных биклик. Если таких вариантов несколько, то отдаем предпочтение множествам X_i и X_j максимальной мощности. Естественно, желателен вариант $X_i \cap X_j = \varnothing$. Примем пару (B_i, B_j) за начальное значение пары биклик, которая должна покрывать граф $G(f)$, и обозначим ее (B_1, B_2). Конъюнкций минимального ранга в ДНФ может быть несколько, и есть возможность выбора варианта, лучшим образом удовлетворяющего указанным условиям.

Дальнейший процесс представляет собой последовательное расширение тех долей биклик B_1 и B_2, которые в начальных значениях были одноэлементными, за счет вершин, являющихся центрами рассматриваемых звездных биклик. Соответственно меняются множества X_1 и X_2. Пусть, например, $B_1 = (V_1^1, V_1^0)$, $B_2 = (V_2^1, V_2^0)$ и $V_1^1 \cup V_2^1 = V^0$, а множество V' состоит из вершин графа $G(f)$, которые не принадлежат ни V^0, ни одному из V_1^0 и V_2^0. Выбираются вершина $v_k \in V'$, являющаяся центром некоторой звездной биклики B_k, и множество V_i^0, $i \in 1, 2$, такие, что мощность множества $X_i \cup X_k$ отличается от мощности множества X_i или X_k на минимальную величину. Множество V_i^0 меняется на $V_i^0 \cup \{v_k\}$, а вершина v_k удаляется из V'. Процесс заканчивается, когда множество V' окажется пустым. Пара (B_1, B_2) представит искомое покрытие.

Далее рассмотрим на примерах построение комбинационных схем в базисах двухвходовых элементов ИЛИ-НЕ и И, ИЛИ.

3. Синтез комбинационных схем в базисе ИЛИ-НЕ. Пусть требуется построить логическую сеть из двухвходовых элементов ИЛИ-НЕ, реализующую систему булевых функций, представ-

ленную следующими матрицами (на любом наборе значений аргументов, не совпадающем ни с какой строкой матрицы **X**, значения функций не определены):

$$\mathbf{X} = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & \mathbf{0} & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array}, \quad \mathbf{Y} = \begin{array}{c} \begin{matrix} f_1 & f_2 & f_3 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array}.$$

Каждую из функций f_1, f_2 и f_3 зададим следующими матрицами с сохранением нумерации строк матриц **X** и **Y**:

$$\mathbf{M}^1(f_1) = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}, \quad \mathbf{M}^1(f_2) = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 6 \\ 7 \\ 9 \end{array},$$

$$\mathbf{M}^1(f_3) = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 3 \\ 4 \\ 6 \\ 7 \\ 9 \\ 10 \end{array},$$

$$\mathbf{M}^0(f_1) = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 4 \\ 5 \\ 6 \\ 8 \end{array}, \quad \mathbf{M}^0(f_2) = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 5 \\ 8 \\ 10 \end{array},$$

$$\mathbf{M}^0(f_3) = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 5 \\ 8 \end{array}.$$

Двудольные графы $G(f_i) = (V^1, V^0, E)$, $i = 1, 2, 3$, представим матрицами $\mathbf{G}(f_i)$, подобными матрице смежности. Строки матрицы $\mathbf{G}(f_i)$ соответствуют вершинам из множества V^1 (строкам матрицы $\mathbf{M}^1(f_i)$), а столбцы – вершинам из множества V^0 (строкам матрицы $\mathbf{M}^0(f_i)$). На пересечении строки k и столбца l матрицы $\mathbf{G}(f_i)$ (в качестве обозначений строк и столбцов и вершин графа используются номера соответствующих строк исходных матриц $\mathbf{X}$ и $\mathbf{Y}$) находится элементарная дизъюнкция или одиночная переменная, приписанные ребру между вершинами k и l:

$$\mathbf{G}(f_1) = \begin{array}{c} \begin{matrix} 4 & 5 & 6 & 8 \end{matrix} \\ \left[\begin{array}{cccc} x_2 \vee x_3 & x_1 \vee x_2 \vee x_3 \vee x_4 & x_1 \vee x_2 \vee x_3 & x_1 \vee x_2 \vee x_3 \vee x_4 \vee x_5 \\ x_3 & x_1 \vee x_3 \vee x_4 & x_1 \vee x_3 & x_1 \vee x_3 \vee x_4 \vee x_5 \\ x_4 & x_1 & x_1 \vee x_4 & x_1 \vee x_5 \\ x_1 \vee x_5 & x_4 \vee x_5 & x_5 & x_4 \\ x_1 \vee x_3 & x_3 \vee x_4 & x_3 & x_3 \vee x_4 \vee x_5 \\ x_1 \vee x_3 \vee x_4 \vee x_5 & x_3 \vee x_5 & x_3 \vee x_4 \vee x_5 & x_3 \end{array}\right] \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \\ 7 \\ 9 \\ 10 \end{array},$$

$$\mathbf{G}(f_2) = \begin{array}{c} \begin{matrix} 5 & 8 & 10 \end{matrix} \\ \left[\begin{array}{ccc} x_1 \vee x_2 \vee x_3 \vee x_4 & x_1 \vee x_2 \vee x_3 \vee x_4 \vee x_5 & x_1 \vee x_2 \vee x_4 \vee x_5 \\ x_1 \vee x_3 \vee x_4 & x_1 \vee x_3 \vee x_4 \vee x_5 & x_1 \vee x_4 \vee x_5 \\ x_1 & x_1 \vee x_5 & x_1 \vee x_3 \vee x_5 \\ x_1 \vee x_4 & x_1 \vee x_4 \vee x_5 & x_1 \vee x_3 \vee x_4 \vee x_5 \\ x_4 & x_4 \vee x_5 & x_3 \vee x_4 \vee x_5 \\ x_4 \vee x_5 & x_4 & x_3 \vee x_4 \\ x_3 \vee x_4 & x_3 \vee x_4 \vee x_5 & x_4 \vee x_5 \end{array}\right] \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 6 \\ 7 \\ 9 \end{array},$$

$$\mathbf{G}(f_3) = \begin{array}{c} \begin{matrix} 1 & 2 & 5 & 8 \end{matrix} \\ \left[\begin{array}{cccc} x_2 \vee x_3 \vee x_4 & x_3 \vee x_4 & x_1 & x_1 \vee x_5 \\ x_2 \vee x_3 & x_3 & x_1 \vee x_4 & x_1 \vee x_4 \vee x_5 \\ x_1 \vee x_2 \vee x_3 & x_1 \vee x_3 & x_4 & x_4 \vee x_5 \\ x_1 \vee x_2 \vee x_3 \vee x_5 & x_1 \vee x_3 \vee x_5 & x_4 \vee x_5 & x_4 \\ x_1 \vee x_2 & x_1 & x_3 \vee x_4 & x_3 \vee x_4 \vee x_5 \\ x_1 \vee x_2 \vee x_4 \vee x_5 & x_1 \vee x_4 \vee x_5 & x_3 \vee x_5 & x_3 \end{array}\right] \end{array} \begin{array}{c} 3 \\ 4 \\ 6 \\ 7 \\ 9 \\ 10 \end{array}.$$

Получим реализацию функции f_1. Биклики $B_1 = (V_1^1, V_1^0)$ и $B_2 = (V_2^1, V_2^0)$, покрывающие граф $G(f_1)$, должны иметь одну общую долю: согласно приведенной выше таблице, для выбранного базиса ИЛИ-НЕ имеем $V_1^0 = V_2^0 = V^1$. Звездные биклики с приписанными КНФ (в квадратных скобках) имеют следующий вид:

({4}, {1,2,3,7,9,10}) $[x_3\, x_4\, (x_1 \vee x_5)]$; ({5}, {1,2,3,7,9,10}) $[x_1\, (x_3 \vee x_4)\, (x_4 \vee x_5)\, (x_3 \vee x_5)]$;
({6}, {1,2,3,7,9,10}) $[x_3\, x_5\, (x_1 \vee x_4)]$; ({8}, {1,2,3,7,9,10}) $[x_3\, x_4\, (x_1 \vee x_5)]$.

За начальные значения биклик B_1 и B_2 примем ({4}, {1,2,3,7,9,10}) $[x_3\, x_4\, (x_1 \vee x_5)]$ и ({6}, {1,2,3,7,9,10}) $[x_3\, x_5\, (x_1 \vee x_4)]$. Результатом выполнения следующего шага может быть один из следующих вариантов:

({4,5}, {1,2,3,7,9,10}) $[x_1\, x_3\, x_4]$; ({4,8}, {1,2,3,7,9,10}) $[x_3\, x_4\, (x_1 \vee x_5)]$;
({5,6}, {1,2,3,7,9,10}) $[x_1\, x_3\, x_5]$; ({6,8}, {1,2,3,7,9,10}) $[x_3\, x_4\, x_5]$.

Во всех вариантах, кроме одного, приписанные КНФ совпадают с ДНФ. Следует выбрать вариант ({4,8}, {1,2,3,7,9,10}) $[x_3\, x_4\, (x_1 \vee x_5)]$, поскольку приписанная этой биклике ДНФ состоит из двух элементарных конъюнкций одного ранга 3 и имеется в дальнейшем больше возможностей оптимизировать решение. Таким образом, текущим значением пары биклик (B_1, B_2) является

({4,8}, {1,2,3,7,9,10}) $[x_3\, x_4\, (x_1 \vee x_5)]$, ({6}, {1,2,3,7,9,10}) $[x_3\, x_5\, (x_1 \vee x_4)]$.

Выбор варианта на следующем шаге привел к следующему взвешенному покрытию графа $G(f_1)$:

$B_1 = (\{4,8\}, \{1,2,3,7,9,10\})\ [x_3\, x_4\, (x_1 \vee x_5)]$, $B_2 = (\{5,6\}, \{1,2,3,7,9,10\})\ [x_1\, x_3\, x_5]$.

Таким образом, функция f_1 разлагается на две функции g_1 и g_2, связанные операцией ИЛИ-НЕ, или «стрелка Пирса»: $f_1 = g_1 \uparrow g_2$. Функция g_2, соответствующая биклике B_2, зависит от переменных x_1, x_3, x_5 и ее можно задать матрицами, которые получаются из матриц $\mathbf{M}^1(f_1)$ и $\mathbf{M}^0(f_1)$ и после удаления строки 6, совпадающей со строкой 5, имеют следующий вид:

$$\mathbf{M}^1(g_2) = \begin{matrix} x_1 & x_3 & x_5 \\ \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} & & 5 \end{matrix}, \quad \mathbf{M}^0(g_2) = \begin{matrix} x_1 & x_3 & x_5 \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} & \begin{matrix} 1 \\ 3 \\ 7 \\ 9 \\ 10 \end{matrix} \end{matrix}.$$

Аргументами функции g_1 могут быть x_1, x_3, x_4 или x_3, x_4, x_5, и ее могут задавать матрицы

$$\mathbf{M}^1(g_1) = \begin{matrix} x_1 & x_3 & x_4 \\ \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} & \begin{matrix} 4 \\ 8 \end{matrix} \end{matrix}, \quad \mathbf{M}^0(g_1) = \begin{matrix} x_1 & x_3 & x_4 \\ \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} & \begin{matrix} 1 \\ 3 \\ 7 \\ 9 \\ 10 \end{matrix} \end{matrix}$$

или

$$\mathbf{M}^1(g_1) = \begin{matrix} x_3 & x_4 & x_5 \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \begin{matrix} 4 \\ 8 \end{matrix} \end{matrix}, \quad \mathbf{M}^0(g_1) = \begin{matrix} x_3 & x_4 & x_5 \\ \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} & \begin{matrix} 1 \\ 3 \\ 7 \\ 10 \end{matrix} \end{matrix}.$$

Выберем второй вариант, где функция g_1 определена на меньшей части булева пространства и имеется больше возможностей лучшего доопределения.

Функциям g_1 и g_2 соответствуют графы, которые задаются следующими матрицами:

$$\mathbf{G}(g_1) = \begin{matrix} 1 & 3 & 7 & 10 \\ \begin{bmatrix} x_3 & x_4 & x_5 & x_3 \vee x_4 \vee x_5 \\ x_3 \vee x_4 \vee x_5 & x_5 & x_4 & x_3 \end{bmatrix} & \begin{matrix} 4 \\ 8 \end{matrix} \end{matrix}, \quad \mathbf{G}(g_2) = \begin{matrix} 1 & 3 & 7 & 9 \\ \begin{bmatrix} x_1 \vee x_3 & x_1 & x_5 & x_3 \end{bmatrix} & 5 \end{matrix}.$$

Граф $G(g_1)$ покрывают биклики $(\{1,10\}, \{4,8\})\ [x_3]$ и $(\{3,7\}, \{4,8\})\ [x_4\, x_5]$, а граф $G(g_2)$ – $(\{1,3,9\}, \{5\})\ [x_1\, x_3]$ и (7), (5) $[x_5]$. Эти биклики определяют разложения $g_1 = x_3 \uparrow g_3(x_4, x_5)$ и $g_2 = \overline{g_4}(x_1, x_3) \uparrow x_5$, где функции $g_3 = (x_4 \uparrow \overline{x_5}) \uparrow (\overline{x_4} \uparrow x_5)$ и $g_4 = \overline{x_1} \uparrow x_3$ представлены матрицами, получаемыми из матриц $\mathbf{M}^1(g_1)$ и $\mathbf{M}^0(g_1)$:

$$\mathbf{M}^1(g_3) = \begin{matrix} x_4 & x_5 \\ \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} & \begin{matrix} 3 \\ 7 \end{matrix} \end{matrix}, \quad \mathbf{M}^0(g_3) = \begin{matrix} x_4 & x_5 \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{matrix} 4 \\ 8 \end{matrix} \end{matrix} \text{ и } \mathbf{M}^1(g_4) = \begin{matrix} x_1 & x_3 \\ \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} & \begin{matrix} 1 \\ 3 \\ 9 \end{matrix} \end{matrix}, \quad \mathbf{M}^0(g_4) = \begin{matrix} x_1 & x_3 \\ \begin{bmatrix} 1 & 0 \end{bmatrix} & 5 \end{matrix}.$$

Функция f_1 реализована сетью, представляемой следующей системой уравнений:

$f_1 = g_1 \uparrow g^2$,

$g_1 = x_3 \uparrow g_3$, $g_2 = \overline{g_4} \uparrow x_5$,

$g_3 = g_5 \uparrow g_6$, $g_4 = \overline{x}_1 \uparrow x_3$,
$g_5 = x_4 \uparrow x_5$, $g_6 = \overline{x}_4 \uparrow x_5$.

В графе $G(f_2)$ имеются звездные биклики ({5}, {1,2,3,4,6,7,9}) $[x_1\ x_4]$, ({8}, {1,2,3,4,6,7,9}) $[x_4\ (x_1 \vee x_5)]$ и ({10}, {1,2,3,4,6,7,9}) $[(x_1 \vee x_3 \vee x_5)\ (x_3 \vee x_4)\ (x_4 \vee x_5)]$. Отсюда видно, что функция f_2 может быть реализована функцией от двух переменных – x_1 и x_4, которая задается матрицами

$$\mathbf{M}^1(f_2) = \begin{matrix} x_1 & x_4 \\ \end{matrix}\begin{bmatrix} 1 & 0 \end{bmatrix}\ 5 \text{ и } \mathbf{M}^0(f_2) = \begin{matrix} x_1 & x_4 \\ \end{matrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}\begin{matrix} 1 \\ 3 \\ 6 \end{matrix},$$

т. е. $f_2 = \overline{x}_1 \uparrow x_4$.

Граф $G(f_3)$ содержит следующие звездные биклики:

({1}, {3,4,6,7,9,10}) $[(x_1 \vee x_2)\ (x_2 \vee x_3)]$, ({2}, {3,4,6,7,9,10}) $[x_1\ x_3]$,
({5}, {3,4,6,7,9,10}) $[x_1\ x_4\ (x_3 \vee x_5)]$, ({8}, {3,4,6,7,9,10}) $[x_3\ x_4\ (x_1 \vee x_5)]$.

По ним получим покрытие с биклики ({1}, {3,4,6,7,9,10}) $[(x_1 \vee x_2)\ (x_2 \vee x_3)]$ и ({2,5,8}, {3,4,6,7,9,10}) $[x_1\ x_3\ x_4]$, которые определяют разложение $f_3 = \overline{x}_2 \uparrow g_7(x_1, x_3, x_4)$, где функция g_7 задается матрицами

$$\mathbf{M}^1(g_7) = \begin{matrix} x_1 & x_3 & x_4 \\ \end{matrix}\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}\begin{matrix} 2, \\ 5 \end{matrix}\ \mathbf{M}^0(g_7) = \begin{matrix} x_1 & x_3 & x_4 \\ \end{matrix}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}\begin{matrix} 3 \\ 4 \\ 6 \\ 9 \\ 10 \end{matrix}.$$

Граф $G(g_7)$ представляется следующей матрицей:

$$\mathbf{G}(g_7) = \begin{matrix} 3 & 4 & 6 & 9 & 10 \\ \end{matrix}\begin{bmatrix} x_3 \vee x_4 & x_3 & x_1 \vee x_3 & x_1 & x_1 \vee x_4 \\ x_1 & x_1 \vee x_4 & x_4 & x_3 \vee x_4 & x_3 \end{bmatrix}\begin{matrix} 2. \\ 5 \end{matrix}$$

Искомое покрытие графа $G(g_7)$ составляют биклики ({3,6,9}, {2,5}) $[x_1\ x_4]$ и ({4,10}, {2,5}) $[x_3\ (x_1 \vee x_4)]$, и функция g_7 определяется как $g_7 = g_8(x_1, x_4) \uparrow g_9(x_1, x_3)$, где функции g_8 и g_9 задаются следующими матрицами:

$$\mathbf{M}^1(g_8) = \begin{matrix} x_1 & x_4 \\ \end{matrix}\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}\begin{matrix} 3. \\ 6 \end{matrix}\ \mathbf{M}^0(g_8) = \begin{matrix} x_1 & x_4 \\ \end{matrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{matrix} 2 \\ 5 \end{matrix} \text{ и } \mathbf{M}^1(g_9) = \begin{matrix} x_1 & x_3 \\ \end{matrix}\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}\begin{matrix} 4, \\ 10 \end{matrix}\ \mathbf{M}^0(g_9) = \begin{matrix} x_1 & x_3 \\ \end{matrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{matrix} 2. \\ 5 \end{matrix}$$

Вся система не полностью определенных булевых функций реализуется структурой, описываемой следующей системой уравнений:

$f_1 = g_1 \uparrow g_2$, $f_2 = g_{11}$, $f_3 = \overline{x}_2 \uparrow g_7$,
$g_1 = x_3 \uparrow g_3$, $g_2 = x_5 \uparrow \overline{g}_4$, $g_7 = g_8 \uparrow g_9$,
$g_3 = \overline{g}_5 \uparrow g_6$, $g_8 = g_{10} \uparrow \overline{g}_{11}$, $g_9 = \overline{g}_{12} \uparrow g_4$,
$g_4 = \overline{x}_1 \uparrow x_3$, $g_5 = x_4 \uparrow \overline{x}_5$, $g_6 = \overline{x}_4 \uparrow x_5$, $g_{10} = x_1 \uparrow \overline{x}_4$, $g_{11} = \overline{x}_1 \uparrow x_4$, $g_{12} = x_1 \uparrow \overline{x}_3$.

Комбинационная схема с элементами ИЛИ-НЕ и инверторами, реализующая заданную систему не полностью определенных булевых функций, изображена на рис. 1. Инверторы использованы для упрощения схемы. В качестве них могут быть взяты те же элементы ИЛИ-НЕ, у которых на входы подается один и тот же сигнал.

4. Синтез комбинационных схем в базисе И, ИЛИ. Пусть теперь требуется построить логическую сеть из элементов И и ИЛИ с доступными инверсиями входных сигналов, реализующую систему булевых функций из рассмотренного примера. Согласно приведенной выше таблице в бикликах $B_1 = (V_1^1, V_1^0)$ и $B_2 = (V_2^1, V_2^0)$, покрывающих граф $G(f_i)$, имеем $V_1^1 = V_2^1 = V^1$ для операции И и $V_1^0 = V_2^0 = V^0$ для операции ИЛИ. Для очередного разложения $h = \varphi(g_k, g_l)$, где h – любая из функций, исходная или получаемая в процессе разложения, следует выбрать элемент И или ИЛИ, соответствующий функции φ. Предлагается взять тот элемент, который дает меньшую степень определенности функций g_k и g_l. Если функции g_k и g_l оказываются полностью определёнными, то по матрице $\mathbf{G}(h)$ можно заметить, что для лучшего варианта разложения $h = \varphi(g_k, g_l)$ желательно выбрать для функции ф операцию И, если матрица $\mathbf{M}^0(h)$ имеет больше строк, чем матрица $\mathbf{M}^1(h)$, и, наоборот, операцию ИЛИ, если $\mathbf{M}^1(h)$ имеет строк больше, чем $\mathbf{M}^0(h)$. Кроме того, разложение считается лучшим, если при прочих равных условиях получаемые функции имеют меньшее суммарное число аргументов.

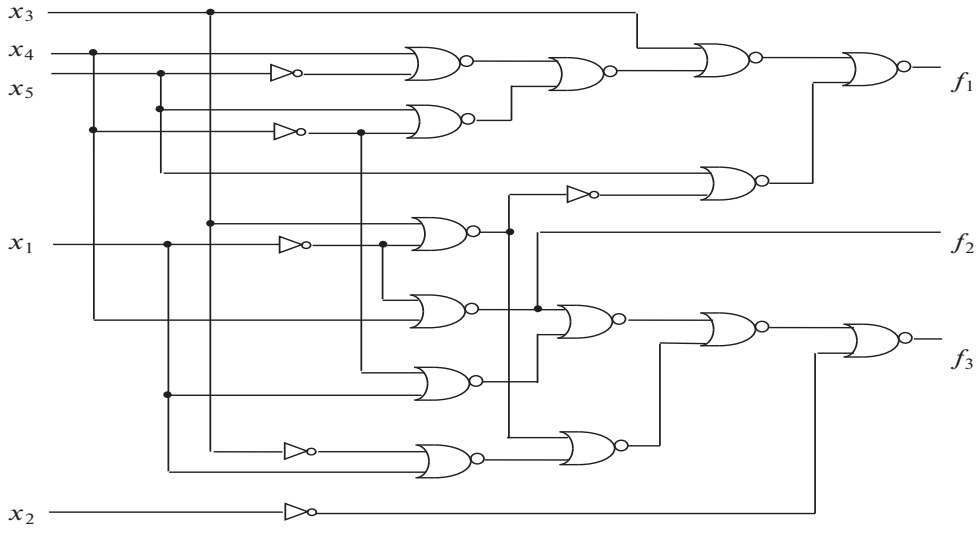

Рис. 1. Схема из элементов ИЛИ-НЕ

Для функции f_1 и операции ИЛИ имеем следующие звёздные биклики, полученные по матрице $\mathbf{G}(f_1)$:

$(\{1\}, \{4,5,6,8\})\ [x_2 \vee x_3]$, $(\{2\}, \{4,5,6,8\})\ [x_3]$, $(\{3\}, \{4,5,6,8\})\ [x_1\, x_4]$,
$(\{7\}, \{4,5,6,8\})\ [x_4\, x_5]$, $(\{9\}, \{4,5,6,8\})\ [x_3]$, $(\{10\}, \{4,5,6,8\})\ [x_3]$.

Описанным выше способом получаем $B_1 = (\{1,2,9,10\}, \{4,5,6,8\},)\ [x_3]$ и $B_2 = (\{3,7\}, \{4,5,6,8\},)\ [x_1\, x_4\, x_5]$, что определяет разложение $f_1 = x_3 \vee g_1(x_1, x_4, x_5)$.

Если взять операцию И для разложения функции f_1, то звездные биклики будут иметь следующий вид:

$(\{1,2,3,7,9,10\}, \{4\})\ [x_3\, x_4\, (x_1 \vee x_5)]$, $(\{1,2,3,7,9,10\}, \{5\})\ [x_1\, (x_3 \vee x_4)\, (x_3 \vee x_5)\, (x_4 \vee x_5)]$,
$(\{1,2,3,7,9,10\}, \{6\})\ [x_3\, x_5\, (x_1 \vee x_4)]$, $(\{1,2,3,7,9,10\}, \{8\})\ [x_3\, x_4\, (x_1 \vee x_5)]$.

В этом случае граф $G(f1)$ покрывают биклики $(\{1,2,3,7,9,10\}, \{4,8\})\ [x_3\ x_4\ (x_1 \vee x_5)]$ и $(\{1,2,3,7,9,10\}, \{5,6\})\ [x_1\, x_3\, x_5]$. Выбираем элемент ИЛИ, так как в разложении $f_1 = g_1\, g_2$ обе функции g_1 и g_2 зависят от трех переменных, тогда как в случае элемента ИЛИ числа аргументов получаемых функций – 1 и 3. Таким образом, имеем разложение $f_1 = x_3 \vee g_1(x_1, x_4, x_5)$, где функция g_1 задается матрицами

$$\mathbf{M}^1(g_1) = \begin{matrix} x_1 & x_4 & x_5 \end{matrix} \begin{matrix} \\ \end{matrix}$$
$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{matrix} 3 \\ 7 \end{matrix} \text{ и } \mathbf{M}^0(g_1) = \begin{matrix} x_1 & x_4 & x_5 \end{matrix}$$
$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{matrix} 4 \\ 5. \\ 6 \\ 8 \end{matrix}$$

Граф $G(g_1)$ представляется матрицей

$$\mathbf{G}(g_1) = \begin{matrix} 4 & 5 & 6 & 8 \\ \left[\begin{matrix} x_4 & x_1 & x_1 \vee x_4 & x_1 \vee x_5 \\ x_1 \vee x_5 & x_4 \vee x_5 & x_5 & x_4 \end{matrix}\right] & \begin{matrix} 3. \\ 7 \end{matrix} \end{matrix}$$

По числу аргументов функций разложения следует выбрать операцию ИЛИ. Непосредственно по матрице $\mathbf{G}(g_1)$ находится покрытие, которое составляют биклики ({3}, {4,5,6,8}) $[x_1\ x_4]$ и ({7}, {4,5,6,8}) $[x_4\ x_5]$, определяющие разложение $g_1 = g_2(x_1, x_4) \vee g_3(x_4, x_5)$. Матрица $\mathrm{M}^1(g_2)$ является минором матрицы $\mathrm{M}^1(g_1)$, образованным строкой 3 и столбцами x_1 и x_4, а матрица матриц $\mathrm{M}^0(g_2)$ – минором матрицы $\mathrm{M}^0(g_2)$, образованным строками 4 – 6, 8 и столбцами x_1 и x_4. По этим матрицам функция g_2 находится как $g_2 = \bar{x}_1\ \bar{x}_4$. Аналогично определяется функция g_3 по минорам тех же матриц: $g_3 = x_4\, x_5$. Таким образом, получено полное разложение функции f_1 на функции не более чем от двух аргументов.

Звездными бикликами графа $G(f_2)$ при операции ИЛИ являются:

({1}, {5,8,10}) $[(x_1 \vee x_2 \vee x_3 \vee x_4)\ (x1 \vee x2 \vee x4 \vee x5)]$,

({2}, {5,8,10}) $[(x1 \vee x3 \vee x4)\ (x1 \vee x4 \vee x5)]$,

({3}, {5,8,10}) $[x1]$,

({4}, {5,8,10}) $[x1 \vee x4]$,

({6}, {5,8,10}) $[x4]$,

({7}, {5,8,10}) $[x4]$,

({9}, {5,8,10}) $[(x3 \vee x4)\ (x4 \vee x5)]$.

Из них легко получается покрытие графа $G(f_2)$ бикликами ({1,2,3,4}, {5,8,10}) $[x_1]$ и ({6,7,9}, {5,8,10}) $[x_4]$, которое определяет реализацию функции f_2: $f_2 = \bar{x}_1 \vee x_4$.

По степени определенности функций g_4 и g_5, на которые разлагается функция f_3, выбираем операцию И, для которой покрытие графа $G(f_3)$ составят биклики ({3,4,6,7,9,10}), {1,2}) $[x_1\ x_3]$ и ({3,4,6,7,9,10}, {5,8}) $[x_1\ x_3\ x_4]$, определяющие разложение $f_3 = g_4(x_1, x_3)\ g_5(x_1\ x_3\ x_4)$. Матрицы, получаемые из матриц $\mathrm{M}^1(f_3)$ и $\mathrm{M}^0(f_3)$ и задающие функции g_4 и g_5, имеют следующий вид:

$$\mathbf{M}^1(g_4) = \begin{matrix} x_1 & x_3 \\ \left[\begin{matrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{matrix}\right] & \begin{matrix} 3 \\ 6 \\ 9 \end{matrix} \end{matrix}. \quad \mathbf{M}^0(g_4) = \begin{matrix} x_1 & x_3 \\ \left[\begin{matrix} 0 & 1 \end{matrix}\right] & 1 \end{matrix}, \quad \mathbf{M}^1(g_5) = \begin{matrix} x_1 & x_3 & x_4 \\ \left[\begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{matrix}\right] & \begin{matrix} 3 \\ 4 \\ 6 \\ 9 \\ 10 \end{matrix} \end{matrix}, \quad \mathbf{M}^0(g_5) = \begin{matrix} x_1 & x_3 & x_4 \\ \left[\begin{matrix} 1 & 0 & 0 \end{matrix}\right] & 5 \end{matrix}.$$

Функция g_4 представляется как $g_4 = x_1 \vee \bar{x}_3$. Для функции g_5, как сказано выше, следует выбрать дизъюнктивное разложение, поскольку матрица $\mathrm{M}^1(g_5)$ имеет больше строк, чем матрица $\mathrm{M}^0(g_5)$. Граф $G(g_5)$, представляемый матрицей

$$\mathbf{G}(g_5) = \begin{matrix} 5 \\ \left[\begin{matrix} x_1 \\ x_1 \vee x_4 \\ x_4 \\ x_3 \vee x_4 \\ x_3 \end{matrix}\right] & \begin{matrix} 3 \\ 4 \\ 6 \\ 9 \\ 10 \end{matrix} \end{matrix},$$

покрывается бикликами ({3,4}, {5}) $[x_1]$ и ({6,9,10}, {5}) $[x_3\ x_4]$, определяющими разложение $g_5 = \bar{x}_1 \vee g_6(x_3, x_4)$, где $g_6(x_3, x_4) = x_3 \vee x_4$.

Таким образом, получено полное разложение функций заданной системы на функции, представляемые операциями И и ИЛИ, и вся заданная система не полностью определенных

булевых функций реализуется структурой, описываемой следующей системой уравнений:

$f_1 = x_3 \vee g_1$, $f_2 = \bar{x}_1 \vee x_4$, $f_3 = g_4 g_5$,

$g_1 = g_2 \vee g_3$, $g_4 = x_1 \vee x_3$, $g_5 = x_1 \vee g_6$,

$g_2 = x_1 x_4$, $g_3 = x_4 x_5$, $g_6 = x_3 \vee x_4$.

Соответствующая комбинационная схема из элементов И и ИЛИ показана на рис. 2.

$x_1\ x_1\ x_3\ x_3\ x_4\ x_4\ x_5$

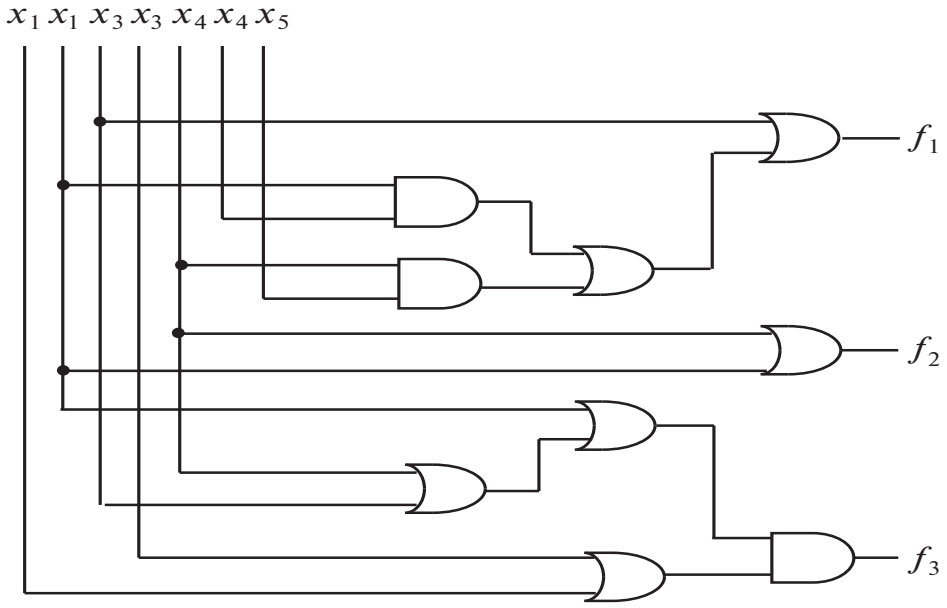

Рис. 2. Схема из элементов И и ИЛИ

Заключение. Показано, как можно применить метод алгебраической декомпозиции для синтеза комбинационных схем. Как отмечено во Введении, алгебраическая декомпозиция дает возможность получения схем с повышенным быстродействием, которое характеризуется числом уровней или глубиной схемы. Особенностью предлагаемого подхода к решению рассматриваемой задачи является применение двудольных неориентированных графов. Язык теории графов обеспечивает хорошую наглядность постановок задач, делает их понимание проще, чем при использовании других понятий. Также проще оказывается и описание методов их решения.

СПИСОК ЛИТЕРАТУРЫ

1. *Perkowski M.A., Grygiel S. A* Survey of Literature on Functional Decomposition, Version IV (Technical Report). Portland, USA: Portland State University, Department of Electrical En-gineering, 1995.
2. *Zakrevskij A.D.* On a Special Kind Decomposition of Weakly Specified Boolean Functions // Second Intern. Conf. on Computer-Aided Design of Discrete Devices (CAD DD'97). Minsk, Belarus: National Academy of Sciences of Belarus, Institute of Engineering Cybernetics, 1997. V. 1. P. 36–41.
3. *Fišer P., Schmidt J.* Small but Nasty Logic Synthesis Examples // Proc. 8th Intern. Workshop on Boolean Problems (IWSBP'8), Freiberg, Germany, 2008. P. 183–190.
4. *Бибило П.Н.* Декомпозиция булевых функций на основе решения логических уравне-ний. Минск: Беларус. навука, 2009.
5. *Choudhury M., Mohanram K.* Bi-decomposition of Large Boolean Functions Using Blocking Edge Graphs // 2010 IEEE/ACM Intern. Conf. on Computer-Aided Design (ICCAD'2010). San Jose: IEEE Press, 2010. P. 586–591.
6. *Cheng D., Xu X.* Bi-decomposition of Logical Mappings via Semi-Tensor Product of Matri-ces // Automatica. 2013. V. 49. N 7. P. 1979–1985.
7. *Steinbach B., Posthoff C.* Vectorial Bi-decomposition for Lattices of Boolean Functions // Further Improvements in the Boolean Domain / Cambridge. Cambridge Scholars Publishing, 2018. P. 175–198.

8. *Jóźwiak L., Chojnacki A.* An Effective and Efficient Method for Functional Decomposition of Boolean Functions Based on Information Relationship Measures // Design and Diagnos-tics of Electronic Circuits and Systams: Proc. of 3rd DDECS Workshop, Smolenice Castle, Slovakia, Bratislava: Institute of Informatics, Slovak Academy of Sciences, 2000. P. 242–249.
9. *Закревский А.Д.* Декомпозиция частичных булевых функций – проверка на раздели-мость по заданному разбиению // Информатика. 2007. № 1 (13). С. 16–21.
10. *Поттосин Ю.В., Шестаков Е.А.* Применение аппарата покрытий троичных матриц для поиска разбиения множества аргументов при декомпозиции булевых функций // Вестн. Томск. гос. ун-та. Управление, вычислительная техника и информатика. 2011. № 3(16). С. 100–107.
11. *Taghavi Afshord S., Pottosin Yu.V., Arasteh B.* An Input Variable Partitioning Algorithm for Functional Decomposition of a System of Boolean Functions Based on the Tabular Method // Discrete Applied Mathematics. 2015. V. 185. P. 208–219.
12. *Поттосин Ю.В.* Метод бидекомпозиции частичных булевых функций // Информа-тика. 2019. Т. 16, № 4. С. 77–87.
13. *Поттосин Ю.В., Шестаков Е.А.* Декомпозиция системы частичных булевых функ-ций с помощью покрытия графа полными двудольными подграфами // Новые инфор-мационные технологии в исследовании дискретных структур. Докл. второй Всерос-сийск. конф. Екатеринбург: УрО РАН, 1998. С. 185–189.
14. *Pottosin Yu.V.* A Method for Bi-decomposition of Partial Boolean Functions // Приклад-ная дискретная математика. 2020. № 47. С. 108–116.
15. *Поттосин Ю.В.* Эвристический метод алгебраической декомпозиции частичных булевых функций // Информатика. 2020. Т. 17, № 3. С. 44–53.
16. *Поттосин Ю.В., Шестаков Е.А.* Параллельно-последовательная декомпозиция си-стемы частичных булевых функций // Прикладная дискретная математика. 2010. № 4(10). С. 55-63.
17. *Cortadella J.* Timing-driven Logic Bi-decomposition // IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems. 2003. V. 22. N 6. P. 675–685.
18. *Mishchenko A., Steinbach B., Perkowski M.* An Algorithm for Bi-decomposition of Logic Functions // Proc. 38th Annual Design Automation Conf. (DAC'2001), Las Vegas, USA, 2001. P. 103–108.
19. *Chang S.C., Marek-Sadowska M., Hwang T.* Technology Mapping for TLU FPGA's Based on Decomposition of Binary Decision Diagrams // IEEE Trans. Computer-Aided De-sign. 1996. V. 15. N 10. P. 1226–1235.
20. *Поттосин Ю.В.* Комбинаторные задачи в логическом проектировании дискретных устройств. Минск: Беларус. навука, 2021.
21. *Закревский А.Д., Поттосин Ю.В., Черемисинова Л.Д.* Логические основы проекти-рования дискретных устройств. М.: Физматлит, 2007.
22. *Евстигнеев В.А., Касьянов В.Н.* Толковый словарь по теории графов в информатике и программировании. Новосибирск: Наука. СО РАН, 1999.

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, с. 77–92

СИСТЕМНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

УДК 519.853,517.977.5

ОПТИМИЗАЦИЯ ПРОИЗВОДСТВЕННЫХ ПРОГРАММ ПРЕДПРИЯТИЯ С УЧЕТОМ НЕОПРЕДЕЛЕННОСТИ[1]

© 2024 г. И. А. Борисов[a, *], О. А. Косоруков[b, **], А. В Мищенко[a, *], В. И. Цурков[c, ****]**

[a]ФГОБУ ВО "Финансовый университет при Правительстве Российской Федерации", Москва, Россия

[b]МГУ им. М. В. Ломоносова, Российская академия народного хозяйства и государственной службы при Президенте РФ, Российский экономический университет имени Г. В. Плеханова, Москва, Россия,

[c]ФИЦ ИУ РАН, Москва, Россия

**e-mail: ilyaborisov2015@yandex.ru*

***e-mail: kosorukovoa@mail.ru*

****e-mail: alnex4957@rambler.ru*

*****e-mail: v.tsurkov@mail.ru*

Поступила в редакцию 12.04.2024 г.
После доработки 21.05.2024 г.
Принята к публикации 15.07.2024 г.

Рассмотрен метод ветвей и границ, используемый для выбора оптимальной производственной программы, который основан на вычислении верхней, нижней и текущих верхних оценок при анализе различных вариантов производственных программ. Дана верхняя оценка количества допустимых решений приведенной задачи. Описаны модели выбора оптимальной производственной программы в условиях расширения производства, а также вопросы анализа устойчивости этих программ при изменении исходных данных модели и критерия оптимальности модели. Применение моделей выбора оптимальной производственной программы в рамках проектного управления на предприятиях обеспечит повышение эффективности мероприятий, в том числе на этапах планирования и реализации проектов, классификации и выбора метода реализации проектов.

Ключевые слова: производственная программа, метод ветвей и границ, анализ устойчивости, расширение производства.

DOI: 10.31857/S0002338824040052 **EDN:** UEGQUQ

OPTIMIZATION OF ENTERPRISE PRODUCTION PROGRAMS TAKEN INTO ACCOUNT OF UNCERTAINTY

I. A. Borisov[a, *], O. A. Kosorukov[b, **], A. V Mishchenko[a, *], V. I. Tsurkov[c, ****]**

[a]FGOBU VO "Financial University under the Government of the Russian Federation", Moscow, Russia

[b]Moscow State University named after. M. V. Lomonosov, Russian Academy of National Economy and Public Administration under the President of the Russian Federation, Russian Economic University named after G. V. Plekhanov, Moscow, Russia,

[c]Federal Research Center "Computer Science and Control," Russian Academy of Sciences, Moscow, Russia

**e-mail: ilyaborisov2015@yandex.ru*

***e-mail: kosorukovoa@mail.ru*

[1] Результаты исследований, представленные в разд. 6, 7, получены за счет средств Российского научного фонда (проект № 24-21-00339).

***e-mail: alnex4957@rambler.ru
****e-mail: v.tsurkov@mail.ru

The branch and bound method used to select the optimal production program is considered, based on the calculation of the upper, lower and current upper estimates when analyzing various options for production programs. An upper bound for the number of feasible solutions to the problem under consideration is given. Models for choosing an optimal production program in conditions of production expansion are considered, as well as issues of analyzing the stability of these programs when changing the initial data of the model and when changing the criterion for the optimality of the model. The use of models for selecting the optimal production program within the framework of project management at enterprises will ensure increased efficiency of activities, including at the stages of planning and implementation of projects, classification and selection of a method for implementing projects.

Keywords: production program, branch and bound method, production expansion, stability of the programs.

Введение. Одной из задач производственного планирования является выбор оптимальной производственной программы предприятия. В условиях динамически изменяющейся внешней среды такой выбор – непростая задача, требующая для решения использования математико-статистических методов и моделей [1, 2].

Статические методы и модели выбора таких программ, в том числе методы оценки риска доходности этих программ и риска упущенной выгоды, рассмотрены в [3,4]. Динамические модели выбора оптимальной производственной программы, методы оптимизации загрузки оборудования при выпуске продукции, заданной производственной программой, оценки устойчивости расписаний при изменении параметров задачи представлены в [5–7]. Динамические и статические модели и методы управления ограниченными ресурсами на транспорте излагаются в [8–10]. В [11–13] приведены точные и приближенные алгоритмы построения оптимальных расписаний для планирования работы многопроцессорной вычислительной техники. Модели и методы управления ограниченными ресурсами, которые сводятся к решению минимаксных задач, описаны в [14–20].

В настоящей статье для выбора оптимальной производственной программы предприятия предлагается использование метода ветвей и границ, основанного на вычислении верхней, нижней и текущих верхних оценок при анализе различных вариантов производственных программ, дана верхняя оценка количества допустимых решений рассматриваемой задачи. Также представлены модели выбора оптимальной производственной программы в условиях расширения производства, вопросы анализа устойчивости этих программ при изменении исходных данных модели и при изменении критерия оптимальности модели.

Предложенные методы и модели могут использоваться в том числе в рамках проектного управления на предприятиях, обеспечивая возможность выбора оптимального метода управления проектом, эффективное выполнение мероприятий на этапах планирования и реализации проектов [21, 22].

1. Постановка задачи и метод решения. Рассмотрим следующую модель выбора оптимальной производственной программы:

$$\sum_{i=1}^{n} a_i x_i - \sum_{i=1}^{n} b_i x_i - Z_{\text{пост}} \to \max, \tag{1.1}$$

$$\sum_{i=1}^{n} l_{ij} x_i \le L_j,\ j = \overline{1, M}, \tag{1.2}$$

$$\sum_{i=1}^{n} t_{il} x_i \le K_l \tau_l,\ l = \overline{1, K}, \tag{1.3}$$

$$x_i \le Pt_i,\ i = \overline{1, n}, \tag{1.4}$$

$$x_i \in Z^+,\ i = \overline{1, n}. \tag{1.5}$$

Здесь использовались следующие обозначения: a_i – цена продукции i -го вида; b_i – переменные издержки на единицу продукции i -го вида; x_i – объем выпуска продукции i –го вида; $Z_{\text{пост}}$ – постоянные издержки; l_{ij} – норма потребления материальных ресурсов j-го вида при выпуске единицы продукции i -го вида; L_j – запасы материальных ресурсов j-го вида; t_{il} – норма времени загрузки оборудования l-го вида при выпуске единицы продукции i -го вида; K_l – число единиц оборудования l-го вида, участвующих в процессе производства; τ_l – эффективное время работы оборудования l-го вида на периоде планирования (0, T); Pt_i – объем спроса на продукцию i -го вида; Z^+ – множество целых неотрицательных чисел.

1.1. Метод ветвей и границ для задачи выбора оптимальной производственной программы. Задача (1.1) – (1.5) является задачей линейной целочисленной оптимизации и для ее решения может быть использован метод ветвей и границ.

Шаг 1. Верхняя оценка задачи (1.1) – (1.5) $F_в$ может быть получена путем замены ограничения (1.5) на ограничение (1.6) следующего вида:

$$x_i \geq 0,\ i = \overline{1,n}\,, \tag{1.6}$$

Значение целевой функции на оптимальном решении задачи (1.1) – (1.4), (1.6) будем считать $F_в$.

Шаг 2. Нижняя оценка задачи (1.1) – (1.5) $F_н$ может находиться путем выбора допустимого решения задачи (1.1) – (1.5) и вычисления значения целевой функции (1.1) на этом решении.

Шаг 3. Вычисление верхних текущих оценок. Если $F_в = F_н$, то решение задачи (1.1) – (1.5) получено. Если $F_в > F_н$, то начинаем формировать очередное допустимое решение с вычислением $F_в^{\text{тек}}(\tilde{x})$. Здесь $\tilde{x} = (x_1,\ldots,x_n)$ – вектор, задающий объемы выпуска продукции, которые уже вошли в производственную программу.

Верхняя текущая оценка выполняется по следующей формуле:

$$F_в^{\text{тек}}\left(\widetilde{x_1},\ldots,\widetilde{x_n}\right) = \sum_{i=1}^{n} a_i\tilde{x}_i - \sum_{i=1}^{n} b_i\tilde{x}_i - Z_{\text{пост}} + F_в\left(\widetilde{L_j},\widetilde{\tau_l},\widetilde{Pt_i}\right), \tag{1.7}$$

$$\widetilde{L_j} = L_j - \sum_{i=1}^{n}\tilde{x}_i l_{ij},\ j = \overline{1,M},$$

$$K_l\widetilde{\tau_i} = K_l\tau_l - \sum_{i=1}^{n}\tilde{x}_i t_{il},\ l = \overline{1,K},$$

$$\widetilde{Pt}_i \leq Pt_i - \tilde{x}_i,$$

где $F_в\left(\widetilde{L_j},\ K_l\widetilde{\tau_l}\right)$ – верхняя оценка задачи (1.1) – (1.5) с учетом того, что объем материальных ресурсов равен $\widetilde{L_j}$, $j = \overline{1,M}$, а эффективное время по каждому виду оборудования равно $K_l\widetilde{\tau_i}$, $l = \overline{1,K}$. Здесь $\widetilde{L_j}$ – остаток материальных ресурсов j-го вида после выпуска продукции в объеме $\widetilde{x} = \left(\widetilde{x_1},\ldots,\widetilde{x_n}\right)$; $K_l\widetilde{\tau_l}$ – остаток эффективного времени для оборудования вида l после выпуска продукции в объеме $\widetilde{x} = \left(\widetilde{x_1},\ldots,\widetilde{x_n}\right)$.

Если $F_в^{\text{тек}}\left(\widetilde{x_1},\ldots,\widetilde{x_n}\right) > F_H$, то формирование производственной программы продолжается путем включения в производственную программу еще одной единицы продукции и дальнейшей вычисленной текущей верхней оценки.

Если $F_в^{\text{тек}}\left(\widetilde{x_1},\ldots,\widetilde{x_n}\right) \leq F_H$, то данная программа не будет оптимальной и исключается из дальнейшего рассмотрения.

Если $F_в^{\text{тек}}\left(\widetilde{x_1},\ldots,\widetilde{x_n}\right) > F_H$ остается до момента, когда ни одну единицу продукции невозможно включить в производственную программу, не нарушив одно из ограничений (1.2) – (1.5), то вычисляется значение целевой функции (1.1) на сформированной производственной программе. Обозначим это значение:

$F^* > F_H$ – значение F_H сдвигается вправо и становится равным F^*;

$F^* = F_в$ – задача (1.1) – (1.5) решена;

$F^* < F_в$ – переходим к анализу очередного допустимого решения.

Решение задачи (1.1) – (1.5) будет получено, если:

а) при очередной корректировке $F_н$ ее значение совпадает с $F_в$;

б) все варианты формирования производственных программ исследованы, тогда в качестве оптимального решения выбирается та программа, которая соответствует последнему (максимальному) значению $F_н$.

1.2. Верхняя оценка числа допустимых производственных программ. Верхняя оценка объема выпуска по каждому виду продукции определяется исходя из ограничений (1.2) – (1.5).

Так, если мы определяем максимальный объем выпуска продукции первого вида, исходя из ограничений на материальные ресурсы, то этот объем $\theta^1_{\max}$ задается следующей формулой:

$$\theta^1_{\max} = \min_{j=\overline{1,M}} \left\{ \frac{L_j}{l_{1j}} \right\}.$$

Максимальный объем выпуска продукции первого вида $r^1_{\max}$ при ограничениях на производственные мощности вычисляется как

$$r^1_{\max} = \min_{l=\overline{1,k}} \left\{ \frac{k_l \tau_l}{t_{1l}} \right\}.$$

Таким образом, максимальный выпуск продукции первого вида $x_1^{\max}$ рассчитывается следующим образом:

$$x_1^{\max} = \min \left\{ r^1_{\max}, \theta^1_{\max}, \ pt_1 \right\}.$$

Аналогично определяется максимальный объем выпуска по другим видам продукции:

$$x_i^{\max} = \min_{i=\overline{1,n}} \left\{ r^i_{\max}, \theta^i_{\max}, \ pt_i \right\}.$$

Таким образом, количество допустимых производственных программ задачи (1.1) – (1.5) не превысит числа N:

$$N = \prod_{i=1}^{n} \left(x_i^{\max} + 1 \right).$$

Наряду с критерием прибыли (целевая функция (1.1)) при выборе производственной программы может использоваться критерий рентабельности следующего вида:

$$\left(\sum_{i=1}^{n} a_i x_i - \sum_{i=1}^{n} b_i x_i - Z_{\text{пост}} \right) / \left(\sum_{i=1}^{n} b_i x_i + Z_{\text{пост}} \right) \to \max. \tag{1.8}$$

Очевидно, критерий (1.8) есть отношение прибыли к затратам.

Как будет показано ниже, при определенных условиях оптимальные производственные программы по критериям (1.1) и (1.8) совпадают.

2. Устойчивость при нелинейном изменении доходности производственной программы от инфляции. Пусть в задаче (1.1) – (1.5) маржинальный доход c_i равен:

$$c_i = a_i + b_i.$$

Здесь a_i – цена продукции i-го вида; b_i – переменные издержки при выпуске продукции i-го вида. Будем полагать, что c_i зависит от уровня инфляции ξ следующим образом:

$$c_i(\xi) = c_i + \varphi_i(\xi),$$

$$\frac{d\varphi_i(\xi)}{d\xi} \geq 0,$$

$$\varphi_i(0) = 0,$$

$$\overline{X} = \left\{ x^1, \ldots, x^\ell, \ldots, x^N \right\}.$$

Множество $\overline{X}$ задает перечень всех производственных программ. Допустим, что x^ℓ – оптимальное решение при $\xi = 0$, которое обозначили через

$$f^j(\xi) = \sum_{i=1}^{n} c_i(\xi) x_i^j,\ j = \overline{1,N}.$$

Тогда переход на новую оптимальную производственную программу x^k при каком-то значении ξ возможен, если:

а) существует интервал $[\xi_1, \xi_2]$, на котором

$$\frac{df^k(\xi)}{d\xi} > \frac{df^l(\xi)}{d\xi};$$

б) существует $\xi^* \in [\xi_1, \xi_2]$, для которого $f^k(\xi^*) = f^l(\xi^*)$;

в) $\dfrac{df^k(\xi)}{d\xi} > \dfrac{df^l(\xi)}{d\xi}$ при $c_i = a_i + b_i$.

Если условие в) не выполняется, то возможна следующая ситуация (рис. 1).

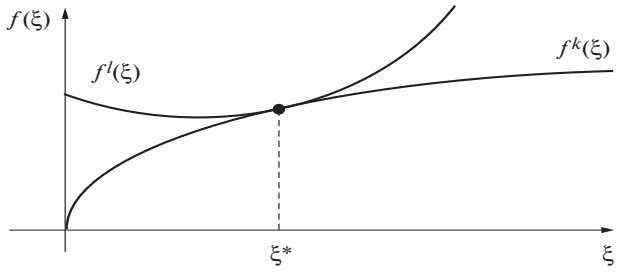

Рис. 1. Отсутствие перехода на новую производственную программу

Не соблюдается условие в) для перехода на новую оптимальную производственную программу. На рис. 1 существует ξ^*, для которого $f^l(\xi^*) = f^k(\xi^*)$, но

$$\frac{df^k(\xi)}{d\xi} \le \frac{df^l(\xi)}{d\xi},$$

поэтому переход на любую оптимальную производственную программу в точке ξ^* не происходит.

В ситуации нелинейного роста $c_i(\xi)$ возможно несколько переходов от одной оптимальной производственной программы к другой (рис. 2).

3. Анализ устойчивости модели к изменению критерия. Пусть есть множество допустимых значен $\underline{X} = \{x^1, ..., x^\ell, ..., x^N\}$ для оптимальных моделей (1.1) – (1.5) и (1.8), (1.2) – (1.5). Очевид-

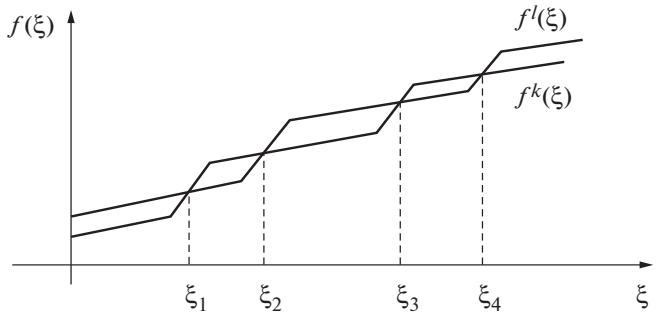

Рис. 2. Несколько переходов от оптимальной производственной программы x^l к оптимальной производственной программе x^k и обратно

но, в силу того, что системы ограничений в моделях (1.1) – (1.5) и (1.8), (1.2) – (1.5) совпадают, множество допустимых решений у этих моделей также будет одно и то же.

Допустим, что x^{ℓ} является оптимальным решением для модели (1.1) – (1.5). Очевидно, что если переменные издержки $b_i = 0$, то x^{ℓ} будет оптимально и для модели (1.8), (1.2) – (1.5), так как в этом случае критерий (1.8) – это целевая функция (1.1), умноженная на константу.

В каком диапазоне можно менять значения показателей b_i, умножая их на $\lambda > 0$ так, чтобы оптимальное решение для критериев (1.1) и (1.8) совпадали, иллюстрирует следующий численный эксперимент.

Пусть существует пекарня, которая выпускает четыре вида продукции (табл. 1).

Постоянные издержки составляют 5000 руб/мес.

Таблица 1. Виды продукции пекарни

Продукт	Цена, руб.	Переменные затраты, руб.
1	400	300
2	50	35
3	200	150
4	150	110

Нормы потребления ресурсов для производства продукции пекарни, их запасы, нормы времени выработки, эффективное время работы оборудования и месячный спрос на продукцию представлены в табл. 2–6 соответственно.

Тогда числовой пример выглядит следующим образом.

1. Прибыль и рентабельность:

Таблица 2. Нормы потребления ресурсов

Ресурс	Продукт			
	1	2	3	4
1	0.4 кг	0.05 кг	0.21 кг	0.17 кг
2	0.1 кг	0.01 кг	0.12 кг	0.08 кг
3	150 кв. см	50 кв. см	90 кв. см	75 кв. см
4	2 шт.	0.25 шт.	1 шт.	0.75 шт.

Таблица 3. Запасы материальных ресурсов

Ресурс			
1	2	3	4
200 кг	120 кг	70 000 кв. см	850 шт.

Таблица 4. Нормы времени выработки

Оборудование	Продукт			
	1	2	3	4
1	25 мин	10 мин	15 мин	12 мин
2	14 мин	6 мин	9 мин	11 мин
3	7 мин	2 мин	3 мин	4 мин

Таблица 5. Эффективное время работы оборудования

Оборудование	Эффективное время работы, ч	K
1	9.09	K1 = 1
2	11.36	K2 = 1
3	7.6	K3 = 1

Таблица 6. Месячный спрос на продукцию, шт.

Продукт			
1	2	3	4
200	250	400	350

$$100x_1 + 15x_2 + 50x_3 + 40x_4 - 5000 \to \max,$$

$$\frac{100x_1 + 15x_2 + 50x_3 + 40x_4 - 5000}{\left(300x_1 + 35x_2 + 150x_3 + 110x_4\right) + 5000} \to \max.$$

2. Ограничение на материальные ресурсы:

$$0.4x_1 + 0.05x_2 + 0.21x_3 + 0.17x_4 \le 200,$$
$$0.1x_1 + 0.01x_2 + 0.12x_3 + 0.08x_4 \le 120,$$
$$150x_1 + 50x_2 + 90x_3 + 75x_4 \le 70\,000,$$
$$2x_1 + 0,25x_2 + 1x_3 + 0,75x_4 \le 850.$$

3. Ограничение на производственную мощность (оборудование). Для расчетов берем 22 рабочих дня:

$$\frac{25}{60}x_1 + \frac{10}{60}x_2 + \frac{15}{60}x_3 + \frac{12}{60}x_4 \le 9.09 \cdot 22,$$

$$\frac{14}{60}x_1 + \frac{6}{60}x_2 + \frac{9}{60}x_3 + \frac{11}{60}x_4 \le 11.36 \cdot 22,$$

$$\frac{7}{60}x_1 + \frac{2}{60}x_2 + \frac{3}{60}x_3 + \frac{4}{60}x_4 \le 7.6 \cdot 22.$$

4. Ограничение на спрос на продукцию:

$$x_1 \le 200,$$
$$x_2 \le 250,$$
$$x_3 \le 400,$$
$$x_4 \le 350.$$

5. Ограничение на решения (целочисленное, положительное):

$$x_1 \in Z^+,$$
$$x_2 \in Z^+,$$
$$x_3 \in Z^+,$$
$$x_4 \in Z^+.$$

Вычислим значения прибыли и рентабельности переходов λ (рис. 3):

$$\lambda_{\min} \in [0;0.33],$$
$$\lambda_1 \in [0.34;0.54),$$
$$\lambda_2 \in [0.54;0.55),$$

$$\lambda_3 \in [0.55;0.86],$$
$$\lambda_4 \in (0.86;1.2),$$
$$\lambda_5 \in [1.2;1.3),$$
$$\lambda_6 \in [1.3;1.43),$$
$$\lambda_7 \in [1.43;2.9),$$
$$\lambda_8 \in [2.9;+\infty).$$

Результаты вычислений представлены в табл. 7:

4. Оптимизация производственной программы в условиях расширения производства. Рассмотрим ситуацию, когда наряду с традиционной продукцией предприятие будет выпускать еще и новые виды продукции: $n+1, \ldots, n_1$. Для этого потребуются дополнительные материальные ресурсы: $M+1, M+2, \ldots, M_1$ и дополнительное оборудование: $K+1, \ldots, K_1$. Для приобретения

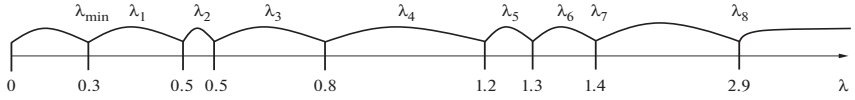

Рис. 3. График расположения переходов λ

Таблица 7. Результаты вычислений для переходов λ

λ	Решения
$0 \leqslant \lambda \leqslant 0.33$	Решения сохраняются: $x^{\text{р}} = (200; 0; 400; 53)$, $x^{\text{пр}} = (200; 0; 400; 53)$
$0.34 \leqslant \lambda < 0.54$	Решения по прибыли не изменяются, решения по рентабельности изменяются: $x^{\text{пр}} = (200; 0; 400; 53)$, $x^{\text{р}} = (200; 0; 152; 350)$
$0.54 \leqslant \lambda < 0.55$	Решения по прибыли не изменяются, решения по рентабельности изменяются: $x^{\text{пр}} = (200; 0; 400; 53)$, $x^{\text{р}} = (200; 0; 155; 347)$
$0.55 \leqslant \lambda \leqslant 0.86$	Решения изменяются по рентабельности: $x^{\text{пр}} = (200; 0; 400; 53)$, $x^{\text{р}} = (200; 1; 152; 350)$
$0.86 < \lambda < 1.2$	Решения изменяются по рентабельности: $x^{\text{пр}} = (200; 0; 400; 53)$, $x^{\text{р}} = (200; 250; 13; 350)$
$1.2 \leqslant \lambda < 1.3$	Решения изменяются по прибыли: $x^{\text{пр}} = (200; 0; 155; 347)$, $x^{\text{р}} = (200; 250; 13; 350)$
$1.3 \leqslant \lambda < 1.43$	Решения изменяются по прибыли: $x^{\text{пр}} = (200; 250; 13; 350)$, $x^{\text{р}} = (200; 250; 13; 350)$
$1.43 \leqslant \lambda < 2.9$	Решение по прибыли становится равным 0: $x^{\text{пр}} = (0;0;0;0)$, $x^{\text{р}} = (200;250;13;350)$
$\lambda \geqslant 2.9$	Решение по рентабельности становится равным 0, далее обе функции принимают значения 0: $x^{\text{пр}} = (0; 0; 0; 0)$, $x^{\text{р}} = (0; 0; 0; 0)$

дополнительного объема материальных ресурсов и дополнительных единиц оборудования вычисляются инвестиции в объемах $I1$ и $I2$ соответственно.

Задача выбора оптимальной производственной программы в этом случае формулируется следующим образом:

$$\sum_{i=1}^{n} a_i x_i - \sum_{i=1}^{n} b_i x_i - Z_{\text{пост}} \to \max, \tag{4.1}$$

$$\sum_{i=1}^{n} l_{ij}\alpha_i \le L_j + Z_j,\ j = \overline{1, M}, \tag{4.2}$$

$$\sum_{i=n+1}^{n_1} l_{ij}\alpha_i \le Z_j,\ j = \overline{M+1, M_1}, \tag{4.3}$$

$$\sum_{i=1}^{n} t_{ie} x_i \le \left(K_e + y_e\right)\tau_e,\ e = \overline{1, K}, \tag{4.4}$$

$$\sum_{i=n+1}^{n_1} t_{ie} x_i \le y_e \tau_e,\ e = \overline{K+1, K_1}, \tag{4.5}$$

$$\sum_{j=1}^{M_1} Z_j \beta_j \le V_1, \tag{4.6}$$

$$\sum_{e=1}^{K_1} y_e \gamma_e \le V_2, \tag{4.7}$$

$$x_i \le Pt_i,\ i = \overline{1, n_1}, \tag{4.8}$$

$$x_i \in Z^+,\ i = \overline{1, n_1}, \tag{4.9}$$

$$y_e \in Z^+, \qquad e = \overline{1, K_1},$$

$$Z_j \ge 0, \qquad j = \overline{1, M_1}\ .$$

5. Анализ устойчивости в условиях расширения производства. Рассмотрим ситуацию роста маржинального дохода $C_i(\xi)$ при росте инфляции ξ в модели (4.1) – (4.9). Будем считать $C_i(\xi) = a_i(\xi) - b_i(\xi)$, а в общем случае:

$$C_i(\xi) = C_i(0) + \varphi_i(\xi),\ i = \overline{1, n}\ . \tag{5.1}$$

Здесь

$$\frac{d\varphi_i(\xi)}{d\xi} \ge 0,\ \varphi_i(0) = 0\ u\ \varphi_i(\xi) > 0\ ;$$

$C_i(0)$ – маржинальный доход в начальный момент времени $t = 0$.

Пусть $\underline{X} = \left\{x^1, \ldots, x^N\right\}$ – множество допустимых производственных программ в модели (4.1) – (4.9), x^ℓ – оптимальная производственная программа.

Изменение целевой функции (1.1) на решении x^ℓ, при росте инфляции можно описать следующей функцией:

$$f^l(\xi) = \sum_{i=1}^{n} C_i(\xi) x_i^\ell + Z_{\text{пост}}. \tag{5.2}$$

Аналогичным образом можно задать значение целевой функции (4.1) на любой другой производственной программе $x^j\ \left(j = \overline{1, N}; j \ne \ell\right)$:

$$f^j(\xi) = \sum_{i=1}^{n} C_i(\xi) \cdot x_i^j + Z_{\text{пост}}. \tag{5.3}$$

Возникает вопрос: если x^l было оптимально при $\xi = 0$, останется ли оно оптимальным при изменении $\xi \in (0,\theta)$?

Продифференцируем $f^l(\xi)$ и $f^j(\xi)$ по ξ с учетом соотношения (5.1). Получаем

$$\frac{df^l(\xi)}{d\xi} = \sum_{i=1}^{n} \frac{d\varphi_i(\xi)}{d\xi} x_i^l, \tag{5.4}$$

$$\frac{df^j(\xi)}{d\xi} = \sum_{i=1}^{n} \frac{d\varphi_i(\xi)}{d\xi} x_i^j.$$

Если $\varphi_i(\xi)$ линейные функции:

$$\frac{df^l(\xi)}{d\xi} \geq \frac{df^j(\xi)}{d\xi}, \ \forall \xi \in (0,\theta), \tag{5.5}$$

то легко понять, что производственная программа x^l остается оптимальной для всех решений $\xi \in (0,\theta)$. Если же существует $K, K = \overline{1,N}$, такое, что

$$\frac{df^k(\xi)}{d\xi} > \frac{df^l(\xi)}{d\xi}, \ \forall \xi \in (0,\theta), \tag{5.6}$$

то при определенном значении ξ^* оптимальной становится программа x^k.

Значение ξ^* определяется из следующего уравнения:

$$f^k(\xi) = f^l(\xi). \tag{5.7}$$

В силу линейности $\varphi_i(\xi)$ уравнение (5.7) будет иметь единственное решение, которое мы обозначим ξ^*.

Таким образом, при $\xi \leq \xi^*$ оптимальным будет решение x^l, при $\xi \geq \xi^*$ – решение x^k, при $\xi = \xi^*$ – решение x^l и x^k. Область изменения $\xi^* \in [0,\xi^*]$ – область устойчивости для решения x^l.

Если существует несколько производственных программ $x^{K_1},\ldots,x^{K_M}$, для которых выполняется условие (5.6), то решается M уравнений, находятся решения $\xi_1,\ldots,\xi_M$ и в качестве точки перехода на новую оптимальную производственную программу выбирается точка $\xi^* = \min\{\xi_1,\ldots,\xi_M\}$.

В силу линейности $f^j(\xi)$ число переходов на новую оптимальную производственную программу будет конечно и не будет превышать числа $N-1$. Если же $\varphi_i(\xi)$ нелинейны, то нелинейны и $f^j(\xi)\ (i = \overline{1,n}; j = \overline{1,N})$. Тогда количество переходов от одной оптимальной производственной программы к другой может быть бесконечным при $\xi \in (0;\infty)$ уже для $N = 2$.

Приведем пример этой ситуации для случая, когда $\varphi_i(\xi)$ кусочно-линейны (рис. 4).

Изменение оптимальной производственной программы может быть связано с ростом стоимости материальных ресурсов и стоимости оборудования при росте инфляции, а также с падением спроса при росте инфляции. Рассмотрим ситуацию, когда цена материальных ресурсов зависит от инфляции следующим образом:

(5.8)

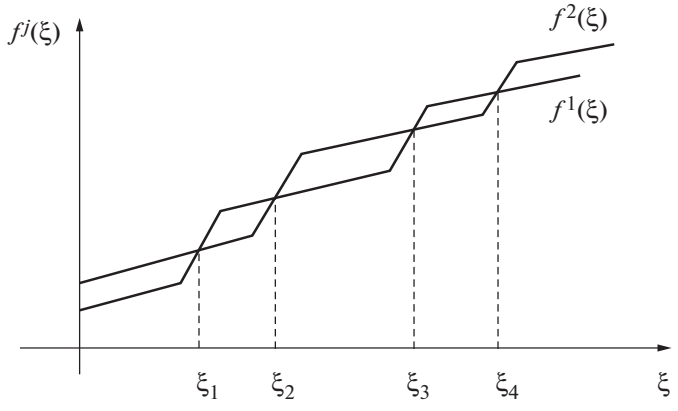

Рис. 4. Несколько точек перехода от одной оптимальной производственной программы к другой в условиях расширения производства

$$\beta_j(\xi) = \beta_j(0) + \psi_j(\xi),$$

где $\psi_j(\xi) \geqslant 0, \forall \xi$ и

$$\frac{d\psi_j(\xi)}{d\xi} > 0.$$

Здесь $\beta_j(0)$ – стоимость материального ресурса j в начальный момент времени $t = 0$. В этом случае неравенство (4.6) можно переписать в следующем виде:

$$\sum_{j=1}^{M_1} Z_j{}^2\beta_j(\xi) \leqslant V_1. \tag{5.9}$$

С учетом того, что с ростом ξ правая часть неравенства (5.9) растет, а объем инвестиций V_1 не меняется, можно вычислить такое ξ_τ, при котором

$$\sum_{j=1}^{M_1} Z_j{}^2\beta_j(\xi) = V_1. \tag{5.10}$$

Дальнейший рост инфляции приведет к тому, что объем затрат на материальные ресурсы (левая часть равенства (5.10)) станет больше, чем объем инвестиций V_1. Следовательно, невозможно обеспечить производственную программу x^l (оптимальную в момент времени t) необходимым объемом материальных ресурсов и, следовательно, предприятие будет выпускать продукцию в меньших объемах, т. е. необходим переход при уровне инфляции ξ_τ от производственной программы x^l к новой программе $x^\tau, \tau = \overline{1, N}$.

Аналогичная ситуация при росте цен на оборудование. Если их изменение происходит по закону, описываемому следующей формулой:

$$\gamma_1(\xi) = \gamma_1(0) + \chi_1(\xi), \tag{5.11}$$

где $\chi_1(\xi) \geqslant 0$, $\chi_1(\xi) = 0$ и

$$\frac{d\chi_l(\xi)}{d\xi} > 0,$$

то существует ξ_j при котором неравенство (4.7) будет иметь вид

$$\sum_{l=1}^{K_1} Y_l\gamma_l(\xi_\gamma) = V_2. \tag{5.12}$$

Следовательно, при $\xi = \xi_\gamma$ также будет переход на новую оптимальную производственную программу X^γ.

Наконец, естественно предположить, что спрос Pt_i будет падать с ростом инфляции. Пусть это падение описывается формулой

$$Pt_i(\xi) = Pt_i(0) - Q_i(\xi), \tag{5.13}$$

где $Q_i(\xi) = 0$ при $\xi = 0$; $Q_i(\xi) \geqslant 0$ при $\xi > 0$;

$$\frac{dQ(\xi)}{d\xi} > 0$$

для $\forall \xi \in (0, \theta)$.

Рассмотрим (4.8) с учетом (5.13):

$$x_i \leq Pt_i(\xi),\ i = \overline{1, n_1}, \tag{5.14}$$

Так как x^l оптимально, при $\xi = 0$ также должно выполняться

$$x_i^l \leq Pt_i(\xi),\ i = \overline{1, n_1}. \tag{5.15}$$

Так как правая часть в (5.15) уменьшается с ростом ξ, то существует ξ_p и существует γ, такие, что

$$x_\gamma^l = Pt_\gamma\left(\xi_p\right). \tag{5.16}$$

Следовательно, при $\xi > \xi_p$ производственная программа x^l перестает быть допустимой и, следовательно, предприятие вынуждено снижать объем выпуска, переходя к другой производственной программе $x^p, p = \overline{1,N}$.

Таким образом, при изменении маржинального дохода $C_i(\xi)$, $i = \overline{1,n_1}$, на материальные ресурсы $\beta_j(\xi)$, $j=\overline{1,M_1}$, цен на оборудование $\gamma_l(\xi)$, $l=\overline{1,K_1}$, и спроса $Pt_i(\xi)$, $i=\overline{1,n_1}$, под влиянием инфляции ξ произойдет переход на новую оптимальную производственную программу. Константы n_1, M_1, K_1 определены нами в начале разд. 4. Уровень инфляции ξ, при котором произойдет этот переход вычисляется по формуле

$$\xi = \min\{\xi_c,\ \xi_\beta,\ \xi_\gamma,\ \xi_p\}.$$

6. Динамическая модель выбора оптимальной производственной программы. Рассмотрим ситуацию, когда материальные ресурсы поступают динамически на вход производственной системы с интенсивностью $L_j(t)$, $j=\overline{1,M}$. В этом случае интенсивность выпуска конечной продукции также будет задана динамически как $x_i(t)$, $i=\overline{1,n}$, и задача выбора оптимальной производственной программы может быть сформулирована следующим образом:

$$\sum_{i=1}^{n}\int_0^T C_i(t)x_i(t)dt \to \max. \tag{6.1}$$

Здесь $C_i(t) = a_i(t) - b_i(t)$

$$\sum_{i=1}^{n} l_{ij}\int_0^t x_i(t')dt' \le \int_0^t L_j(t)dt,\ j = \overline{1,M},\ \forall t \in (0,T), \tag{6.2}$$

$$\sum_{i=1}^{n} t_{il}\int_{t_1}^{t_2} x_i(t')dt' \le \frac{t_2 - t_1}{T}k_l\tau_l,\ l=\overline{1,K}, \tag{6.3}$$

$$\forall t_1, t_2 (t_2 > t_1),\ t_1 \in (0,T),\ t_2 \in (0,T),$$

$$\int_0^T x_i(t)dt \le Pt_i,\ i=\overline{1,n}, \tag{6.4}$$

$$x_i(t) \ge 0,\ i=\overline{1,n}. \tag{6.5}$$

Неравенство (6.2) задает ограничения на потребление материальных ресурсов; неравенство (6.3) – ограничения на производственные мощности с учетом равномерной загрузки оборудования.

Задача (6.1) – (6.5) является задачей оптимального управления. Она может быть сведена к задаче линейной целочисленной оптимизации следующим образом. Разобьем интервал $(0, T)$ на конечное число отрезков времени (дней) и будем полагать, что $x_{i\tau}$ – объем выпуска продукции в день с номером τ, а $L_{j\tau}$ – объем поступления материальных ресурсов j в день τ. Тогда задача (6.2) – (6.5) может быть переписана следующим образом:

$$\sum_{i=1}^{n}\sum_{\tau=1}^{T} c_{i\tau}x_{i\tau} \to \max, \tag{6.6}$$

где $c_{i\tau}$ – маржинальный доход от выпуска одной единицы продукции i в день τ:

$$\sum_{i=1}^{n}\sum_{\tau=1}^{\theta} l_{ij}x_{i\tau} \le \sum_{\tau=1}^{\theta} L_{j\tau},\ j = \overline{1,M}, \theta = \overline{1,T}, \tag{6.7}$$

$$\sum_{i=1}^{n} t_{il} x_{i\tau} \le \frac{1}{T} k_l \tau_l,\ l=\overline{1,K},\ \tau=\overline{1,T}, \tag{6.8}$$

$$\sum_{\tau=1}^{T} x_{i\tau} \le P t_i, \tag{6.9}$$

$$x_{i\tau} \in Z^+,\ i=\overline{1,n},\ \tau=\overline{1,T}. \tag{6.10}$$

Таким образом, задача (6.6) – (6.10) является задачей целочисленной линейной оптимизации, в которой переменные $x_{i\tau}$ задают объем выпуска продукции i в день с номером τ.

7. Устойчивость решений динамической модели. Рассмотрим ситуацию, когда $c_{i\tau}$ могут меняться в зависимости от параметра ξ следующим образом: $c_{i\tau}(\xi) = c_{i\tau} + \varphi_{i\tau}(\xi)$, где $\varphi_{i\tau}(\xi)$ – возрастающая непрерывная функция и $\varphi_{i\tau}(0) = 0$. Рассмотрим ситуацию, когда $\varphi_{i\tau}(\xi)$ линейна и $\varphi_{i\tau}(\xi) = c_{i\tau} * \alpha_{i\tau} * \xi$.

Пусть $\underline{x} = \left\{x^1,\ldots,x^n\right\}$ – множество допустимых решений задачи (6.6) – (6.10) и решение $x^l = x^l_{i\tau}$ оптимально при $\xi = 0, l = \overline{1,N}$, и обозначается через $f^l(\xi)$:

$$f^l(\xi) = \sum_{i=1}^{n}\sum_{\tau=1}^{T} c_{i\tau}(\xi) x^l_{i\tau}.$$

Тогда

$$\frac{df^l(\xi)}{d\xi} = \sum_{\tau=1}^{T}\sum_{i=1}^{n} c_{i\tau}\alpha_{i\tau} x^l_{i\tau}.$$

Для любого другого $j = \overline{1,N},\ j \ne l$;

$$\frac{df^j(\xi)}{d\xi} = \sum_{\tau=1}^{T}\sum_{i=1}^{n} c_{i\tau}\alpha_{i\tau} x^j_{i\tau}.$$

Очевидно, что если существует такое $k,\ k = \overline{1,N},\ k \ne l,$ такое, что

$$\frac{df^k(\xi)}{d\xi} > \frac{df^l(\xi)}{d\xi},$$

то линейное уравнение $f^k(\xi) = f^l(\xi)$ имеет одно положительное решение. Пусть это решение равно ξ^*. Тогда очевидно, что при $\xi \in \left(0,\xi^*\right)$ оптимальной будет производственная программа x^l, а при $\xi > \xi^*$ оптимальной производственной программой будет x^k. При $\xi = \xi^*$ оптимальной станет и программа x^k, и программа x^l, отрезок $\left(0,\xi^*\right)$ назовем областью устойчивости решения x^l.

Если $\varphi_{i\tau}(\xi)$ нелинейны, то соответственно

$$\frac{df^j(\xi)}{d\xi} = \sum_{i=1}^{T}\sum_{\tau=1}^{n} \frac{d\varphi_{i\tau}(\xi)}{d\xi} x^j_{i\tau},\ j=\overline{1,N}.$$

Если в этом случае x^l оптимально при $\xi = 0$, то для перехода на новое оптимальное решение x^k необходимым условием является существование отрезка $[\xi_1,\xi_2]$, на котором

$$\frac{df^k(\xi)}{d\xi} > \frac{df^l(\xi)}{d\xi},\ \forall \xi \in [\xi_1,\xi_2].$$

Достаточным условием выступает выполнение приведенных далее требований:

$$a)\ \exists[\xi_1,\xi_2]; \frac{df^k(\xi)}{d\xi} > \frac{df^l(\xi)}{d\xi}, \forall \xi \in [\xi_1,\xi_2],$$

$$b)\ \exists\xi^*; f^k\left(\xi^*\right) = f^l\left(\xi^*\right), \forall\ \xi^* \in [\xi_1,\xi_2].$$

8. Анализ устойчивости производственной программы, пример. Рассмотрим следующую задачу:

$$\sum_{i=1}^{n} c_{i(\xi)} x_i \to \max, \tag{8.1}$$

$$\sum_{i=1}^{n} l_{ij} x_i \le L_j; j = \overline{1,M}, \tag{8.2}$$

$$\sum_{i=1}^{n} t_{il} x_i \le k_l \tau_l; l = \overline{1,K}, \tag{8.3}$$

$$x_i \le Pt_i; i = \overline{1,n}, \tag{8.4}$$

$$x_i \in Z^+, \tag{8.5}$$

$$c_{i(\xi)} = a_{i(\xi)} - b_{i(\xi)}.$$

Пусть у задачи (8.1) – (8.5) есть два допустимых решения:

$$x^1 = \left(x_1^1, x_2^1\right) = (1,2) \text{ и } x^2 = \left(x_1^2, x_1^2\right) = (2,1),$$

$$C_1(0) = C_1 = 2; C_2(0) = C_2 = 1.$$

Допустим также, что $C_1(\xi)$ и $C_2(\xi)$ – линейные функции:

$$C_i(\xi) = C_i + C_i * q_i * \xi;\ i = 1,2;\ q_1 = 1;\ q_2 = 5,$$

$$f^j(\xi) = \sum_{i=1}^{n} C_i(\xi) * x_i^j; j = 1,2.$$

$$f^j(0) = \sum_{i=1}^{n} C_i\left(x_i^j\right) = f^1(0) = 1 * 2 + 2 * 1 = 4; f^2(0) = 2 * 2 + 1 * 1 = 5.$$

Следовательно, при $\xi = 0$ оптимальна программа x^2.

Рассчитаем $f^1(\xi) = 1(2 + 2 * 1 * \xi) + 2(1 + 1 * 5 * \xi) = 4 + 12\xi$,

$$f^2(\xi) = 2(2 + 2 * 1 * \xi) + 1(1 + 1 * 5 * \xi) = 5 + 9\xi,$$

$$\frac{df^1(\xi)}{d\xi} = 12;\ \frac{df^2(\xi)}{d\xi} = 9,$$

т.е.

$$\frac{df^1(\xi)}{d\xi} \ge \frac{df^2(\xi)}{d\xi}.$$

Тогда есть $\xi^* \ge 0$, при котором $f^2\left(\xi^*\right) = f^1\left(\xi^*\right)$, начиная с которого $f^1(\xi) > f^2(\xi)$, $\xi > \xi^*$. Следовательно, при $\xi \ge \xi^*$ оптимальной будет программа $x^1 = (1,2)$. Интервалом устойчивости в этом случае для x^1 станет $(0,\ \xi^*)$. Для данного примера ξ^* вычисляется путем решения уравнения:

$$f^1(\xi) = f^2(\xi)$$

$$4 + 12\,\xi = 5 + 9\,\xi$$

$$\xi = 1/3, \text{ т.е. } \xi^* = 1/3.$$

Пусть при $\xi = 2$ промежуточное изменение коэффициентов q_1 и q_2 равно $q_1 = 1$ и $q_2 = 5$. Рассчитаем $f^1(\xi)$ и $f^2(\xi)$ для $\xi = 2$:

$$f^1(\xi) = 4 + 12 * 2 = 4 + 24 = 28,$$

$$f^2(\xi) = 5 + 9 * 2 = 5 + 18 = 23.$$

При изменении q_1 и q_2 как $\xi = 2$ получили следующее задание $f^1(\xi)$ и $f^2(\xi)$:
для $f^1(\xi)$:

$$28 = 9 * 2 + b \Rightarrow b = 10,$$

$$f^1(\xi) = 9\xi + 10 \text{ для } \xi > 2,$$

для $f^2(\xi)$:

$$23 = 12 * 2 + b \Rightarrow b = -1,$$

$$f^2(\xi) = 12\xi - 1,$$

$$\frac{df^2(\xi)}{d\xi} > \frac{df^1(\xi)}{d\xi},$$

а следовательно,возникает точка перехода. Решаем уравнение:

$$12\xi - 1 = 9\xi + 10,$$

$$\xi = 11 : 3 = 3\frac{2}{3}.$$

Начиная с $3\frac{2}{3}$ оптимальным будет снова решение x^2.

Таким образом, область устойчивости для решения x^1 – это интервал $\left[\frac{1}{3};3\frac{2}{3}\right]$, а область устойчивости для x^2 – интервал $\left[0;\frac{1}{3}\right]$ и $\left[3\frac{2}{3};\infty\right]$ (рис. 5).

Заключение. Предложено использование метода ветвей и границ, основанного на вычислении верхней, нижней и текущих верхних оценок при анализе различных вариантов производственных программ, который обеспечивает выбор оптимальной производственной программы предприятия. Представлена верхняя оценка числа допустимых производственных программ. Показано, что наряду с критерием прибыли при выборе производственной программы может использоваться критерий рентабельности. Показано, что при определенных условиях оптимальные производственные программы по критериям прибыли и рентабельности совпадают. Проведен анализ устойчивости производственных программ при изменениях исходных данных модели и критерия оптимальности модели, в том

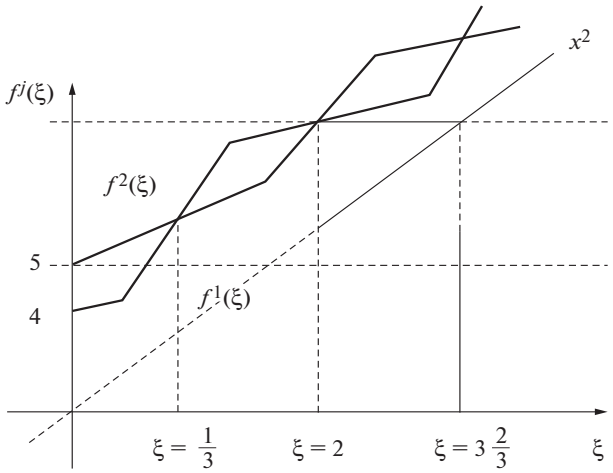

Рис. 5. Области устойчивости для решений задачи

числе при нелинейном изменении доходности производственной программы от инфляции в условиях расширения производства.

Представлены различные модели выбора оптимальной производственной программы. В рамках рассмотрения динамической модели выбора, описывающей ситуацию, при которой материальные ресурсы поступают динамически на вход производственной системы с интенсивностью выпуска конечной продукции, которая также задана динамически, сформулирована задача выбора оптимальной производственной программы.

Предлагаемые методы и модели можно использовать при реинжиниринге производственных и управленческих процессов, в рамках проектного управления на предприятиях для повышения эффективности производственных и управленческих процессов. Применение моделей позволит руководству предприятия более эффективно планировать реализацию проектов, оценивать объем необходимых для реализации проектов ресурсов, увидеть направления совершенствования процессов управления проектами, оптимизировать проектную деятельность.

СПИСОК ЛИТЕРАТУРЫ

1. *Brucker P., Jurisch B., Jurisch M.* Open Shop Problems with Unit Time Operations, ZOR // Methods and Models of Operations Research. 1993. V. 37. P. 59–73.
2. *Coffman E.G., Nozari A., Yannakakis M.* Optimal Scheduling of Products with Two Subassemblies on a Single Machine // Oper. Res. 1989. V. 37. P. 426–436.
3. *Данилин В.И.* Финансовое и операционное планирование в корпорации РАНХиГС. М., 2014.
4. *Мищенко А.В., Халиков М.А.* Распределение ограниченных ресурсов в задаче оптимизации производственной деятельности предприятия // Изв. АН СССР. Техн. кибернетика. 1991. № 6.
5. *Мищенко Л.В., Пилюгина Л.В.* Динамические модели управления научно-производственными системами // Вестн. МГТУ им. Н.Э. Баумана. Сер. Приборостроение. 2019. № 2.
6. *Мищенко А.В., Сушков Б.Г.* Задача оптимального распределения ресурсов на сетевой модели при линейных ограничениях на время выполнения работ // ЖВМ и МФ. 1980. Т. 10. № 5.
7. *Мищенко А.В., Когаловский В.М.* Проблемы устойчивости задач производственного планирования в машиностроении // Экономика и мат. методы. 1992. № 3.
8. *Мищенко А.В.* Устойчивость решений в задаче перераспределения транспортных средств в случае экстренного закрытия движения на участке метрополитена // Изв. АН СССР. Техн. кибернетика. 1990. № 3.
9. *Мищенко А.В.* Задача распределения транспортных средств по автобусным маршрутам при неточно заданной матрице корреспонденций пассажиропотока // Изв. АН СССР. Техн. кибернетика. 1992. № 2.
10. *Катюхина О.А., Мищенко А.В.* Динамические модели управления транспортными ресурсами на примере организации работы автобусного парка // Аудит и финансовый анализ. 2016. № 2. С. 156–167.
11. *Косоруков Е.О., Фуругян М.Г.* Некоторые алгоритмы распределения ресурсов в многопроцессорных системах // Вестн. МГУ. Сер. 15. Вычисл. математика и кибернетика. 2009. № 4. С. 34–37.
12. *Фуругян М.Г.* Планирование вычислений в многопроцессорных АСУ реального времени с дополнительным ресурсом // АиТ. 2015. № 3.
13. *Косоруков Е.О., Фуругян М. Г.* Алгоритмы распределения ресурсов в многопроцессорных системах с нефиксированными параметрами // Некоторые алгоритмы планирования вычислений и организации контроля в системах реального времени. М.: ВЦ РАН, 2011. С. 40–51.
14. *Mironov A.A., Tsurkov V.I.* Transport-type Problems with a Criterion // АиТ. 1995. №12. С. 109–118.
15. *Миронов А.А., Цурков В.И.* Наследственно минимаксные матрицы в моделях транспортного типа / / Изв. РАН. ТиСУ. 1998. № 6. С. 104–121.
16. *Mironov A.A., Levkina TA., Tsurkov V.I.* Minimax Estimations of Expectates of Are Weights in Integer Networks with Fixed Node Degrees // Applied and Computational Mathematics. 2009. V. 8. № 2. P. 216–226.
17. *Mironov A.A., Tsurkov V.I.* Class of Distribution Problems with Minimax Criterion // Doklady Akademii Nauk. 1994. V. 336. № 1. P. 35–38.
18. *Tizik A.P., Tsurkov V.I.* Iterative Functional Modification Method for Solving a Transportation Problem // Automation and Remote Control. 2012. V. 73. № 1. P. 134–143.
19. *Mironov A.A., Tsurkov V.I.* Hereditarily Minimax Matrices in Models of Transportation Type // J. Computer and Systems Sciences International. 1998. V. 37. № 6. P. 927–944.
20. *Mironov A.A., Tsurkov V.I.* Minimax in Transportation Models with Integral Constraints. I // J. Computer and Systems Sciences International. 2003. V. 42. № 4. P. 562–574.
21. *Борисов И.А.* Методика сравнительного анализа и оптимального выбора варианта управления проектами // Альманах «Крым». 2023. № 38–4.
22. *Борисов И.А.* Кластеризация проектов в целях повышения эффективности процессов проектного управления в ФНС России // Экономика и управление: проблемы, решения. 2023. № 8. Т. 3. С. 153–160.

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, С. 92–120

СИСТЕМНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

УДК 378.1

ТЕХНОЛОГИЯ УВЕРЕННЫХ СУЖДЕНИЙ ПРИ ПРИНЯТИИ РЕШЕНИЙ В СИСТЕМЕ ОБРАЗОВАНИЯ

© 2024 г. В. В. Малышев[a, *], С. А. Пиявский[b, **]

[a]Московский авиационный институт (научно-исследовательский университет), Москва, Россия
[b]Самарский филиал Московского педагогического университета, Самара, Россия
**e-mail: veniaminmalyshev@mail.ru*
***e-mail: spiyav@mail.ru*

Поступила в редакцию 15.07.2023 г.
После доработки 25.09.2023 г.
Принята к публикации 02.10.2023 г.

Стремительная цифровизация всех сторон жизни общества приводит к кардинальным изменениям в сфере образования. Для того чтобы в этих условиях формировать и реализовывать принципиально новые методы образования и методики обучения, необходимо применять современные наукоемкие методы подготовки и принятия сложных решений, в частности активно работать с информацией, представляемой не только в числовых, количественных, но и в порядковых шкалах типа "лучше — хуже", "более важные — менее важные". С этих позиций используется предложенный авторами метод уверенных суждений лица, принимающего решение. Показано, что его применение при отборе наиболее достойных абитуриентов для продолжения образования в высшей школе расширит полномочия вузов, позволит более точно учесть индивидуальные особенности и предпочтения школьников и их родителей. В рамках целостной системы выявления и многолетнего развития творчески одаренной молодежи в сфере науки и техники этот метод позволяет создать объективный измеритель (творческий рейтинг) для всесторонней оценки степени творческого роста каждого молодого исследователя, а также научно управляемую систему поддержки его направляемого развития. В результате в организации научной деятельности вуза появится целостная система оптимального планирования деятельности его институтов, исходящая из целевых ориентиров и общей стратегии развития вуза и в то же время учитывающая специфические особенности общей удовлетворенности трудом коллектива каждого института.

Ключевые слова: принятие решений, метод уверенного суждения, система образования, числовые, порядковые шкалы, интегральная система, творческий рейтинг, научная деятельность, оптимальное планирование

DOI: 10.31857/S0002338824040062 **EDN:** UEFMUE

TECHNOLOGY OF CONFIDENT JUDGMENT WHEN DECISION MAKING IN THE EDUCATION SYSTEM

V. V. Malyshev[a, *], S. A. Piyavsky[b, **]

[a]Moscow aviation institute (scientific research university), Moscow
[b]Samara branch of Moscow pedagogical university, city Samara
**e-mail: vemiaminmalyshev@mail.ru*
***e-mail: spiyavsky@mail.ru*

The rapid digitalization of all aspects of society is leading to fundamental changes in the field of education. In order to form and implement fundamentally new methods of education and teaching, it is necessary to use modern knowledge-intensive methods of preparing and making complex decisions, in particular, to actively work with information presented not only in numerical, quantitative, but also in ordinal scales such as “better worse”, “more important — less important”. From these positions, the article uses the method proposed by the authors confident judgments of the decision maker. It is shown that its use in selecting the most worthy applicants to continue their education in higher education will expand the powers of universities and will more accurately take into account the individual characteristics and preferences of schoolchildren and their parents;

when used within the framework of an integral system for identifying and long-term development of creatively gifted youth in the field of science and technology, it will make it possible to create an objective meter (creative rating) for a comprehensive assessment of the degree of creative growth of each young researcher and create on this basis a scientifically managed system for supporting his guided development; in the organization of scientific activities of a university, an integral system of optimal planning of the activities of its institutes will appear, based on target guidelines and the general development strategy of the university and at the same time taking into account the specific features of the general satisfaction with the work of the staff of each institute.

Keywords: decision making, confident judgment method, education system, numerical, ordinal scales, integral system, creative rating, scientific activities, optimal planning

Введение. Стремительная цифровизация всех сторон жизни общества приводит к кардинальным изменениям таких социальных процессов, которые, казалось бы, прочно устоялись в течение многих столетий. Так, сфера образования с XVI – XVII вв. до последнего времени твердо держалась на трех китах: классно-урочной системе, диктате обучающего над обучающимся, разделении сфер общего образования, высшего образования, производства. Это давало возможность упорядочивать, систематизировать и формализовать весьма сложные явления, упрощая их и подгоняя под простые, но надежно действующие схемы. Сегодня же процесс образования все более индивидуализируется, обучаемый все раньше становится более субъектным, различные сферы человеческой деятельности все более тесно взаимодействуют и пересекаются с образовательной сферой. Система становится слабоформализуемой, ее уже нельзя "обстрогать" под простые схемы. Для того чтобы в этих условиях формировать и реализовывать принципиально новые методы образования и методики обучения, всем акторам образовательной среды приходится использовать более совершенные и наукоемкие методы подготовки и принятия сложных решений, причем, в частности работать с информацией, представляемой не только в числовых, количественных, но и в порядковых шкалах типа "лучше – хуже", "более важные – менее важные". Проблема при этом состоит в том, что существующий мощный аппарат математического моделирования и оптимизации сложных решений и лежащих в их основе закономерностей базируется на количественных, а не порядковых шкалах, и потому не может быть использован напрямую. Следовательно, необходимо применение надежно обоснованных, но в то же время достаточно понятных для понимания не имеющих специальной подготовки пользователей, методов, которые позволяли бы решить эту проблему.

В статье предлагаются некоторые возможные направления совершенствования отдельных образовательных систем с помощью двух подобных методов: нового метода уверенных суждений (МУС) лица, принимающего решение (ЛПР) и менее обоснованного и удобного в применении, но зато имеющего более чем полустолетнюю историю использования метода аналитической иерархии (метод AHP, analytic hierarchy process).

1. О решении проблемы применения порядковых шкал в сфере образования. Наиболее значимым примером рассматриваемой проблемы является многокритериальная оценка эффективности принимаемых сложных решений. Необходимость использования не единственного, а нескольких частных критериев эффективности отражает разнообразие и различную ценность как отдельных составляющих принимаемого решения, так и последствий его реализации. Однако даже в том случае, когда значения частных критериев могут быть выражены в допускающих точную оценку количественных шкалах, напрямую сопоставить их сравнительную значимость в стремлении построить на их основе единственную комплексную количественную оценку ожидаемой эффективности решения невозможно. Это может сделать только ЛПР (или коллегиальный орган), который в силу своих полномочий является носителем максимально полного и потому неформализуемого представления о целостном назначении и возможных последствиях принимаемого решения. Именно это его понимание в конечном счете должно отразиться в алгоритме расчета комплексного показателя эффективности решения в зависимости от значений частных критериев.

В подавляющем большинстве случаев такой алгоритм прост и состоит в следующем. Пусть $f_j(y_i),\ i=\overline{1,n},\ j=\overline{1,m}$, – значение j-го частного критерия для варианта решения y_i из множества рассматриваемых вариантов решения Y, $y_i \in Y$. Будем считать, что значения критериев пронормированы по каждому из них в пределах их значений на множестве Y, например от нуля до единицы, и приведены к единому желаемому направлению оптимизации – на минимум (т.е. представлены в виде относительных потерь) или на максимум (представлены в виде относительных выигрышей).

Тогда комплексная эффективность $F(y_i)$ каждого варианта решения рассчитывается как средневзвешенная сумма значений его частных критериев:

$$F(y_i) = \sum_{j=1}^{m} \pm_j f_j(y_i),\ i = \overline{1,n},$$

$$\pm_j \geq 0,\ j = \overline{1,m},\ \sum_{j=1}^{m} \pm_j = 1.$$

Здесь весовые коэффициенты, т.е. коэффициенты сравнительной важности частных критериев, отражают их относительный "вклад" в оценку общей эффективности варианта решения. Интегрирующая функция ЛПР состоит в том, чтобы в конечном результате его действий появились количественные значения этих коэффициентов, что позволит в дальнейшем использовать для глубокой проработки вариантов решений весь мощный аппарат количественной математики.

Но человек, за исключением, может быть, профессиональных математиков, в своих раздумьях не мыслит в количественных шкалах, поэтому ЛПР не может "напрямую" уверенно задать числовые значения коэффициентов сравнительной важности частных критериев. Зачастую ЛПР уходят от этой обязанности, перекладывая ее на коллектив привлеченных экспертов, а затем принимая их согласованные в результате некоторой процедуры значения коэффициентов. Однако такой путь нельзя считать плодотворным, поскольку переместить в головы экспертов уникальное СВОЕ НЕФОРМАЛИЗУЕМОЕ понимание всей ситуации, в которой происходит выработка решения, ЛПР не может по определению самого его статуса. Поэтому хотя советы экспертов, безусловно, полезны, процедуру формирования значений коэффициентов сравнительной важности частных критериев ЛПР должен выполнить самостоятельно, тем более, что современные математические методы поддержки принятия решений предлагают ему инструментарий, облегчающий эту деятельность.

Человек имеет возможность производить сравнительную оценку любых объектов, однако не в количественных, а в порядковых шкалах типа "лучше – хуже – намного хуже" и т.п. Эта оценка субъективна, т.е. несет на себе "отпечаток" конкретного производящего ее человека, однако ЛПР, по сути своих полномочий, именно и обязан производить свою, субъективную, оценку. После того, как эта оценка выражена ЛПР, стоит задача сформулировать некоторые достаточно естественные (несущие на себе в минимальной степени "отпечаток" индивидуальности предложившего их человека) гипотезы и построить на их основе уже чисто математическим путем алгоритма, переводящий оценки, выраженные в порядковой шкале, в их количественные эквиваленты – искомые коэффициенты сравнительной важности частных критериев.

Одним из наиболее плодотворных подходов к такому преобразованию порядковой информации в числовую является предложенный Т. Саати (1926–2017 гг.) в 70–80-х годах прошлого века метод АНР [1–4]. Базовой частью этого метода является аргументированный переход от сравнительной оценки некоторых объектов с помощью порядковой шкалы к их сравнительной оценке по числовой шкале, т.е. к замене порядковых оценок их количественными эквивалентами.

Поясним основную идею метода АНР на следующем примере. Рассмотрим 5-уровневую порядковую шкалу:

<Уровень 1, Уровень 2, ..., Уровень 5>,

например, при сравнении превосходства одного объекта над другим:

<равное, умеренное, существенное, большое, очень большое>.

Относительно ее уровней известно лишь, что оценка некоторого объекта в этой порядковой шкале уровнем с большим номером означает большую его предпочтительность по сравнению с объектом, оцененным по этой шкале уровнем с меньшим номером. Используем метод АНР для того, чтобы обоснованно (в соответствии с аксиомами метода АНР) заменить порядковые оценки объектов их количественными эквивалентами, что даст возможность в дальнейшем использовать для исследования этих объектов аппарат количественной математики.

Для этого построим табл. 1. В столбцах 2–6 таблицы отразим отношение предпочтения при попарном сравнении между собой уровней исходной порядковой шкалы, перечислен-

ных в столбце 1. При этом будем пользоваться вместо названия уровня его количественным эквивалентом из стандартной табл. 2, являющейся основной аксиоматической частью метода АНР. Заполнение столбца 2 табл. 1 очевидно. Также очевидно, что по главной диагонали столбцов 3–6 табл. 1 будут стоять единицы. Несомненно, и заполнение ячеек табл. 1, расположенных ниже главной диагонали, а симметричные им ячейки, которые находятся выше главной диагонали, заполним обратными им значениями: если, например, в исходной ячейке табл. 1 стоит 3, то в симметричную ей ячейку внесем значение 0.333 (это еще одна аксиома метода АНР). В столбце 7 рассчитаем среднее геометрическое значений столбцов 2–6 (третья аксиома метода). Значения столбца 7 пронормируем так, чтобы в зависимости от решаемой задачи либо их сумма была равна единице (столбец 8, вариант А), либо единице был равен количественный эквивалент наивысшего уровня порядковой шкалы (столбец 9, вариант Б). Полученные результаты и есть искомые количественные эквиваленты уровней исходной порядковой шкалы.

Таблица 1. Пример использования метода АНР для расчета количественных эквивалентов уровней 5-уровневой порядковой шкалы

Уровень порядковой шкалы	1	2	3	4	5	Среднее геометрическое	Количественный эквивалент	
							А	Б
1	1	0.333	0.2	0.143	0.111	0.254	0.033	0.065
2	3	1	0.333	0.2	0.143	0.491	0.064	0.125
3	5	3	1	0.333	0.2	1	0.13	0.254
4	7	5	3	1	0.333	2.036	0.263	0.517
5	9	7	5	3	1	3.936	0.51	1
Сумма						7.717	1	—

Таблица 2. Шкала относительной важности в методе АНР (по [4])

Уровень важности	Количественное значение
Равная важность	1
Умеренное превосходство	3
Существенное или сильное превосходство	5
Значительное (большое) превосходство	7
Очень большое превосходство	9

Метод АНР позволяет решать значительно более широкий круг задач, чем рассмотренная в приведенном примере, однако алгоритм их решения структурно одинаков. При использовании этого метода функция ЛПР состоит в том, чтобы на основании своего целостного понимания ситуации принятия решения попарно сравнить между собой частные критерии эффективности принимаемого решения по их важности для комплексной оценки этого решения и выразить свою оценку в терминах порядковой шкалы, указанной в первом столбце 1 табл. 2. Всю остальную работу по получению количественной комплексной оценки эффективности различных рассматриваемых вариантов решения выполнят математика и компьютер.

Единственным осложняющим фактором при этом является трудоемкость такой оценки для ЛПР. На практике, в том числе и в сфере образования, возникают задачи с достаточно большим количеством частных критериев. Например, в Самарской областной системе мер по выявлению, развитию и вовлечению в инновационную деятельность творчески одаренной молодежи в сфере науки и техники, успешно функционирующей с 2015 г., для оценки на-

учно-исследовательских работ используются 15 частных критериев. Чтобы воспользоваться методом АНР, ЛПР (научному руководству этой системы) необходимо было бы провести 15 × 14 / 2 = 105 попарных сравнений важности частных критериев и, возможно, изменить их для каждой из 20 секций, а впоследствии, в зависимости от уровня развития каждого участника системы, для каждого из них – индивидуально. Вызывает сомнения сама неизбежность выполнения такого огромного количества парных сравнений, которая вытекает не из объективной необходимости, а из субъективной гипотезы, положенной в основу метода его автором.

Эта проблема успешно решается с помощью другой гипотезы, положенной в основу, идейно близкого к АНР метода, также направленного на преобразование порядковых шкал в количественные – МУС [5, 6].

Основная гипотеза метода МУС более естественна и проста, чем в методе АНР. Она состоит в следующем. ЛПР вместо попарного сравнения критериев относит каждый из них к одной из немногих стандартных групп критериев различной важности, например, "важные", "более важные", "наиболее важные" и т.п. Если частные критерии отнесены ЛПР к различным группам важности, то равно допустимыми могут быть любые количественные сочетания коэффициентов сравнительной важности этих критериев, лишь бы выполнялось естественное условие: если группа важности, к которой ЛПР отнесен частный критерий А, ниже, чем та группа, к которой им отнесен частный критерий Б, то значение количественного коэффициента важности критерия А должно быть меньше, чем критерия Б, а если критерии А и Б отнесены ЛПР к одной и той же группе важности, то их количественные коэффициенты сравнительной важности должны быть одинаковы. Исходя из этой гипотезы, в методе МУС автоматически формируется (полностью или с точной оценкой степени полноты) набор любых допустимых в соответствии с ней сочетаний значений количественных коэффициентов важности частных критериев, такой, что сумма значений равна единице. Обозначим: k – количество элементов этого набора, $q = \overline{1,k}$ – номер текущего элемента набора, $\pm_j^q$, $q = \overline{1,k}$, $j = \overline{1,m}$, – количественное значения коэффициентов важности критериев в этом наборе. Их будем рассматривать как количественные эквиваленты номеров уровней сравнительной важности критериев в этом наборе.

Среднее арифметическое по всему набору этих значений, рассчитанное по каждому критерию для всех элементов набора, принимается за искомые количественные коэффициенты важности критериев, отвечающие их сравнительной важности, указанной ЛПР:

$$\pm_j = \frac{1}{k}\sum_{q=1}^{k} \pm_j^q, \quad j = \overline{1,m}.$$

Соответствующий расчет в каждой задаче можно осуществить в компьютере простейшим методом статистических испытаний или сразу получить из универсальных таблиц, рассчитанных по строгим математическим алгоритмам [7, 8]. Для пояснения приведем пример такого расчета. Пусть рассматривается случай задачи принятия решения с тремя частными критериями, в которой ЛПР указано, что первый критерий более важен с позиций комплексной оценки эффективности решения, чем каждый из двух остальных.

В табл. 3 показаны допустимые сочетания числовых коэффициентов сравнительный важности этих критериев, равномерно с шагом 0.1 представляющие все бесконечное множество таких допустимых сочетаний. Сумма этих коэффициентов в каждом сочетании, естественно, равна единице. В предпоследней строке табл. 3 показаны средние значения этих коэффициентов, которые в соответствии с методом МУС принимаются за количественные эквиваленты групп сравнительной важности критериев в рассматриваемом примере. Они несколько отличаются от показанных в последней строке табл. 3 точных значений, взятых из универсальных таблиц Приложения к миографии [6], в которой они носят название универсальных коэффициентов важности (УКВ) критериев. Отличие объясняется тем, что перебор допустимых вариантов сочетаний значений коэффициентов, представленный в табл. 3, производился с шагом 0.1, что не полностью покрывает все возможное множество допустимых сочетаний значений этих коэффициентов. Результат будет тем точнее, чем меньшим будет шаг перебора.

Отметим, что для пользователя, применяющего метод МУС при решении своей конкретной задачи, нет надобности производить подобные расчеты, поскольку заранее рассчитаны соответствующие таблицы, приведенные в [6], которые содержат количественные значения коэффициентов сравнительной важности критериев для любого конкретного их распределения по уровням порядковой шкалы важности.

Для примера в табл. 4 приведены значения количественных эквивалентов групп коэффициентов важности критериев для множества, включающего четыре сравниваемых объекта. Ниж-

Таблица 3. Перечень допустимых сочетаний коэффициентов важности для трех критериев, удовлетворяющих условию, что критерии 1 важнее, чем каждый из критериев 2 и 3

Номер сочетания	Коэффициент важности критерия		
	1	2	3
1	0.4	0.4	0.2
2	0.4	0.3	0.3
3	0.4	0.2	0.4
4	0.5	0.5	0
5	0.5	0.4	0.1
6	0.5	0.3	0.2
7	0.5	0.2	0.3
8	0.5	0.1	0.4
9	0.5	0	0.5
10	0.6	0.4	0
11	0.6	0.3	0.1
12	0.6	0.2	0.2
13	0.6	0.1	0.3
14	0.6	0	0.4
15	0.7	0.3	0
16	0.7	0.2	0.1
17	0.7	0.1	0.2
18	0.7	0	0.3
19	0.8	0	0.2
20	0.8	0.1	0.1
21	0.8	0.2	0
22	0.9	0	0.1
23	0.9	0.1	0
24	1	0	0
Среднее значение коэффициентов	0.633	0.183	0.183
Стандартное значение количественных эквивалентов групп важности критериев	0.611	0.194	0.194

няя строчка табл. 4 отвечает задаче, в которой все четыре объекта имеют различную важность, а предпоследняя – задаче, в которой два объекта наиболее важны (третий уровень важности B3), один – менее важен (второй уровень важности B2), и один – еще менее важен (уровень важности B1). В четырех столбцах слева показано всевозможное распределение объектов по группам важности, а в правой группе столбцов – соответствующие количественные коэффициенты сравнительной важности объектов (заметим, что сумма произведений этих коэффициентов, умноженная на количество соответствующих им объектов, всегда равна единице).

Сводные таблицы, отвечающие возможным вариантам распределения до 10 критериев при использовании двух и трех групп сравнительной важности представлены в конце настоящей статьи.

Итак, при использовании метода МУС функция ЛПР состоит в том, чтобы на основании своего понимания ситуации принятия решения выбрать количество групп важности, позво-

Таблица 4. Количественные эквиваленты групп сравнительной важности от 2 до 4 критериев различной важности

Число частных критериев	Количество критериев в каждой группе важности				УКВ критериев			
	Группа важности критериев							
	B1	B2	B1	B1	B1	B1	B3	B4
2	2				0.5			
	1	1			0.25	0.75		
3	3				0.333			
	2	1			0.194	0.611		
	1	2			0.111	0.444		
	1	1	1		0.111	0.278	0.611	
4	4				0.25			
	3	1			0.16	0.521		
	2	2			0.104	0.396		
	1	3			0.063	0.313		
	2	1	1		0.104	0.271	0.521	
	1	2	1		0.063	0.208	0.521	
	1	1	2		0.063	0.146	0.396	
	1	1	1	1	0.063	0.146	0.271	0.521

ляющее ему выразить свое неформальное понимание ситуации принятия решения, а затем отнести каждый частный критерий эффективности решения к одной из этих групп – нужные количественные значения коэффициентов относительной важности частных критериев ЛПР найдет в стандартных ранее рассчитанных таблицах или легко получит в компьютере.

2. Иллюстративный пример и сравнение методов АНР и МУС. Пусть принимается решение о зачислении одного из шести абитуриентов на вакантное место в вуз и учитывается, что оценка абитуриента по математике, с учетом специфика направления обучения, важнее (группа важности B2), чем по физике или русскому языку (группы важности B1). В строках 5–10 и столбцах 2–5 табл. 5 указаны исходные сведения об абитуриентах. Для принятия хорошо аргументированного решения необходимо для каждого абитуриента обоснованно рассчитать количественный показатель – вступительный балл, комплексно оценивающий его успешность в обучении в школе и учитывающий предпочтение, высказанное ЛПР.

Использование для решения этой задачи метода АНР показано в табл. 6, где отражены результат попарного сравнения частных критериев в соответствии с мнением ЛПР и дальнейшая математическая обработка по описанной выше методике АНР. Рассчитанные в ней искомые коэффициенты сравнительной важности частных критериев приведены в последнем столбце табл. 6 и перенесены во вторую строку табл. 5.

Аналогичные им коэффициенты из метода МУС просто берутся из ранее составленных универсальных таблиц, в данном примере – из табл. 4 (строка 3). В последних двух столбцах табл. 4 показаны конечные результаты решения примера обоими методами – средний вступительный балл, который рассчитывается как линейная свертка частных критериев с весовыми коэффициентами, рассчитанными каждым из методов.

Как видим, несмотря на различные базовые гипотезы, в иллюстративном примере методы АНР и МУС дают весьма близкие результаты. Однако это имеет место лишь при небольшом числе критериев, а для двух критериев различной важности шкалы результаты вообще совпадают: 0.75 и 0.25. В табл. 7, 8 приведены значения количественных эквивалентов 5- и 10-уровневых порядковых шкал, рассчитанные обоими методами. Видно, что различие возрастает с увеличением количества уровней порядковой шкалы. Это имеет в реальных задачах большое значение.

Таблица 5. Иллюстративный пример зачисления в вуз

Номер строки	Частные критерии	Балл ЕГЭ			Средний вступительный балл по	
		Математика	Физика	Русский язык	АНР	МУС
1	Группы важности по МУС	B2	B1	B1		
2	Коэффициент важности по АНР	0.600	0.200	0.200		
3	Коэффициент важности по МУС	0.611	0.194	0.194		
4	Абитуриенты					
5	A1	91	71	68	82.40	82.57
6	A2	89	63	75	81.00	81.15
7	A3	57	82	77	66.00	65.67
8	A4	81	89	52	76.80	76.85
9	A5	83	85	62	79.20	79.23
10	A6	74	78	70	74.00	73.93

Таблица 6. Расчет коэффициентов сравнительной важности критериев в иллюстративном примере АНР

Критерий (балл ЕГЭ)	Балл ЕГЭ			Среднее геометрическое	Коэффициенты сравнительной важности критериев
	Математика	Физика	Русский язык		
Математика	1	3	3	2.080	0.600
Физика	0.333	1	1	0.693	0.200
Русский язык	0.333	1	1	0.693	0.200
Сумма				3.466	1

Таблица 7. Количественные эквиваленты уровней 5-уровневой (5-балльной) порядковой шкалы в методах АНР и МУС

Уровень порядковой шкалы	Числовой эквивалент уровня		Отношение МУС/АНР
	АНР	МУС	
0	0	0	—
1	0.065	0.085	1.32
2	0.125	0.193	1.55
3	0.254	0.337	1.33
4	0.517	0.557	1.08
5	1.000	1.000	1

Таблица 8. Количественные эквиваленты 10 уровней (10-балльной) порядковой шкалы в методах АНР и МУС

Уровень порядковой шкалы	Числовой эквивалент		Отношение МУС/АНР
	АНР	МУС	
0	0	0	—
1	0.017	0.034	1.98
2	0.026	0.072	2.79
3	0.040	0.115	2.85
4	0.064	0.164	2.55
5	0.089	0.220	2.47
6	0.147	0.289	1.97
7	0.241	0.374	1.55
8	0.393	0.488	1.24
9	0.634	0.659	1.04
10	1.000	1.000	1

Проведем сравнение обоих методов по важнейшей характеристике: их аргументированности, т.е. по степени доверия у ЛПР и возможных заинтересованных лиц к базовым гипотезам. Одна гипотеза – использование линейной свертки – у обоих гипотез общая и практически общепринятая. Кроме нее, метод МУС содержит лишь одну гипотезу, описанную выше и вполне естественную: при равновероятных оценках некоторой величины принимать за ее значение среднее арифметическое этих оценок.

Метод же АНР включает четыре дополнительные гипотезы. Первая из них – это стандартная табл. 1, вторая – использование среднего геометрического попарных сравнительных оценок важности критериев. При 3, 4 критериях достоинством метода АНР является возможность более "дробного" попарного сравнения важности частных критериев по сравнению с их отнесением к различным группам важности в методе МУС. Однако с увеличим числа частных критериев это достоинство обращается в недостаток. Так, например, при использовании 15 критериев, как это происходит в Самарском областном конкуре исследовательских работ учащихся [9], для расчета объективно обоснованных коэффициентов важности отдельных критериев было бы необходимо сравнить (обсудив!) между собой $15 \times 14 / 2 = 105$ пар критериев, что нереально. С позиций аргументируемого решения перед заинтересованными лицами этот недостаток является решающим, так как осмыслить результаты столь масштабного попарного сравнения и тем более проверить значения полученных коэффициентов для специально не подготовленного пользователя практически невозможно.

Неоспоримым же преимуществом метода МУС будет именно легкость получения готовых значений коэффициентов сравнительной важности критериев из стандартных таблиц, как только высказано мнение ЛПР об их сравнительной важности в порядковых шкала, причем понять это мнение и его аргументацию может любой человек.

С учетом сказанного считаем предпочтительным метод МУС и в настоящей статье будем использовать именно его.

3. Возможности применения метода МУС при зачислении в вуз. Рассмотрим на примере порядка зачисления в вуз, насколько использование современных гибких и хорошо аргументируемых методов принятия решений способно повысить эффективность образовательной системы.

В основе существующей системы зачисления в вуз лежит простой метод, основанный на учете результатов по 100-балльной шкале ЕГЭ (единый государственный экзамен) по трем дисциплинам (например, математика, физика, русский язык) с небольшой добавкой за успехи во внеучебной деятельности (в целом до 10 баллов). Право вуза, содержательно заинтересованного в отборе наиболее перспективных студентов из числа абитуриентов, сведено к смехотворному распределению возможных дополнительных 10 баллов по видам внеучебной

деятельности. Рассмотрим, насколько более эффективной может стать эта система при использовании одного из хорошо аргументируемых методов принятия решений МУС.

П р и м е р 1 (учет приоритетности дисциплин). Ученый совет вуза получает от учредителя право устанавливать различную степень важности перечисленных выше четырех составляющих вступительного балла абитуриента для математического и физического факультетов (табл. 9). Аргументированность отнесения различных направлений деятельности к группам важности обосновывается для каждого факультета в правилах приема в вуз, а соответствующие коэффициенты важности – отсылкой к методу МУС, который может быть без труда хорошо освоен учителями, учениками и их родителями в школе в самом начале подготовки к ЕГЭ. В нашем примере помимо предпочтения "титульной" дисциплины демонстрируем возможное различие во взглядах членов советов математического и физического факультетов.

Таблица 9. Установленные Ученым советом вуза приоритеты учета различных направлений успешности абитуриента в школе (пример 1)

Факультет	Балл ЕГЭ			Дополнительный балл, пересчитанный к 100-балльной шкале
	Математика	Физика	Русский язык	
Группа важности по решению Ученого совета вуза				
Математический	B4	B3	B2	B1
Физический	B2	B3	B1	B1
Коэффициент сравнительной важности (по методу МУС)				
Математический	0.521	0.271	0.146	0.063
Физический	0.271	0.521	0.104	0.104

В табл. 10, 11 показаны результат расчета вступительного балла для условных шести абитуриентов, из которых в вуз могут быть зачислены только три имеющих более высокий вступительный балл. При этом в табл. 11 во избежание двойного учета сравнительной важности дополнительный балл абитуриента будет увеличен в 10 раз для того, чтобы быть измеренным в той же шкале максимально возможных результатов деятельности абитуриента по различным направлениям, что и результаты его учебной деятельности.

Таблица 10. Зачисление в вуз по традиционной схеме

Абитуриент	Балл ЕГЭ			Дополнительный балл	Вступительный балл	Номер в списке на зачисление
	Математика	Физика	Русский язык			
	91	71	68	3	233	1–2
A2	89	63	75	4	231	3
A3	57	82	77	6	222	6
A4	81	89	52	4	226	5
A5	83	85	62	3	233	1–2
A6	74	78	70	7	229	4

В рассматриваемом примере демонстрируется положительный эффект, который может принести предлагаемое изменение правил приема. Если при традиционных правилах, подав свои документы на оба факультета, в вуз были бы зачислены только трое: A1, A2, A5, причем не был бы зачислен даже A4, имеющий наивысший балл по физике, то при измененных правилах – пятеро: все, кроме A3.

П р и м е р 2 (учет успешности абитуриента по дополнительным дисциплинам). Рассмотрим более полную возможность повысить гибкость отбора абитуриентов при зачислении

Таблица 11. Зачисление в вуз при праве вуза устанавливать приоритеты различных направлений деятельности абитуриента в школе

Абитуриент	Балл ЕГЭ			Дополнительный балл	Вступительный балл на математический факультет	Номер в списке на зачисление	Вступительный балл на физический факультет	Номер в списке на зачисление
	Математика	Физика	Русский язык					
A2	89	63	75	40	76.912	3	68.902	6
A3	57	82	77	60	66.941	6	72.417	4
A4	81	89	52	40	76.432	4	77.888	1
A5	83	85	62	30	77.22	2	76.346	2
A6	74	78	70	70	74.322	5	75.252	3

в вуз. Представим, что вуз имеет право учесть, помимо перечисленных четырех предметов, дополнительно представленные абитуриентом, по его желанию, результаты ЕГЭ по информатике и химии.

В этом случае Ученый совет вуза, опираясь на мнения советов отдельных факультетов, определяет сравнительную важность успешности деятельности абитуриента уже в четырех вариантах, в зависимости от представленных им дополнительных документов (табл. 12). Обратим внимание на то, что вступительный балл при использовании различного комплекса документов рассчитывается по-разному, однако все результаты характеризуют абитуриента в единой шкале. Поэтому при проведении конкурса на зачисление правильным будет использовать максимальный балл, рассчитанный по представленных документам. Следовательно, школьнику, заблаговременно задумывающемуся о поступлении в такой вуз, необходимо решить, какой состав документов ему целесообразно представить. Если он полагает, оценивая свои способности и интересы, что дополнительно включенные дисциплины охарактеризуют его более высоко, он будет готовиться и сдавать по ним ЕГЭ, в противном случае он не использует эту возможность. В табл. 12–15 приведены исходные данные и результаты соответствующих расчетов.

Таблица 12. Установленные Ученым советом вуза приоритеты учета различных направлений успешности абитуриента в школе (пример 2)

Вариант комплекта документов	Балл ЕГЭ			Дополнительный балл	Балл ЕГЭ	
	Математика	Физика	Русский язык		Информатика	Химия
Группа важности по решению Ученого совета вуза						
Математический факультет						
М1	В4	В3	В2	В1	—	—
М2	В4	В3	В2	В1	В3	—
М3	В4	В3	В2	В1	—	В2
М4	В4	В2	В1	В1	В3	В2
Физический факультет						
Ф1	В2	В3	В1	В1	—	—
Ф2	В2	В3	В1	В1	В2	—
Ф3	В2	В3	В1	В1	—	В2
Ф4	В3	В4	В1	В2	В3	В2

Таблица 13. Коэффициенты учета результатов ЕГЭ по отдельным дисциплинам, установленные Ученым советом вуза

Вариант комплекта документов	Балл ЕГЭ			Дополнительный балл	Балл ЕГЭ	
	Математика	Физика	Русский язык		Информатика	Химия
Коэффициент сравнительной важности (по МУС)						
Математический факультет						
М1	0.521	0.271	0.146	0.063	0	0
М2	0.46	0.206	0.089	0.039	0.206	0
М3	0.46	0.256	0.122	0.039	0	0.122
М4	0.417	0.128	0.043	0.043	0.242	0.128
Физический факультет						
Ф1	0.271	0.521	0.104	0.104	0	0
Ф2	0.206	0.46	0.064	0.064	0.206	0
Ф3	0.206	0.46	0.064	0.064	0	0.206
Ф4	0.199	0.417	0.026	0.079	0.199	0.079

Таблица 14. Расчет вступительного балла для зачисления на математический факультет

Абитуриент	Балл ЕГЭ						Вариант				Вступительный балл (максимальный из вариантов)
	Математика	Физика	Русский язык	Дополнительный балл	Информатика	Химия	М1	М2	М3	М4	
А1	91	71	68	30	—	—	78.47	—	—	—	78.47
А2	89	63	75	40	97	—	76.91	82.07	—	—	82.07
А3	57	82	77	60	—	90	66.94	—	69.89	—	69.89
А4	81	89	52	40	97	99	76.43	81.73	79.98	85.18	85.18
А5	83	85	62	30	—	83	77.22	—	78.82	—	78.82
А6	74	78	70	70	99	74	74.32	79.47	74.26	80.25	80.25

Таблица 15. Расчет вступительного балла для зачисления физический факультет

Абитуриент	Балл ЕГЭ						Вариант				Вступительный балл
	Математика	Физика	Русский язык	Дополнительный балл	Информатика	Химия	Ф1	Ф2	Ф3	Ф4	
А1	91	71	68	30	—	—	71.84	—	—	—	71.84
А2	89	63	75	40	97	—	68.90	74.59	—	—	74.59
А3	57	82	77	60	—	90	72.42	—	76.71	—	76.71
А4	81	89	52	40	97	99	77.89	83.46	83.84	84.81	84.81
А5	83	85	62	30	—	83	76.35	—	79.22	—	79.22
А6	74	78	70	70	99	74	75.25	80.49	75.24	80.13	80.49

В табл. 12 показаны результаты решения Ученого совета. Теперь каждый абитуриент, планируя поступление в вуз, может определить, насколько целесообразно ему представлять свои результаты в расширенных вариантах. Из табл. 12, используя универсальные таблицы коэффициентов важности метода МУС, легко определятся коэффициенты сравнительной важности соответствующих критериев при расчете средневзвешенного вступительного балла абитуриента. Результаты приведены в табл. 13.

В приемной комиссии каждому абитуриенту по поданным им документам автоматически вычисляется вступительный балл (см. табл. 13), по которому и проводится зачисление (табл. 14, 15). При этом рассчитываются варианты по всем комбинациям поданных абитуриентом документов и в качестве входного балла берется наибольший из баллов, полученных по этим вариантам. Другими словами, в любом случае учитывается вариант 1 (обязательный комплект документов), и если подаются результаты ЕГЭ по информатике и/или по химии, то учитываются и остальные варианты.

Предлагаемое в рассматриваемом примере изменение правил приема в вуз имеет ряд преимуществ перед существующей системой.

Во-первых, при нем существенно расширяются возможности наиболее успешной и активной части школьной молодежи, а именно в таких студентах заинтересованы и вузы, и государство. У них появляется свобода выбора тех дисциплин, к которым они имеют интерес или чувствуют особую склонность, и этот выбор закрепляется тем, что способствует поступлению в вуз.

Во-вторых, появление в вузе студентов с более широким спектром интересов и возможностей стимулирует их привлечение к проведению вначале студенческих, а затем и серьезных междисциплинарных научных исследований, что впоследствии создает предпосылки для формирования в вузах междисциплинарных научных школ с крепким молодежным резервом.

4. Возможности применения метода МУС при выявлении и развитии творчески одаренной научной молодежи. В вузы ежегодно приходит большое количество творчески мотивированных абитуриентов, успешно попробовавших свои силы в проектно-исследовательской деятельности, предусмотренной действующим Федеральным образовательным стандартом в качестве одного из обязательных компонентов обучения. Многие из них уже в школе выполнили учебно-исследовательские работы, получившие высокую оценку на различных конференциях и конкурсах молодежных научно-исследовательских работ регионального, российского и международного масштаба. Вовлечение после поступления в вуз в научно-исследовательскую деятельность кафедр является закономерным продолжением их творческого развития. В условиях расширяющейся информатизации общества работа по выявлению одаренных в научной сфере школьников и студентов и целенаправленному развитию их способностей должна приобрести систематический, свободный от территориальных и ведомственных барьеров характер.

В связи с этим общепринятое понимание учебно-исследовательской деятельности должно быть соответствующим образом развито. Традиционно оно описывается следующим образом (например, [10]).

Учебно-исследовательская деятельность – это "такая форма организации учебно-воспитательной работы, которая связана с решением учениками творческой, исследовательской задачи с заранее неизвестными результатами и предполагающая наличие основных этапов, характерных для научного исследования: постановка проблемы, изучение теории, посвященной данной проблематике, подбор методик исследования и практическое овладение ими, сбор собственного материала, его анализ и обобщение, научный комментарий, собственные выводы".

Представляется, что в реалиях современного информационного общества для наиболее мотивированной и потенциально творчески одаренной молодежи в сфере науки и техники – это определение является узким и не раскрывает весь потенциал ее дальнейшего развития. Для этой категории молодых исследователей предлагаем заменить понятие "учебно-исследовательская деятельность" на более широкое понятие "продвинутая учебно-исследовательская деятельность".

Продвинутая учебно-исследовательская деятельность представляет собой форму системной, ориентированной на ряд лет организационной и научно-направляемой самостоятельной деятельности мотивированного школьника или студента, которая:

направлена на удовлетворение его познавательных интеллектуальных и связанных с ними иных потребностей и эффективное развитие своего творческого потенциала;

поддерживается специально организованной единой развивающей научно-образовательной средой в рамках вуза, региона, отрасли, страны;

связана с последовательным решением исследовательских задач с заранее неизвестными результатами, возрастающей новизной, актуальностью и сложностью;

предполагает научно-обоснованную унифицированную оценку как развивающего эффекта, так и научной значимости этой деятельности.

Одним из хорошо знакомых авторам прообразов развивающей научно-образовательной среды, направленной на реализацию продвинутой учебно-исследовательской деятельности, выступает успешно функционирующая с 2015 г. Самарская областная система мер по выявлению, развитию и вовлечению в инновационную деятельность творчески одаренной молодежи в сфере науки, техники и технологий (ЕСМ, единая самарская система мер). Ее Концепция [11, 12] устанавливает ряд принципов построения, основанных на теории управляемого развития творческих способностей молодежи [13] и учитывающих первоочередные направления развития Самарской области, ее высокий научно-технический потенциал, накопленный опыт координации работы с творчески одаренной научной молодежью, в том числе с использованием телекоммуникационных технологий и интеллектуальных информационных систем:

поэтапность формирования;

координация и интеграция действующих механизмов работы с творчески одаренной молодежью на платформе персонального мониторинга ее развития;

развивающая продуктивная деятельность творчески одаренной молодежи;

индивидуальное научное руководство исследованиями и воодушевляющая перспективная тематика;

многолетнее целенаправленное дифференцированное индивидуальное управление развитием творчески одаренной молодежи;

базовая развивающая программа и индивидуальные планы развития молодых исследователей;

формирование положительных ценностных ориентиров молодежи.

Идея концепции состоит в том, что инфокоммуникационные технологии территориально, информационно и статусно существенно сближают город и деревню, сферы среднего и высшего образования, обучения и труда. В сфере творческого развития появляются качественно новые возможности реализации концепции "зон ближайшего развития" (по Выготскому), в которых в качестве наставников (научных консультантов при наличии научных руководителей) для школьников выступают преподаватели из вузов, а для студентов – ученые и творческие специалисты из сферы труда. Такое объединение полезно не только для молодых исследователей, но и для их научных руководителей, поскольку за счет их поверхностного, зато необременительного, взаимодействия с научными консультантами существенно повышается уровень "интеллектуализации" и перспективности как руководимых ими исследований, так и используемых при этом интеллектуальных и материальных инструментов.

Наряду с привычными для учебно-исследовательской деятельности парами "ученик + учитель" (У + У) и "студент – преподаватель" (С + П) в ЕСМ появились в качестве предпочитаемых им элементов тройки У + У + К "ученик – учитель – консультант" и С + П + К "студент – преподаватель – консультант", а также разновозрастные исследовательские коллективы (типа студенческих конструкторских бюро и научных кружков при кафедрах), ведущие исследования по тематике (не обязательно оплачиваемой), предложенной и консультируемой заинтересованными организациями – лидерами научно-технического прогресса. Причем коммуникационная составляющая ЕСМ позволяет (пока, к сожалению, в небольшом числе случаев) рекомендовать эту схему как для городских, так и для сельских школьников.

Вторая идея состоит в том, что необходимо обеспечить высокий развивающий уровень выполняемых из года в год молодыми участниками ЕСМ исследовательских работ, чтобы их деятельность при всем, возможно, практически полезном ее эффекте не была в творческом отношении "топтанием на месте" или совершенствованием чисто ремесленных (в широком и благородном понимании этого термина) навыков. Для этого при всем разнообразии научно-технических направлений и индивидуального содержания выполняемых исследований развивающий эффект должен ежегодно объективно оцениваться и измеряться для каждого молодого исследователя в единообразной творческой шкале с тем, чтобы при планировании его ближайших исследований и после их завершения, и у него, и у его наставников, и у организаторов ЕСМ была четкая количественная оценка того, насколько возрос его творческий уровень. Средством для решения этой задачи стал ежегодный Объединенный губернский конкурс исследовательских работ, на котором работы объективно оценивались бы слепым методом двумя высококвалифицированными экспертами по единой научно обоснованной систем частных критериев и на этой основе рассчитывался бы творческий рейтинг самой работы, а далее с учетом ряда дополнительных факторов и психологических особенностей – структура

творческих компетенций творческий профиль и творческий рейтинг автора в интересующих его сферах будущей профессиональной деятельности. Эти понятия будут далее в статье рассмотрены более подробно.

Третья идея состоит в том, что с учетом первых двух идей нелепо ограничиваться простой констатацией того, как складывается развитие молодого исследователя, не пытаясь дать ему и его наставникам инструмент, позволяющий моделировать ожидаемые последствия различных вариантов планируемой деятельности с тем, чтобы спланировать ее оптимальным образом. При этом и им, и организаторам ЕСМ должно быть понятно, что в столь тонкой сфере, как творческое развитие одаренной личности, ожидать от подобного моделирования прецизионной точности невозможно, однако это не причина не пытаться воспользоваться, в консультационном плане, тем, что предлагает в этом направлении современная наука. Фигурально говоря, хоть стрелка простенького наручного компаса дрожит, лучше воспользоваться им, чем просто наугад бродить по незнакомому лесу. Поэтому составной частью ЕСМ является консультационное научно-методическое обеспечение деятельности всех участников этой системы.

Целостная реализация описанных идей организуется на единой платформе интеллектуальной инфокоммуникационной системы ИИС АСТРА (https://vzlet.asurso.ru/).

Все функционирование ЕСМ, связанное с измерением динамики развития творческих способностей, мониторингом и мягким управлением развивающей деятельностью всех акторов системы построено на использовании метода МУС.

Прежде всего это связано с количественной оценкой текущего уровня творческих способностей. Они представляют собой, в частности, степень сформированности у молодого исследователя следующих основных исследовательских функций:

1) поиск проблемы,
2) постановка (осознание) темы исследования,
3) формирование ключевой идеи и плана решения проблемы,
4) выбор, освоение и реализация необходимого обеспечения,
5) реализация отдельных элементов исследования,
6) синтез решения,
7) оформление решения,
8) ввод в научный обиход, защита и сопровождение решения,
9) внутренний критический анализ решения.

Степень развития исследовательских функций проявляется в выполненной молодым исследователем работе, конечно с учетом степени его самостоятельности. Это вытекает из определения одаренности, сформулированного коллективом авторитетных российских психологов в [13]: "Одаренность – это системное, развивающееся в течение жизни качество психики, которое определяет возможность достижения человеком более высоких, незаурядных результатов в одном или нескольких видах деятельности по сравнению с другими людьми".

Отсюда следует, что наиболее достоверный и естественный способ оценки одаренности личности – оценка выполненной работы как результата ее деятельности. В ЕСМ работа оценивается объективными экспертами на региональном конкурсе по специально разработанной многоплановой системе критериев (табл. 16).

Таблица 16. Система критериев оценки молодых ученых

Критерий оценки НИР	Группа важности	Коэффициент сравнительной важности (округленно)
1. Тип результатов (насколько они носят исследовательский характер)	3	0.137
2. Результаты являются частью НИР руководителя, научной группы кафедры, вуза	2	0.042
3. Результаты относятся к перспективному направлению науки, техники. технологий	1	0.008
4. Направлена (подготовлена) публикация в научной печати	1	0.008
5. Результаты внедрены или подготовлены к внедрению в сторонних организациях	2	0.042
6. Представлен глубокий обзор научной проблематики	2	0.042

Окончание таблицы 16

Критерий оценки НИР	Группа важности	Коэффициент сравнительной важности (округленно)
7. Используются теоретические методы (математические, понятийный аппарат социально-гуманитарного научного познания)	3	0.137
8. Получены новые научные результаты	3	0.137
9. Имеются собственные оригинальные идеи участника	2	0.042
10. Имеется глубокий анализ литературы (по авторам и времени)	2	0.042
11. Используются/разработаны специальные технологии проведения исследований	2	0.042
12. Масштабность предполагаемых последствий полной реализации работы	3	0.137
13. Масштабность проведенного исследования	3	0.137
14. Качество оформления представленных результатов	2	0.042
15. Качество доклада и ответов на вопросы при защите работы	1	0.008
Сумма		1.003

Эксперты оценивают каждый критерий по расшифровывающей его 5-уровневой порядковой шкале, примеры которой для двух критериев приведены в табл. 17. Уровни шкалы соотнесены с последовательными этапами функционального развития молодого исследователя в соответствующем критерию направлении. Таким образом, оценка по этой системе критериев последовательно выполняемых молодым исследователем работ позволяет структурно оценить прогресс в его развитии. Для количественной же оценки степени прогрессирования автора работ снова используется метод МУС. В столбцах 2 и 3 табл. 16 представлено распределение критериев по группам важности (в творческом отношении) и показаны соответствующие весовые коэффициенты оценок каждого критерия в комплексном творческом рейтинге.

Таблица 17. Шкала оценки первых двух критериев

Характеристика частного результата и его структурные уровни
1. Тип результатов
0. Не носят исследовательского характера
1. Носят исследовательский характер, т.е. получен результат, который был неочевиден до ее выполнения
2. Кроме 1, автор сопоставляет полученный им результат с известными аналогичными результатами
3. Кроме 2, знает по литературе о научных школах соответствующего направления
4. Кроме 3, работа содержит выдвижение собственных новых идей
5. Кроме 4, предложена новая формализованная постановка задачи
2. Результаты являются частью НИР руководителя, научной группы кафедры, вуза
0. Не являются
1. Связаны с НИР руководителя
2. Связаны с НИР разновозрастного исследовательского коллектива, в который входит автор
3. Результаты использованы в публикациях в научной печати с указанием фамилии автора и научного руководителя
4. Автор является оплачиваемым участником ведущихся исследовательских работ
5. Автор является оплачиваемым участником работ по грантам РФФИ или отраслевым программам

Метод МУС также позволяет на основе оценок многоплановой системы критериев рассчитать текущий творческий рейтинг автора, демонстрируемый выполненной им работой. При этом в качестве критериев следует использовать все 9 показателей, упомянутые ранее при оценке степени сформированности у молодого исследователя основных исследовательских функций. Была разработана табл. 18, отражающая сравнительную значимость отдельных составляющих исследовательской квалификации автора для творческого уровня выполненной им работы. На ее основе автоматически создана аналогичная таблица, в которой названия групп важности критериев в каждом столбце заменены их количественными эквивалентами из метода МУС. После этого степень сформированности различных компонентов исследовательской функциональной квалификации рассчитывается как средневзвешенная сумма оценок различных критериев. Аналогичным образом вводится поправка, которая учитывает степень самостоятельности молодого исследователя при выполнении работы.

Таблица 18. Сравнительная значимость отдельных составляющих исследовательской квалификации автора НИР (В1 — отчасти влияет, В2 — влияет, В3 — значительно влияет)

Частные критерии оценки НИР	1	2	3	4	5	6	7	8	9
1. Тип результатов	В3	В2	В3	В1		В1		В1	В2
2. Результаты являются частью НИР руководителя, научной группы кафедры, вуза	В1	В1		В2	В1	В3	В1	В2	
3. Результаты относятся к перспективному направлению науки, техники, технологий	В2	В1	В3	В1	В1			В2	
4. Направлена (подготовлена) публикация в научной печати	В1		В1			В1	В2	В3	
5. Результаты внедрены или подготовлены к внедрению в сторонних организациях	В1		В1			В1	В2	В3	
6. Представлен глубокий обзор научной проблематики	В3	В2		В1				В1	
7. Используются теоретические методы (математические, понятийный аппарат социально-гуманитарного научного познания)		В2	В3	В3	В2	В1			
8. Получены новые научные результаты	В2	В3	В3	В2	В1	В2		В1	В3
9. Имеются собственные оригинальные идеи участника	В2	В1	В3	В1	В3	В2			В1
10. Имеется глубокий анализ литературы (по авторам и времени)				В3	В2		В1		
11. Используются/разработаны специальные технологии проведения исследований			В2	В1	В3	В3			В2
12. Масштабность предполагаемых последствий полной реализации работы	В1	В1	В2		В1		В2		В3
13. Масштабность проведенного исследования		В2			В1		В2		В3
14. Качество оформления представленных результатов							В3	В2	
15. Качество доклада и ответов на вопросы при защите работы							В2	В3	

5. Возможности применения метода МУС при оценке и оптимальном планировании научной деятельности вуза. Покажем на условном примере возможности повышения эффективности оценки и планирования результатов научной деятельности подразделений вуза благодаря использованию метода МУС.

П р и м е р 3 (оценка эффективности научной деятельности институтов вуза). Будем рассматривать условный вуз, состоящий из пяти институтов, результаты научной деятельности которых показаны в табл. 19. Для простоты ограничимся лишь семью показателями:

подготовленные и защищенные докторские диссертации (Д),
подготовленные и защищенные кандидатские диссертаций (К),
опубликованные статьи (С),
подготовленные и опубликованные учебники и монографии (М),
зарегистрированные программы и программное обеспечение (ПО),
полученные патенты (П),
объем научно-исследовательских работ, млн руб. (НИР).

Таблица 19. Результаты научной деятельности институтов вуза в году

Институт	Вид результатов научной деятельности						
	Д	К	С	М	ПО	П	НИР
1	2	3	4	5	6	7	8
И1	1	2	230	7	10	1	730
И2	1	7	140	9	15	5	260
И3	0	8	250	11	15	6	210
И4	2	3	230	9	7	2	110
И5	1	4	400	4	5	1	120
Всего по вузу	5	24	1250	40	52	15	1430

Эти же результаты представлены в относительной шкале как доля общих результатов вуза в табл. 20 (строки 2–8).

Таблица 20. Относительные результаты научной деятельности институтов и варианты оценки эффективности их научной деятельности

Институт	Относительный вклад институтов в различные виды результатов научной деятельности							Общий вклад института	Место при сравнении	Средневзвешенный вклад института	Место при средневзвешенной оценке
	Д	К	С	М	ПО	П	НИР				
1	2	3	4	5	6	7	8	9	10	11	12
И1	0.200	0.083	0.184	0.175	0.192	0.067	0.510	1.412	3	0.253	1
И2	0.200	0.292	0.112	0.225	0.288	0.333	0.182	1.632	2	0.213	2-3
И3	0.000	0.333	0.200	0.275	0.288	0.400	0.147	1.644	1	0.180	4
И4	0.400	0.125	0.184	0.225	0.135	0.133	0.077	1.279	4	0.213	2-3
И5	0.200	0.167	0.320	0.100	0.096	0.067	0.084	1.033	5	0.139	5
Сумма	1.000	1.000	1.000	1.000	1.000	1.000	1.000	—	—	—	—
Уровни сравнительной важности	3	2	2	3	1	2	3	—	—	—	—
Количественные эквиваленты уровней	0.255	0.071	0.071	0.255	0.02	0.071	0.255	—	—	—	—

При традиционной схеме оценки сравнительной эффективности научной деятельности институтов сравнительная "ценность" различных видов результатов научной деятельности принимается одинаковой. В этом случае комплексно оценить общий вклад института по всем видам деятельности можно по сумме его относительных вкладов в различные виды деятельности. Соответствующие значения приведены в столбце 10 табл. 20. Тогда места институтов по этому комплексному показателю определяются, как показано в столбце 10 табл. 20.

Однако с учетом стратегии развития вуза и его положения в социуме разные виды научных результатов все-таки имеют различную "ценность", и защищенная докторская диссертация ценится больше, чем опубликованная научная статья. Невозможно указать, насколько "больше" в количественном выражении, но это легко сделать, используя порядковую шкалу "больше – меньше". Соответственно руководство вуза путем коллективного обсуждения на Ученом совете вуза имеет возможность установить стратегические направления развития вуза в виде сравнительной важности для него получения различных видов научных результатов.

В рассматриваемом примере для этого выбрана трехуровневая порядковая шкала сравнительной оценки видов научных результатов:

$$<1 – \text{обычные}, 2 – \text{важные}, 3 – \text{наиболее важные}>.$$

Принятая стратегия развития условного вуза показана в этой шкале в предпоследней строке табл. 20. В последней строке таблицы приведены количественные эквиваленты этих приоритетов. Они соответствуют распределению семи видов НИР по группам возрастающей важности по схеме 7 = 1, 3, 3, т.е. из семи направлений одно отнесено к обычной группе важности, три – к более важной группе и еще три – к наиболее важной группе. Соответствующие количественные эквиваленты этих уровней при такой схеме сравнительной важности равны соответственно 0.02; 0.071; 0.255.

В этом случае комплексная оценка сравнительной эффективности научной деятельности институтов вуза будет рассчитываться как средневзвешенная сумма его относительных вкладов в отельные виды результатов научной деятельности с весами, показанными в последней строке табл. 20. Результаты представлены в столбцах 11–12 этой таблицы. Как видим, места институтов при учете соответствия ценности различных результатов научной деятельности принятой стратегии развития института существенно поменялись, а следовательно, поменялась вся стратегия ориентации коллективов институтов на основные направления научной деятельности и методика стимулирования их усилий.

Заметим, что необходимо при сравнении эффективности институтов учесть численность коллективов институтов, усилиями которых были достигнуты результаты. Чем ниже численность коллектива, которым был достигнут результат, тем, очевидно, выше его заслуга. Для того чтобы учесть этот фактор (и для дальнейшего использования в настоящей статье), необходимо включить в рассмотрение трудовой потенциал институтов, отражающий их трудовые затраты на получение результатов. Сделаем его упрощенный расчет, отталкиваясь от нормативов нагрузки второй половины дня, принятой при планировании учебно-педагогической нагрузки во многих вузах. Для иллюстрации примем условные значения трудоемкости получения различных видов результатов научной деятельности, содержащиеся в табл. 21.

Таблица 21. Условные нормативы трудоемкости отдельных результатов научной деятельности

Результат научной деятельности	Д	К	С	М	ПО	П	НИР
Единица измерения	1	1	1	1	1	1	10
Нагрузка второй половины дня преподавателя, усл.ч/год	400	200	70	400	70	100	150

Трудовой потенциал институтов, характеризуемый трудоемкостью полученных ими результатов научной деятельности (столбцы 2–8), и комплексный показатель эффективности институтов, рассчитанный с его учетом (столбец 12), приведены в табл. 22. Трудовой потенциал институтов показан в столбце 9 этой таблицы, в столбце 10 он пересчитан в относительную шкалу делением на суммарный трудовой потенциал всех институтов, а значения столбца 11 этой таблицы получены делением соответствующих элементов столбца 11 табл. 20 на соответствующие значения элементов столбца 10 табл. 22.

Таблица 22. Расчет трудового потенциала институтов и его учет при сравнительной оценке эффективности их научной деятельности

Институт	Вид научной деятельности в рассматриваемом году							Общая трудоемкость, усл.ч/год	Относительная трудоемкость	Средневзвешенный вклад института	Место при сравнительной оценке
	Д	К	С	М	ПО	П	НИР				
И1	400	400	16 100	2800	700	100	10 950	31 450	0.230	1.103	3
И2	400	1400	9800	3600	1050	500	3900	20 650	0.151	1.411	1
И3	0	1600	17 500	4400	1050	600	3150	28 300	0.207	0.869	4
И4	800	600	16 100	3600	490	200	1650	23 440	0.171	1.244	2
И5	400	800	28 000	1600	350	100	1800	33 050	0.241	0.576	5
Всего	2000	4800	87 500	16 000	3640	1500	21 450	136 890	—	—	—

Как видим, более полный учет факторов, характеризующих эффективность деятельности институтов привел к изменению их сравнительной оценки (столбцы 10 и 12 табл. 20, столбец 12 табл. 22). Таким образом, рекомендуется использование именно такой, более полной, методики оценки сравнительной эффективности научной деятельности институтов с учетом общей стратегии вуза и численности научных коллективов.

П р и м е р 4 (оптимальное планирование результатов научной деятельности институтов). Рассмотрим возможность использования метода МУС для оптимального планирования научной деятельности институтов вуза на предстоящий год. В качестве трудового ресурса будем предполагать трудовой потенциал институтов, замеренный по результатам научной деятельности прошлого года. Под оптимальностью естественно полагать наиболее эффективное продвижение вуза в направлении стратегии его развития. Стратегия развития вуза характеризуется, как предложено в примере 3, различной сравнительной важностью тех или иных результатов научной деятельности (две последние строчки табл. 20). Последняя строчка этой таблицы задает как бы направляющий вектор оптимального продвижения вуза в семимерном пространстве, координатами которого являются различные виды научной деятельности. Для того чтобы придать естественный смысл перемещению по направлению этого вектора, необходимо нормировать реальные результаты планируемой деятельности вуза в некоторой объективно обоснованной относительной шкале. В качестве базы для такой шкалы можно применить или результаты научной деятельности вуза в прошлом году или целевой ориентир развития вуза на некоторый стратегический период.

Используем как наиболее отвечающий, по нашему мнению, современным требованиям, второй вариант. Под целевым ориентиром развития вуза будем понимать максимальные значения результатов научной деятельности, которые запланированы руководством вуза на некоторый обозримый, но достаточно длительный промежуток развития, например, на 5 лет. Этот целевой ориентир опирается на рекордные достижения вузов-лидеров, установки и перспективные плановые документы Минобрнауки России и других органов и т.п.

Поясним сказанное. В первой строке этой табл. 23 представлены результаты научной деятельности вуза в году, предшествующем планируемому году. Предположим, что, анализируя стратегию развития вуза и его возможности, учитывая требования времени и вышестоящих инстанций, руководство вуза установило на некоторый стратегический период развития целевые показатели, представленные во второй строке этой таблицы,

которые вуз стремится достичь на конец стратегического периода. Они являются нормирующими значениями для оценки ценности соответствующих научных результатов в любой момент оценки деятельности вуза. Относительные значения показаны в строках 5–7 табл. 23. Они получены делением строк 1–3 этой таблицы на строку 2. Тогда столбец 10, рассчитываемый как сумма произведений элементов столбцов 3–9 в соответствующих строках на элементы строки 4, показывает положение вуза на направляющем векторе его стратегии развития. Видно, что в прошлом году вуз находился на этом векторе на относительном расстоянии 0.673 от некоторой начальной точки, а по достижении результатов, запланированных на предстоящий год, окажется от этой точки на относительном расстоянии 0.715. Сам же целевой ориентир отстоит от начальной точки на расстоянии, равном одной относительной единице (собственно он и принят за эту единицу).

Таким образом, задача оптимального планирования состоит в разработке такого плана научной деятельности вуза, который при использовании его известного трудового потенциала максимально близко продвинет его в направлении стратегии развития. Для решения такой задачи оптимального планирования разработаем достаточно простую модель линейного программирования.

Обозначим через $i = \overline{1,n}$ номер института, $j = \overline{1,m}$ – номер вида результата научной деятельности, a_{ij} – j-й результат i-го института в предшествующем году, a_j – результат вуза по j-му виду научной деятельности в предшествующем году:

$$a_j = \sum_{i=1}^{n} a_{ij}, \quad j = \overline{1,m},$$

где t_j – норматив трудоемкости j-го вида научной деятельности, k_{ij} – количественный эквивалент уровня мотивированности i-го института в выполнении j-го вида научной деятельности.

Таблица 23. К формированию оптимального плана развития вуза

Название?	Вид научной деятельности							Коэффициент успешности продвижения
	Д	К	С	М	ПО	П	НИР	
Результаты вуза прошлого года	5	24	1250	40	52	15	1430	—
Целевые ориентиры вуза	10	30	1500	50	80	35	2000	—
Запланированные результаты вуза	5	26	1149	44	57	16	1573	—
Количественные эквиваленты уровней важности отдельных видов результатов научной деятельности вуза с соответствии со стратегией его развития	0.255	0.071	0.071	0.255	0.02	0.071	0.255	—
Относительные результаты вуза в прошлом году к целевым ориентирам	0.500	0.867	0.766	0.80	0.713	0.457	0.787	0.673
Относительные результаты целевого ориентира	1	1	1	1	1	1	1	1.000
Относительные запланированные результаты вуза к целевым ориентирам	0.500	0.867	0.787	0.880	0.713	0.457	0.787	0.715

Тогда общая трудоемкость результатов i-го института в предшествующем году T_i составляет

$$T_i = \sum_{j=1}^{m} a_{ij} t_{ij}, \quad i = \overline{1,n}.$$

Степень мотивированности выполняемых i-м институтом научных исследований определим как средневзвешенную трудоемкость их выполнения M_i

$$M_i = \sum_{j=1}^{m} a_{ij} t_{ij} k_{ij}, \quad i = \overline{1,n}.$$

Соответствующие показатели для вуза в целом зададим через T и M:

$$T = \sum_{i=1}^{n} T_i,$$

$$M = \sum_{i=1}^{n} M_i.$$

Аналогичные переменные и показатели для планируемого года будем обозначать чертой над символом соответствующей переменной или показателя.

Тогда управляемыми переменными в обсуждаемой оптимизационной задаче линейного программирования будут целочисленные переменные:

$$\bar{a}_{ij}, \quad i = \overline{1,n}, \quad j = \overline{1,m}.$$

Их значения могут изменяться в определенных не очень широких относительных пределах, отклоняясь от соответствующих значений в предшествующем году:

$$a_{ij}(1 - \text{»}) \le \bar{a}_{ij} \le a_{ij}(1 + \text{»}), \quad i = \overline{1,n}, \quad j = \overline{1,m}. \tag{5.1}$$

Однако трудовой потенциал институтов, по условию задачи, из года в год должен оставаться неизменным, поэтому

$$\overline{T}_i = T_i, \quad i = \overline{1,n}.$$

Критерием оптимальности будет степень мотивируемости планируемых вузом работ:

$$\overline{M} \to \max.$$

Эта задача легко решается с использованием надстройки EXCEL "Поиск решения".

Проведем расчета по указанной модели без учета ограничений (5.1) (так называемый "Предельный вариант") для того, чтобы определить максимально возможную скорость приближения вуза к целевому ориентиру. Результаты показаны в табл. 24. Конечно, он не может быть реализован, так как предусматривает полный слом сложившейся в вузе структуры научной деятельности, когда любой член коллектива может в рамках, отведенных на это нормативных часов хоть защищать диссертацию, хоть писать статьи, хоть вести договорные работы с предприятиями. Однако он представляет интерес, поскольку показывает предельно возможную скорость движения вуза к целевым ориентирам:

$$(0.715–0.673) / (1–0.673) = 0.042 / 0.327 = 0.128,$$

т.е. примерно за год на 1/8 часть оставшегося пути до достижения целевого ориентира.

Таблица 24. Оптимальная структура планируемых результатов научной деятельности институтов вуза в "Предельном варианте"

Институт	Д	К	С	М	ПО	П	НИР	Трудовой потенциал институтов
И1	0	0	112	0	0	0	1569	31 450
И2	1	0	289	0	0	0	0	20 650

Окончание таблицы 24

Институт	Д	К	С	М	ПО	П	НИР	Трудовой потенциал институтов
И3	0	0	404	0	0	0	0	28 300
И4	0	0	334	0	0	0	4	23 440
И5	4	26	42	44	57	16	0	33 050
Вуз в целом	5	26	1149	44	57	16	1573	136 890

Построим оптимальный годичный план научной деятельности вуза и его институтов, вернув предусмотренное в математической модели ограничение (5.1) на возможность изменения структуры деятельности коллективы по сравнению с предыдущим годом не более чем на некоторую относительную величину. Для примера в табл. 25 приведены результаты расчета оптимального плана научной деятельности при возможном отклонении в 15%. При этом из варьируемых моделью параметров исключены планируемые защиты диссертаций. Предполагается, что они заданы самими институтами на базе традиционного содержательного анализа своих возможностей. Такое исключение вызвано тем, что представленная выше математическая оптимизационная модель является весьма простой и служит лишь для пояснения основных идей использования метода МУС. Ясно, что при реальном воплощении методика оптимального планирования и соответствующая математическая модель будут значительно более совершенными.

В табл. 26 показано изменение показателя эффективности развития научной деятельности вуза при различном допустимом изменении структуры деятельности институтов.

Видно, что с изменением структуры деятельности институтов для ускоренного развития вуза достаточно быстро наступает насыщение, и потому незначительное ускорение продвижения к целевому ориентиру не искупает неизбежной напряженности в работе сотрудников и повышает возможность срыва плана. В рассматриваемом примере вполне достаточно менее чем 15%-ного изменения структуры деятельности институтов для того, чтобы добиться практически максимально быстрого развития вуза.

Таблица 25. Оптимальная структура планируемых результатов научной деятельности институтов вуза при допустимых изменениях в структуре научных результатов в пределах 15% сравнительно с прошлым годом

Люди / Критерий	Д	К	С	М	ПО	П	НИР	Коэффициент успешности продвижения вуза к целевому ориентиру	Коэффициент мотивированности труда коллектива
1	2	3	4	5	6	7	8	9	10
П1	0	3	202	8	11	1	839	—	0.668
П2	1	7	122	10	17	5	299		0.424
П3	2	7	239	12	17	6	185		0.820
П4	1	4	226	10	7	2	115		0.837
П5	2	5	388	4	5	1	135		0.946
По оптимальному плану с допустимым изменением структуры в 15%	6	26	1177	44	57	15	1573	0.740	0.759

П р и м е р 5 (оптимальное планирование результатов научной деятельности институтов и вуза с учетом общей удовлетворенности коллективов своим трудом). Удовлетворенность работника трудом – одно из ключевых понятий экономики труда. Под этим понимают психологическое и моральное удовлетворение, испытываемое человеком в процессе трудовой деятельности. Различают общую (работа в целом) и частичную удовлетворенность трудом (условия

Таблица 26. Оптимальная структура планируемых результатов научной деятельности вуза при различных допустимых изменениях в структуре научной деятельности институтов сравнительно с прошлым годом

Условия планирования	Д	К	С	М	ПО	П	НИР	Коэффициент успешности продвижения вуза к целевому ориентиру
Результаты прошлого года	5	24	1250	40	52	15	1430	0.673
Изменения в 5%	6	26	1220	40	52	15	1500	0.711
Изменения в 10%	6	26	1195	41	55	15	1573	0.725
Изменения в 15%	6	26	1177	44	57	15	1573	0.740
Изменения в 20%	6	26	1177	44	57	16	1573	0.742
Изменения в 25%	6	26	1177	44	57	16	1573	0.742

труда, оплата труда, возможности профессионального роста и др.). В настоящей статье мы будем рассматривать общую удовлетворенность, в дальнейшем этого не оговаривая.

Степень удовлетворенности научного работника различными видами научной деятельности (написание статей, подготовка диссертаций, изобретательская деятельность) невозможно выразить количественно, однако можно описать в терминах адекватной порядковой шкалы, например:

<1 – тружусь неохотно, 2 – успешно справляюсь, 3 – работаю с увлечением>.

Весьма полезно для планирования научной деятельности сотрудников вуза предложить им оценить по такой порядковой шкале степень своей удовлетворенности различными видами научной деятельности, а затем на этой основе сформировать спектр удовлетворенности трудом для коллективов, объединенных близким по профессиональной направленности научного направления видом деятельности. Индивидуальные особенности этого направления в сочетании с индивидуальными склонностями, особенностями подготовки и интересами каждого члена такого сравнительно небольшого коллектива во многом однородны и могут характеризоваться специфической структурой коллективной удовлетворенности различными видами научной деятельности.

В табл. 27 представлены средние предпочтения коллективов институтов к различным видам научной деятельности, оцененные в трехуровневой порядковой шкале:

<1 – тружусь неохотно, 2 – успешно справляюсь, 3 – работаю с увлечением>.

В табл. 28 они пересчитаны в их количественные эквиваленты с использованием МУС. В последнем столбце табл. 27 приведена структура возможностей, необходимая для определения количественных эквивалентов:

количество критериев = количество В1, количество критериев В2, количество критериев В3.

Таблица 27. Спектр удовлетворенности коллективов институтов вуза различными видами научной деятельности (1 — тружусь неохотно, 2 — успешно справляюсь, 3 — работаю с увлечением)

Институт	Вид научной деятельности							Структура возможностей
	Д	К	С	М	ПО	П	НИР	
1	2	3	4	5	6	7	8	
И1	1	1	2	2	2	1	3	7 = 3, 3, 1
И2	1	2	1	2	2	3	3	7 = 2, 3, 2
И3	1	3	3	2	2	1	2	7 = 2, 3, 2
И4	3	3	3	2	3	1	1	7 = 2, 1, 4
И5	1	3	3	3	2	1	1	7 = 3, 1, 3

Таблица 28. Количественные эквиваленты удовлетворенности различными видами научной деятельности институтов вуза

Институт	Вид научной деятельности						
	Д	К	С	М	ПО	П	НИР
И1	0.123	0.123	0.494	0.494	0.494	0.123	1
И2	0.107	0.375	0.107	0.375	0.375	1	1
И3	0.107	1	1	0.375	0.375	0.107	0.375
И4	1	1	1	0.298	1	0.147	0.147
И5	0.169	1	1	1	0.412	0.169	0.169

Однако, как уже указывалось, количественные эквиваленты порядковой шкалы нормированы условием, что их сумма должна равняться единице. Такая нормировка применяется, если порядковая шкала устанавливает сравнительную важность объектов, которые в совокупности выполняют некоторую общею функцию, и каждый уровень шкалы указывает на соответствующий ему вклад объекта в выполнение этой общей функции. Применительно же к количественной оценке степени удовлетворенности трудом требуется использовать иной вид нормировки, при котором за единицу принимается количественный эквивалент высшего уровня порядковой шкалы, с ней сравниваются остальные шкалы, поэтому количественные эквиваленты других уровней отражает соответствующую долю единицы. Перейти от первого вида нормировки ко второму весьма просто: поделить значение количественного эквивалента в первом виде нормировки на значение количественного эквивалента, отвечающего при первом виде нормировки наибольшему уровню порядковой шкалы. Результаты такого пересчета показаны в табл. 28.

На основе столбцов 2–8 табл. 22 и 28 легко рассчитать данные, приведенные в столбцах 9 и 10 табл. 29, а именно степень удовлетворенности коллективов институтов и вуза в целом выполненными научными исследованиями в течение прошлого года. Столбец 2 табл. 29 дублирует столбец 9 табл. 22, т.е. трудовой потенциал институтов и вуза в целом при выполнении научных исследований прошлого года (он же, по условию задачи, сохраняется и в планируемом году). Столбец 3 табл. 29 представляет собой сумму произведений времени, затраченного на выполнение НИР по определенному направлению на соответствующий коэффициент из табл. 28, отражающей степень мотивированности коллектива в выполнении НИР по этому направлению. В столбце 4 табл. 29 представлено отношение соответствующих чисел из столбцов 2 и 4, т.е. взвешенный с учетом удовлетворенности трудом человеческий потенциал институтов и вуза при рассмотренном выше оптимальном плане с ограничением на изменение структуры труда в пределах 15%. Этот же результат приведен в последнем столбце табл. 25, представляющей собой разработанный оптимальный план научной деятельности вуза.

Таблица 29. Расчет трудового потенциала институтов и его учет при сравнительной оценке эффективности их научной деятельности

Институт	Трудовой потенциал, суммарная трудоемкость, усл.ч/год	Трудовой потенциал, взвешенный с учетом удовлетворенности трудом коллективов институтов и вуза	Коэффициент мотивированности труда коллективов институтов и вуза при оптимальном плане с ограничением 15%
1	2	3	4
И1	31 450	20 749	0.668
И2	20 650	7758	0.424
И3	28 300	22 386	0.820
И4	23 440	19 335	0.837
И5	33 050	30 932	0.946
Вуз в целом	136 890	101 159	0.759

Для информации в конце статьи приводятся табл. 30, 31, в которых, согласно МУС, рассчитаны количественные эквиваленты групп сравнительной важности частных критериев при различном их распределении по двум и трем группам важности (В1 – обычная важность, В2 – значительная важность, В3 – наибольшая важность).

Таблица 30. Количественные эквиваленты групп сравнительной важности частных критериев при различном их распределении по двум группам важности (В1 — обычная важность, В2 — значительная важность)

Количество критериев	Распределение критериев по группам важности		Количественные эквиваленты групп важности критериев		Количество критериев	Распределение критериев по группам важности		Количественные эквиваленты групп важности критериев	
	В1	В2	В1	В2	*n*	В1	В2	В1	В2
2	1	1	0.250	0.750	8	1	7	0.016	0.141
3	1	2	0.111	0.444	8	2	6	0.025	0.158
3	2	1	0.194	0.611	8	3	5	0.032	0.181
4	1	3	0.063	0.313	8	4	4	0.042	0.208
4	2	2	0.104	0.396	8	5	3	0.057	0.239
4	3	1	0.160	0.521	8	6	2	0.075	0.275
5	1	4	0.040	0.240	8	7	1	0.098	0.315
5	2	3	0.065	0.290	9	1	8	0.012	0.123
5	3	2	0.093	0.360	9	2	7	0.019	0.137
5	4	1	0.138	0.450	9	3	6	0.025	0.154
6	1	5	0.028	0.194	9	4	5	0.032	0.174
6	2	4	0.044	0.228	9	5	4	0.042	0.198
6	3	3	0.061	0.272	9	6	3	0.055	0.224
6	4	2	0.086	0.328	9	7	2	0.071	0.253
6	5	1	0.121	0.394	9	8	1	0.089	0.285
7	1	6	0.020	0.163	10	1	9	0.010	0.110
7	2	5	0.032	0.187	10	2	8	0.016	0.121
7	3	4	0.043	0.218	10	3	7	0.020	0.134
7	4	3	0.059	0.255	10	4	6	0.025	0.150
7	5	2	0.080	0.299	10	5	5	0.032	0.168
7	6	1	0.108	0.350	10	6	4	0.041	0.188
					10	7	3	0.053	0.210
					10	8	2	0.066	0.234
					10	9	1	0.082	0.261

Таблица 31. Количественные эквиваленты групп сравнительной важности частных критериев при различном их распределении по трем группам важности (В1 — обычная важность, В2 — значительная важность, В3 — наибольшая важность)

Количество критериев	Распределение критериев по группам важности			Количественные эквиваленты групп важности критериев			Количество критериев	Распределение критериев по группам важности			Количественные эквиваленты групп важности критериев		
	В1	В2	В3	В1	В2	В3	*n*	В1	В2	В3	В1	В2	В3
3	1	1	1	0.111	0.278	0.611	9	7	1	1	0.071	0.221	0.285

Окончание таблицы 31

Количество критериев	Распределение критериев по группам важности			Количественные эквиваленты групп важности критериев			Количество критериев	Распределение критериев по группам важности			Количественные эквиваленты групп важности критериев		
4	2	1	1	0.104	0.271	0.521	9	6	2	1	0.055	0.193	0.285
4	1	2	1	0.063	0.208	0.521	9	5	3	1	0.042	0.168	0.285
4	1	1	2	0.063	0.146	0.396	9	4	4	1	0.032	0.147	0.285
5	3	1	1	0.093	0.27	0.45	9	3	5	1	0.025	0.128	0.285
5	2	2	1	0.065	0.21	0.45	9	2	6	1	0.019	0.113	0.285
5	1	3	1	0.04	0.17	0.45	9	1	7	1	0.012	0.1	0.285
5	2	1	2	0.065	0.15	0.36	9	6	1	2	0.055	0.165	0.253
5	1	2	2	0.04	0.12	0.36	9	5	2	2	0.042	0.142	0.253
5	1	1	3	0.04	0.09	0.29	9	4	3	2	0.032	0.122	0.253
6	4	1	1	0.086	0.261	0.394	9	3	4	2	0.025	0.105	0.253
6	3	2	1	0.061	0.211	0.394	9	2	5	2	0.019	0.091	0.253
6	2	3	1	0.044	0.172	0.394	9	1	6	2	0.012	0.08	0.253
6	1	4	1	0.028	0.144	0.394	9	5	1	3	0.042	0.119	0.224
6	3	1	2	0.061	0.161	0.328	9	4	2	3	0.032	0.1	0.224
6	2	2	2	0.044	0.128	0.328	9	3	3	3	0.025	0.085	0.224
6	1	3	2	0.028	0.106	0.328	9	2	4	3	0.019	0.073	0.224
6	2	1	3	0.044	0.094	0.272	9	1	5	3	0.012	0.063	0.224
6	1	2	3	0.028	0.078	0.272	9	4	1	4	0.032	0.082	0.198
6	1	1	4	0.028	0.061	0.228	9	3	2	4	0.025	0.068	0.198
7	5	1	1	0.08	0.248	0.35	9	2	3	4	0.019	0.057	0.198
7	4	2	1	0.059	0.207	0.35	9	1	4	4	0.012	0.049	0.198
7	3	3	1	0.043	0.173	0.35	9	3	1	5	0.025	0.054	0.174
7	2	4	1	0.032	0.146	0.35	9	2	2	5	0.019	0.045	0.174
7	1	5	1	0.02	0.126	0.35	9	1	3	5	0.012	0.039	0.174
7	4	1	2	0.059	0.167	0.299	9	2	1	6	0.019	0.035	0.154
7	3	2	2	0.043	0.136	0.299	9	1	2	6	0.012	0.031	0.154
7	2	3	2	0.032	0.112	0.299	9	1	1	7	0.012	0.026	0.137
7	1	4	2	0.02	0.095	0.299	10	8	1	1	0.066	0.208	0.261
7	3	1	3	0.043	0.105	0.255	10	7	2	1	0.053	0.184	0.261
7	2	2	3	0.032	0.085	0.255	10	6	3	1	0.041	0.163	0.261
7	1	3	3	0.02	0.071	0.255	10	5	4	1	0.032	0.144	0.261
7	2	1	4	0.032	0.065	0.218	10	4	5	1	0.025	0.128	0.261
7	1	2	4	0.02	0.054	0.218	10	3	6	1	0.02	0.113	0.261
7	1	1	5	0.02	0.044	0.187	10	2	7	1	0.016	0.101	0.261
8	6	1	1	0.075	0.234	0.315	10	1	8	1	0.01	0.091	0.261

Таким образом, рассмотрена процедуру планирования научной деятельности вуза, при которой благодаря использованию МУС и традиционных методов математического моделирования и многокритериальной оптимизации гармонично учитываются стратегия развития и целевые ориентиры вуза в сочетании со сложившейся структурой и мотивационными предпочтениями научного коллектива. Это обеспечит высокую эффективность и реализуемость разработанного плана.

Заключение. Применение в сфере образования современных количественных методов подготовки и принятия решений, в частности с использованием метода МУС, допускающего оценку предпочтений в порядковых шкалах, откроет новые возможности в повышении эффективности важных научно-образовательных процессов и реализации передовых образовательных концепций, например [14, 15]. При отборе наиболее достойных абитуриентов для продолжения образования в высшей школе это расширит полномочия вузов, позволит более точно учесть индивидуальные особенности и предпочтения школьников и их родителей, даст им большую возможность сознательно развивать широкий спектр наиболее выраженных индивидуальных задатков детей, зная, что это разнообразие повысит их возможность поступления в предпочитаемый ими вуз. При использовании в рамках целостной системы выявления и многолетнего развития творчески одаренной молодежи в сфере науки и техники применение рассмотренного в статье подхода позволит создать объективный измеритель (творческий рейтинг) для всесторонней оценки степени творческого роста каждого молодого исследователя и создать на этой основе научно управляемую систему поддержки его развития. В организации научной деятельности вуза появится целостная система оптимального планирования деятельности его институтов, исходящая из целевых ориентиров и общей стратегии развития вуза и в то же время учитывающая специфические особенности общей удовлетворенности трудом коллектива каждого института.

Простота освоения и применения метода МУС неподготовленными пользователями будет способствовать его широкому применению для реализации описанных выше и еще более широких перспектив.

СПИСОК ЛИТЕРАТУРЫ

1. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. ISBN 0-07-054371-2. McGraw-Hill, 1980.
2. *Саати Т.* Принятие решений. Метод анализа иерархий / Пер. с англ. Р.Г. Вачнадзе. М.: Радио и связь, 1993. 278 с.
3. *Саати Т.* Об измерении неосязаемого. Подход к относительным измерениям на основе главного собственного вектора матрицы парных сравнений // Cloud of Science. 2015. V. 2. No. 1. http://cloudofscience.ru
4. *Ларичев О.И.* Теория и методы принятия решений. М.: Логос, 2002. 392 с.
5. *Малышев В.В., Пиявский С.А.* Метод "уверенных суждений" при выборе многокритериальных решений // Изв. РАН. ТиСУ. 2015. № 5. С. 90—101.
6. *Пиявский С.А., Малышев В.В.* Новые методы принятия многокритериальных решений в цифровой среде. М.: Наука, 2022. 391 с.
7. *Пиявский С.А.* Как "нумеризовать" понятие "важнее" // Онтология проектирования. 2016. Т.6. № 4. С. 414—435.
8. *Пиявский С.А.* Формулы для вычисления универсальных коэффициентов при принятии многокритериальных решений // Онтологии проектирования. 2019. Т. 9. № 2. С. 282—298.
9. *Пиявский С.А., Кирюков С.Р., Кузнецов А.С.* Формирование творческих компетенций одаренной молодежи в телекоммуникационной развивающей научно-образовательной среде // Информатизация образования и науки. 2020. № 2 (46). С. 127—143.
10. *Кекух О.Л.* Организация творческой исследовательской деятельности учащихся как средство повышения качества образования (эл. Ресурс). URL: https://nsportal.ru/shkola/rodnoy-yazyk-i-literatura/library/2012/10/31/organizatsiya-tvorcheskoy-issledovatelskoy (дата обращения: 23.06.2023).
11. Концепция Единой Самарской областной системы мер по выявлению и развитию творчески одаренной молодежи в сфере науки, техники и технологий и инновационному развитию Самарской области. URL: https://samara.mgpu.ru/files/laboratorii/01/nauka/001.pdf. (дата обращения: 23.06.2023).
12. *Пиявский С.А.* Исследовательская деятельность студентов в инновационном вузе: учебник Самара: СГАСУ. 2011. 198 с.
13. Рабочая концепция одаренности / Под ред. Д.Б. Богоявленской, В.Д. Шадрикова. М.: Магистр, 1998. 68 с.
14. *Асмолов А.Г., Шехтер Е.Д., Черноризов А.М.* Антропологический поворот: восхождение к сложности. Человек как открытая целостность. Новосибирск: Институт философии Российской академии наук, 2022. С. 33—53. URL: https://elibrary.ru/download/elibrary_50164637_44009225.pdf (дата обращения: 23.06.2023).
15. *Берберян А.С., Корнилова О.А.* Экзистенциально-гуманистическая психология как фундаментальное основание развития личности и позитивного мышления // Методология современной психологии. 2021. № 13. С. 39—50.

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, с. 121–128

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

УДК 519.711.2

УПОРЯДОЧИВАНИЕ ГИПОТЕЗ В МОДЕЛЯХ ПЕРЕВОДА С ИСПОЛЬЗОВАНИЕМ ЧЕЛОВЕЧЕСКОЙ РАЗМЕТКИ

© 2024 г. К. В. Воронцов[a], *, Н. А. Скачков[a], **

[a] *ВЦ ФИЦ ИУ РАН, Москва, Россия*

**e-mail: vokov@forecsys.ru*

***e-mail: nikolaj-skachkov@ya.ru*

Поступила в редакцию 31.01.2024 г.

После доработки 23.04.2024 г.

Принята к публикации 22.07.2024 г.

Современные системы машинного перевода обучаются на больших объемах параллельных данных, полученных с помощью эвристических методов обхода интернета. Низкое качество этих данных приводит к систематическим ошибкам перевода, которые могут быть достаточно заметными для человека. Для борьбы с такими ошибками предлагается интегрирование человеческих оценок гипотез переводной модели в процесс обучения системы перевода. Показано, что использование человеческих разметок позволяет не только вырастить общее качество перевода, но и заметно снизить количество систематических ошибок перевода. Кроме того, относительная простота человеческой разметки и ее применения для улучшения качества модели открывает новые возможности в области доменной адаптации моделей перевода под новые домены, что удалось показать на примере переводов заголовков товаров из интернет-магазинов.

Ключевые слова: машинный перевод, нейронная сеть, стохастический градиентный спуск, контрастное обучение, дообучение модели, обучение с негативными примерами.

DOI: 10.31857/S0002338824040074 **EDN:** UEFMST

HYPOTHESES RE-RANKING IN TRANSLATION MODELS USING HUMAN MARKUP

K. V. Vorontsov[a], *, N. A. Skachkov[a], **

[a] *Moscow, CC FRC CSC RAS*

**vokov@forecsys.ru*

***nikolaj-skachkov@ya.ru*

Modern machine translation systems are trained on large volumes of parallel data obtained using heuristic methods of the Internet bypassing. The poor quality of the data leads to systematic translation errors, which can be quite noticeable from the human point of view. To fix such errors a human based models hypotheses re-ranking is introduced in this work. In this paper the use of human markup is shown not only to increase the overall quality of translation, but also to significantly reduce the number of systematic translation errors. In addition, the relative simplicity of human markup and its integration in the model training process opens up new opportunities in the field of domain adaptation of translation models for new domains like online retail.

Keywords: Machine translation, neural network, stochastic gradient descent, contrastive learning, model fine-tuning, negative examples.

Введение. Создание систем автоматического перевода является одной из сложных задач анализа текстов естественного языка. Обучение алгоритмов, лежащих в основе таких систем, требует большого количества параллельных текстов на разных языках и существенно зависит от качества этих данных и степени их выравненности. Ввиду высокой стоимости работы профессиональных переводчиков данные для обучения алгоритмов перевода собираются автоматически с помощью эвристических алгоритмов. При этом высокая степень выравненности

отдельных обучающих примеров не гарантируется, что позволяет собирать большие объемы параллельных текстов. [1]

На собранных параллельных данных обучаются нейросетевые модели [2, 3], которые учатся восстанавливать каждое слово перевода при условии входного текста и предыдущих слов. Обученные таким образом модели, при должном размере модели, способны хорошо восстанавливать языковые закономерности и генерировать достаточно гладкие с человеческой точки зрения переводы. Однако качество перевода естественным образом связано с объемом и качеством собранных параллельных данных. Так из-за невозможности находить в интернете больших объемов хорошо выравненных переводов модели перевода страдают такими типовыми ошибками, как недостаточные переводы [4].

Для борьбы с типовыми ошибками существует способ интеграции негативных примеров в процесс обучения [4]. При использовании обучения с негативными примерами возникает возможность передать сети информацию, какие переводы являются неприемлемыми с точки зрения качества. В работе [4] предлагалось портить переводы из обучающего корпуса с помощью выкидывания случайных слов. В результате такого обучения модель перевода становится более внимательной к информации, содержащейся во входном предложении, и количество ошибок с пропуском слов в переводе уменьшается. Однако генерация негативных примеров по простым шаблонам несет в себе некоторые ограничения. Так, негативные примеры, полученные с помощью генерации по шаблону, могут оказаться слишком простыми и не покрывать все возможные случаи внутри одного класса ошибок. Кроме того, для каждого отдельного класса ошибок требуется описывать свой шаблон для генерации негативных примеров. Если для недостаточных переводов может подойти выкидывание случайных слов из правильного перевода, то для таких ошибок как неверное согласование текста с точки зрения языка выбрать правильный шаблон становится сложнее. Сама по себе необходимость придумывать шаблоны для различных видов ошибок делает процесс улучшения качества машинного перевода более трудоемким и менее масштабируемым.

К проблеме выбора негативных примеров можно подойти с другой стороны. Для каждого текста на входном языке можно получить несколько переводов модели, например с помощью процедуры разнообразного поиска в ширину [5] или сэмплирования с температурой. При этом в разнообразных переводах могут случайным образом содержаться или не содержаться типовые ошибки данной модели. Тогда, если выбрать лучший по качеству перевод среди имеющихся, то его можно использовать в качестве позитивного примера, а все остальные относительно него будут негативными. Каждый негативный пример, полученный таким образом, оказывается сложным с точки зрения модели, так как он оказался достаточно вероятным, чтобы она его сгенерировала. Соответственно такое обучение представляет из себя упорядочивание гипотез модели с точки зрения некой метрики качества.

Остается вопрос: как выбрать метрику качества для ранжирования сгенерированных моделью переводов? В работе представлен подход к обучению моделей машинного перевода с помощью упорядочивания гипотез модели на основе человеческих оценок. Показано, что применение данного подхода позволяет не только улучшить качество перевода, но и заметно снизить долю типовых ошибок перевода, которыми модель перевода изначально страдала. Кроме того, благодаря тому, что модель перевода в подходе не учится с нуля, дообучение с упорядочиванием гипотез не требует большого количества данных человеческой разметки.

Более того, обучение с упорядочиванием гипотез на основе человеческих оценок не требует наличия параллельных данных или эталонных переводов. Это открывает возможность для улучшения качества модели перевода на тех доменах, где нет данных для доменной адаптации. В статье исследуется применимость предложенной процедуры для улучшения качества переводов заголовков товаров из домена электронной коммерции. Этим текстам свойственна специфичная структура текстов и лексика, что является причиной их сложности для систем автоматического перевода.

1. Постановка задачи. Теперь рассмотрим вероятностные модели, лежащие в основе систем автоматического перевода текстов естественного языка, а также опишем подход с упорядочиванием гипотез перевода на базе человеческих оценок.

1.1. Постановка задачи машинного перевода. Перейдем к математической постановке задачи машинного перевода. Пусть задано множество параллельных данных, состоящее из пар текстов $\{(x_i. y_i)\}_{i=1}^{N}$. Тексты x_i написаны на языке входа, а тексты y_i — на целевом языке и являются переводами соответствующих текстов x_i. Тогда обучение переводной модели будет заключаться в максимизации правдоподобия переводов при условии входных текстов

для всех объектов выборки. Основываясь на методе максимального правдоподобия и переходя к логарифму правдоподобия, получим

$$\sum_{i=1}^{N} \log P_\theta\left(y_i \mid x_i\right) \to \max_\theta, \tag{1.1}$$

где θ — параметры обучаемой переводной модели, а $P_\theta(\cdot)$ — функция правдоподобия модели. Обученная с помощью максимизации правдоподобия (1.1) модель умеет оценивать вероятность перевод y для входного x для любого y. Однако для модели перевода этого недостаточно, так как она должна быть способна генерировать переводы, а не только оценивать их вероятность. Для генерации перевода с помощью такой оценивающей модели необходимо перебрать все возможные тексты на выходном языке и, оценив каждый из них с помощью модели, выбрать наиболее вероятный перевод. Так как число текстов на выходном языке бесконечно, генерация с помощью такой модели невозможна. Для решения этой проблемы используют авторегрессионные модели перевода [2]. В данном подходе вводится дополнительное ограничение, что t-е слово перевода зависит только от предыдущих слов. Правдоподобие модели перевода с этим ограничением записывается следующим образом:

$$\log P_\theta\left(y \mid x\right) = \sum_{t=1}^{|y|} \log P_\theta\left(y^t \mid y^{<t}, x\right), \tag{1.2}$$

где y^t — t-е слово перевода, $y^{<t}$ – префикс перевода для t-го слова, а $P_\theta(y^t \mid y^{<t}, x)$ — моделируемая вероятность t-го слова перевода при условии префикса и входного текста.

Для обученной с авторегрессионной функцией потерь (1.2) модели генерация перевода может осуществляться при помощи выбора наиболее вероятного слова y^t.

1.2. К о н т р а с т н о е о б у ч е н и е с н е г а т и в н ы м и п р и м е р а м и. Для борьбы с систематическими ошибками перевода, которые возникают из-за низкого качества выравнивания в параллельных данных, в работе [4] предложен подход по обучению с негативными примерами. Для генерации негативных примеров, т.е. заведомо неправильных переводов, переводы портятся по некоторому шаблону. Так, для параллельной пары (x, y) строится испорченный перевод $y\_$, полученный из y с помощью шаблона t. Пример y далее будем называть положительным примером, а испорченный перевод $y\_$ — отрицательным. Для борьбы с ошибками пропуска слов в переводе авторы предлагают выбрасывать случайные слова в переводе. Тогда шаблон t можно описать следующим образом:

$$t(y) = y_1 .. y_{t-1} y_{t+1} \cdots y_{|y|}, \quad t \sim \mathcal{U}\left[1, \ldots, |y|\right],$$

где y состоит из слов $y_1 \ldots y_{|y|}$, $|y|$ — длина текста, а t выбирается случайно из целых чисел от 1 до $|y|$.

Для интеграции в процесс обучения негативных примеров авторы после обучения с авторегрессионной функции потерь (1.2) использовали дообучение модели с контрастной функцией потерь:

$$L_\alpha\left(x, y, y\_, \theta\right) = \max\left(0, \log P_\theta\left(y\_ \mid x\right) - \log P_0\left(y \mid x\right) + \alpha\right), \quad y\_ = t\left(y\right), \tag{1.3}$$

при которой увеличивается вероятность положительного перевода y текста x относительного более плохого перевода $y\_$. Таким образом, при обучении с контрастной функцией потерь (1.3) модель учится правильно упорядочивать положительный и отрицательный примеры. Для тех обучающих примеров, у которых положительный пример более вероятен, чем отрицательный на значение, превышающее отступ α, контрастная функция потерь (1.3) обращается в ноль и эти примеры не участвуют в обучении.

Можно заметить, что при дообучении с контрастной функцией потерь (1.3) модель может «забыть» задачу авторегрессионной генерации и, следовательно, потерять возможность итеративно генерировать перевод. Это объясняется тем, что контрастная функция потерь (1.3) действует только на уровне предложения и воздействие на пословные предсказания $P_\theta(y_t \mid y_{<t}, x)$, из которых складывается вероятность всего перевода, может быть достаточно шумным и непредсказуемым. Для исключения возможности возникновения такой проблемы в данной работе дообучение происходит с функцией потерь, являющейся линейной комбинацией авторегрессионной функции потерь (1.2) и контрастной функции потерь (1.3):

$$L_{\alpha,\beta} = \beta \log P_\theta\left(y \mid x\right) + \max\left(0, \log P_\theta\left(y\_ \mid x\right) - \log P_\theta\left(y \mid x\right) + \alpha\right). \tag{1.4}$$

Подбор параметров α и β описан в разд. 2.4, где будет показано, что дообучение только на контрастную функцию потерь (1.3) действительно приводит к более плохому результату, чем

обучение на линейную комбинацию (1.4). Далее обучение с функцией потерь (1.4) будем называть контрастным обучением.

1.3. Выбор лучшего перевода с помощью человеческой разметки. Вместо генерации негативных примеров по шаблону в работе предлагается генерировать несколько переводов из одной модели с помощью разнообразного поиска в ширину [5] и упорядочивать их по качеству с помощью человеческой разметки. Как уже было описано, таким образом модель получает в качестве негативных примеров достаточно вероятные с точки зрения этой же модели переводы. Благодаря этому факту полученные негативные примеры по построению будут достаточно сложными для модели.

Для нахождения лучшего перевода среди двух переводов модели необходимо разработать инструкцию и шаблон разметки, чтобы унифицировать представления о задании у размечающих. В используемом задании предлагалось для представленного текста на входном языке выбрать один из двух переводов в качестве лучшего либо указать, что представленные переводы одинакового качества. Кроме того, для упрощения разметки различия в переводах подсвечивались. Пример задания разметки можно увидеть на рисунке. В инструкции к заданию указывались различные критерии оценки, такие, как точность передаваемого смысла, грамматическая корректность перевода, правильность расстановки пунктуации и выбора капитализации. Указанные категории ошибок в инструкции не упорядочивались по грубости, а лишь указывались в качестве напоминания. Это было сделано для того, чтобы оставить свободу выбора более грубой категории ошибок самому размечающему. Результатом такой человеческой разметки считаем упорядоченные пары переводов исходного текста, в каждой паре размечающий указывает, какой перевод лучше.

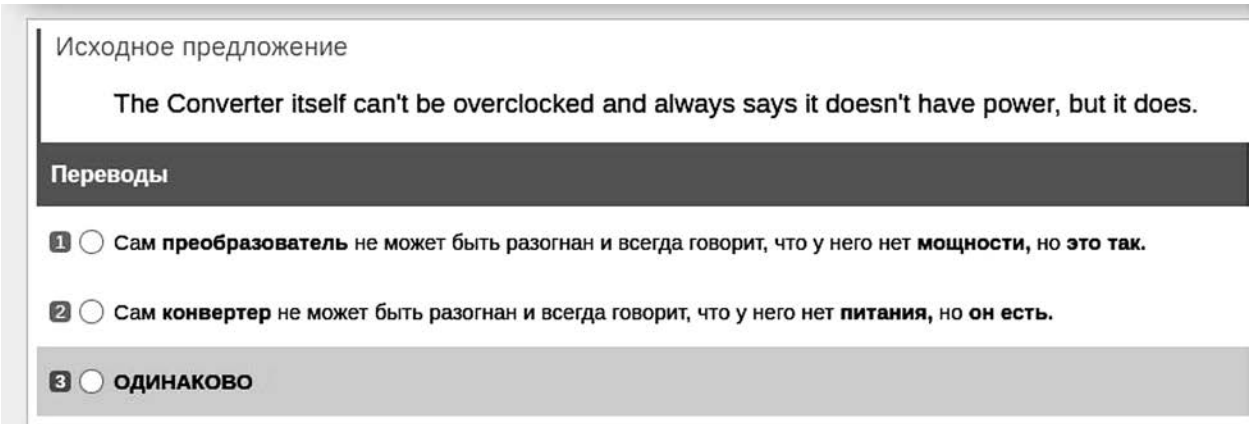

Рис. 1. Пример задания для человеческих разметчиков

Для уменьшения количества некачественной разметки был составлен экзамен для допуска к разметке. Во время разметки задания размечающим периодически показывались вопросы с заготовленным ответом. При ошибках на таких вопросах размечающие не допускались к продолжению разметки. Последнее помогало бороться с усталостью размечающих в процессе оценки переводов и потерей внимательности.

1.4. Оценка качества перевода. Для оценки качества переводов дообученных моделей будем использовать автоматическую метрику BLEU [6]. Эта метрика требует наличия эталонных переводов для тестового корпуса и показывает достаточно высокую корреляцию с оценками людей. BLEU рассчитывается на основе пересечения n-грамм в автоматическом и эталонном переводах одного предложения. Для каждой n-граммы длины n рассчитывается P_n – отношение частоты n-граммы среди всех кандидатов к частоте n-граммы среди всех эталонных переводов. При этом частота в кандидате ограничивается значением в эталоне, чтобы отношение частот было ограничено сверху единицей:

$$P_n = \frac{\sum\limits_{\text{n-gram}\in C} \text{Count}_{\text{clip}}\left(\text{n-gram}\right)}{\sum\limits_{\text{n-gram}'\in C'} \text{Count}\left(\text{n-gram}'\right)},$$

где C — переводы тестового корпуса моделью, C' — эталонные переводы тестового корпуса, а $\text{Count}_{\text{clip}}$ — частота n-граммы в переводе, ограниченное частотой этой же n-граммы в эталоне. Далее сама метрика BLEU считается как геометрическое среднее значений P_n, умноженное на константу BP (brevity penalty). Константа BP и нормализация в формуле BLEU при этом выбираются эмпирически:

$$\text{BP} = \min\left(\exp(1 - r/c), 1\right),$$

$$\text{BLEU} = \text{BP} \sum_{n=1}^{4} \frac{1}{n} \log P_n,$$

где r — суммарная длина эталонных переводов тестового корпуса; c' — суммарная длина переводов тестового корпуса моделью. Умножение на константу BP предлагается авторами метрики для уменьшения штрафа для более коротких предложений. Это необходимо из-за того, что более длинные переводы в среднем содержат больше случайных пересечений по n-граммам с эталонными текстами.

При проведении экспериментов метрика BLEU вычисляется с эталонными переводами, подготовленными профессиональными переводчиками. Для экспериментов использовались тестовые корпуса для русско-английского направления, подготовленные к конференции WMT- 2019 [7]. Размер тестового корпуса составляет 3000 предложений, исходные предложения выбирались из новостных статей.

Кроме автоматической метрики BLEU в данной работе для сравнения обученных моделей перевода будет применяться человеческая разметка на подобии описанной в разд. 1.3 (рисунок). Чтобы снизить эффект от переобучения под предпочтения размечающих, в оценке переводов будут использоваться только разметчики, не участвовавшие в разметке обучающих данных. Оценка на основе людей необходима для того, чтобы количественно оценить изменения в модели, а также получить возможность оценивать качество перевода текстов, не имеющих эталонных переводов.

Кроме сравнения переводов различных моделей, большой интерес представляет то, как предложенная процедура дообучения с упорядочиванием гипотез на основе человеческой разметки уменьшает долю систематических ошибок перевода. Для оценки этого эффекта часть переводов показывалась размечающему с вопросом «является ли перевод недостаточным».

2. Эксперименты. Перейдем к описанию экспериментов и условий их проведения. Кроме того, приведем основные результаты, полученные при контрастном дообучении модели на человеческую разметку (1.4).

2.1. Архитектура модели. Обучаемые в данной работе модели перевода основаны на архитектуре Transformer [3]. Данная архитектура подразумевает, что модель состоит из кодировщика и декодировщика, каждый из которых в свою очередь состоит из нескольких блоков одинаковой структуры, выполняющихся последовательно.

Каждый блок имеет свой набор параметров и применяется к векторным представлениям слов предложения, полученных от предыдущего блока, и возвращает новые обогащенные векторные представления слов. Внутри самого блока происходит обогащение векторного представления контекстной информацией с помощью механизма внимания, а также к векторному представлению применяется нелинейное преобразование. В блоках декодировщика кроме контекстной информации векторные представления обогащаются еще и информацией о входном предложении с помощью механизма внимания на выход кодировщика. При этом контекстная информация в декодировщике ограничивается только левым контекстом. Такая архитектура позволяет осуществлять итеративную генерацию для моделей, обученных с авторегрессионной функцией потерь (1.2).

2.2. Условия экспериментов. В экспериментах проводится дообучение модели с различными функциями потерь и с помощью размеченных человеком переводов. В качестве предобученной модели в экспериментах использовалась модель из библиотеки fairseq, являющаяся победителем на направлении с русского на английский WMT-2019 [7]. Модель соответствует архитектуре Transformer-big и обладает размерностью векторных представлений:

1024 для внутренних представлений,

4096 для представлений внутри FFN,

16 голов внимания.

Дообучение модели производилось с помощью оптимизатора Adam [8] с нагревом в течение 1000 шагов и охлаждением по sqrt-расписанию в течение оставшихся 4000 шагов. Обучались модели на сервере с 4 GPU Tesla M40.

2.3. Данные для обучения. Эксперименты, как уже было описано, проводятся на направлении с русского языка на английский. Предобученная модель обучалась на данных, предоставленных для соревнования WMT-2019 [7], которые состоят из данных Paracrawl v3, Common Crawl, News Commentary и других корпусов. Для дообучения используются данные разметки, осуществленной на основе переводов текстов из датасета News Commentary.

Для разметки выбраны случайно 10000 текстов из указанного обучающего набора. Для каждого из них составлены по два перевода предобученной модели с помощью процедуры разнообразного поиска в ширину [5]. Далее тексты, у которы длина входного текста и перевода не превышает три слова, а также перевод содержит больше половины английских слов, отфильтровываются. Из оставшихся переводов выбирается лучший на основе процедуры, указанной в разд. 1.3.

Из полученных данных разметки выбираются только те примеры, где размечающие выбрали какой-либо перевод в качестве лучшего. Оказалось, что из всех обучающих примеров в 33% случаев по оценке человека перевод, который был менее вероятен с точки зрения модели, оказывался лучше, чем более вероятный. В 15% случаев с точки зрения размечающих качество оказывалось одинаковым. Эти примеры не используются для дообучения моделей, так как их не удалось упорядочить по качеству.

2.4. Эксперименты с контрастной функцией потерь. Для выбора гиперпараметров отступа α и веса β в контрастной функции потерь с человеческой разметкой (1.4), а также гиперпараметра отступа α в контрастном обучении с сгенерированными по шаблону негативными примерами (1.3) были обучены модели с перебором гиперпараметров по сеткам. При отборе моделей осуществлялся выбор лучшей конфигурации по метрике BLEU на датасете WMT-17 с более старого соревнования по машинному переводу. Отбор моделей по WMT-19 не проводился, чтобы избежать переобучения под тестовую выборку.

Для контрастного обучения с человеческой разметкой (1.4) оптимальные значения гиперпараметров оказались $\alpha = 0.3$, $\beta = 0.1$. Для обучения с негативными примерами, сгенерированными по шаблону (1.3), оптимальное значение α оказалось равным 1.0.

3. Сравнение подходов. Теперь обучим модели, представляющие описанные подходы к обучению и дообучению машинного перевода, и оценим их качество. Наибольший интерес представляют следующие модели:

базовая – модель, взятая из библиотеки fairseq, которая дообучалась в остальных экспериментах;

дообученная на параллельные данные – модель, которая училась столько же шагов, сколько и остальные дообученные модели с авторегрессионной функцией потерь (1.2);

дообученная на победителя разметки – модель, которая дообучалась с авторегрессионной функцией потерь (1.2) на тот перевод, который победил в разметке;

дообученная с шаблонными негативными примерами – модель, которая дообучалась с контрастной функцией потерь (1.3) с негативными примерами, которые генерировались по шаблону;

дообученная на человеческую разметку – модель, которая обучалась с контрастной функцией потерь (1.4) на упорядоченные по качеству с помощью человеческой разметки переводы.

Результаты оценки обученных моделей по BLEU на тестовом наборе WMT-19 можно увидеть в табл. 1. Как можно заметить, что обе модели, которые дообучались на данные, полученные с помощью человеческой разметки на качество, имеют значительный прирост на тестовом наборе. Причем модель, которая дообучалась с контрастной функцией потерь (1.4), превосходит по BLEU модель, обучавшуюся только на победителя без негативных примеров. Отдельный интерес представляет вопрос: насколько обучение с человеческой разметкой помогает справиться с систематическими ошибками перевода? Размечая переводы моделей на предмет того, является ли перевод недостаточным, т.е. в нем отсутствуют какие-то части, представленные во входном предложении, удалось выяснить, что дообучение на человеческую разметку наиболее полно решает данную проблему. Это можно объяснить тем, что негативные примеры, полученные из самой же модели, оказываются существенно более сложными, чем негативные примеры, сгенерированные по шаблону. Так, доля недостаточных переводов при контрастном дообучении с человеческой разметкой (1.4) падает с 4 до 1%, тогда как дообучение с шаблонными негативными примерами (1.3) понижает долю таких ошибок лишь до 2%.

Рассмотрим теперь, что происходит с качеством перевода с точки зрения человеческих оценок. Для этого разметим и сравним переводы разных моделей с помощью разметки, описанной в разд. 1.3. При этом размечающие выбираются только те, кто не участвовал в разметке обучающих данных. В итоге получается, что контрастное дообучение на человеческую разметку (1.4) улучшает перевод в 15% случаев, тогда как обучение с шаблонными негативами улучшает перевод лишь в 3% случаев. Примеры, где дообучение с человеческой разметкой улучшает перевод с точки зрения размечающих, можно увидеть в табл. 2.

Таблица 1. Сравнение дообученных моделей на тестовом наборе WMT19 на направлении с английского на русский

Модель	Функция потерь	BLEU	Недостаточные переводы, %
Базовая	Авторегрессионная (1.2)	35.6	4
Дообучение на переводные данные	Авторегрессионная (1.2)	35.7	–
Дообучение на победителя разметки	Авторегрессионная (1.2)	36.7	–
Дообучение с шаблонными негативами	Контрастная (1.3)	35.8	2
Дообучение на разметку	Контрастная (1.4)	**37.3**	1

Таблица 2. Примеры переводов моделей.

Вход:	а в одиночку как-то скучно
Базовая:	and alone as it is boring
Дообученная:	and it’s kind of boring alone
Вход:	все время загорается красным
Базовая:	it’s always red
Дообученная:	it lights up red all the time
Вход:	у меня немного опыта путешествий
Базовая:	I don’t have much experience
Дообученная:	I have a little experience in traveling

Входом обозначены тексты, подаваемые на вход моделям. Базовой называется модель до дообучения на разметку (1.4). Дообученной обозначена модель после контрастного дообучения на человеческую разметку (1.4)

3.1. Эксперименты с доменной адаптацией. Рассмотрим теперь применимость подхода с контрастным обучением на человеческую разметку (1.4) к задаче доменной адаптации. Как уже было сказано, при использовании человеческой разметки нет необходимости в параллельных данных, что открывает возможности по улучшению качества перевода на тех доменах, где сложно найти параллельные данные достаточно хорошего качества.

Для экспериментов было выбрано направление перевода с английского на русский на домене заголовков товаров из интернет-магазинов. В качестве предобученной модели использовалась все так же обученная с авторегрессионной функцией потерь (1.2) модель перевода. Для доменной адаптации были размечены переводы моделью из 10000 заголовков с помощью процедуры описанной в разд. 1.3.

Стоит заметить, что заголовки товаров представляют из себя достаточно сложный домен для машинного перевода. Это связано с тем, что такие тексты обладают нестандартной структурой: в них зачастую отсутствует сказуемое, а также присутствует большое количество определений и перечислений. В табл. 3 можно увидеть пример того, насколько плохо справляется модель с переводом входного текста до процедуры дообучения.

Таблица 3. Пример переводов моделей при доменной адаптации.

Вход:	car temporary parking card luminous calling phone number cards with sucker plate
Базовая:	автомобильная временная парковочная карта, светящиеся карточки с номером телефона для звонков с присоской
Дообученная:	светящиеся карточки с номерами телефонов для временной парковки автомобилей с присоской

Входом обозначены тексты, подаваемые на вход моделям. Базовой называется модель до дообучения на разметку (1.4). Дообученной обозначена модель после контрастного дообучения на человеческую разметку (1.4)

Так как для данного домена отсутствуют качественные тестовые наборы, оценка качества проводилась с помощью разметки. По итогам разметки переводов моделями из 1000 заголовков товаров получилось, что качество после дообучения на контрастную функцию потерь с человеческой разметкой (1.4) выросло на 40% заголовков. В табл. 3 можно также увидеть то, как улучшился перевод заголовков после процедуры дообучения. В среднем стоит отметить, что после дообучения переводы стали более гладкими с точки зрения русского языка и части заголовков стали переводиться согласованно друг с другом.

Заключение. Проведено исследование по возможности использования человеческой разметки для улучшения качества машинного перевода. Удалось показать, что упорядочивание переводов модели по качеству с помощью разметки позволяет заметно усилить эффект от обучения перевода с негативными примерами. Благодаря сложности получаемых из разметки примеров модель после дообучения показывает более высокие результаты как по тестовым наборам, как и с точки зрения человеческих оценок. Также данная процедура толкает модель исправлять свои же систематические ошибки, доля таких ошибок, как недостаточные переводы, заметно падает. Более того, модель, обученная с негативными примерами, сгенерированными по шаблону специально для борьбы с данным типом ошибок, чаще допускает недостаточные переводы. Данный эффект объясняется тем, что шаблон не способен описать достаточно сложные негативные примеры и модель, учащаяся на разметке своих же ошибок, исправляет их лучше.

Кроме того, предложенная процедура открывает возможности для более эффективной доменной адаптации. Для тех доменов, где есть недостаток качественных параллельных данных, вместо привлечения профессиональных переводчиков появляется возможность улучшения качества с помощью разметки переводов модели. Так, на домене заголовков товаров из интернет-магазинов с помощью дообучения модели на разметку переводов удалось заметно поднять качество модели с точки зрения человеческих оценок.

СПИСОК ЛИТЕРАТУРЫ

1. *Bañón M., Chen P., Haddow B. et. al.* ParaCrawl: Web-Scale Acquisition of Parallel Corpora // Proc. 58th Annual Meeting of the Association for Computational Linguistics. Seattle, 2020. P. 4555–4567.
2. *Stahlberg F.* Neural Machine Translation: A Review // J. Artific. Intelligence Res. 2020. № 69. P. 343–418.
3. *Vaswani A., Shazeer N., Parmar N. et. al.* Attention is All You Need // Proc. 31st Intern. Conf. on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook. N.Y., 2017. P. 6000–6010.
4. *Yang Z., Cheng Y., Liu Y. et. al.* Reducing Word Omission Errors in Neural Machine Translation: A Contrastive Learning Approach // Proc. 57th Annual Meeting of the Association for Computational Linguistics. Florence, 2019. P. 6191–6196.
5. *Vijayakumar A.K., Cogswell M., Selvaraju R.R. et. al.* Diverse Beam Search: Decoding Diverse Solutions from Neural Sequence Models // ArXiv. 2016. abs/1610.02424.
6. *Papineni K., Roukos S., Ward T. et. al.* Bleu: a Method for Automatic Evaluation of Machine Translation // Proc. 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, 2002. P. 311–318.
7. *Barrault L., Bojar O.R., Costa-jussà M. et al.* Findings of the Conf. on Machine Translation (WMT19) // Proc. Fourth Conf. on Machine Translation. Florence, 2019. V. 2: Shared Task Papers.
8. *Kingma D.P., Ba J.* Adam: A Method for Stochastic Optimization // 3rd Intern. Conf. on Learning Representations (ICLR). San Diego, CA, 2015.

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, с. 129–138

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

УДК 004.93

ПОИСК ПОЧТИ ДУБЛИКАТОВ ИЗОБРАЖЕНИЙ РУКОПИСНЫХ ТЕКСТОВ ДЛЯ ВЫСОКОНАГРУЖЕННЫХ СЕРВИСОВ

© 2024 г. К. В. Варламова[a, b, *], М. С. Каприелова[a, b, c, **], И. О. Потяшин[a, b, *], Ю. В. Чехович[a, ****]**

[a]*Компания «Антиплагиат», Москва, Россия*
[b]*Московский физико-технический институт, Москва, Россия;*
[a]*ФИЦ ИУ РАН, Москва, Россия*
**e-mail: kvarlamova@ap-team.ru*
***e-mail: kaprielova@ap-team.ru*
****e-mail: potyashin@ap-team.ru*
*****e-mail: chehovich@ap-team.ru*

Поступила в редакцию 22.05.2024 г.
После доработки 27.05.2024 г.
Принята к публикации 15.07.2024 г.

Решение задачи поиска заимствований в рукописных текстах становится год от года более актуальным. Одним из видов заимствований является почти дублирование рукописной работы – съемка того же рукописного текста в других условиях или использование различных аугментаций. Существующие подходы к обнаружению почти дубликатов не приспособлены к работе с большими коллекциями, что существенно ограничивает их использование на практике. Представлен метод на основе машинного обучения, который позволяет производить обнаружение почти дубликатов изображений рукописных текстов среди больших коллекций потенциальных источников. Процесс включает в себя три основных этапа: перевод изображения в векторное представление, поиск кандидатов и последующий отбор источника дублирования среди кандидатов. Приведены результаты экспериментов по оценке качества и производительности разработанной системы: достигнуты 59 и 80% полноты и 5.5 и 4.8% доли ложноположительных срабатываний приближенных к реальным и синтетических данных соответственно, время работы метода составляет 5.5 с/запрос при размере коллекции около 10 тыс. изображений. Результаты показали, что созданный метод может быть использован для решения задач, требующих проверки рукописных документов по большому количеству потенциальных источников заимствований.

Ключевые слова: компьютерное зрение, поиск почти дубликатов, анализ рукописных документов, большие базы данных, русский рукописный текст.

DOI: 10.31857/S0002338824040085 **EDN:** UEFADS

HANDWRITTEN DOCUMENTS NEAR-DUPLICATE SEARCH FOR DATA INTENSIVE APPLICATIONS

K. Varlamova[a, b, *], M. Kaprielova[a, b, c, **], I. Potyashin[a, b, *], Yu. Chekhovich[a, ****]**

[a]*Antiplagiat Company, Moscow, Russian Federation*
[b]*Moscow Institute of Physics and Technology, Moscow, Russian Federation*
[c]*FRC CSC RAS, Moscow, Russian Federation*
**e-mail: kvarlamova@ap-team.ru*
***e-mail: kaprielova@ap-team.ru*
****e-mail: potyashin@ap-team.ru*
*****e-mail: chehovich@ap-team.ru*

The problem of cheating in handwritten academic essays has become more significant over last several years. One of the cheating cases is submitting the same paper, photographed in different environment (for

example, from another angle, in different light or in lower quality), or changed by means of automatic augmentation. The existing methods are not designed to work on large collections of handwritten documents. The proposed approach consists of three stages. The first stage is embedding generation, the second one is finding closest candidates in the collection of handwritten documents and the final one is similarity estimation between query image and each of candidates obtained at previous step. Our solution showed Recall@1 80% and 59% with FPR 4.8% and 5.5% on Synthetic and Real data respectively. The search latency is 5.5 seconds per query for the collection of 10 000 images. The results showed that the developed method is robust enough to work on large collections of handwritten documents.

Keywords: computer vision; near-duplicate detection; handwritten document analysis; large collections; Russian cursive.

Введение. Поиск заимствований в академических и учебных работах является актуальной задачей. Уже существуют системы, позволяющие обнаруживать много типов нарушений академической этики в текстах [1], такие, как переводные заимствования [2,3], парафраз, машинная генерация [4,5] и др. Однако проблеме поиска заимствований в рукописных текстах уделяется гораздо меньше внимания. С бурным развитием онлайн-образования и необходимостью повышения автоматизации проверок работ школьников проблема заимствований в рукописных работах становится все более актуальной [6,7]. Требуется повышение уровня автоматизации проверки работ, причем с возможностью ее осуществления для больших коллекций. В частности, необходимость поиска заимствований в рукописных текстах школьников обусловлена важностью обучения принципам работы с информацией и формированием правильных представлений о правовых и этических нормах использования материалов, находящихся в открытом доступе.

Заимствования в рукописных работах можно разделить на две крупные категории. К первой категории относится переписывание текста работы с возможным изменением его части. Второй категорией заимствований является визуальное изменение той же работы с помощью различных манипуляций, таких, как фотографирование под другим углом, освещением, изменение качества фотографии или применение различных аугментаций по отношению к одному и тому же изображению рукописного текста. Второй тип заимствования, рассматриваемый в текущей статье, назовем *почти дублированием* изображения рукописного текста. Представлен новый метод детекции почти дублированных изображений рукописных текстов, написанных на русском языке.

1. Краткий обзор литературы. Задачу поиска почти дублированных изображений рукописного текста можно назвать сложной и при этом критически важной для образовательной системы [6,7]. В настоящее время становится все более востребованно работать с большими объемами данных, что усложняет задачу, так как в литературе не было описано достаточно эффективных для крупных коллекций методов. В [8] представлен подход к задаче детекции заимствований изображений, специализирующийся на поиске по крупным коллекциям. Однако рукописные тексты составляют специфичный домен данных, поэтому требуется более детализированный по отношению к сравнению изображений подход.

Подход к сравнению двух рукописных документов представлен в [9]. Он основан на сопоставлении рамок, ограничивающих слова (bounding box), с помощию сверточной нейронной сети. В [8] описан метод, полагающийся на сегментацию слов с последующим анализом их длин. Ряд подходов, связанных с анализом рукописного текста, базируется на его распознавании [10,11]. В сфере распознавания рукописного текста были достигнуты значительные результаты [12,13]. Это делает потенциально возможным использование систем распознавания рукописного текста (optical character recognition (OCR)) совместно с существующими методами поиска заимствований в текстах [14]. Однако такой подход имеет существенный недостаток: для обучения модели распознавания рукописного текста требуется большой объем данных для разметки. Для кириллических языков такого количества открытых размеченных данных нет, что приводит к невысокому качеству моделей распознавания рукописного текста на русском языке. Для моделей OCR также могут быть критичны изменения цвета, качества, поворот изображения и т.д., поэтому разные варианты съемки одного и того же рукописного текста могут привести к различным выходам модели. В связи с этим, в рамках задачи поиска почти дубликатов пока нельзя назвать эффективными подходы, основанные на распознавании рукописного текста.

В статье ключевая задача – разработка метода детекции почти дублированных изображений рукописных текстов. Основными достоинствами предлагаемого метода является то, что он не требует большого количества размеченных данных и достигает высокой эффективности при поиске по большим коллекциям. Решение состоит из трех логических частей: генерация векторных представлений, быстрый поиск кандидатов на источник заимствования и отбор наиболее релевантных кандидатов путем подсчета уровня сходства между изображениями. Проведены эксперименты на двух коллекциях, состоящих из рукописных сочинений на русском языке.

2. Постановка задачи. Будем называть *почти дубликатами* такие изображения, которые полностью или почти копируют оригинал: используется та же фотография работы либо работа сфотографирована под другим углом, с другим фоном, тенью, фотография изменена по качеству, яркости и пр.

Задачу детекции почти дубликатов можно сформулировать следующим образом.
Существует два множества:

1) множество изображений-запросов Q,

2) множество изображений-источников S, из которого могли происходить почти дублирования, содержащиеся в Q; S также имеет нулевой элемент s_0.

Существует отображение $FindSourse : Q \to S$. Для всех запросов q^{orig} из Q, не являющихся почти дубликатами изображений из S, выполняется

$$FindSourse(q^{orig}) = s_0.$$

Множество таких q^{orig} обозначим как $Ker(FindSourse)$, его мощность – как $M = |Ker(FindSourse)|$. Введем модель поиска почти дубликатов $f(q)$, $q \in Q$, выходом которой является множество $\{s1, \ldots, s_K, s_k \in S, k = 1,K\}$, $K \in \mathbb{N}$ – заранее фиксированное число кандидатов на источник почти дублирования.

Основными метриками качества будем считать метрику полноты $Recall@K$ на выбранном числе кандидатов K и метрику доли ложноположительных срабатываний (false positive rate (FPR)):

$$Recall@K = \frac{100\%}{|Q|}\sum_{i=1}^{|Q|}\left|f(q_i)@K \cap \left\{FindSourse(q_i)\right\}\right|, q_i \in Q \tag{2.1}$$

$$FPR = \frac{100\%}{|M|}\sum_{i=1}^{M} f \left| \left(q_i^{orig}\right)@\cap S\backslash s_0 \right|, q_i^{orig} \in Ker(FindSourse). \tag{2.1}$$

Задачей является поиск модели $\hat{f}$, наилучшей в смысле полноты, при ограничении на долю ложноположительных срабатываний:

$$\begin{cases} \hat{f} = \arg\max\limits_{f \in F} Recall@1(f, FindSourse, Q, S), \\ FPR(\hat{f}, FindSourse, Q, S) < \alpha, \end{cases}$$

где α – установленный порог для ограничения ложных срабатываний, в данной работе рассматривался $\alpha = 7\%$; F – пространство моделей.

3. Предлагаемый подход. Решение задачи, описанной в разд. 2, разделено на три последовательные части: генерация векторных представлений для изображений; векторный поиск кандидатов-источников заимствования для изображения-запроса; выбор кандидатов с помощью подсчета уровня сходства между кандидатом и изображением-запросом.

3.1. О п и с а н и е д а н н ы х. В работе используется три набора данных. Первый датасет – закрытый. Он состоит из 1 млн фотографий рукописных школьных сочинений, написанных на русском языке. Для взятых работ были искусственно сгенерированы почти дублирования с помощью наложений таких аугментаций, как изменения по цвету, углу, геометрических преобразований. Пример пары изображение – сгенерированный дубликат (рис. 1). Коллекцию, полученную из этого набора данных с помощью таких преобразований, будем называть *Synthetic*. Она использовалась для обучения моделей, входящих в итоговое решение.

Второй используемый датасет HWR200 [15] специализирован для задачи обнаружения заимствований из рукописных текстов, но обладает структурой, подходящей и для задачи поиска почти дубликатов. Каждое изображение коллекции – фотография или скан страницы школьного сочинения на русском языке. При этом каждая страница представлена в трех ви-

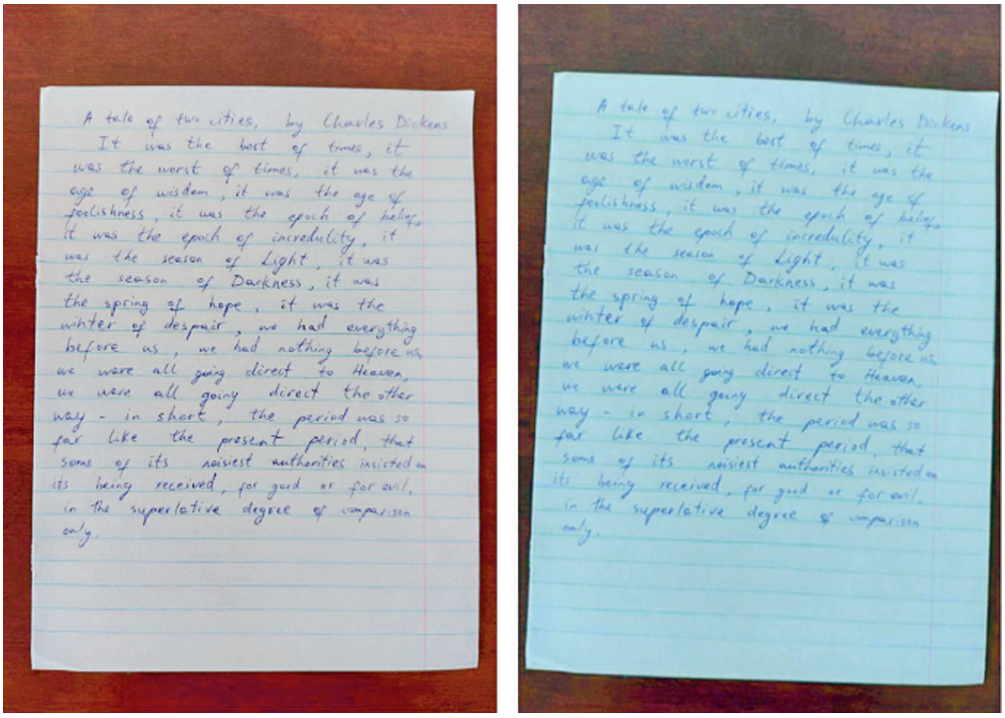

Рис. 1. Пример изображений из датасета с синтетическими почти дубликатами: оригинал и сгенерированный почти дубликат

дах: скан, светлая фотография и темная фотография. Светлая фотография сделана под хорошим освещением и с хорошим качеством. Темная фотография обладает меньшей яркостью, а также чаще содержит посторонние предметы – объекты, помимо самой страницы сочинения содержащиеся на фотографии. Пример элемента датасета можно увидеть на рис. 2. Таким образом, эти три вида изображений для одной и той же страницы являются друг для друга почти дубликатами. Коллекция содержит около 30 тыс. элементов по 10 тыс. изображений каждого типа. Она применялась для экспериментов, описанных ниже.

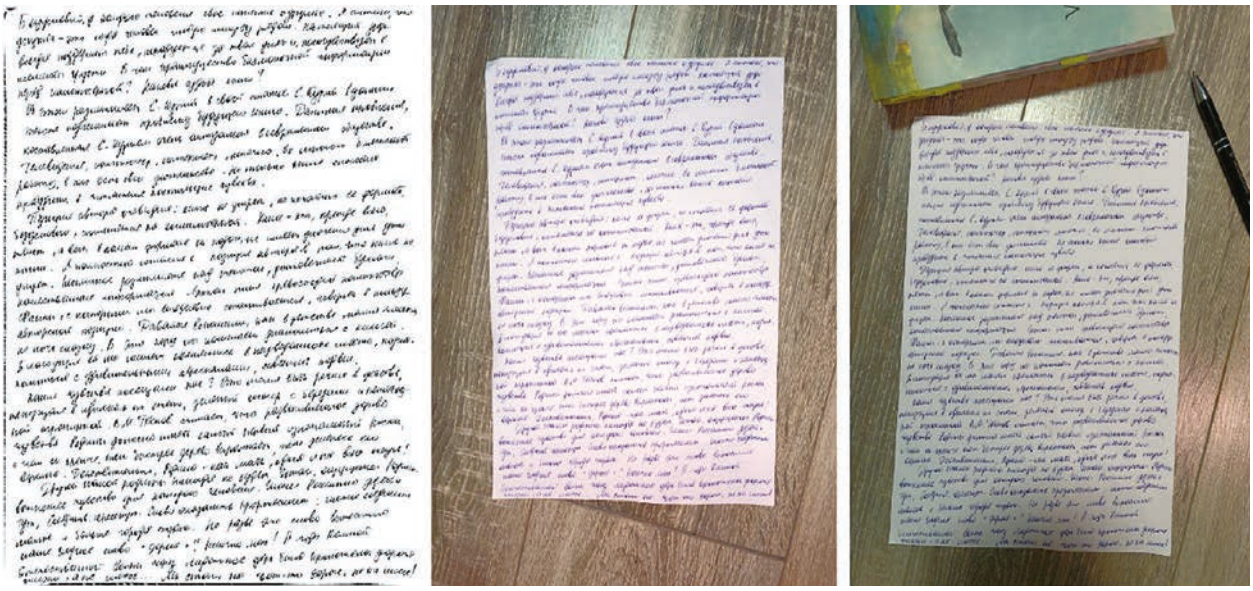

Рис. 2. Типы изображений из датасета HWR200 слева направо: скан, светлая фотография и темная фотография

Третий набор данных, полученный из открытых источников [16,17], и с помощью генерации почти дубликатов исходных изображений, также использовался для экспериментов.

3.2. Генерация векторных представлений. Для перевода пространства изображений с рукописными текстами в пространство векторных представлений применялась архитектура *Deep Ranking*, описанная в [18]. Авторы работы предложили архитектуру, целью которой было производить ранжирование не просто на основании классов, к которым принадлежат объекты, а на основании каких-либо характеристик объектов, которыми могут обладать экземпляры класса. Таким образом, в пространстве векторных представлений изображения, имеющие большее сходство, будут расположены ближе друг к другу.

Модель обучалась на *триплетах* – множествах из трех изображений $\{h,h_{pos},h_{neg}\}$, где h – входное изображение, h_{pos} – изображение из того же *положительного* класса, h_{neg} – изображение из другого *отрицательного* класса. Каждое изображение подавалось на вход сверточной нейронной сети для получения векторного представления. Затем векторные представления триплета подавались на вход нейросети, где в качестве функции ошибок использовалась функция *Triplet Loss*, рассмотренная в [19]:

$$L\left(e,e_{pos},e_{neg}\right)=\max\left\{0,dist\left(e,e_{pos}\right)-dist\left(e,e_{neg}\right)+g\right\}$$

где e,e_{pos},e_{neg} – векторные представления h,h_{pos},h_{neg} соответственно, *dist* – Евклидово расстояние, g – параметр сдвига (margin) между положительными и отрицательными классами. В нашем случае для обучения использовалось 100 тыс. случайно выбранных работ из датасета *Synthetic*, описанного в разд. 3.1, при этом для каждого изображения генерировалось пять почти дубликатов.

Таким образом, положительным классом считались сгенерированные почти дубликаты, отрицательным – другие изображения из взятых оригиналов. В качестве сверточной нейросети применялась ResNet-50 [20] с выбранной размерностью пространства векторных представлений 1024. Более детально процесс обучения описан в [18].

3.3. Векторный поиск. Имея обученную модель архитектуры *Deep Ranking*, можно попарно сравнивать изображения по L2-расстоянию между их векторными представлениями. Однако с увеличением коллекций до многотысячных и многомиллионных наборов изображений такой подход перестает быть реализуемым, так как имеет высокую вычислительную сложность. В связи с этим требуется найти подход, работающий более эффективно и способный осуществлять примерный поиск источников среди большого количества изображений.

Для приближенного решения задач поиска ближайших объектов существует широкий спектр методов, основанных на построении индекса, который дает возможность оптимизировать поиск и ускорить вычислительные процессы. Готовые библиотеки и фреймворки, такие, как Annoy [21], Pinecone [22] и Faiss [23], представляют собой библиотеки, специально разработанные для реализации методов поиска ближайших объектов. Они предлагают различные варианты построения индекса и алгоритмов поиска, позволяя настраивать параметры для достижения оптимальной производительности и точности результатов. Использование подобных фреймворков ускоряет процесс поиска ближайших объектов и повышает эффективность вычислительных операций при работе с большими объемами данных.

Для решения задачи был выбран IVF индекс библиотеки Faiss [23], так как библиотека имеет открытый код. Коллекция изображений, среди которых планируется искать источники заимствований, помещается в индекс. Идея IVF индекса заключается в создании заданного количества N кластеров, на которые делится пространство векторов. Каждый вектор принадлежит одному из кластеров. Во время поиска вектор-запрос сначала сравнивается с центральными элементами кластеров, из них отбирается *nprobe* ближайших. Дальнейший поиск производится только по векторам этих ближайших кластеров. В нашей работе были выбраны значения $N = 65536$, *nprobe* = 128. Также использована *Product Quantization* (PQ32), описанная в [23], для уменьшения размерности индекса и более эффективного поиска. Таким образом, коллекция источников индексируется, и по индексу производится поиск для каждого запроса. Результатом поиска является заданное количество ближайших векторов – кандидатов на источники дублирования.

3.4. Оценка схожести изображений. Для отбора источников из кандидатов, полученных на предыдущем шаге, предлагается искать некую меру схожести между изображением-запросом и изображениями, соответствующими найденным кандидатам. Построение этой части модели оказалось наиболее трудоемким. Далее описаны несколько подходов к решению данной части задачи и подбор наиболее релевантного из них.

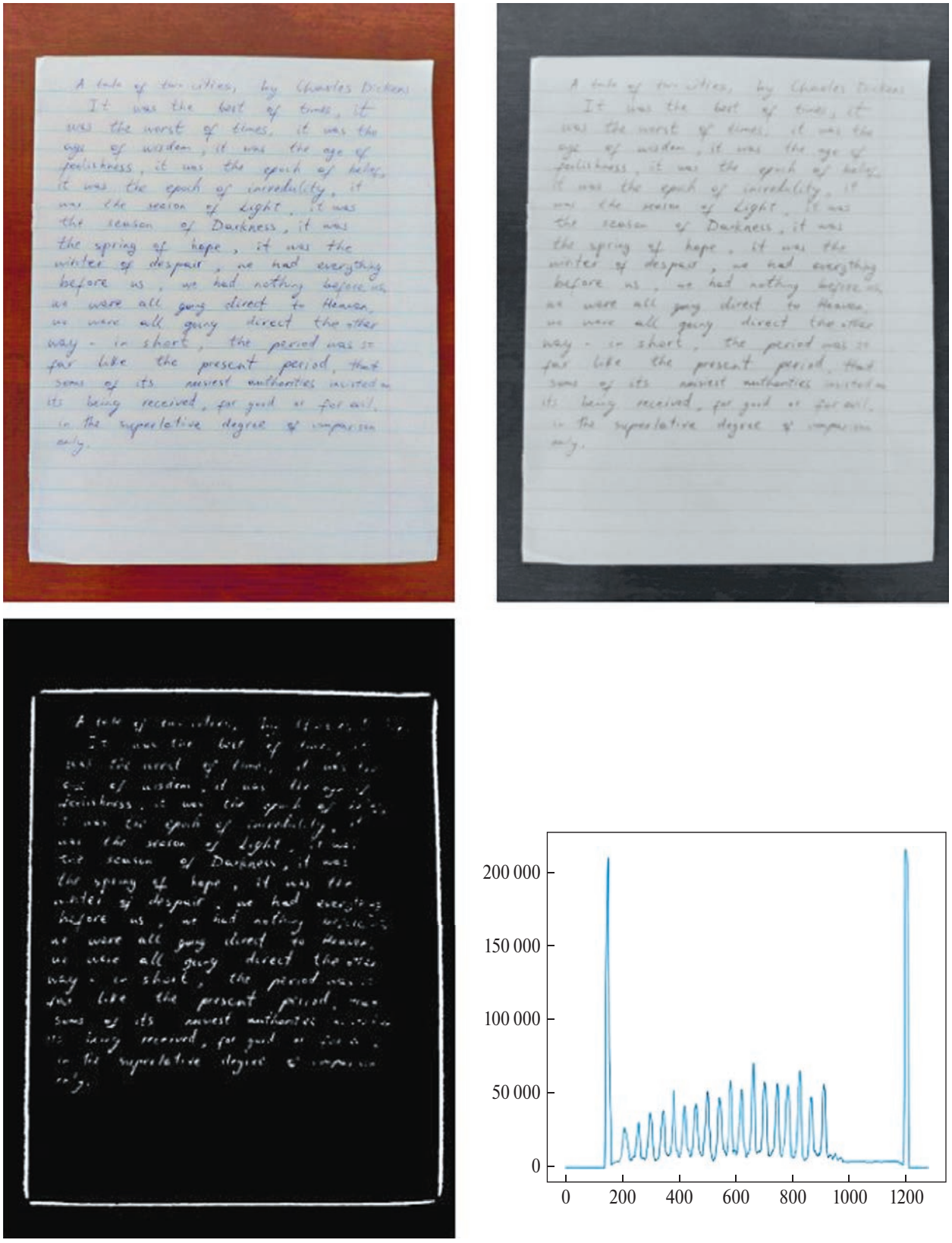

Рис. 3. Процесс извлечения сигнала из изображения: перевод в оттенки серого, применение адаптивного порога, выделение сигнала

Одним из распространенных подходов к поиску схожести между изображениями является использование Сиамских нейросетей [24]. Они позволяли достичь высоких результатов для случая, когда коллекция содержит большое количество изображений из разных доменов [25]. Однако в нашей задаче все изображения коллекции фактически принадлежат одному домену – фотографиям рукописных текстов. Согласно проведенному исследованию, применение Сиамских нейронных сетей не обеспечило приемлемого качества.

Другим подходом к данной части задачи выступают методы, используемые для анализа временных рядов. Так как почти дубликатом работы служит изображение, где текст расположен на том же (или почти том же) регионе, то текст входного изображения работы можно определить как некоторый сигнал. В рамках такого рассмотрения задачи в нашей работе извлекался сигнал из изображения (рис. 3). Для получения сигнала сначала цветное изображение преобразовывалось в черно-белое, затем выделялся рукописный текст с помощью фильтрации по яркости. Затем матрица изображения суммировалась по строкам. Далее из сигнала извлекались максимумы как наиболее информативная часть. После этого подсчитывалась схожесть между сигналом работы запроса и работы-кандидата. Схожесть определялась как Евклидово расстояние между последовательностями. Для вычисления Евклидова расстояния был использован алгоритм динамической трансформации временной шкалы (dynamic time wrapping (DTW)). Одним из практических применений алгоритма DTW является онлайн проверка подписи [26]. В работе был использован алгоритм FastDTW [27] – аппроксимация DTW, имеющая линейную вычислительную сложность вместо квадратичной у полной DTW. Для классификации изображения как почти дубликат или оригинал найденная схожесть между запросом и кандидатом сравнивалась с подобранным порогом. При превышении порога изображение считалось почти дубликатом, а кандидат отбирался как источник почти дублирования. Распространенными методами решения задачи сравнения изображений являются методы извлечения опорных точек (keypoint-extraction) от подходов классического компьютерного зрения: SIFT [28], ORB [29], до нейросетевых [30, 31]. Различные алгоритмы извлечения ключевых точек (SIFT, ORB или нейронные сети) используются для выбора ключевых точек и вычисления их дескрипторов для каждого изображения-кандидата и изображения-запроса. Дескриптор – это объект, содержащий некоторую информацию о ключевой точке, на основе которой можно сделать вывод о сходстве или различии между двумя ключевыми точками. Подсчитывалось и нормировалось количество схожих ключевых точек. При превышении установленного порога запрос считался почти дубликатом, а кандидат отбирался как итоговый кандидат на оригинал.

Долгое время разница в результатах между классическими и глубинными методами keypoint-extraction была незначительна [32]. Большинство нейросетевых подходов к сопоставлению ключевых точек требовало двух отдельных архитектур. Одна нейронная сеть определяла ключевые точки, а затем другая сравнивала их. Однако с появлением в компьютерном зрении архитектур типа Трансформер [33] возникли решения, объединяющие эти два этапа. Авторы LoFTR [34] представили новую архитектуру, основанную на архитектуре Трансформера. Вдохновленные их работой, мы модифицировали архитектуру LoFTR [34] в соответствии с конкретными требованиями нашей задачи. Исследования, описанные ниже, показали, что наиболее подходящим решением является использование именно этой архитектуры.

4. Эксперименты. Для выбора модели в последней части метода был проведен эсперимент по сравнению нескольких подходов: LoFTR, SIFT, Signal+FastDTW (разд. 3.4). Для этого был создан специальный набор данных, состоящий из 1000 элементов коллекции *Synthetic* (разд. 3.1). Для каждого почти дубликата сравнение проводилось по 511 кандидатам, помимо истинного оригинала. Кандидаты были получены с помощью поиска, чтобы приблизить эксперимент к реальному использованию. По результатам (представлены в табл. 1) был выбран подход LoFTR ввиду явного преимущества в качестве.

Таблица 1. Качество итогового отбора кандидатов при использовании различных подходов

Метод	*Recall*@1, %
SIFT	30.9
Signal+FastDTW	78.8
LoFTR modification	98.4

Эксперимент по измерению качества всего метода проводился на двух коллекциях: с синтетическими и реальными данными. В качестве метрик применялись *Recall*@1 (2.1) и FPR (2.2). Минимизация доли ложноположительных стабатываний представляла особую важность как минимизация ложных обвинений в дублировании.

Для первой коллекции, *Synthetic*, использовались открытые данные IAM [16] и READ2016 [17]. Были взяты все train и test примеры (747 IAM train, 336 IAM test; 350 Read2016 train, 50 Read2016 test). На тестовых примерах из обоих датасетов были сгенерированы почти дубликаты по два на каждое изображение. Для оценки FPR взяли все validation примеры (116 IAM; 50 Read2016).

С целью проверки метода на данных, приближенных к реальным (*Real*), применялась коллекция, составленная из датасета HWR200 (разд. 3.1). В качестве источника заимствования и почти дубликата поочередно использовались *светлые* (*RealLight*) и *темные* (*RealDark*) изображения. Значения метрик качества вычислялись как среднее для этих двух экспериментов. Для оценки доли ложноположительных стабатываний применялась отложенная выборка из объектов того же датасета, для которых не индексировались почти дубликаты.

Результаты эксперимента показаны в табл. 2. Видно, что даже на данных из HWR200 уровень качества остается достаточно высоким для использования в высоконагруженных системах обнаружения заимствований при достижении низкого количества ложных срабатываний. При этом модель обучалась только на синтетических данных.

Таблица 2. Качество работы полного метода

Метрика, %	Коллекция	
	Synthetic	*Real*
Recall@1	80	59
FPR	4.8	5.5

Также был проведен эксперимент на производительность, которая особенно важна при использовании больших коллекций данных. Производительность измерялась на машине с процессором AMD Ryzen 9 3900XT 12-Core Processor с помощью 8 ядер и 64 Gb RAM. Для модели-модификации LoFTR применялся графический процессор NVIDIA GeForce RTX 3090. Для оценки производительности было подано 500 изображений-запросов из датасета HWR200. Время работы полного метода поиска почти дубликатов составило в среднем 5.5 с на один запрос при общем размере коллекции гипотетических источников около 10 тыс. изображений. Стоит отметить, что основные временные затраты относятся к работе заключительного модуля метода, который обрабатывает уже фиксированное количество кандидатов. Следовательно, при увеличении коллекции скорость работы этого модуля изменяться не будет. Таким образом, увеличение коллекции не приведет к значительному изменению скорости работы системы.

Заключение. Представлен метод обнаружения почти дубликатов изображений рукописного текста, способный обеспечивать высокую производительность при работе с большими коллекциями документов. Подход состоит из нескольких этапов. Первый этап — построение векторных представлений изображений. Второй этап, поиск кандидатов, включает в себя поиск объектов, который сужает число возможных источников почти дублирования. Последний этап – сравнение изображений, в результате которого остается небольшое количество кандидатов, с большой долей вероятности являющихся источником почти дублирования. Были проведены эксперименты на двух коллекциях изображений рукописных текстов. В первом наборе данных содержатся изображения рукописных текстов, дубликаты которых были созданы синтетически, а второй набор состоит из фотографий рукописных текстов, имеющих реальные дубликаты. Было достигнуто 59 и 80% полноты и 5.5 и 4.8% доли ложноположительных срабатываний для набора рукописных сочинений, приближенного к реальным данным, и синтетического набора соответственно. При этом время обработки составило в среднем 5.5 с на запрос. Полученные результаты свидетельствуют о том, что предложенный подход может быть использован в качестве решения для высоконагруженных систем обнаружения заимствований.

Дальнейшие исследования могут быть направлены на работу с изображениями рукописных текстов низкого качества, а также на обработку изображений рукописей на других языках, например, имеющих иероглифическую письменность.

СПИСОК ЛИТЕРАТУРЫ

1. *Bakhteev O., Ogaltsov A., Khazov A., Safin K., Kuznetsova R.* CrossLang: the System of Cross-lingual Plagiarism Detection // Workshop on Document Intelligence at NeurIPS. Vancouver, 2019.
2. *Avetisyan K., Gritsay G., Grabovoy. A.* Cross-Lingual Plagiarism Detection: Two Are Better Than One // Programming and Computer Software. 2023. V. 49. P. 346–354.
3. *Kuznetsova M., Bakhteev O., Chekhovich Y.* Methods of Cross-lingual Text Reuse Detection in Large Textual Collections // Informatika I Ee Primeneniya [Informatics and Its Applications]. 2021. V. 15. P. 30–41.
4. *Gritsay G., Grabovoy A., Kildyakov A., Chekhovich Y.* Artificially Generated Text Fragments Search in Academic Documents // Doklady Rossijskoj Akademii Nauk. Matematika, Informatika, Processy Upravlenia. 2023. V. 108. P. 308–317.
5. *Gritsay G., Grabovoy A., Chekhovich Y.* Automatic Detection of Machine Generated Texts: Need More Tokens // Ivannikov Memorial Workshop (IVMEM). Kazan, 2022. V. 108. P. 20–26.
6. *Ma H.J., Wan G., Lu E.Y.* Digital Cheating and Plagiarism in Schools // Theory Into Practice. 2008. V. 47. P. 197–203.
7. *Wrigley S.* Avoiding 'de-plagiarism': Exploring the Affordances of Handwriting in the Essay-writing Process // Active Learning in Higher Education. 2019. V. 20. P. 167–179.
8. *Bakhteev O., Kuznetsova R., Khazov A., Ogaltsov A., Safin K., Gorlenko T., Suvorova M., Ivahnenko A., Botov P. et. al.* Near-duplicate Handwritten Document Detection Without Text Recognition // Intern. Conf. on Computational Linguistics and Intellectual Technologies. Moscow, 2021. P. 47–57.
9. *Krishnan P., Jawahar C.V.* Matching Handwritten Document Images // Europ. Conf. on Computer Vision. Amsterdam, 2016. P. 766–782.
10. *Rowtula V., Bhargavan V., Kumar M., Jawahar C.V.* Scaling Handwritten Student Assessments with a Document Image Workflow System // IEEE Conf. on Computer Vision and Pattern Recognition Workshops. Salt Lake City, 2018. P. 2307–2314.
11. *Pandey O., Gupta I., Mishra B.S.P.* A Robust Approach to Plagiarism Detection in Handwritten Documents // Intern. Sympos. on Visual Computing. San Diego, 2020. P. 682–693.
12. *Coquenet D., Chatelain C., Paquet T.* End-to-end Handwritten Paragraph Text Recognition Using a Vertical Attention Network // ArXiv 2021. ArXiv Preprint ArXiv:2012.03868.
13. *Voigtlaender P., Doetsch P., Ney H. H*andwriting Recognition With Large Multidimensional Long Short-term Memory Recurrent Neural Networks // 15th Intern. Conf. on Frontiers in Handwriting Recognition (ICFHR). Shenzhen, 2016. P. 228–233.
14. *Khritankov A., Botov P., Surovenko N., Tsarkov S., Viuchnov D., Chekhovich Y.* Discovering Text Reuse in Large Collections of Documents: A Study of Theses in History Sciences // Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conf. (AINLISMW FRUCT). St. Petersburg, 2015. P. 26–32.
15. *Potyashin I., Kaprielova M., Chekhovich Y., Kildyakov A., Seil T., Finogeev E., Grabovoy A.* HWR200: New Open Access Dataset of Handwritten Texts Images in Russian // Intern. Conf. on Computational Linguistics and Intellectual Technologies. Moscow, 2023.
16. *Grieggs S., Shen B., Rauch G., Li P., Ma J., Chiang D., Price B., Scheirer W.J.* Measuring Human Perception to Improve Handwritten Document Transcription // ArXiv 2019. ArXiv Preprint ArXiv:1904.03734 .
17. *Toselli A., Romero V., Villegas M., Vidal E., Sanchez J.* HTR Dataset // Intern. Conf. on Frontiers in Handwriting Recognition (ICFHR). Shenzhen, 2016. P. 630635.
18. *Wang J., Song Y., Leung T., Rosenberg C., Wang J., Philbin J., Chen B., Wu Y.* Learning Fine-grained Image Similarity With Deep Ranking // IEEE Conf. on Computer Vision and Pattern Recognition. Columbus, 2014. P. 1386–1393.
19. *Balntas V., Riba E., Ponsa D., Mikolajczyk K.* Learning Local Feature Descriptors With Triplets and Shallow Convolutional Neural Networks // The British Machine Vision Conference (BMVC). 2016. V. 1. №2. P. 3.
20. *He K., Zhang X., Ren S., Sun J.* Deep Residual Learning for Image Recognition // IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016.
21. Annoy // https://github.com/spotify/annoy.
22. Pinecone // https://github.com/pinecone-io.
23. *Johnson J., Douze M., J'egou H.* Billion-scale Similarity Search With GPUs // IEEE Transactions on Big Data. 2019. V. 7. P. 535–547.
24. *Melekhov I., Kannala J., Rahtu E.* Siamese Network Features for Image Matching // 23rd Intern. Conf. on Pattern Recognition, ICPR. Cancun, 2016. P. 378–383.
25. *Bakhteev O., Chekhovich Y., Finogeev E., Gorlenko T., Kaprielova M., Kildyakov A., Ogaltsov A.* Image Reuse Detection in Large-scale Document Scientific Collection // ENAI Conf., Concurrent Sessions 12. Porto, 2022. P. 107.
26. *Patil B. V., Patil P. R.* An Efficient DTW Algorithm for Online Signature Verification // Intern. Conf. On Advances in Communication and Computing Technology (ICACCT). Painpat, 2018. P. 1–5.
27. *Salvador S., Chan P.* Toward Accurate Dynamic Time Warping in Linear Time and Space // Intellectual Data Analysis. 2007. V. 11. P. 561–580.
28. *Lowe D.G.* Distinctive Image Features from Scale-invariant Keypoints // Intern. J. of Computer Vision. 2004. V. 60. P. 91–110.

29. *Rublee E., Rabaud V., Konolige K., Bradski G.* ORB: An Efficient Alternative to SIFT or SURF // Intern. Conf. on Computer Vision. Barcelona, 2011. P. 25642571.
30. *DeTone D., Malisiewicz T., Rabinovich A.* Superpoint: Self-supervised Interest Point Detection and Description // IEEE Conf. on Computer Vision and Pattern Recognition Workshops. Salt Lake City. 2018, P. 224–236.
31. *Barroso-Laguna A., Riba E., Ponsa D., Mikolajczyk K.* Key. net: Keypoint Detection by Handcrafted and Learned cnn Filters // IEEE/CVF Intern. Conf. on Computer Vision. Seoul, 2019. P. 5836–5844.
32. *Mishkin D.* Local Features: from Paper to Practice // Computer Vision and Pattern Recognition (CVPR) Workshops. Seattle, 2020.
33. *Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Polosukhin I. et. al.* Attention Is All You Need // ArXiv 2017. ArXiv Preprint ArXiv:1706.03762.
34. *Sun J., Shen Z., Wang Y., Bao H., Zhou X.* LoFTR: Detector-Free Local Feature Matching With Transformers // ArXiv 2021. ArXiv Preprint ArXiv:2104.00680. P. 8922–8931.

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, с. 139–147

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

УДК 004.932.2

МЕТОД ОПРЕДЕЛЕНИЯ ДВИЖЕНИЯ В КАДРЕ И ИДЕНТИФИКАЦИИ КРУПНОГАБАРИТНОГО ПЛОЩАДНОГО ОБЪЕКТА

© 2024 г. В. В. Лопатина[a], *

[a]ФИЦ ИУ РАН, Москва, Россия

**e-mail: int00h@mail.ru*

Поступила в редакцию 05.06.2024 г.

После доработки 16.06.2024 г.

Принята к публикации 16.10.2024 г.

Приведен метод определения движения в кадре и идентификации крупногабаритного площадного объекта. Работа метода иллюстрируется на примере из морской транспортной отрасли, в задаче контроля положения автономного морского крупнотоннажного судна относительно причала при выполнении погрузо-разгрузочных работ и швартовных операций. Описана структура измерительного комплекса на базе оптических измерителей, его принцип действия, основанный в том числе на методе определения движения в кадре и идентификации крупногабаритного площадного объекта. Описан порядок анализа подвижных участков изображения. Приведена схема алгоритма определения движения в кадре и идентификации крупногабаритного площадного объекта. Выполнена оценка производительности программной реализации алгоритма определения движения в кадре и идентификации крупногабаритного площадного объекта.

Ключевые слова: компьютерное зрение, анализ изображений, обнаружение движения, оптические измерители, автономный транспорт, морские автономные надводные суда.

DOI: 10.31857/S0002338824040097 **EDN:** UDYVBJ

METHOD FOR MOTION DETECTING IN THE FRAME AND LARGE-SIZED OBJECT IDENTIFICATION

V. V. Lopatina[a], *

[a]Federal Research Center "Computer Science and Control," Russian Academy of Sciences, Moscow, 119333 Russia

**e-mail: int00h@mail.ru*

A method for motion detecting in a frame and large-sized object identification is described in the article. The use-case of the method is illustrated by the example from the maritime transport industry. The example shows the solution of the task of monitoring the position of an autonomous marine large-tonnage ship relative to the berth when performing loading and unloading operations and mooring operations. The paper incudes description of the structure of a measuring complex which includes optical meters. An operating principle of the complex is based on the method of motion detecting in a frame and large-sized object identification. A diagram of the algorithm for motion detecting in the frame and large-sized object identification is presented in the paper. The performance of the software implementation of the algorithm for motion detecting in the frame and large-sized object identification has been assessed in the article.

Keywords: computer vision, image analysis, motion detecting, optical meters, autonomous transport, maritime autonomous surface ships.

Введение. Высокоточные системы проводки и позиционирования используются для стабилизации положения подвижных объектов в различных транспортных системах, для обеспечения контроля положения и, как следствие, повышения безопасности. Такие высокоточные системы применяются, например, для контроля положения автономного морского судна от-

носительно причала при выполнении погрузо-разгрузочных работ и швартовных операций; для контроля стоянки грузового автомобиля относительно складского грузового терминала; позиционирования вагонов на путях необщего пользования в процессах налива цистерн или загрузки сыпучим грузом.

Повышение безопасности эксплуатации транспортной системы позволяет увеличить интенсивность транспортного потока и пропускную способность транспортной сети, что повышает экономическую эффективность всей отрасли.

Высокоточные системы позиционирования пространственно-распределенных объектов применимы не только на транспорте, но и в промышленности. Например, при сцепке деталей, когда точность совмещения деталей влияет на прочность, герметичность и надежность функционирования промышленного изделия. Одна деталь – база, другая – распределенный пространственный объект, который нужно расположить относительно базы, обеспечив минимальное усилие или минимальную нагрузку на элементы конструкции в процессе совмещения деталей.

Представленный в статье метод определения движения в кадре и идентификации крупногабаритного площадного объекта иллюстрируется на примере из морской транспортной отрасли, в задаче контроля положения автономного морского крупнотоннажного судна относительно причала при выполнении погрузо-разгрузочных работ и швартовных операций. Задача сводится к отслеживанию морского судна в процессах подхода и швартовки к причалу, а также к определению его пространственно-скоростных параметров.

1. Постановка задачи. Целевой средой исполнения программной реализации рассмотренного метода является программно-аппаратный комплекс высокоточного определения положения объектов относительно стационарной базы. Комплекс включает оптические измерители, каждый из которых состоит из камеры компьютерного зрения и лазерного дальномера. Если в комплексе задействуется два и более измерителя, компьютер включается в состав только управляющего измерителя, оборудование управляемого измерителя подключается к компьютеру управляющего. Измерители устанавливаются на неподвижное основание, в прямой видимости измерителей находится объект измерений.

В задаче контроля положения автономного морского судна относительно причала, измерители устанавливаются на причале на неподвижном основании в непосредственной близости от края причала (рис. 1), но, как правило, не более 30 см от края, чтобы не допустить перекрытия луча лазера и объектива камеры работниками порта, транспортными средствами и причальными сооружениями. В прямой видимости находится измеряемый объект – морское крупнотоннажное судно.

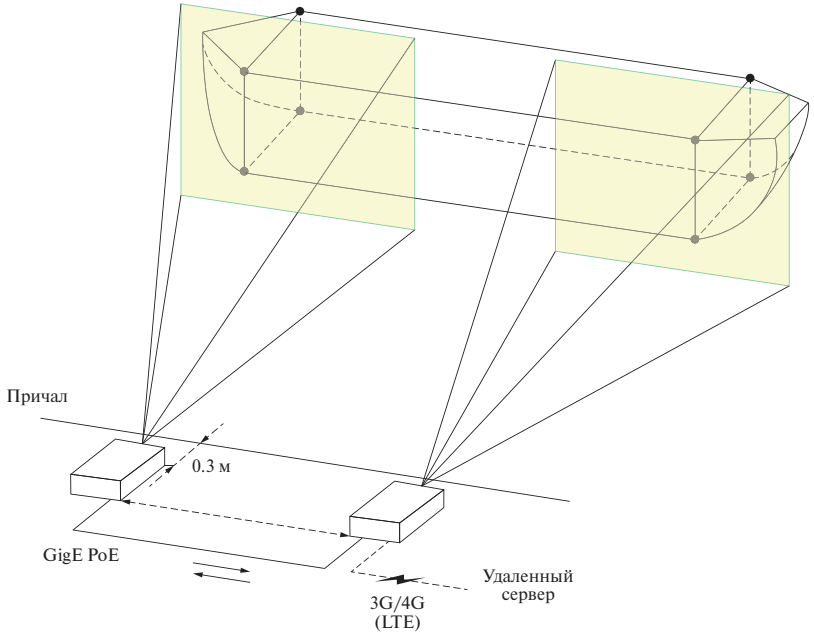

Рис. 1. Схема установки комплекса из двух оптических измерителей на причале

Для выполнения независимых измерений для носа и кормы судна устанавливается комплекс из двух измерителей. Границы области отслеживания судна относительно причала составляют от 2 до 500 м в продольном, поперечном и вертикальном направлениях, что обусловлено требованиями к дистанции последнего маневра морского судна.

Согласно постановлениям по морским портам России на 2024 г. скорость движения судна во время швартовных операций не превышает 3 узлов (5.556 км/ч). Назначение измерителей – это высокоточные измерения продольного (горизонтального), поперечного и вертикального смещения объектов, измерения скоростей движения, определение типа движения (например, смещение, поворот), прогнозирование будущих пространственно-скоростных параметров измеряемого объекта.

Измерение поперечного смещения выполняет лазерный дальномер, установленный в оптический измеритель. Метод измерения основан на сравнении фаз сигнала лазера и сигнала, отраженного от объекта. Задержка при распространении волны создает сдвиг фаз, который измеряется. Лазер работает постоянно, его излучение амплитудно модулируется сигналом определенной частоты [1–3]. Фаза отраженного сигнала сравнивается с фазой опорного сигнала [4, 5].

Измерения вертикального и горизонтального смещения выполняются методами компьютерного зрения. Метод измерения включает анализ кадров видеоряда с камеры оптического измерителя для определения подвижных областей изображения, оценки их характеристик, определения характера движения и выбора объектов (участков изображения), которые потенциально могут принадлежать корпусу морского судна или его палубным конструкциям. Выбранные объекты отслеживаются в реальном времени [6], рассчитывается скорость и разность их смещения, определяются траектории движения.

По смещению объектов на изображении (отслеживаемых участков изображения) рассчитываются пространственно-скоростные параметры отслеживаемого крупногабаритного площадного объекта – морского крупнотоннажного судна. Изображение судна на разных дистанциях от камеры занимает разную часть кадра. Если на 500 м судно полностью помещается в кадр, то по мере приближения судна к причалу фрагмент судна занимает большую часть кадра или весь кадр (рис. 2).

В зависимости от типа причала (например, пирс или набережная стенка), его расположения, особенностей выполнения швартовных операций в кадр могут попадать буксиры, другие суда, причальные сооружения. Отслеживание смещения судна относительно причала начинается на этапе подхода судна к причалу, продолжается во время швартовки судна, стоянки, выполнения погрузо-разгрузочных работ и отшвартовки. Границы области отслеживания судна относительно причала составляют от 2 до 500 м в продольном, поперечном и вертикальном направлениях. Скорость движения судна во время швартовных операций не превышает 3 узлов (5.556 км/ч), но комплекс начинает работать еще на этапе подхода судна к причалу, когда судно находится на расстоянии 300–500 м, на этом этапе скорость судна может достигать 5 узлов.

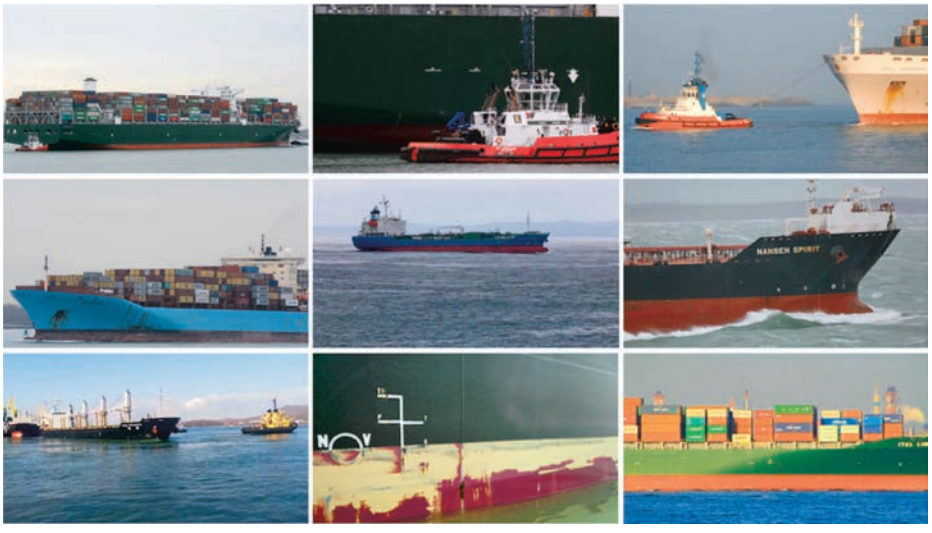

Рис. 2. Морские крупнотоннажные суда на различных расстояниях от причала

2. Предлагаемый подход.

2.1. О с н о в н а я и д е я. Метод определения движения в кадре и идентификации крупногабаритного площадного объекта предполагает два этапа обработки и анализа изображений. На первом этапе определяется движение в кадре путем оценки разницы между соседними кадрами. Вторым этапом является выделение связных компонент (контуров), их анализ и составление карты движущихся участков изображения.

2.2. О ц е н к а р а з н и ц ы м е ж д у с о с е д н и м и к а д р а м и. Под объектами понимаются части кадра, содержащие изображения таких объектов, как суда, причальные сооружения, небо, вода и быстро движущиеся объекты, случайно попавшие в кадр (птицы, насекомые и т.д). Далее по тексту алгоритм означает систему последовательных операций реализующих метод определения движения в кадре и идентификации крупногабаритного площадного объекта. Движение в кадре определяется с целью сегментации объектов кадра, выбора объекта отслеживания, сопровождения этого объекта и последующей работы с ним.

На протяжении всего рабочего цикла оптического измерителя поиск движения в кадре выполняется постоянно, но с разной частотой. На первой последовательности кадров обнаружение движения в кадре необходимо для того, чтобы собрать информацию о присутствующих в кадре движущихся объектах и выбрать объект отслеживания, подходящий по критериям соответствия морскому крупнотоннажному судну. По мере работы оптического измерителя выбранный для отслеживания объект уточняется, поскольку некоторые части объекта выходят за границы кадра, скорость объекта меняется, возможно частичное перекрытие объекта, изменение его формы, так как по мере приближения судна к причалу судно может выполнить разворот. При этом на заднем плане может появится схожий объект с похожими скоростными параметрами отслеживаемого.

На первом этапе делается оценка разницы между двумя соседними кадрами. Из текущего кадра попиксельно вычитается предыдущий кадр. Перед этим значения яркости пикселей обоих кадров нормализуются в диапазон [0;1] (рис. 3) и сглаживаются прямоугольным фильтром низких частот со скользящим окном размером 5×5. Это позволяет сгладить на изображении мелкие незначащие на этом этапе детали (рис. 4):

$$k=\frac{1}{5}\begin{bmatrix}1&1&1&1&1\\1&1&1&1&1\\1&1&1&1&1\\1&1&1&1&1\\1&1&1&1&1\end{bmatrix}.$$

Матрица коэффициентов k – это маска фильтра, применяемая для вычисления пространственной корреляции с целью получения необходимого уровня сглаживания. Вычисляется сумма произведений значений элементов маски и значений пикселей, на которые попадают соответствующие элементы маски, для всех точек изображения.

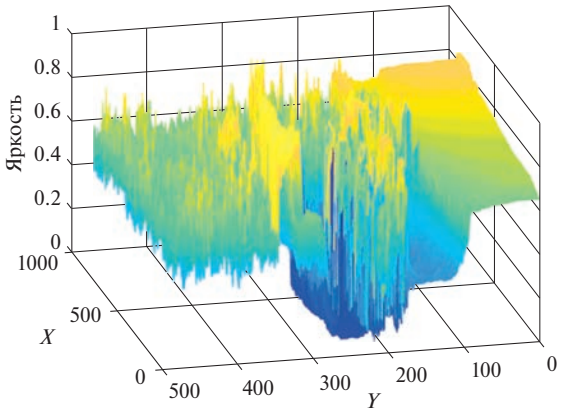

Рис. 3. Визуализация яркости пикселей изображения, нормализованных в диапазоне [0;1]; X, Y – продольная и вертикальная координаты пикселей

Для оптимизации времени исполнения программной реализации алгоритма метод, реализующий отдельную операцию, может быть заменен на аналогичный. Например, для низкочастотной фильтрации в зависимости от требований к производительности программной реализации алгоритма можно выбрать либо метод, представленный в статье, либо его более оптимизированный аналог.

К изображению применяется пороговое преобразование: яркость всех пикселей, имеющих значение больше m, приравнивается к m, значения яркости остальных пикселей не меняются (рис. 5).

Значение параметра m подбирается в зависимости от условий работы. По результатам анализа тестовой выборки видеофайлов с различными вариантами подходов и швартовок судов к целевому причалу было установлено, что оптимальное значение $m = 1$, которое является средним значением яркости предобработанного изображения (глобальный порог). Под целевым причалом понимается причал, на который устанавливаются оптические измерители.

Использование готовых программных реализаций алгоритмов порогового преобразования осложняется подбором параметров, а также непредсказуемостью результатов. Из текущего кадра попиксельно вычитается предыдущий кадр (рис. 6, 7). Пикселы отличные от нуля – это подвижные участки изображения. Далее они приравниваются к единице. Таким образом, результатом первого этапа обработки становится бинаризованное изображение.

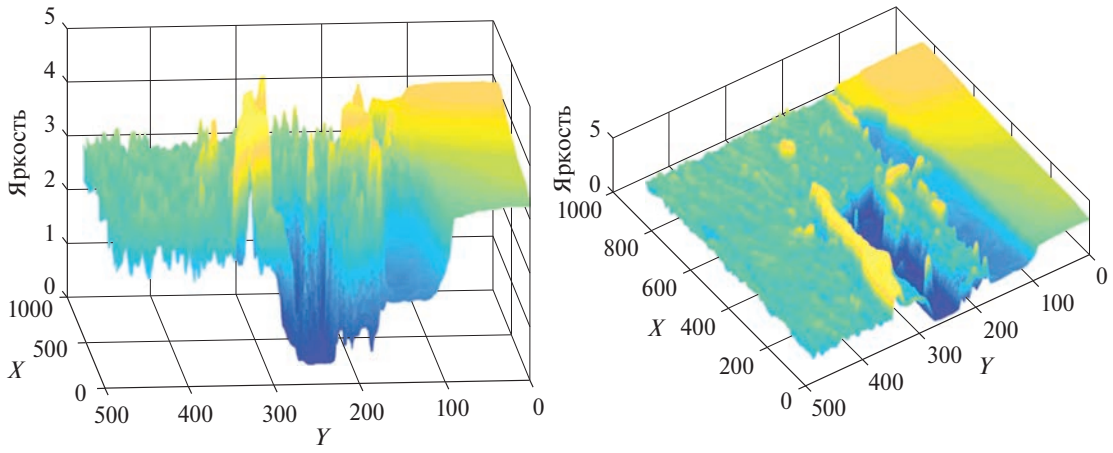

Рис. 4. Визуализация в разных проекциях яркости пикселей изображения после применения прямоугольного фильтра низких частот; X, Y – продольная и вертикальная координаты пикселей

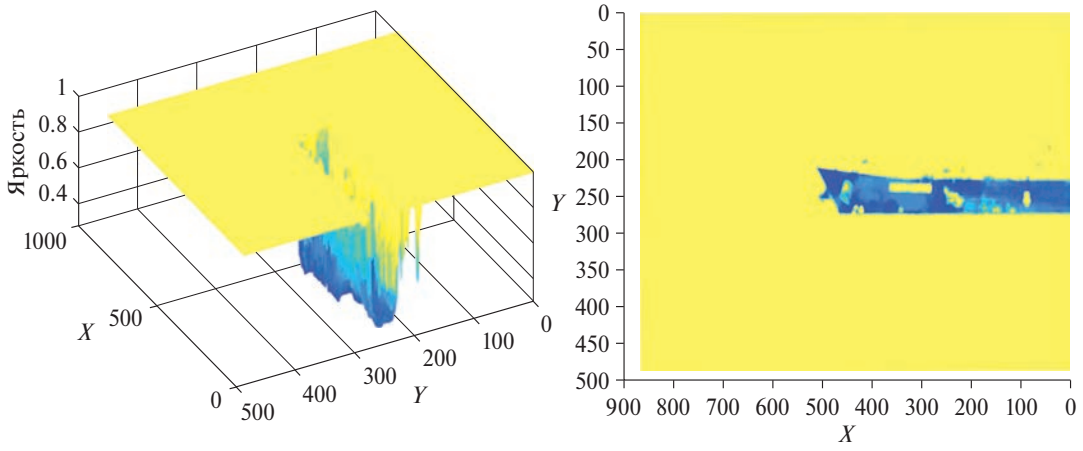

Рис. 5. Визуализация в разных проекциях яркости пикселей изображения после применения порогового преобразования; X, Y – продольная и вертикальная координаты пикселей

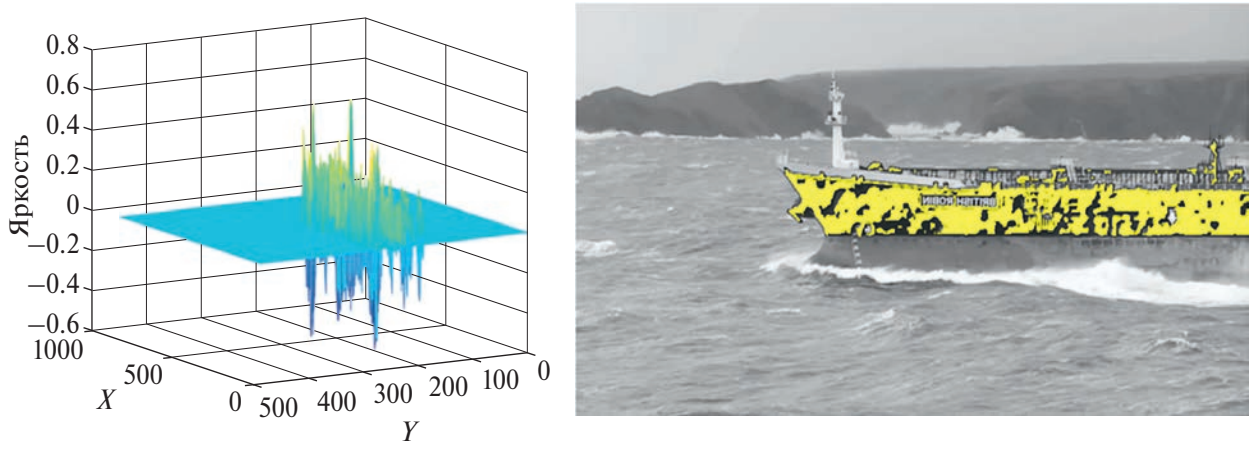

Рис. 6. Визуализация разности двух соседних кадров; X, Y – продольная и вертикальная координаты пикселей

Рис. 7. Визуализация разности соседних кадров на примере судов разного размера, с разной скоростью движения и на различных расстояниях от причала

Группы пикселей, выделенные на первом этапе, двигаются с разной скоростью: часть пикселей не сдвигается дальше своего участка изображения (как правило, эти области принадлежат таким объектам, как бакены, буи, волны) и характер их движения определяется как движение в окрестности заданного центра; некоторые пиксели пропадают от кадра к кадру и появляются снова со смещением; и пиксели, которые смещаются постоянно и имеют четко отслеживаемую траекторию движения. Последняя группа пикселей в соответствии с характером движения может принадлежать судну или его палубным конструкциям.

2.3. Анализ подвижных участков изображения. Вторым этапом является выделение связных компонент (контуров), их анализ и составление карты движущихся участков изображения. Для определения контуров используется алгоритм топологического структурного анализа [7]. Программная реализация алгоритма взята из библиотеки OpenCV 4.7.0. Под контуром подразумевается кривая, соединяющая все непрерывные точки вдоль границы белого объекта на черном фоне. Извлекаются только внешние контуры и сохраняются конечные точки линий, образующих контур (рис. 8). Внутренние контуры не задействуются.

Найденные контуры сортируются от большего к меньшему, и первые 10 из них сохраняются для дальнейшего анализа. Для каждого контура из 10 сохраненных рассчитывается средневзвешенное значение яркости пикселей внутри этого контура, а также центр масс [8], ширина и высота контура (на основе минимальной прямоугольной области, охватывающей контур). На каждом последующем кадре список из 10 контуров анализируется: рассчитывается допустимая величина сдвига по X, Y центра масс контура, величина сдвига составляет 40% от ширины и высоты минимальной прямоугольной области, охватывающей контур: если контур попал в область, вычисленную, на предыдущем шаге, сравниваются размеры контуров, допустимая разница в размерах контуров составляет 30%; иначе контур идентифицируется как новый.

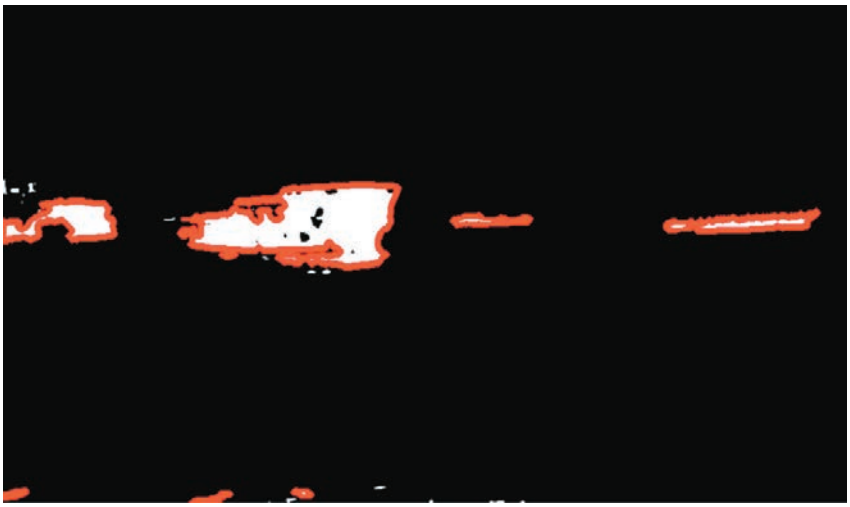

Рис. 8. Визуализация контурного анализа

Процентные соотношения и другие величины, подобраны для использования в задаче позиционирования морского крупнотоннажного судна относительно причала. В задачах позиционирования других крупногабаритных площадных объектов (например, железнодорожных вагонов, грузовых автомобилей) следует выполнить подбор значений указанных величин в соответствии со спецификой задачи (размерами объекта, условиями работы измерителей).

С момента определения контуров в кадре составляется карта движущихся участков изображения и обновляется с каждой новой итерацией программной реализации алгоритма (далее алгоритма). В каждой итерации учитывается количество кадров, прошедшее с момента запуска алгоритма, когда контур идентифицировался в кадре, что позволяет исключить случайно попавшие в кадр движущиеся объекты (птицы, люди). Общая схема алгоритма представлена на рис. 9.

Работа алгоритма проверялась на выборке данных, включающей варианты подхода и швартовок морских крупнотоннажных судов к целевому причалу, на который устанавливаются оптические измерители. Среднее время подхода и швартовки судна (один рабочий цикл оптического измерителя) составляет 25 мин, визуализация работы алгоритма представлена на рис. 10.

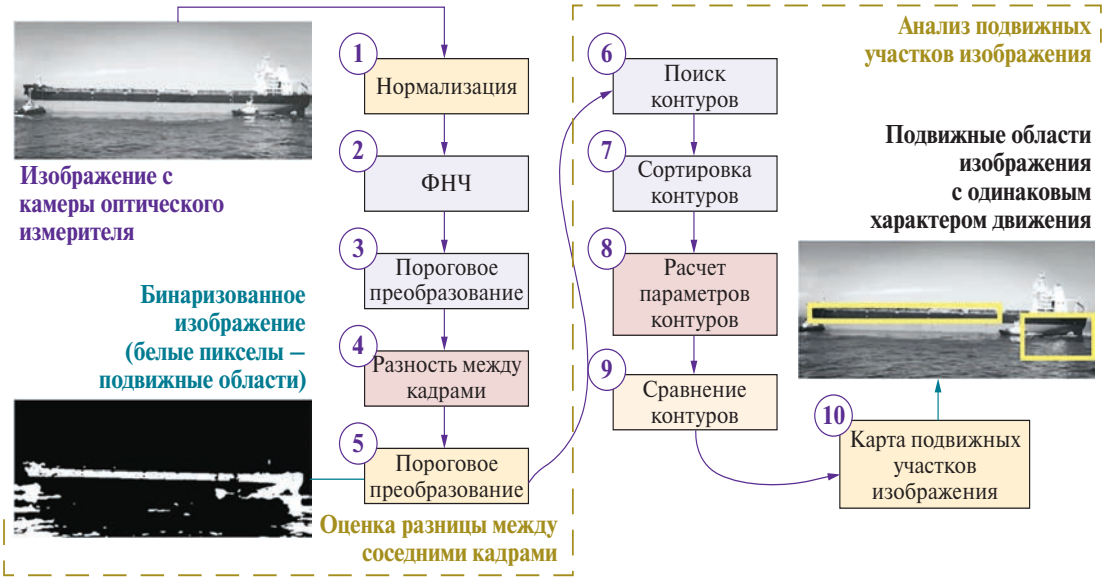

Рис. 9. Схема алгоритма определения движения в кадре и идентификации крупногабаритного площадного объекта. ФНЧ – фильтр низких частот

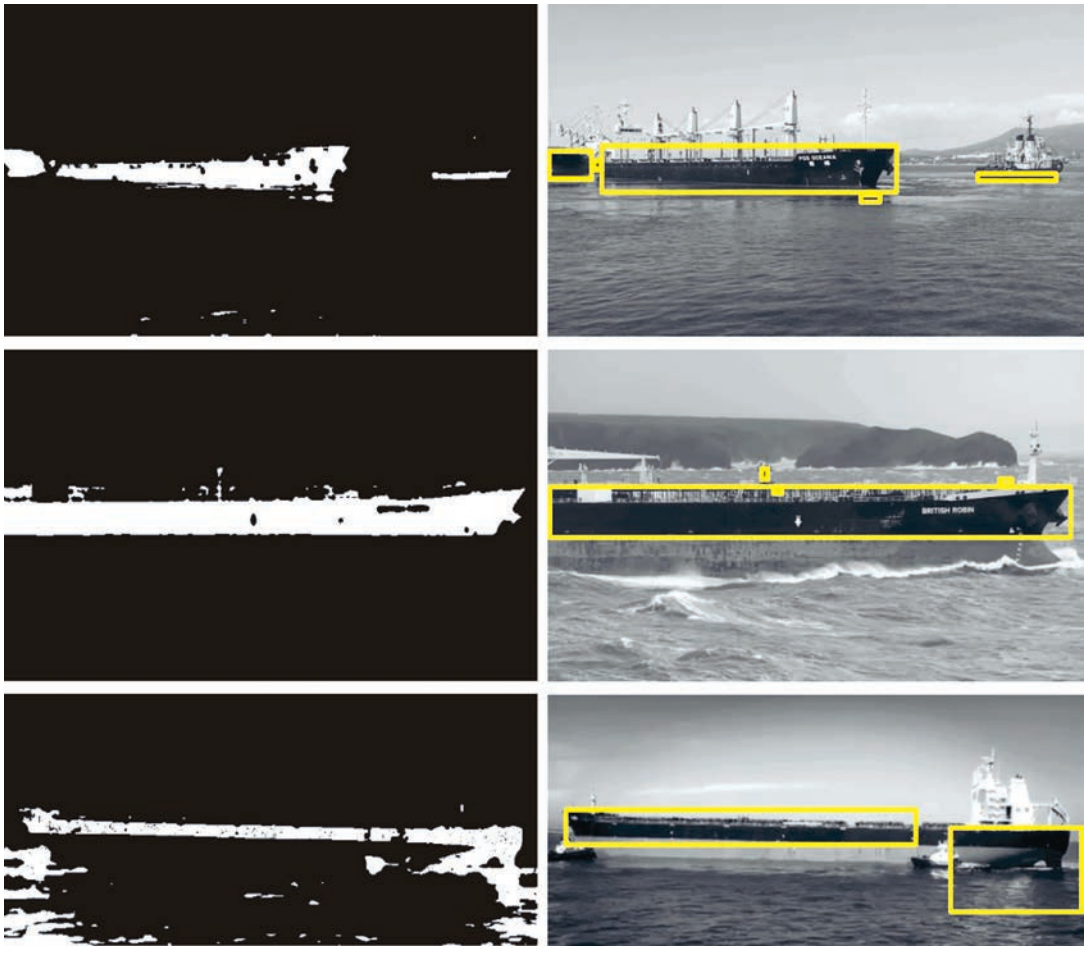

Рис. 10. Примеры движущихся областей изображения (слева) и движущихся областей, потенциально принадлежащих судну и его палубным конструкциям (справа)

Время работы программной реализации алгоритма определения движения в кадре и идентификации крупногабаритного площадного объекта на языке Python версии 3.10.10 составляет в среднем 10.1 мс на 1000 кадров (640×360) при измерении времени функцией Win32 QueryPerformanceCounter, которая возвращает часы, прошедшие с момента первого вызова этой функции, в виде числа с плавающей запятой.

Заключение. Результаты измерений высокоточного измерительного комплекса в соответствии с требованиями стандартов метрологии [9] должны быть детерминированными, т.е. каждый шаг алгоритма должен быть последователен и заранее определен, а используемые методы – верифицируемыми.

Использование в алгоритмическом аппарате высокоточного измерительного комплекса готовых программных реализаций некоторых алгоритмов [10–14] (например, алгоритмы оптического потока, кластеризации, порогового преобразования, поиска границ) или алгоритмов, в основе которых лежит случайный выбор объектов отслеживания, осложняется подбором параметров, непредсказуемостью результатов работы и ложными срабатываниями. Это было подтверждено в ходе испытаний на тестовой выборке видеофайлов, содержащих более чем 50 вариантов подходов и швартовок судов к причалу.

Представленный метод определения движения в кадре и идентификации крупногабаритного площадного объекта позволяет найти координаты подвижных регионов изображения, их размеры, характеристики, необходимые для последующей обработки и вычислений. Полученные подвижные регионы могут быть использованы для выбора объектов отслеживания (подвижных участков изображения) и расчета по смещению этих объектов пространственно-скоростных параметров крупногабаритного площадного объекта – морского крупнотоннажного судна.

Способ решения задачи определения движения в кадре и идентификации крупногабаритного площадного объекта является новым. Реализация отдельных элементов метода выполняется с применением известных подходов, например, поиск связных компонент с помощью топологического структурного анализа. Новой также будет как последовательность действий, обеспечивающая результат, так и отдельные компоненты метода.

Метод определения движения в кадре и идентификации крупногабаритного площадного объекта позволяет определить области изображения с постоянным движением и исключить случайно попавшие в кадр подвижные объекты. Он имеет высокий потенциал применения для построения систем компьютерного зрения, обеспечивающих обратную связь для высокоточного позиционирования крупногабаритных объектов в пространстве, что является востребованным для повышения эффективности многих отраслей экономики.

СПИСОК ЛИТЕРАТУРЫ

1. *Poujouly S., Journet B.* A Twofold Modulation Frequency Laser Range Finder // J. Optics A: Pure and Applied Optics. 2002. № 4. P. 356–363.
2. *Zheng X.Y., Zhao C., Zhang H.Y., Zheng Z., Yang H.Z.* Coherent Dual-frequency Lidar System Design for Distance and Speed Measurements // Intern. Conf. on Optical Instruments and Technology: Advanced Laser Technology and Applications. International Society for Optics and Photonics. Beijing, China, 2018. V. 10619.
3. *Jia F.X., Yu J.Y., Ding Z.L., Yuan F.* Research on Real-time Laser Range Finding System // Applied Mechanics and Materials. 2013. V. 347.
4. *Beraldin J.A., Steenaart W.* Overflow Analysis of a Fixed-Point Implementation of the Goertzel Algorithm // IEEE Transactions on Circuits and Systems. 1989. V. 36. № 2. P. 322–324.
5. *Finlayson D.M., Sinclair B.* Advances in Lasers and Applications // Boca Raton, Florida, USA. CRC Press, 1998. P. 346.
6. *Lopatina V.V.* Method of Fragment Based Tracking of Displacement of a Large Areal Object in Images // J. Phys.: Conf. Ser. 2021. V. 2061. P. 012113.
   https://doi.org/10.1088/1742-6596/2061/1/012113.
7. *Suzuki S., Keiichi A.* Topological Structural Analysis of Digitized Binary Images by Border Following // Comput. Vis. Graph. Image Process. 1985. V. 30. P. 32–46.
8. OpenCV 4.7.0. Open Source Computer Vision Library, 2022.
9. ГОСТ Р 8.1030-2024.
10. *Lucas B.D., Kanade T.* An Iterative Image Registration Technique with an Application to Stereo Vision // Intern. Joint Conf. on Artificial Intelligence. Vancouver, B.C., Canada, 1981.
11. *Farnebäck G.* Two-Frame Motion Estimation Based on Polynomial Expansion // Image Analysis (SCIA). Lecture Notes in Computer Science, Eds J. Bigun, T. Gustavsson. Berlin, Heidelberg: Springer, 2003. V. 2749.
12. *Jain A., Murty M., Flynn P.* Data Clustering: A Review // ACM Computing Surveys. 1999. V. 31. P. 264–323.
13. *Otsu N.* A Threshold Selection Method from Gray-level Histograms // IEEE Trans. Sys., Man., Cyber. J. 1979. V. 9. P. 62–66.
14. *Гонсалес Р., Вудс Р.* Цифровая обработка изображений. Изд. 3-е, исправ. и доп. М.: Техносфера, 2019. 1104 с. ISBN 978-5-94836-331-8.

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, с. 148–153

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

УДК 519.8, 510.71, 510.22

МЯГКИЕ МНОЖЕСТВА (ОБЗОР)

© 2024 г. В. Н. Бобылев[a], *, Е. К. Егорова[a], **, В. Ю. Леонов[a], ***

[a]ФИЦ ИУ РАН, Москва, Россия

**e-mail: vbobylev@frccsc.ru*

***e-mail: eegorova@frccsc.ru*

****e-mail: vleonov@frccsc.ru*

Поступила в редакцию 29.02.2024 г.

После доработки 25.03.2024 г.

Принята к публикации 13.05.2024 г.

Рассматриваются так называемые мягкие множества. По сути дела, речь идет об обобщении нечетких множеств Л. Заде, которые формируют, в частности, математический аппарат искусственного интеллекта. С другой стороны, отказ от понятия инфинитезимальности зарождает основы нового математического анализа. Впоследствии появилось много статей по мягким множествам, организовывались конференции, имеются публикации о приложениях в различных областях. Приведены основные определения и термины теории мягких множеств, даны ссылки на практические приложения данной теории.

Ключевые слова: мягкие множества, операции над мягкими множествами, мягкий предел, мягкий интеграл, мягкий дифференциал, рациональный анализ.

DOI: 10.31857/S0002338824040102 **EDN:** UDWLOK

A SOFT SETS REVIEW

V. N. Bobylev[a],*, E. K. Egorova[a], **, V. Yu. Leonov[a], ***

[a]Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russia

**e-mail: vbobylev@frccsc.ru*

***e-mail: eegorova@frccsc.ru*

****e-mail: vleonov@frccsc.ru*

In this review we consider the so-called soft sets. In fact, it is a generalization of L. Zadeh's fuzzy sets, which form the mathematical apparatus of artificial intelligence. On the other hand, the rejection of the notion of infinitesimality originates the foundations of a new mathematical analysis. Subsequently, many papers on soft sets have appeared, conferences have been organized, and there are publications on applications in various fields. The paper gives the basic definitions and terms of soft sets theory, and references to its practical applications.

Keywords: Soft Set, Soft Sets Operations, Soft Limit, Soft Integral, Soft Differential, Rational Analysis.

Введение. С развитием вычислительной техники и выделением таких областей, как теория игр, теория принятия решений, теория оптимизации, появилась потребность в расширении аппарата классической теории множеств. Применение данных теорий на практике связано с использованием данных, содержащих разного рода неопределенности. Классическая теория множеств подразумевает определение для элементов универсального множества функции принадлежности, которая принимает значение из множества $\{0,1\}$ в зависимости от того, принадлежит рассматриваемый элемент множеству или нет. Расширить понятие функции принадлежности предложил Л. Заде в 1965 г. [1]. В нечетком множестве функция может принимать значение из замкнутого интервала $[0,1]$.

В дальнейшем появились несколько десятков различных обобщений и вариаций понятия нечеткого множества. Одним из таких расширений стали мягкие множества — термин, впервые введенный Д. А. Молодцовым (1948–2020) в [2]. Его подход отличается от подхода Заде, но, как показано в [2], нечеткие множества являются одним из представлений мягких множеств.

Цель статьи – ознакомление читателей с аппаратом данной теории, который находит свое применение во многих областях.

1. Основные определения. Мягкие множества были введены Д.А. Молодцовым для разрешения непределенностей в задачах теории игр. В первых публикациях в данной области [3–5] было показано, что устойчивости в играх с передачей информации можно добиться введением параметризующего множества для описания решения, которое получило название *принцип оптимальности* [3].

О п р е д е л е н и е 1 [3]. Под принципом оптимальности понимается отображение

$$R : m \to 2^C,$$

где m – модель операции, C-множество выборов оперирующей стороны, в общем случае зависящее от m. Позднее идея такого множества была обобщена и получила название *мягкое множество.*

О п р е д е л е н и е 2 [2]. Пара (S, A) называется мягким множеством над U, если S – отображение из A в множество всех подмножеств U, т.е. $S : A \to 2^U$.

О п р е д е л е н и е 3 [6]. Пару (F, A) будем называть мягким отображением (мягкой функцией) из M в U (где M – множество моделей), если F является отображением из множества $M \times A$ в множество подмножеств универсального множества U, т. е. $F : M \times A \to 2^U$.

Наиболее цитируемая работа [2], обобщающая полученные результаты, вышла в 1999 г. Она была опубликована на английском языке и положила начало широкому распространению аппарата теории мягких множеств для решения различных задач. Кроме определения мягких множеств и операций над ними в статье вводится понятие мягкой функции и рассматриваются свойства таких функций. Предлагается использование мягких функций в качестве функций выбора в задачах, связанных с неопределенностью стратегий в исследовании игр и теории операций, что должно упростить решение задачи в условиях расплывчатой и неопределенной информации. В [7] функция выбора определяется как

$$R : 2^P \times E \to 2^P,$$

где E – множество параметров ε, P – множество ситуаций, 2^P – множество всех подмножеств множества P, т. е. если $X \subseteq P$ – допустимое подмножество ситуаций, то $R(X, \varepsilon)$ множество – ε-птимальных ситуаций. При наличии неопределенных факторов мягкое множество стратегий задается следующим способом:

$$Q(C, \varepsilon) = \left\{ c \in C \middle| \pi(c) \in R(\pi(c), \varepsilon) \right\},$$

где C – множество стратегий лица принимающего решение, $\pi : C \to \mathcal{P}(P)$. Определяется понятие гладкость, являющееся аналогом непрерывности функции. (По утверждению автора, каждая мягкая функция порождает свою топологию и переход к мягкой функции делает проблему устойчивой.) Обычно под устойчивостью функции понимается малое изменение значения функции при малом изменении значения ее аргумента. Но отсутствие устойчивости характерно для многих явлений в физике, экономике и других областях. Непрерывность мягких отображений, аналогично классической непрерывности отображений, позволяет обосновывать замену задачи поиска решений по приближенной информации и применение приближенных численных методов.

Но следует отметить, что быстрое развитие аппарата теории мягких множеств повлекло за собой некоторые сложности с корректным описанием операций над мягкими множествами. Изначально операции с мягкими множествами были определены как

$$\Theta\big((S', A), (S'', B)\big) = (S, A \times B),$$

где $S(\alpha, \beta) = \Theta\big(S'(\alpha), S''(\beta)\big)$, $\alpha \in A$, $\beta \in B$. При этом получается результирующее мягкое множество, параметризованное парой параметров α, β. Различные их виды вводились в [8–10]. Но некоторые утверждения, приведенные в [9], оказываются неверны [10, 11]. Поэтому в [10] вводятся новые понятия: ограниченное объединение, ограниченное пересечение и др., причем показывается, что для новых понятий законы де Моргана справедливы.

Возвращение к этой теме и определение того, какие операции над мягкими множествами являются корректными, было предпринято Молодцовым в [12].

О п р е д е л е н и е 4 [12]. Если задано мягкое множество (S, A), то задано семейство $\mathfrak{S}(S,A) = \{S(a) | a \in A\}$. Два мягких множества (S, A), (S', A'), определенных над универсальным множеством X называются эквивалентными тогда и только тогда, когда $\mathfrak{S}(S,A) = \mathfrak{S}(S',A')$. Эквивалентные мягкие множества запишем как $(S,A) \cong (S',A')$.

Сформулировано понятие корректности операций над мягкими множествами.

О п р е д е л е н и е 5 [12]. Унарная операция Φ азывается корректной, если для любой пары эквивалентных мягких множеств (S, A), (S', A'), заданных над универсальным множеством U, выполнено $\Phi(S,A) \cong \Phi(S',A')$.

О п р е д е л е н и е 6 [12]. Бинарная операция Θ называется корректной, если для любой четверки попарно эквивалентных мягких множеств $(S,A) \cong (S',A')$, $(T,B) \cong (T',B')$, заданных над универсальным множеством U, выполнено

$$\Theta\big((S,A),(T,B)\big) \cong \Theta\big((S',A',),(T',B')\big).$$

2. Мягкий анализ. На основе мягких множеств разработан мягкий анализ. В [2] предпринимается попытка сформулировать аппарат анализа, основанный на теории мягких множеств: мягкие верхний и нижний пределы, мягкое приближение (аналог дифференциала), мягкие аналоги интеграла. Особый интерес представляет доказательство идентичности мягких аналогов интеграла по Риману и по Перрону.

О п р е д е л е н и е 7 [2]. Мягким верхним (ε,τ)-пределом функции f в точке x называется множество

$$\overline{\mathrm{Softlimit}}\,[f,\varepsilon,\tau](x) = \{v \in \mathbb{R} | f(y) \le v + \varepsilon, \forall y \in \tau(x)\},$$

а мягким нижним (ε,τ)-пределом функции f в точке x – множество

$$\underline{\mathrm{Softlimit}}\,[f,\varepsilon,\tau](x) = \{v \in \mathbb{R} | f(y) \ge v - \varepsilon, \forall y \in \tau(x)\}.$$

Множество

$$\mathrm{Softlimit}\,[f,\alpha,\beta,\tau](x) = \{v \in \mathbb{R} | v - \alpha \le f(y) \le v + \beta, \forall y \in \tau(x)\}$$

называется мягким (α, β, τ)-пределом функции f в точке x.

О п р е д е л е н и е 8 [2]. Множество

$$\bar{D}[f,\alpha,\beta,\tau](x) = \{v \in \mathbb{R} | f(y) \le f(x) + (v + \alpha(x))(y - x) + \beta(x), \forall y \in \tau(x)\}$$

называется верхним (α, β, τ)-приближением функции f в точке x, а множество

$$\underline{D}[f,\alpha,\beta,\tau](x) = \{v \in \mathbb{R} | f(y) \ge f(x) + (v - \alpha(x))(y - x) - \beta(x), \forall y \in \tau(x)\}$$

– нижним (α, β, τ)-пределом функции f в точке x.

Набор верхних и нижних (α, β, τ)-приближений образует верхние и нижние мягкие приближения. Под мягким приближением D подразумевается пересечение верних и нижних мягких приближений:

$$D[f,\alpha,\beta,\gamma,\delta,\tau] = \bar{D}[f,\alpha,\beta,\tau](x) \cap \underline{D}[f,\gamma,\delta,\tau](x).$$

В дальнейшем разработанный аппарат анализа был применен для формулировки оптимизационных задач [13].

3. Рациональный анализ. Обобщением полученных результатов в области мягкого анализа является серия работ, посвященная началам рационального анализа. Так как при численном решении задач найденные значения являются лишь приближениями к иррациональностям, то логичным шагом стало использование мягких множеств – аппарата, предназначенного для работы с неопределенностями, – для построения анализа на базе рациональных чисел.

Так, в работе [14] вводится понятие рационального числа.

О п р е д е л е н и е 9 [14]. Мягким рациональным числом называется пара (S, A), где S – отображение $S : A \to 2^{\mathbb{Q}}$.

Как отмечено еще в [2], каждое мягкое отображение описывает свою собственную топологию. Далее последовательно даются определения в терминах мягких множеств верхней и нижней граней множества рациональных чисел, минимального и максимального элементов, окрестности множества. На этой основе формулируется понятие мягкого предела рациональной функции и мягкой непрерывности.

О п р е д е л е н и е 10 [14]. Отображение $\tau : \mathbb{Q} \to 2^{\mathbb{Q}}$, для которых $Dom(\tau) = \mathbb{Q}$, называется отображением близости. Значение отображения близости интерпретируется как множество точек, τ-близких к точке x. Множество отображений близости обозначается $\mathbb{P} = \left\{\tau | \tau \in \left(\mathbb{Q}, 2^{\mathbb{Q}}\right)\right\}$.

О п р е д е л е н и е 11 [14]. Обратное отображение $\tau^{\leftarrow} : \mathbb{Q} \to 2^{\mathbb{Q}}$ определяется как

$$\tau^{\leftarrow}(y) = \{x \in \mathbb{Q} | y \in \tau(x)\}.$$

О п р е д е л е н и е 12 [14]. Рассмотрим функцию $f \in \Phi$ с областью определения $Dom(f) \subseteq \mathbb{Q}$ и μ, τ – отображения близости. Число $y \in \mathbb{Q}$ называется мягким (τ,μ)-пределом функции f в точке $x \in \mathbb{Q}$, если $f \circ \tau(x) \subseteq \mu(y)$.

О п р е д е л е н и е 13 [14]. Функция f называется (τ,μ)-непрерывной в точке $x \in Dom(f)$, если справедливо включение $f \circ \tau(x) \subseteq \mu \circ f(x)$.

О п р е д е л е н и е 14 [14]. Функция f называется (τ,π)-непрерывной на множестве $X \subseteq Dom(f)$, если для любого $x \in X$ верно включение $f \circ \tau(x) \subseteq \pi[x] \circ f(x)$, где π – мягкое отображение близости $\pi : \mathbb{Q} \to \mathbb{P}$. Для того, чтобы отличить аргумент-параметр от аргумента-точки, в дальнейшем он будет указываться в квадратных скобках.

Дальнейшее развитие рационального анализа продолжилось в [15], где предлагаются два подхода к построению мягкой производной рациональной функции.

Первый способ основан на идее, что производная функции f в точке x должна характеризовать скорость изменения функции на множестве $\tau(x)$. Вторая идея заключается в подходе к производной как к угловому коэффициенту линейной функции, приближающей исходную функцию. Свойства полученных мягких производных и дифференциалов рассматриваются применительно к лемме Ферма о локальном экстремуме и показывается, что мягкий аналог леммы устойчив к возмущениям. Также в этой работе вводится понятие мягкого интеграла.

О п р е д е л е н и е 15 [14]. Последовательность чисел $x = (x_1,\ldots,x_n)$, $n > 1$, называется τ-путем, если для любых $i = 1,\ldots,n-1$ выполнено $x_{i+1} \in \tau(x_i)$. Множество τ-путей с начальной точкой x и конечной точкой y обозначается $Path(x,y,\tau)$.

О п р е д е л е н и е 16 [15]. Множество

$$\mathcal{I}_x^y[\varphi,\tau,\mu] = \bigcap_{z \in Path(x,y,\tau)} \sum_{i=1}^{n-1} (z_{i+1} - z_i)\mu^{\leftarrow}[z_i](\varphi(z_i))$$

называется (τ,μ)-интегралом функции φ от x до y.

Для (μ,τ)-интеграла сформулированы достаточные условия существования и рассматриваются аналоги некоторых свойств, в частности свойства дифференцируемости.

Последней работой по данной тематике стала [16], в которой определен градиент функции в многомерном случае. При помощи мягкого градиента формулируются условия для приближенного локального экстремума. При использовании мягкого градиента предлагается построить различные аналоги производной классического анализа, например производной по направлению. При этом показано, что задача численного нахождения мягкого градиента сводится к решению конечной системы линейных неравенств.

4. Мягкие дифференциальные уравнения. Еще одним приложением стала формулировка мягких аналогов определенных дифференциальных уравнений первого порядка [17], разрешенных относительно производной. Для одного из типов уравнений приводится соответствующее мягкое интегральное уравнение. Находится решение мягкой задачи Коши для вещественной и интервальной функций.

О п р е д е л е н и е 17 [17]. Множество

$$\mathcal{D}[y,x;h,\varepsilon] = \left\{v \in E | y(x) + v\Delta x - \varepsilon \le y(x+\Delta x) \le y(x) + v\Delta x + \varepsilon, \forall \Delta x \in \left(0, h(x)\right)\right\}$$

называется (h,ε)-приближенным дифференциалом функции y в точке x. Здесь h, ε – вещественные функции, которые играют роль параметров, описывающих приближенное понятие. Функция h определяет близкие к x справа точки, а функция ε – точность аппроксимации. При фиксированных значениях y и x приближенный дифференциал можно рассматривать как мягкое множество над вещественной прямой.

По аналогии с обыкновенным дифференциальным уравнением первого порядка, разрешенным относительно производной, строятся два типа мягких дифференциальных уравнений.

О п р е д е л е н и е 18 [17]. Мягкое дифференциальное уравнение типа A имеет вид

$$f(x,y) \subseteq \mathcal{D}[y,x;h,\varepsilon].$$

Здесь f – функция двух вещественных аргументов, значениями которой являются подмножества вещественной оси, в частности вещественные числа.

О п р е д е л е н и е 19 [17]. Мягкое дифференциальное уравнение типа B записывается как

$$f(x,y) \supseteq \mathcal{D}[y,x;h,\varepsilon].$$

Здесь f – функция двух вещественных аргументов, значениями которой являются подмножества вещественной оси.

Для задачи исследуется вопрос существования мягких решений, изучается зависимость решения от начальных условий. При рассмотрении мягкой задачи Коши определяется и соответствующее интегральное уравнение. Показано, что достаточные условия существования решения мягкой задачи Коши также являются достаточными условиями существования мягкого интегрального уравнения.

5. Обобщения и применение мягких множеств. Подробное обсуждение моделей поведения человека и формулировки математических постановок задач с использованием принципа оптимальности приведено в [7]. Рассматриваются задачи на максимум в случае независимых и связанных ограничений, на поиск равновесия, задачи оптимизации в игровой обстановке, иерархические игры. Автор отмечает, что сведение сложных задач вариационного типа к экстремальным задачам на исходных множествах удалось при единственном предположении об ограниченности целевых функций, не потребовалось ни непрерывности, ни компактности, ни дополнительных условий регулярности. Отдельно постановка задачи на максимин приводится в [6]. В работе рассматриваются способы ослабления условий устойчивости таких задач. Примечательно, что устойчивость мягкого отображения не требует ограничения на непрерывность или полунепрерывность функций модели.

Предложения по применению аппарата теории мягких множеств в задачах теории принятия решений [18] и дальнейшее рассмотрение мягких множеств в области алгебры [19–22] позволили развить методы многокритериального принятия решений, которые используются в медицинских целях для постановки диагнозов.

В [23] предложена концепция нейросети на основе мягких множеств, а также мягких нечетких множеств.

Заключение. Несмотря на то, что теория мягких множеств создавалась как инструмент для решения задач теории игр и в основном применяется при решении оптимизационных задач, ее потенциал намного шире. Это, в частности, показывает разработанный на базе мягких множеств аппарат рационального анализа. Главной тенденцией в области оптимизации выступает дальнейшая модификация мягких множеств и появление гибридных моделей.

СПИСОК ЛИТЕРАТУРЫ

1. *Zadeh L.A.* Fuzzy Sets // Inf. Control. 1965. V. 8. № 3. P. 338–353. ISSN 0019-9958. https://doi.org/10.1016/S0019-9958(65)90241-X.
2. *Molodtsov D.A.* Soft Set Theory – First Results // Computers & Mathematics with Applications. 1999. V. 37. № 4/5. P. 19–31. ISSN 0898-1221, 1873-7668. https://doi.org/10.1016/s0898-1221(99)00056-5.
3. *Молодцов Д.А.* Устойчивость и регуляризация принципов оптимальности // ЖВМиМФ. 1980. Т. 20. № 5. С. 25–38. ISSN 0041-5553. https://doi.org/10.1016/0041-5553(80)90086-5.
4. *Молодцов Д.А.* Аппроксимация принципов оптимальности в задаче нахождения кратного максмина // Докл. АН СССР. 1985. Т. 32. С. 426–428. ISSN 0197-6788.
5. *Молодцов Д.А.* Структура регуляризующих принципов оптимальности // Докл. АН СССР. 1985. Т. 32. С. 82—85. ISSN 0197-6788.
6. *Молодцов Д.А., Ковков Д.В.* Устойчивость и аппроксимация максиминных задач // АиТ. 2014. Т. 75. № 3. С. 447–457. ISSN 0005-1179, 1608-3032. https://doi.org/ 10.1134/S0005117914030035.
7. *Молодцов Д.А.* Принципы оптимальности как математическая модель поведения человека // Математическое моделирование. 1991. Т. 3. № 5. С. 29–48. ISSN 0234-0879.
8. *Ma Z.M., Yang W., Hu B.Q.* Soft Set Theory Based on Its Extension // Fuzzy Information and Engineering. 2010. V. 2. № 4. P. 423–432. ISSN 1616-8658. https://doi.org/10.1007/s12543-010-0060-7.

9. *Maji P.K., Biswas R., Roy A.R.* Soft Set Theory // Computers & Mathematics with Applications. 2003. V. 45. № 4. P. 555–562.
ISSN 0898-1221. https://doi.org/10.1016/S0898-1221(03)00016-6.
10. *Ali M.I., Feng F., Liu X., Min W.K., Shabir M.* On Some New Operations in Soft Set Theory // Computers & Mathematics with Applications. 2009. V. 57. № 9. P. 1547–1553.
ISSN 0898-1221, 1873-7668. https://doi.org/10.1016/j.camwa.2008.11.009.
11. *Yang C.F.* A Note on "Soft Set Theory" [Comput. Math. Appl. 45 (4–5) (2003) 555–562] // Computers & Mathematics with Applications. 2008. V. 56. № 7. P. 1899–1900.
ISSN 0898-1221. https://doi.org/10.1016/j.camwa.2008.03.019.
12. *Молодцов Д.А.* Структура мягких множеств // Нечеткие системы и мягкие вычисления. 2017. Т. 12. № 1. С. 5–18. ISSN 1819-4362.
13. *Kovkov D.V., Kolbanov V.M., Molodtsov D.A.* Soft Sets Theory-based Optimization // J. Computer and Systems Sciences International. 2007. V. 46. № 6. P. 872–880.
ISSN 1064-2307, 1555-6530. https://doi.org/10.1134/S1064230707060032.
14. *Молодцов Д.А.* Начала рационального анализа – непрерывность функций // Нечеткие системы и мягкие вычисления. 2019. Т. 2. С. 126–141.
ISSN 18194362. https://doi.org/10.26456/fssc57.
15. *Молодцов Д.А.* Начала рационального анализа – производные и интегралы // Нечеткие системы и мягкие вычисления. 2020. Т. 15. № 1. С. 5–25.
ISSN 1819-4362. https://doi.org/10.26456/fssc70.
16. *Acharjee S., Molodtsov D.A.* Soft Rational Line Integral // Vestnik Udmurtskogo Universiteta. Matematika. Mekhanika. Komp'yuternye Nauki. 2021. V. 31. № 4. P. 578–596.
ISSN 2076-5959, 1994-9197. https://doi.org/10.35634/vm210404.
17. *Молодцов Д.А.* Мягкое дифференциальное уравнение // ЖВМиМФ. 2000. Т. 40. № 8. С. 1116–1128. ISSN 0965-5425.
18. *Maji P.K., Roy A.R., Biswas R.* An Application of Soft Sets in a Decision Making Problem // Computers & Mathematics with Applications. 2002. V. 44. № 8. P. 1077–1083.
ISSN 0898-1221. https://doi.org/10.1016/S0898-1221(02)00216-X.
19. *Aktas H., Cagman N.* Soft Sets and Soft Groups // Information Sciences. 2007. V. 177. № 13. P. 2726–2735.
ISSN 0020-0255, 1872-6291. https://doi.org/10.1016/j.ins.2006.12.008.
20. *Park C.H., Jun Y.B., Ozturk M.A.* Soft WS-algebras // Communications of the Korean Mathematical Society. 2008. V. 23. № 3. P. 313–324.
ISSN 1225-1763. https://doi.org/10.4134/CKMS.2008.23.3.313 ; Publisher: Korean Mathematical Society.
21. *Jun Y.B., Park C.H.* Applications of Soft Sets in Ideal Theory of BCK/BCI-algebras // Information Sciences. 2008. V. 178. № 11. P. 2466– 2475.
ISSN 0020-0255. https://doi.org/10.1016/j.ins.2008.01.017.
22. *Ma X., Zhan J., Xu Y.* Lattice Implication Algebras Based on Soft Set Theory // Computational Intelligence. World Scientific, 2010. P. 535–540.
ISBN 978-981-4324-69-4. https://doi.org/10.1142/9789814324700 0080.
23. *Liu Z., Alcantud J.C.R., Qin K., Xiong L.* The Soft Sets and Fuzzy Sets-Based Neural Networks and Application // IEEE Access. 2020. V. 8. P. 41615–41625.
https://doi.org/10.1109/ACCESS.2020.2976731.

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, с. 154–166

НАВИГАЦИОННЫЙ СИСТЕМЫ

УДК 629.783

НАХОЖДЕНИЕ ОПТИМАЛЬНОГО ВЕКТОРА ПРИЗНАКОВ ДЛЯ ОПРЕДЕЛЕНИЯ КОНТЕКСТА ОКРУЖАЮЩЕЙ СРЕДЫ ПО ДАННЫМ ГЛОБАЛЬНЫХ НАВИГАЦИОННЫХ СПУТНИКОВЫХ СИСТЕМ

© 2024 г. А. И. Болкунов[a], В. В. Кульнев[a], Е. В. Кульнев[a], Е. О. Наконечный[a,*], В. И. Яремчук[a]

[a]АО «ЦНИИмаш», Королёв, Россия

**e-mail: nakonechnyieo@tsniimash.ru*

Поступила в редакцию 16.02.2024 г.

После доработки 10.03.2024 г.

Принята к публикации 13.05.2024 г.

В глобальных навигационных спутниковых системах показатели качества позиционирования зависят как от условий окружающей среды, так и от поведения потребителя. Окружающая среда влияет на качество приема радиосигналов, которые доступны для позиционирования. Для работы в различных условиях окружающей среды требуется адаптивное навигационное решение, которое будет определять тип окружающей среды и применять различные методы для навигационного решения. Рассматриваются признаки, формируемые по данным принимаемых навигационных сигналов, которые могут быть использованы для определения типа окружающей среды. Настоящая статья посвящена нахождению оптимального вектора признаков для определения типа окружающей среды по информации от глобальных навигационных спутниковых систем. Собраны экспериментальные навигационные данные для различных типов окружающей среды. Рассмотрены критерии и методы определения оптимального вектора признаков с помощью алгоритмов из математической статистики. Предложен оптимальный вектор признаков, который вносит наибольший вклад в определение различных типов окружающей среды.

Ключевые слова: ГНСС, навигационные сигналы, контекст окружающей среды, вектор признаков

DOI: 10.31857/S0002338824040118 **EDN:** UDUYPP

FINDING THE OPTIMAL FEATURE VECTOR FOR DETECTING THE ENVIRONMENTAL CONTEXT FROM GLOBAL NAVIGATION SATELLITE SYSTEMS DATA

A. I. Bolkunov[a], V. V. Kulnev[a], E. V. Kulnev[a], E. O. Nakonechnyi[a,*], V. I. Yaremchuk[a]

[a]JSC «TsNIImash», Korolev, Russia

**e-mail: nakonechnyieo@tsniimash.ru*

In global navigation satellite systems, positioning quality indicators depend on both environmental conditions and user behaviour. The environment affects the reception quality of the radio signals that are available for positioning. An adaptive navigation solution is required to operate in different environmental conditions, which will detect the type of environment and apply different methods for navigation solution. The features formed from the received navigation signal data that can be used to determine the type of environment are discussed. This paper is devoted to finding an optimal feature vector for determining the type of environment from information from global navigation satellite systems. Experimental navigation data for different types of environment are collected. Criteria and methods for determining the optimal feature vector using algorithms from mathematical statistics are considered. The optimal feature vector that contributes the most to the determination of different types of environment is proposed.

Keywords: GNSS, navigation signals, environmental context, feature vector

Введение. Глобальные навигационные спутниковые системы (ГНСС) являются основой современной навигации благодаря высокой точности, охвату и низкой стоимости абонентских терминалов. В последние годы спутниковое позиционирование стало наиболее эффективным в связи с увеличением количества космических аппаратов (КА), совершенствованием навигационной аппаратуры потребителя (НАП) и разработкой новых алгоритмов, которые повышают точность навигации [1].

Однако характеристики внешних условий и поведение потребителя обычно отличаются от средних показателей, учтенных в алгоритмах НАП, что может привести к ухудшению точности местоопределений в НАП, например из-за приема сигналов вне зоны прямой видимости или эффекта многолучевости [1]. Данные эффекты особенно актуальны в плотной городской застройке, в которой ошибки позиционирования могут составлять значения до десятков метров [1]. Основная проблема заключается в том, что большинство алгоритмов, направленных на уменьшение влияния указанных выше факторов, эффективно работают только в условиях определенной окружающей среды (контексте окружающей среды) и алгоритмы навигации требуется адаптировать под реальные условия окружающей среды [2]. В контексте окружающей среды обычно выделяют классы окружающей среды (например, открытая местность, закрытая и др.) и типы окружающей среды (морская поверхность, воздушное пространство и т.п.) [2]. В связи с этим важным становится понимание окружающего контекста (ОК) потребителя и выбор подходящего алгоритма для увеличения качества навигации [2]. ОК обычно определяется путем анализа степени соответствия характерных признаков анализируемой среды возможным классам и типам окружающей среды. Характерные признаки среды могут быть найдены и проанализированы различными методами.

В настоящие момент часто применяемым подходом для распознавания реального ОК выступает техническое зрение (путем обработки изображений с камер или лидаров), которое широко используется в автономных транспортных средствах и роботизированной промышленности. Однако применение средств технического зрения совместно с ГНСС требует высокого затрат энергопотребления, вычислительных ресурсов, большой массы, габаритов и стоимости оборудования и не всегда возможно для массового потребителя. Другим подходом, который позволяет устранить указанные выше недостатки, является непосредственное использование данных от ГНСС [2]. Благодаря принимаемым с КА навигационным сигналам можно сформировать характерные признаки (вектор признаков), которые изменяются в зависимости от ОК.

В [3] рассмотрен комплексный подход для контекстно-адаптивной навигации, заключающийся в применении ГНСС, Wi-Fi и инерциальных навигационных систем. Для определения ОК по данным ГНСС рассчитывалось среднее значение отношения мощности несущей к мощности шума на единицу полосы пропускания C/No дБГц. В [4] демонстрируется определение ОК с помощью модуля ГНСС в смартфоне. Для определения внешней или внутренней среды из данных смартфона извлекается C/No и количество видимых спутников, затем для классификации применяется схема обнаружения с помощью скрытой модели Маркова. В [5] для определения типа ОК по данным от НАП предлагается формировать многомерный вектор признаков, состоящий из среднего значения мощности сигнала C/No, среднеквадратичного отклонения (СКО) мощности сигнала C/No, коэффициента блокировки спутников, суммарного геометрического фактора, расширенного геометрического фактора, количества видимых спутников, которые извлекаются из принимаемых данных навигационного протокола NMEA. Предложенный алгоритм учитывает измерения C/No для открытой среды в других рассматриваемых средах, а также скорость движения потребителя, тем самым дополняя вектор признаков. В [6] представлен метод определения ОК по следующим признакам: среднее значение C/No, СКО мощности сигнала C/No, медиана C/No, верхние и нижние квартили C/No, количество видимых спутников, геометрический фактор по местоположению, горизонтали и вертикали. В [7] описана мобильная наземная платформа для сбора необходимых данных с увеличенной частотой измерений информации, в которой используются следующие признаки: количество видимых спутников, среднее значение маски для количества видимых спутников по C/No, среднее значение для C/No для предыдущих измерений спутников, остаток ошибки псевдодальности и среднее значение угла места.

В рассмотренных выше статьях представлены основные признаки для определения ОК, однако во всех данных работах отсутствует комплексный подход для формирования оптимального вектора признаков, который бы содержал только наиболее значимые признаки. Кроме того, для НАП массового применения нахождение оптимального вектора признаков важно также для минимизации используемых вычислительных ресурсов и энергопотребления. Предлагаемый в работе оптимальный вектор признаков для определения ОК в НАП может быть рассчитан на этапе вторичной обработки сигналов в навигационном чипе или в постобработ-

ке данных на процессоре навигационного устройства, например, в приложении мобильного телефона, однако в таком случае оперативность определения ОК будет ниже.

В разд. 1 сформирован вектор признаков для последующей оптимизации, а также добавлены новые признаки, не рассмотренные ранее в перечисленных работах. В разд. 2 описаны критерии и методы определения оптимального вектора признаков на основе алгоритмов математической статистики. В разд. 3 рассмотрены различные классы и типы ОК в исследовании. В разд. 4 показаны результаты исследований и предложены оптимальные векторы признаков.

1. Формирование вектора признаков. В статье [5] для определения типа ОК были предложены следующие признаки:

$$\gamma(t) = \left[\mu(t), \sigma(t), \alpha(t), \lambda(t), n(t), GDOP(t)\right], \tag{1.1}$$

где $\mu(t)$ – среднее значение мощности сигнал C/No, $\sigma(t)$ – СКО средней мощности сигнала C/No, $\alpha(t)$ – коэффициент блокировки спутников, $\lambda(t)$ – расширенный геометрический фактор, $n(t)$ – количество видимых спутников, $GDOP(t)$ – геометрический фактор.

Значения указанных признаков могут быть вычислены по следующим зависимостям.

Коэффициент блокировки $\alpha(t)$ спутников [5]:

$$\alpha(t) = 1 - \frac{N_{\text{видимый}}(t)}{N_{\text{всего}}(t)}, \tag{1.2}$$

где $N_{\text{видимый}}(t)$ – количество видимых спутников, $N_{\text{всего}}(t)$ – общее количество всех спутников, которые должны быть видны потребителю.

Геометрический фактор точности определения местоположения с учетом поправок показаний часов потребителя (GDOP) рассчитывается по следующей формуле [1, 5]:

$$GDOP = \sqrt{tr(H)}, \tag{1.3}$$

$$H = \left(G^T G\right)^{-1}, \tag{1.4}$$

$$G = \begin{bmatrix} -\cos\theta^{(1)}\sin\varphi^{(1)} & \cos\theta^{(1)}\cos\varphi^{(1)} & -\sin\theta^{(1)} & 1 \\ -\cos\theta^{(2)}\sin\varphi^{(2)} & \cos\theta^{(2)}\cos\varphi^{(2)} & -\sin\theta^{(2)} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ -\cos\theta^{(N)}\sin\varphi^{(N)} & \cos\theta^{(N)}\cos\varphi^{(N)} & -\sin\theta^{(N)} & 1 \end{bmatrix}, \tag{1.5}$$

где θ – угол возвышения спутника, φ – азимут спутника.

Расширенный геометрический фактор $\lambda(t)$ [5]:

$$\lambda(t) = \frac{GDOP(t)}{GDOP^0(t)}, \tag{1.6}$$

где $GDOP^0(t)$ – геометрический фактор по всем видимым спутникам.

В [7] для определения типа ОК рассмотрен признак ограничения количества видимых спутников по C/No:

$$M(t) = \sum_{k=1}^{N} Mask_k(t), \tag{1.7}$$

$$Mask_k = \begin{cases} 1, & \dfrac{C}{No} > \varepsilon \\ 0, & \text{иначе} \end{cases}, \tag{1.8}$$

где $Mask_k$ – маска для спутников по граничному значению ε для C/No.

Подход среднего значения $\zeta(t)$ для всех видимых спутников в момент измерения с учетом сглаживающего окна $C\ /\ No_t$ по каждому спутнику приведен в [7]:

$$\zeta(t) = \frac{1}{N}\sum_{t=1}^{N} C/No_t \,. \tag{1.9}$$

Здесь N – количество измерений по каждому КА в единицу времени, C/No_t – среднее значение C/No для k предыдущих измерений видимого спутника (сглаживающее окно для видимого спутника):

$$C \,/\, No_t = \frac{1}{W}\sum_{kt=0}^{W} C/No_{kt} \,. \tag{1.10}$$

где W – длительность сглаживающего окна для видимого спутника, C/No_{kt} – предыдущее k-e измерение значения C/No для видимого спутника относительно текущего момента измерения t.

СКО $\xi(t)$ сглаживающего окна для среднего значения C/No [7]:

$$\xi(t) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(C \,/\, No_t - \zeta(t)\right)^2}. \tag{1.11}$$

Средний угол возвышения для видимых спутников $E(t)$ [7]:

$$E(t) = \frac{1}{N}\sum_{k=1}^{N} elev_k(t), \tag{1.12}$$

где N – количество измерений по каждому КА в единицу времени, $elev_k(t)$ – угол возвышения по k КА.

Для увеличения состава определяемых типов ОК и повышения точности определения предложено расширить состав вектора и добавить следующие новые признаки: геометрический фактор точности определения местоположения потребителя по горизонтали, расширенный геометрический фактор точности по горизонтали $\chi(t)$, средний коэффициент многолучевости $\omega(t)$, средняя оценка ошибки псевдодальности $\mathrm{p}(t)$ и СКО оценки ошибки псевдодальности $\varphi(t)$. Выражения для расчета значений новых признаков приведены ниже.

Геометрический фактор точности определения местоположения потребителя по горизонтали рассчитывается по следующей формуле [1]:

$$HDOP = \sqrt{D_{11} + D_{22}}, \tag{1.13}$$

где D – диагональные элементы матрицы H, расширенный геометрический фактор точности $\chi(t)$ определения местоположения потребителя по горизонтали вычисляется аналогично из (1.6).

Средний коэффициент многолучевости $\omega(t)$ определяется как

$$\omega(t) = \frac{1}{N}\sum_{k=1}^{N} m_k(t), \tag{1.14}$$

где N – количество измерений по каждому КА в единицу времени, m_k – коэффициент многолучевости по k КА [8, 9].

Средняя оценка ошибки псевдодальности (t) равна

$$\mathrm{p}(t) = \frac{1}{N}\sum_{k=1}^{N} R_k(t), \tag{1.15}$$

где N – количество измерений по каждому КА в единицу времени, R_k – оценка ошибки псевдодальности по k КА [8, 9].

СКО оценки ошибки псевдодальности $\varphi(t)$ запишем как

$$\varphi(t) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(R_k(t) - (t)\right)^2}, \tag{1.16}$$

где N – количество измерений по каждому КА в единицу времени, R_k – оценка ошибки псевдодальности по k КА [8], $\mathrm{p}(t)$ – средняя оценка ошибки псевдодальности.

Таким образом итоговый расширенный вектор признаков для оптимизации с помощью алгоритмов из математической статистики может быть представлен в виде:

$$\gamma(t)=\left[\begin{array}{c}\mu(t),\sigma(t),\alpha(t),\lambda(t),n(t),GDOP(t),HDOP(t),\chi(t),\\ \omega(t),(t),\varphi(t),M(t),\zeta(t),\xi(t),E(t)\end{array}\right]. \tag{1.17}$$

Учитывая различную информационную ценность характерных признаков, их число и состав вектора может быть оптимизирован.

2. Критерии и методы оптимизации вектора признаков. Для отбора наиболее значимых признаков и нахождения оптимального вектора рассмотрим часто используемые критерии и методы в математической статистике: корреляционный анализ, оценка дисперсии, хи-квадрат, критерий Фишера, взаимную информацию, L1 – регуляризацию, метод главных компонент, экстремальный градиентный бустинг и случайный лес.

Помимо критериев, которые применяются непосредственно в методах, описанных выше, целесообразно добавить критерий точности определения типа ОК.

Дополнительный критерий точности определяется по формуле [10]:

$$A(y,\hat{y})=\frac{1}{n}\sum_{i=1}^{n-1}1(y=\hat{y})_i, \tag{2.1}$$

где y – истинное значение ОК, $\hat{y}$ – предсказанное значение ОК, 1 (x) – индикатор функции.

Для расчета точности определения ОК будем использовать метод логистической регрессии совместно с перекрестной энтропией как наиболее простой линейный классификатор [11]:

$$P(y=1|x)=\frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_pX_p)}}, \tag{2.2}$$

где y – вероятность наступления события, X_p – признаки модели, β_p – коэффициенты регрессии модели.

Для корреляционного анализа запишем коэффициент корреляции Пирсона и нелинейную корреляцию через расстояние. Коэффициент корреляции Пирсона [12]

$$\rho(X,Y)=\frac{cov(X,Y)}{\sigma_X\sigma_y}, \tag{2.3}$$

где X, Y – случайные величины, $\sigma_X\sigma_y$ – стандартные отклонения, cov – ковариация. Корреляция через расстояние определяется как [13]

$$\nu_n{}^2(x,y)=\frac{1}{n^2}\sum_{i,j=1}^{n}A_{i,j}B_{i,j}, \tag{2.4}$$

$$A_{i,j}=a_{i,j}-\frac{1}{n}\sum_{l=1}^{n}a_{i,l}-\frac{1}{n}\sum_{k=1}^{k}a_{i,k}-\frac{1}{n^2}\sum_{k,l=1}^{k}a_{k,l},$$

$$B_{i,j}=b_{i,j}-\frac{1}{n}\sum_{l=1}^{n}b_{i,l}-\frac{1}{n}\sum_{k=1}^{k}b_{i,k}-\frac{1}{n^2}\sum_{k,l=1}^{k}b_{k,l},$$

где $a_{i,j}=\|x_i-x_j\|$ и $b_{i,j}=\|y_i-y_j\|$ – попарные Евклидовые расстояния x и y.

Из (2.4) получаем корреляцию расстояния:

$$R_n{}^2(x,y)=\begin{cases}\dfrac{\nu_n{}^2(x,y)}{\sqrt{\nu_n{}^2(x,x)\nu_n{}^2(y,y)}}, & \sqrt{\nu_n{}^2(x,x)\nu_n{}^2(y,y)}>0,\\ 0, & \sqrt{\nu_n{}^2(x,x)\nu_n{}^2(y,y)}=0.\end{cases} \tag{2.5}$$

Наиболее простым методом отбора признаков является расчет дисперсии, который сводится к тому, что признаки, имеющие наименьшую дисперсию, отбрасываются:

$$D=E\left[\left(X-E(X)\right)^2\right], \tag{2.6}$$

где X – случайная величина, E – математическое ожидание.

Критерий согласия Пирсона или хи-квадрат может быть вычислен следующим образом [14]:

$$\chi^2 = n\sum_{i=1}^{k}\frac{(\frac{n_i}{n} - P_i(\Theta))^2}{P_i(\Theta)}, \tag{2.7}$$

где n – общее количество измерений, n_i – количество измерений для выбранного интервала, $P_i(\Theta)$ – вероятность для выбранного интервала.

Критерий Фишера в дисперсионном анализе определяет, согласно [15]:

$$F = \frac{\sum_{i=1}^{K}\frac{n_i(\bar{Y}_i - \bar{Y})^2}{K-1}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{(Y_{ij} - \bar{Y}_i)^2}{N-K}}, \tag{2.8}$$

где $\bar{Y}_i$ – выборочное среднее выборки, $\bar{Y}$ – общее среднее выборки, K – количество групп, Y_{ij} – j наблюдение в i выборке, n_i – количество наблюдение в i выборке, N – размер выборки.

Запишем метод взаимной информации, который показывает количество информации, содержащееся в одной случайной величине относительно другой с помощью энтропии [16]:

$$I(X,Y) = H(X) - H(Y), \tag{2.9}$$

где $H(X)$ – энтропия события X, $H(Y)$ – энтропия события X при условии события Y. Распишем подробнее (2.9):

$$I(X,Y) = \sum_{y\in Y}\sum_{x\in X}P_{(X,Y)}(x,y)\log\frac{P_{(X,Y)}(x,y)}{P_{(X)}(x)P_{(Y)}(y)}, \tag{2.10}$$

где $P_{(X,Y)}$ – совместная вероятность событий X и Y, $P_{(X)}$ и $P_{(Y)}$ – вероятности событий X и Y.

Метод экстремального градиентного бустинга использует повышающиеся деревья решений, которые последовательно исправляют ошибки, в качестве оценщика применяется среднее усиление по всем выборкам [17]:

$$\text{Усиление} = \frac{1}{2}\left[\frac{G_L^2}{H_L+\alpha} + \frac{G_R^2}{H_R+\alpha} - \frac{(G_L+G_R)^2}{H_R+H_L+\alpha}\right] - \sigma, \tag{2.11}$$

где G_L, G_R – коэффциенты в функции потерь для 1-го члена ряда Тейлора, H_L, H_R – коэффициенты в функции потерь для 2-го члена ряда Тейлора, α,σ – коэффициенты регуляризации.

Метод случайного леса заключается в использовании ансамбля решающих деревьев [18], в качестве оценщика признаков применяется примесь Джини:

$$I_G(p) = 1 - \sum_{i=1}^{J}p_i^2, \tag{2.12}$$

где J – количество классов, P_i – вероятность элемента выбранного класса i.

Метод регуляризации L1 (совместно с логистической регрессией) [19] равен

$$L_1 = \sum_{i=1}^{N}\left(y_i - f(y_i)\right)^2 + \lambda\sum_{i=1}^{N}|a_i|, \tag{2.13}$$

где y_i – целевой признак, $f(y_i)$ – предсказание модели, λ, a_i – весовые коэффициенты.

Метод главных компонент сводится к нахождению подпространства меньшей размерности, в ортогональной проекции в которых СКО максимально [20].

3. Условия проведения эксперимента и рассматриваемые классы ОК. В работах [3–9] число определяемых классов и типов ОК варьируется от трех до пяти (высотная городская застройка (каньон), средневысотная городская застройка, пригород, открытая местность и бульвар(лес)).

Учитывая более широкое разнообразие реальных сред функционирования массовых образцов НАП ГНСС, для расширения числа классов и типов ОК, рассмотренных ранее, добавляются типы сред внутри помещения и высотная городская застройка со зданиями, расположенными со всех сторон от НАП (колодец).

Для получения экспериментальных данных в реальных условиях ОК использовалась НАП массового применения: U-blox NEO-M8N.

При проведении эксперимента были собраны приблизительно двухчасовые измерения (табл. 1) с помощью НАП U-blox NEO-M8N (частота измерений 1 Гц) антенны Harxon HX–CSX601A по открытым сигналам L1OF ГЛОНАСС и L1OC GPS для стационарного потребителя в разных ОК (помещение, полузакрытое пространство (окно, балкон), средневысотная городская застройка, лес (бульвар), высотная городская застройка (городской каньон), городской колодец и открытая местность). Из табл. 1 видно, что среднеквадратичная ошибка определения местоположения по системе ГЛОНАСС и GPS для рассматриваемых классов и типов ОК различная.

Таблица 1. Экспериментальные данные различных типов ОК

Типы ОК	Открытое небо	Средневысотная Городская застройка	Полузакрытое помещение (окно, балкон)	Внутри помещения	Лесопарк (бульвар)	Городской каньон	Городской колодец
Общее количество измерений, %	6775 / 14.8	6306 / 13.7	7036 / 15.3	6619 / 14.4	6200 / 13.5	6779 / 14.8	6197 / 13.5
СКО ошибки определения местоположения по ГЛОНАСС, м	1.7	10.20	14.06	7.79	8.34	11.47	26.61
Количество навигационных решений от общего количества ГЛОНАСС, %	6775 / 100	5836 / 92.5	5433 / 77.2	336 / 5	5834 / 94	4430 / 65.3	4126 / 66.5
СКО ошибки определения местоположения по GPS, м	1	5.71	13.37	36.82	7.34	14.03	34.26
Количество навигационных решений от общего количества GPS, %	6775 / 100	4418 / 70	4136 / 58.7	1341 / 20.2	5602 / 90.3	5546 / 81.8	981 / 15.8

Примечание. СКО ошибки местоположения рассчитаны для угла места больше 15°, решение навигационной задачи весовым методом наименьших квадратов.

4. Результаты нахождения оптимального вектора признаков. На рис. 1, 2 (для удобства отображения обозначим $G - GDOP$, $H - HDOP$) построена карта корреляций Пирсона и расстояний между признаками сформированного вектора (1.17) для системы GPS. Корреляционный анализ вектора признаков (1.17) по системе GPS показывает наиболее сильную корреляцию ОК со следующими признаками: СКО мощности сигнала δ, количество видимых спутников n, средний коэффициент многолучевости ω, маска M по C/No для количества видимых КА, средний угол места E и сильную нелинейную связь с СКО мощности сигнала δ и маской M по C/No для количества видимых КА.

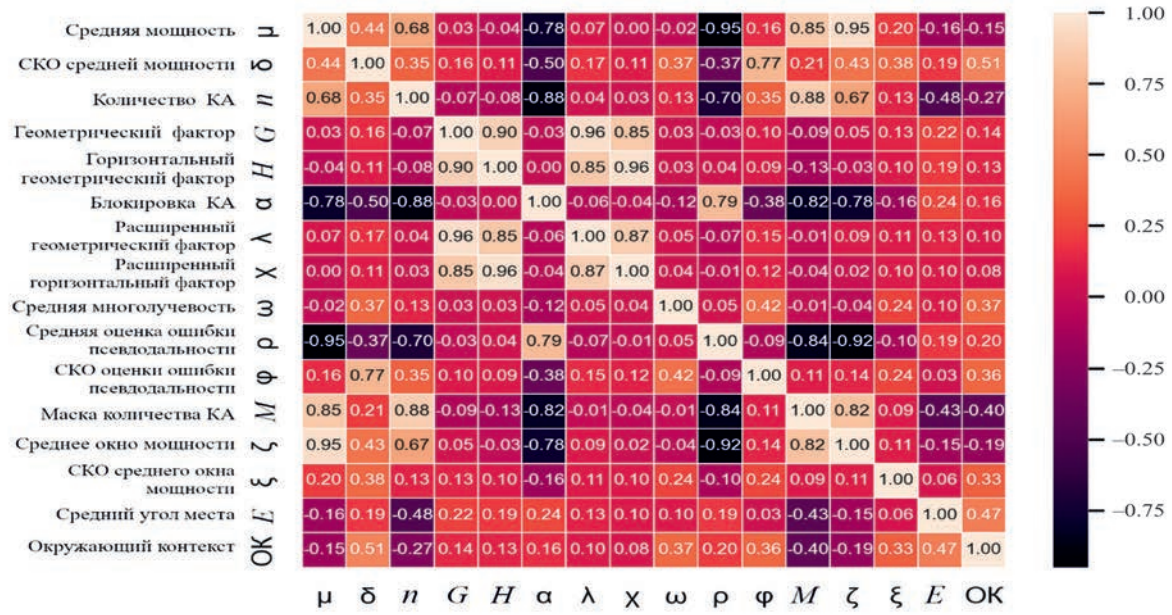

Рис. 1. Корреляция Пирсона для GPS

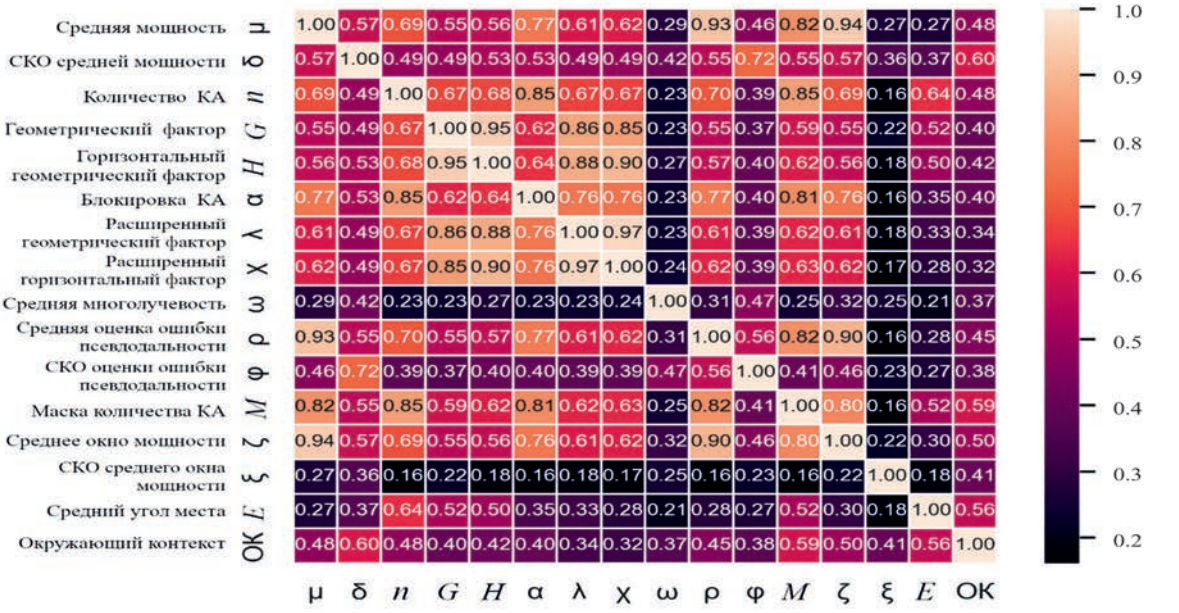

Рис. 2. Корреляция расстояния для GPS

На рис. 3, 4 построена карта корреляций Пирсона и расстояний между признаками сформированного вектора (1.17) для системы ГЛОНАСС. Корреляционный анализ вектора признаков (1.17) по системе ГЛОНАСС показывает наиболее сильную линейную и нелинейную корреляции ОК со следующими признаками: средний коэффициент многолучевости ω, СКО мощности сигнала δ, средний угол места E, маска M по C/No для количества видимых КА и сглаживающего окна ζ для среднего значения C/No, также видно отсутствие связи с геометрическими факторами $GDOP$, $HDOP$, λ, χ, что связано с отсутствием измерений из-за малого количества видимых КА. Оставшиеся признаки для ГЛОНАСС и GPS сильно коррелируют с упомянутыми выше признаками, поэтому их использование нецелесообразно и может привести к неправильному определению ОК.

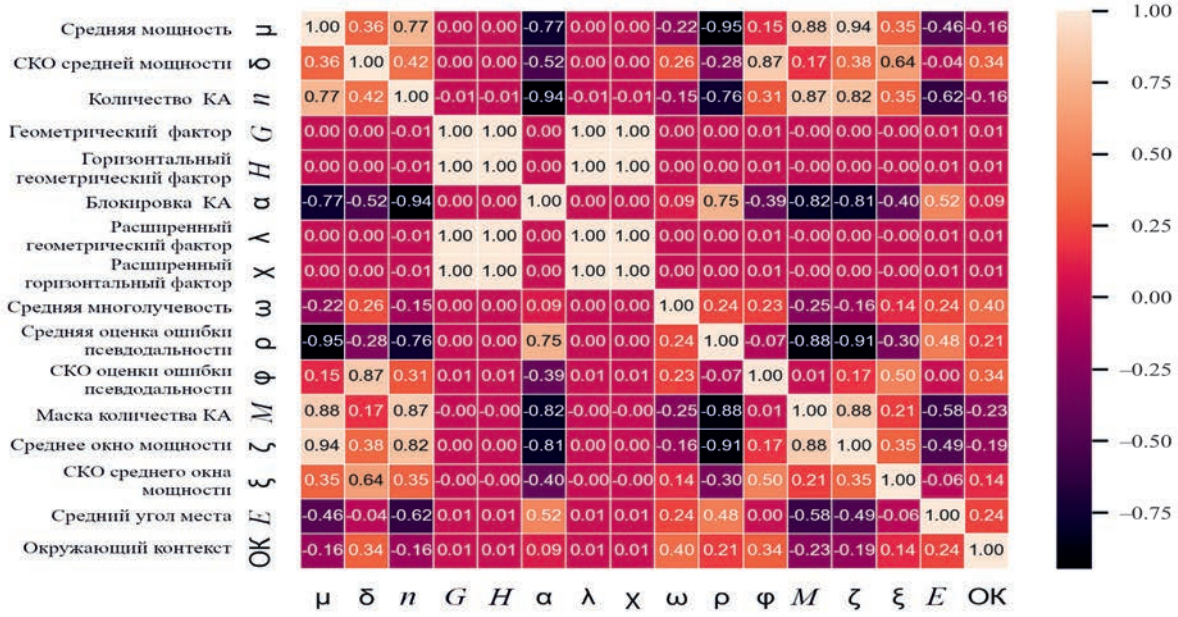

Рис. 3. Корреляция Пирсона для ГЛОНАСС

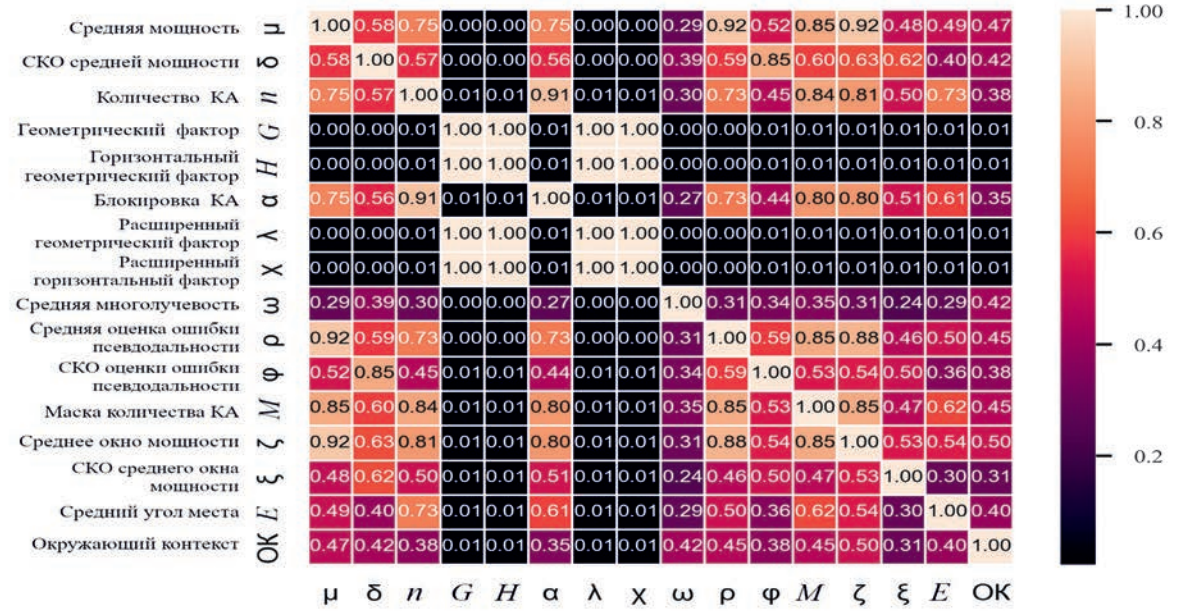

Рис. 4. Корреляция расстояния для ГЛОНАСС

В табл. 2 и 3 показаны результаты выбора наиболее важных признаков по критериям и методам, описанным в разд. 2 для навигационных систем ГЛОНАСС и GPS. Из табл. 2 и 3 видно, что каждый признак может вносить вклад (вплоть до 0.3–2% в методах экстремальный градиентный бустинг и случайный лес) в определение ОК, однако корреляционный анализ показал, что некоторые из этих признаков имеют сильную связь между друг другом и их использование может привести к неверной работе модели в реальных условиях навигации. Также видно, что сглаживающее окно для среднего значения мощности сигнала, СКО мощности сигнала и маска по *C*/*No* для количества видимых КА влияют наиболее сильно, что соответствует выбранным в [7], однако эти признаки могут повлечь за собой потерю информации.

Таблица 2. Отбор признаков для системы GPS

Название признаков	Среднее значение мощности сигнала	СКО мощности сигнала	Количество спутников	Суммарный геометрический фактор	Горизонтальный геометрический фактор	Коэффициент блокировки спутников	Расширенный суммарный геометрический фактор	Расширенный горизонтальный геометрический фактор	Средний коэффициент многолучевости	Средняя оценка ошибки псевдодальности	СКО оценки ошибки псевдодальности	Маска количества спутников для $\varepsilon = 30$	Среднее значение мощности сигнала для окна $k = 3$	СКО мощности сигнала для окна	Средний угол места	Точность на логистической регрессии
	μ	δ	n	G	H	α	λ	χ	ω	ρ	φ	M	ζ	ξ	E	A
Дисперсия	+	+	+	+	+	-	+	+	-	+	+	+	+	+	+	0.849
Хи-квадрат	1708	1823	1890	45	50	8511	17	13	933	3990	2020	3596	1910	324	259	0.839
Критерий Фишера	53412	14962	19452	252	153	23244	120	47	2467	30721	6245	35813	58314	3090	2928	0.882
Взаимная информация	1.23	0.87	0.7	1.57	1.61	0.65	0.99	0.99	0.41	0.93	0.91	0.85	1.29	0.67	1.1	0.885
Рекурсивное исключение логистической регрессией	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	0.903
L1-регуляцизация	+	+	+	+	-	+	+	-	+	-	+	+	+	+	+	0.908
Главные компоненты	+	+	+	+	+	+	+	-	+	+	+	+	-	+	+	0.889
Экстремальный градиентный бустинг, %	5.2	4.8	4.8	1.7	1.4	0.7	0.7	0.6	4.4	0.3	1.9	53.4	9.6	5.2	5.3	0
Случайный лес, %	12.8	5.8	5.5	5.8	5.2	2.6	1.8	1.8	2.6	6.4	2.5	8.2	18.6	11.4	8.8	0

Примечание. 0 означает, что расчеты отсутствуют.

Таблица 3. Отбор признаков для системы ГЛОНАСС

Название признаков	Среднее значение мощности сигнала	СКО мощности сигнала	Количество спутников	Суммарный геометрический фактор	Горизонтальный геометрический фактор	Коэффициент блокировки спутников	Расширенный суммарный геометрический фактор	Расширенный горизонтальный геометрический фактор	Средний коэффициент многолучевости	Средняя оценка ошибки псевдодальности	СКО оценки ошибки псевдодальности	Маска количества спутников для $\varepsilon = 30$	Среднее значение мощности сигнала для окна $k = 3$	СКО мощности сигнала для окна	Средний угол места	Точность на логистической регрессии
	μ	δ	n	G	H	α	λ	χ	ω	ρ	φ	M	ζ	ξ	E	A
Дисперсия	+	+	+	+	+	-	+	+	-	+	+	+	+	+	+	0.784
Хи-квадрат	1485	2096	2329	6	6	9194	6	6	485	3323	2420	3254	1950	526	245	0.803
Критерий Фишера	34837	13409	27440	0	0	23952	0	0	1320	21078	7648	27514	111629	5065	2716	0.857
Взаимная информация	0.98	0.8	0.69	1.03	1.28	0.66	0.8	0.79	0.56	0.82	0.81	0.84	1.26	0.88	1.21	0.832
Рекурсивное исключение логистической регрессией	+	+	+	-	-	+	-	-	+	+	+	+	+	+	+	0.869
L1-регуляцизация	+	+	+	-	-	-	-	-	+	+	-	+	+	-	-	0.807
Главные компоненты	+	-	+	+	+	+	+	-	+	-	+	+	+	+	+	0.842
Экстремальный градиентный бустинг, %	4.1	10.8	9.8	4.5	9.7	1.4	1.5	2.2	6.1	1.4	1.1	4.2	28.3	9.6	5.2	0
Случайный лес, %	8.9	6.3	8.3	0	0	5.3	0	0	2.9	5.3	2.3	8.5	22.6	13.9	11.3	0

Примечание. 0 означает, что расчеты отсутствуют.

На основе анализа таблиц и карт корреляций можно сформировать следующий оптимальный вектор признаков для систем ГЛОНАСС и GPS:

$$\gamma_{\text{опт1}}(t) = \left[\sigma(t), HDOP(t), \omega(t), M(t), \zeta(t), \xi(t), E(t)\right], \tag{4.1}$$

где $\sigma(t)$ – СКО средней мощности сигнала *C/No*, *HDOP*(*t*) – геометрический фактор точности по горизонтали, $\omega(t)$ – средний коэффициент многолучевости, *M*(*t*) – маска количества видимых КА по *C/No*, $\zeta(t)$ – сглаживающее окно для среднего значения *C/No*, $\xi(t)$ – СКО сглаживающее окно для среднего значения *C/No*, $E(t)$ – среднее значение угла места.

В случае если ОК изменяется быстро и скорость движения потребителя высокая, а также вычислительные ресурсы ограничены, можно сформировать следующий оптимальный вектор признаков для систем ГЛОНАСС и GPS:

$$\gamma_{onm2}(t) = \left[\mu(t), \sigma(t), n(t), HDOP(t), \omega(t), E(t)\right], \tag{4.2}$$

где $\mu(t)$ – среднее значение мощности сигнала *C/No*, $\sigma(t)$ – СКО средней мощности сигнала *C/No*, *n*(*t*) – количество видимых спутников, *HDOP*(*t*) – геометрический фактор точности по горизонтали, $\omega(t)$ – средний коэффициент многолучевости, $E(t)$ – среднее значение угла места.

Заключение. При использовании ГНСС позиционирование зависит как от условий окружающей среды, так и от поведения потребителя. Окружающая среда влияет на качество приема радиосигналов, которые доступны для позиционирования. Для работы в различных условиях навигации требуется адаптивное навигационное решение, которое будет определять и учитывать контекст окружающей среды в НАП. Для этого можно формировать признаки из информации от навигационных сигналов, приходящей с каждого КА. Такие исследования активно ведутся в последние годы, однако во всех известных работах отсутствует комплексный подход для формирования оптимального вектора признаков.

В ходе проведенной работы получены следующие результаты: добавлены новые классы и типы ОК (внутри помещения и городской колодец), новые признаки для определения ОК (средний коэффициент многолучевости $\omega(t)$, средняя оценка ошибки псевдодальности $\rho(t)$, СКО оценки ошибки псевдодальности $\varphi(t)$), проведен линейный и нелинейный корреляционный анализ вектора признаков для системы ГЛОНАСС и GPS, отобраны признаки с помощью различных методов и критериев для системы ГЛОНАСС и GPS, найден оптимальный вектор признаков $\gamma_{onm1}(t)$ для системы ГЛОНАСС и GPS, включающий: СКО средней мощности сигнала, средний коэффициент многолучевости, геометрический фактор точности по горизонтали, маску количества видимых КА по *C/No*, сглаживающее окно для среднего значения *C/No*, СКО сглаживающего окна для среднего значения *C/No* и среднее значение угла места. Предложен вектор признаков $\gamma_{onm2}(t)$ для системы ГЛОНАСС и GPS в случае высокодинамичного потребителя и быстрого изменения ОК.

СПИСОК ЛИТЕРАТУРЫ

1. *Перов А.И.* ГЛОНАСС. Модернизация и перспективы развития. Монография. М.: Радиотехника, 2020.
2. *Кульнев В.В., Кульнев Е.В., Наконечный Е.О.* Методы определения контекста по данным навигационной аппаратуры глобальных навигационных спутниковых систем // Космонавтика и ракетостроение. 2023. № 4. С. 43–53.
3. *Groves P.D, Martin H., Voutsis K., Walter D., Wang L.* Context Detection, Categorization and Connectivity for Advanced Adaptive Integrated Navigation. London UK: University College London, 2013.
4. *Gao H., Groves P.D.* Environmental Context Detection for Adaptive Navigation using GNSS Measurements from a Smartphone. London UK: University College London, 2018.
5. *Wang Y., Liu P., Liu Q., Adeel M., Qian J., Jin X., Ying R.* Urban Environment Recognition Based on the GNSS Signal Characteristics // Wiley ION. 2018. V. 66.
6. *Liu H., Zhang M., Pei L., Wang W., L. Li, Pan C., Li Z.* Environment Classification for Global Navigation Satellite Systems Using Attention // Shanghai Key Laboratory of Navigation and Location Based Services. Shanghai: Shanghai Jiao Tong University, 2021. February.
7. *Feriol F., Watanabe Y., Vivet D.* GNSS-based Environmental Context Detection for Navigation // IEEE Intelligent Vehicles Symposium (IV). Aachen. Germany, 2022. Jun.
8. U-blox M8 Receiver description. UBX-13003221. 2022. August.
9. Digital Cellular Telecommunications System (Phase 2+) (GSM) Location Services (LCS) Mobile Station (MS). Serving Mobile Location Centre (SMLC) Radio Resource LCS Protocol (RRLP) (3GPP TS 44.031 Version 17.0.0 Release 17). ETSI TS 144 031 V17.0.0. 2022.

10. BS ISO 5725-1. Accuracy (Trueness and Precision) of Measurement Methods and Results – Part 1: General Principles and Definitions. Geneva, 1994. P.1.
11. *Cramer J.S.* The Origins of Logistic Regression (PDF) // Tinbergen Institute. 2002. V. 119. P. 167–178.
12. *Pearson K.* Notes on Regression and Inheritance in the Case of Two parents. London: Proceedings of the Royal Society of London, 1895. June.
13. *Székely G.J., Rizzo M.L., Bakirov N.K.* Measuring and Testing Dependence by Correlation of Distances // The Annals of Statistics. 2007. V. 35. № 6. P. 2769–2794.
14. *Никулин М.С.* Критерий хи-квадрат для непрерывных распределений с параметрами сдвига и масштаба // Теория вероятностей и ее применение. 1973. Т. XVIII. Вып. 3.
15. *Box G.E.P.* Non-Normality and Tests on Variances // Biometrika. 1953. V. 40. P. 318–335.
16. *Kraskov A., Stogbauer H., Grassberger P.* Estimating Mutual Information // Physical Review E. 2004. V. 69.
17. *Friedman J.* Greedy Function Approximation: A Gradient Boosting // The Annals of Statistics. 2001. V. 29.
18. *Hastie T., Tibshirani R., Friedman J.* Chapter 15. Random Forests // The Elements of Statistical Learning. 2nd ed. Springer-Verlag, 2009.
19. *Boyd S., Vandenberghe L.* Convex Optimization. London UK: Cambridge University Press, 2004. 716 p.
20. *Tipping M.E., Bishop C.M.* Probabilistic Principal Component Analysis // J. Royal Statistical Society: Series B (Statistical Methodology). 1999. V. 61(3). P. 611–622.

ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2024, № 4, с. 167–176

НАВИГАЦИОННЫЙ СИСТЕМЫ

УДК 62-50

СТАБИЛИЗАЦИЯ ИНТЕГРАТОРА 3-ГО ПОРЯДКА ОБРАТНОЙ СВЯЗЬЮ В ВИДЕ ВЛОЖЕННЫХ САТУРАТОРОВ

© 2024 г. Ю. В. Морозов[a], *, А. В. Пестерев[a], **

[a] *Институт проблем управления им. В. А. Трапезникова РАН, Москва, Россия*

**e-mail: tot1983@inbox.ru*

***e-mail: alexanderpesterev.ap@gmail.com*

Поступила в редакцию 26.03.2024 г.

После доработки 11.06.2024 г.

Принята к публикации 15.07.2024 г.

Рассматривается задача стабилизации интегратора 3-го порядка с фазовым ограничением с помощью непрерывного ограниченного управления при дополнительном условии выполнения фазового ограничения. Применение обратной связи в виде вложенных сатураторов приводит к исследованию устойчивости системы с переключениями. Установлены необходимые условия на коэффициенты обратной связи, при выполнении которых система локально устойчива. Построена функция Ляпунова, с помощью которой доказано, что необходимые условия являются и достаточными для глобальной асимптотической устойчивости замкнутой системы. Изложение иллюстрируется численными примерами.

Ключевые слова: стабилизация цепочки трех интеграторов, глобальная асимптотическая устойчивость, вложенные сатураторы, функция Ляпунова.

DOI: 10.31857/S0002338824040121 **EDN:** TRFTDB

STABILIZATION OF A CHAIN OF THREE INTEGRATORS BY A FEEDBACK IN THE FORM OF NESTED SATURATORS

Yu. V. Morozov[a], *, A. V. Pesterev[a], **

[a] *V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia*

**e-mail: tot1983@inbox.ru*

***e-mail: alexanderpesterev.ap@gmail.com*

The problem of stabilizing a chain of three integrators subject to a phase constraint by a continuous constrained control is considered. The application of a feedback in the form of nested saturators results in study of a switching system. Necessary conditions of local stability are established. A Lyapunov function is constructed by means of which it is proved that the necessary conditions are sufficient for global stability of the closed-loop system. The discussion is illustrated by numerical examples.

Keywords: stabilization of a third-order integrator, nested saturators, global asymptotic stability, Lyapunov function.

Введение. Задача стабилизации цепочек интеграторов широко обсуждалась в литературе по управлению в течение нескольких последних десятилетий (например, [1–4] и приведенные там ссылки). Интерес к данной проблематике объясняется тем, что во многих приложениях исходные модели заданы в виде цепочек интеграторов, а управления, разработанные для цепочек интеграторов, легко обобщаются на более широкие классы систем. Большое распространение в последние десятилетия получил подход, основанный на применении специальных непрерывных управлений в виде вложенных негладких функций насыщения – сатураторов. Интерес к обратной связи в виде вложенных сатураторов объясняется рядом замечательных свойств полученной замкнутой системы. Такие обратные связи, в частности, позволяют автоматически учесть ограниченность ресурса управления и при этом гарантируют выполнение определенных фазовых ограничений, а также обеспечивают экспоненциальную скорость убывания отклонения вблизи положения равновесия. Преимущества об-

ратных связей в виде вложенных сатураторов, а также актуальность задачи стабилизации цепочек интеграторов обсуждаются во многих публикациях, например в [2–11]. Применение вложенных сатураторов, однако, приводит к достаточно сложной нелинейной системе с переключениями, анализ устойчивости которой представляет нетривиальную задачу. Доказать глобальную устойчивость удается преимущественно для систем второго порядка [5, 6] и в редких случаях обратных связей специального вида — для систем третьего [9, 10] или четвертого [11] порядков. Общий случай n-мерного интегратора обсуждается, например, в [2, 3]. Но глобальная устойчивость системы замкнутой обратной связи в виде n вложенных сатураторов была доказана только для случая, когда предельные значения вложенных функций насыщения удовлетворяют определенным, на практике редко выполнимым, неравенствам [2, Theorem 2.1].

В статье рассматривается задача стабилизации в нуле цепочки трех интеграторов. В качестве стабилизирующего предлагается непрерывное управление в виде вложенных сатураторов, гарантирующее выполнение фазового ограничения на третью переменную состояния. Обсуждаются преимущества предлагаемой обратной связи и устанавливаются необходимые и достаточные условия глобальной устойчивости замкнутой системы.

1. Постановка задачи. В работах [7, 8] для стабилизации интегратора 2-го порядка

$$\dot{w}_1 = w_2, \dot{w}_2 = U_1(w_1, w_2) \tag{1.1}$$

предложено использовать обратную связь в виде вложенных сатураторов:

$$U_1(w_1, w_2) = -k_4\mathrm{sat}\Big(k_3\big(w_2 + k_2\mathrm{sat}\big(k_1w_1\big)\big)\Big), \tag{1.2}$$

где $\mathrm{sat}(\cdot)$ – негладкая функция насыщения: $\mathrm{sat}(w) = w$ при $|w| \leqslant 1$ и $\mathrm{sat}(w) = \mathrm{sign}(w)$, когда $|w| > 1$. Правая часть (1.2) задает разбиение фазовой плоскости на множества D_1, D_2 и D_3 (рис. 1): область D_1 включает все точки, в которых оба сатуратора не насыщены (наклонная полоса, ограниченная пунктирными линиями на рис. 1); множество $D_2 = D_2^- \cup D_2^+$ — точки, в которых насыщен только внутренний сатуратор; $D_3 = D_3^- \cup D_3^+$ — все точки, в которых насыщен внешний сатуратор (более подробно см. [8, 11, 12]). Система (1.1), (1.2) представляет собой систему с переключениями, состоящую из пяти линейных систем, переключения между которыми зависят от состояния и происходят при пересечении границ между областями.

Главное достоинство управления (1.2) состоит в том, что оно позволяет стабилизировать систему (1.1) из любого начального положения при любых положительных коэффициентах. Доказательство глобальной устойчивости системы (1.1), (1.2) для частного случая выбора коэффициентов обратной связи из однопараметрического семейства можно найти в [11];

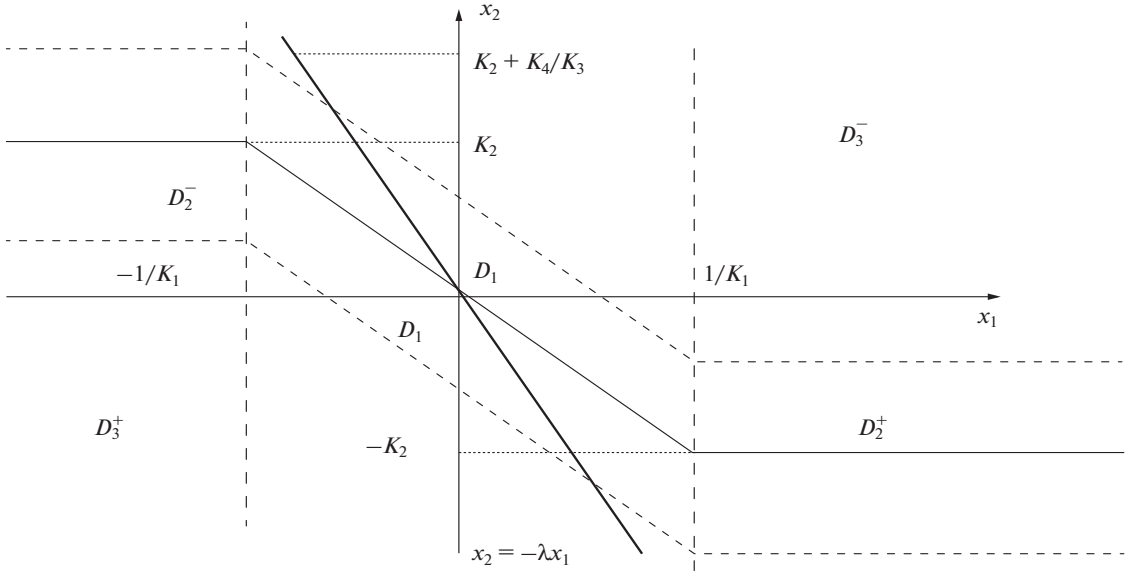

Рис. 1. Разбиение фазовой плоскости на множества D_1, D_2, D_3

доказательство для общего случая произвольных положительных коэффициентов обратной связи приведено в [12]. К другим преимуществам обратной связи вида (1.2) относятся ее непрерывность и ограниченность, а также возможность обеспечить желаемые характеристики переходного процесса для любой конкретной системы простой настройкой коэффициентов. Действительно, коэффициент k_4 задает ресурс управления; k_2 ограничивает максимальную скорость приближения к положению равновесия ($|x_2(t)| \leqslant k_2$); с помощью же коэффициентов k_1 и k_3 выбираются желаемые тип (узел, фокус) положения равновесия и экспоненциальная скорость убывания отклонения вблизи нуля [8, 11, 12].

Рассмотрим теперь интегратор 3-го порядка:

$$\dot{x}_1 = x_2, \dot{x}_2 = x_3, \dot{x}_3 = U(x),\ x \equiv [x_1, x_2, x_3]^{\mathrm{T}} \tag{1.3}$$

и применим к нему обратную связь:

$$U(x) = -k_5\left(x_3 - U_1(x_1, x_2)\right), \tag{1.4}$$

где U_1 (x_1,x_2) определено формулой (1.2). В разд. 1 будет доказана глобальная асимптотическая устойчивость системы (1.3), (1.4). Здесь же приведем обоснование такого выбора обратной связи.

1. Проекция траектории на плоскость (x_1,x_2) стремится к траектории интегратора 2-го порядка с начальными условиями $w_1(0) = x_1(0)$, $w_2(0) = x_2(0) : x_1(t) \to w_1(t)$ и $x_2(t) \to w_2(t)$ при $t \to \infty$. Другими словами, выбор коэффициентов k_1, ... ,k_4, обеспечивающих “желаемое” поведение отклонения $x_1(t)$ и скорости $x_2(t)$, можно сделать на основе решения более простой задачи стабилизации интегратора 2-го порядка.

2. Решение системы (1.3), (1.4) удовлетворяет фазовому ограничению

$$|x_3(t)| \leqslant k_4, \tag{1.5}$$

исходя из того, что $|x_3(0)| \leqslant k_4$. Действительно, допустим фазовое ограничение выполняется в начальный момент. Из третьего уравнения системы (1.3) следует, что переменная $x_3(t)$ достигает локального экстремума на траектории при $U(x) = 0$, а из формулы (1.4) видно, что управление равно нулю, когда $x_3 = -k_4\mathrm{sat}(k_1 x_1)$. Следовательно, $|x_3(t)|$ не может быть больше k_4.

Необходимость выполнения фазового ограничения на третью переменную естественным образом возникает во многих приложениях, например при стабилизации механической системы, где переменными состояния служат позиция, скорость и тяга (ускорение), а в качестве управления берется скорость изменения тяги (например, с помощью шагового мотора). Так как тяга в реальных системах ограничена, максимальное значение $|x_3(t)|$ также должно быть ограничено. Положив k_4 в (1.2) равным максимальной тяге, получим управление, гарантирующее выполнение фазового ограничения.

3. К преимуществам управления (1.4) относятся также а) экспоненциальная скорость убывания отклонения вблизи положения равновесия [8, 11, 12] и б) ограниченность при любых, сколь угодно больших отклонениях от положения равновесия при выполнении фазового ограничения в начальной точке (см. разд. 1).

2. Представление системы в безразмерном виде. Применим следующую замену переменных и времени [10]:

$$\tilde{t} = \frac{k_4 t}{k_2}, \tilde{x}_1 = \frac{k_4 x_1}{k_2^2}, \tilde{x}_2 = \frac{x_2}{k_2}, \tilde{x}_3 = \frac{x_3}{k_4}.$$

Подставляя новые переменные в систему (1.3), (1.4) и переходя к дифференцированию по безразмерному времени, получим безразмерную модель с коэффициентами:

$$\tilde{k}_1 = \frac{k_1 k_2^2}{k_4}, \tilde{k}_2 = 1, \tilde{k}_3 = \frac{k_2 k_3}{k_4}, \tilde{k}_4 = 1, \tilde{k}_5 = \frac{k_2 k_5}{k_4}. \tag{2.1}$$

Таким образом, каковы бы ни были ресурс управления (k_4) и максимальная разрешенная скорость (k_2), задача сводится к исследованию устойчивости модели с единичными коэффициентами k_2 и k_4. Всюду далее будем полагать все переменные и параметры безразмерными и использовать для них прежнее обозначение (без тильды). В безразмерной модели обратная связь (1.4) принимает вид:

$$U(x) = -k_5(x_3 + \mathrm{sat}\left(k_3\left(x_2 + \mathrm{sat}\left(k_1x_1\right)\right)\right)). \tag{2.2}$$

Задача исследования – найти условия на коэффициенты, при которых предлагаемая обратная связь является стабилизирующей во всем пространстве, т.е. установить условия глобальной асимптотической устойчивости системы (1.3), (2.2).

3. Необходимые условия устойчивости.

Л е м м а 1. *Для того, чтобы система (1.3), (2.2) была устойчива, необходимо, чтобы коэффициенты обратной связи были положительны и выполнялось неравенство* $k_5 > k_1$.

Д о к а з а т е л ь с т в о л е м м ы 1. Для устойчивости системы (1.3), (2.2) необходимо, чтобы была устойчива определенная в D_1 (см. рис. 1) линейная система:

$$\dot{x}_1 = x_2, \dot{x}_2 = x_3, \dot{x}_3 = -k_5x_3 - k_5k_3x_2 - k_5k_3k_1x_1. \tag{3.1}$$

Применим критерий Гурвица [13] к последней и выпишем составленный из коэффициентов характеристического полинома определитель Гурвица:

$$\Delta = \begin{vmatrix} k_5 & k_1k_3k_5 & 0 \\ 1 & k_3k_5 & 0 \\ 0 & k_5 & k_1k_3k_5 \end{vmatrix}.$$

Условие положительности всех главных миноров определителя Гурвица дает искомые необходимые условия устойчивости системы (1.3), (2.2):

$$k_5 > k_1 > 0, k_3 > 0. \tag{3.2}$$

Л е м м а 1 доказана.

З а м е ч а н и е 1. Подставляя правые части формул (2.1) для безразмерных коэффициентов в (3.9), получим необходимые условия, которым должны удовлетворять размерные коэффициенты: $k_i > 0$, $i = 1, \ldots, 5$, $k_5 > k_1k_2$.

4. Достаточные условия глобальной устойчивости.

Т е о р е м а 1. Пусть выполнены необходимые условия устойчивости (3.9), т.е. $k_i > 0$, $i = 1, 3, 5$ и $k_5 > k_1$. Тогда система (1.3), (2.2) глобально асимптотически устойчива.

Д о к а з а т е л ь с т в о т е о р е м ы 1. Пусть коэффициенты k_1, k_3 и k_5 положительны. Рассмотрим функцию

$$V(x) = k_5^2\int_0^{x_1}\mathrm{sat}\left(k_3\mathrm{sat}\left(k_1s\right)\right)ds + k_5\int_0^{x_2}\mathrm{sat}\left(k_3\left(s + \mathrm{sat}\left(k_1x_1\right)\right)\right)ds + \frac{1}{2}(x_3 + k_5x_2)^2 \tag{4.1}$$

и докажем, что она является функцией Ляпунова системы (1.3), (2.2). Доказательство опирается на следующие два неравенства, непосредственно вытекающие из монотонности функции насыщения и ее равенству нулю в нуле:

$$\int_0^x \mathrm{sat}\,(s)ds > 0 \ \forall x \neq 0, \tag{4.2}$$

$$[\mathrm{sat}\,(s + s_0) - \mathrm{sat}(s_0)]s \geqslant 0 \ \ \forall s \neq 0, \forall s_0. \tag{4.3}$$

Обозначим через Φ_1 и Φ_2 первое и второе интегральные слагаемые в (4.1) соответственно и покажем, что их сумма, а значит, и функция $V(x)$, положительна в R^3.

Преобразуем Φ_2, применив замену $\tilde{s} = s + \mathrm{sat}(k_1x_1)$:

$$\Phi_2 = \int\limits_{\mathrm{sat}(k_1x_1)}^{x_2+\mathrm{sat}(k_1x_1)} k_5\mathrm{sat}\left(k_3\tilde{s}\right)d\tilde{s} = \int\limits_{0}^{x_2+\mathrm{sat}(k_1x_1)} k_5\mathrm{sat}\left(k_3\tilde{s}\right)d\tilde{s} - \int\limits_{0}^{\mathrm{sat}(k_1x_1)} k_5\mathrm{sat}\left(k_3\tilde{s}\right)d\tilde{s}.$$

Первое слагаемое положительно в силу (4.2). Обозначив второй интеграл в последней формуле как Φ_{22} и сделав в нем еще одну, неявную замену переменной интегрирования $\tilde{s} = \mathrm{sat}(k_1 s)$, получим

$$\Phi_{22} = \int\limits_{0}^{x_1} k_5\mathrm{sat}(k_3\mathrm{sat}(k_1 s))k_1\mathrm{sat}'(k_1 s))ds,$$

где штрих означает дифференцирование по аргументу, а значение производной функции насыщения при $k_1 s = \pm 1$ произвольно доопределено (например, нулем). Сумма интегральных слагаемых, таким образом, равна

$$\Phi_1 + \Phi_2 = k_5\int\limits_{0}^{x_1}\mathrm{sat}\left(k_3\mathrm{sat}\left(k_1 s\right)\right)\left[k_5 - k_1\mathrm{sat}'\left(k_1 s\right)\right]ds + k_5\int\limits_{0}^{x_2+\mathrm{sat}\left(k_1x_1\right)}\mathrm{sat}\left(k_3\tilde{s}\right)d\tilde{s}.$$

Так как производная функции насыщения равна единице или нулю и по условию теоремы $k_5 > k_1 > 0$, имеем

$$k_5 - k_1\ \ \mathrm{sat'(k1s)} > 0\ \ \forall s, \tag{4.4}$$

откуда следует положительность суммы $\Phi_1 + \Phi_2$ и функции $V(x)$ при всех $x \neq 0$.

Очевидно также, что $V(x)$ стремится к бесконечности при $||x|| \to \infty$. Далее, дифференцируя $V(x)$ в силу системы (1.3), (2.2) и опуская аргумент k_1x_1 функций sat и sat′ для сокращения записи, получим

$$\dot{V} = k_5^2\mathrm{sat}\left(k_3\mathrm{sat}\left(\cdot\right)\right)x_2 + k_5x_2\int\limits_{0}^{x_2}\mathrm{sat}\ \left(k_3\left(s + \mathrm{sat}\left(\cdot\right)\right)\right)k_3\mathrm{sat}'\left(\cdot\right)k_1ds +$$

$$+ k_5\mathrm{sat}(k_3(x_2 + \mathrm{sat}(\cdot)))x_3 + (x_3 + k_5x_2)[-k_5(x_3 + \mathrm{sat}(k_3(x_2 + \mathrm{sat}(\cdot))) + k_5x_3] =$$

$$= k_5^2\mathrm{sat}\left(k_3\mathrm{sat}\left(\cdot\right)\right)x_2 - k_5^2\mathrm{sat}\left(k_3\left(x_2 + \mathrm{sat}\left(\cdot\right)\right)\right)x_2 +$$

$$+ k_1k_3k_5\mathrm{sat}'\left(\cdot\right)x_2\int\limits_{0}^{x_2}\mathrm{sat}\ \left(k_3\left(s + \mathrm{sat}\left(\cdot\right)\right)\right)ds.$$

Преобразуем интеграл в правой части последнего выражения:

$$\int\limits_{0}^{x_2}\mathrm{sat}\ \left(k_3\left(s + \mathrm{sat}\left(\cdot\right)\right)\right)ds = \int\limits_{\mathrm{sat}(\cdot)}^{x_2+\mathrm{sat}(\cdot)}\mathrm{sat}\ \left(k_3\tilde{s}\right)\tilde{s} =$$

$$= \frac{1}{k_3}\mathrm{sat}\left(k_3\left(x_2 + \mathrm{sat}\left(\cdot\right)\right)\right) - \frac{1}{k_3}\mathrm{sat}\left(k_3\mathrm{sat}\left(\cdot\right)\right).$$

Подставляя полученное выражение в формулу для $\dot{V}\left(x\right)$, имеем

$$\dot{V}\left(x\right) = k_5\mathrm{sat}\left(k_3\mathrm{sat}\left(\cdot\right)\right)x_2\left(k_5 - k_1\mathrm{sat}'\left(\cdot\right)\right) - k_5\mathrm{sat}\left(k_3\left(x_2 + \mathrm{sat}\left(\cdot\right)\right)\right)x_2\left(k_5 - k_1\mathrm{sat}'\left(\cdot\right)\right) =$$

$$= -k_5\left(k_5 - k_1\mathrm{sat}'\left(\cdot\right)\right)\left[\mathrm{sat}\left(k_3\left(x_2 + \mathrm{sat}\left(\cdot\right)\right)\right) - \mathrm{sat}\left(k_3\mathrm{sat}\left(\cdot\right)\right)\right]x_2.$$

С учетом (4.3) и (4.4) $\dot{V}\left(x\right) \leqslant 0\ \forall x$. При $k_3 < 1$ выражение в квадратных скобках, а с ним и производная, обращаются в ноль только на множестве $x_2 = 0$, которое не содержит ни одной целой траектории, кроме $x = 0$. Если $k_3 \geqslant 1$, производная дополнительно обращается в ноль на подмножествах областей D_3^+ и D_3^-, в которых оба слагаемых в квадратных скобках одновременно равны +1 и −1 соответственно. Докажем, что множества D_3^+ и D_3^- (см. рис. 1) также не могут содержать целых траекторий.

Пусть для определенности $\left(x_1, x_2\right) \in D_3^-$, где $U_1(x_1,x_2) \equiv -1$. Третье уравнение системы (1.3) при этом принимает вид

$$\dot{x}_3 = -k_5(x_3 + 1),$$

и его решение $x_3(t) \to -1$ при $t \to \infty$. Отсюда, каково бы ни было начальное значение $x_3(0)$ и число c, $-1 < c < 0$, начиная с некоторого конечного момента времени $t_* \geqslant 0$, будет выполнено неравенство $x_3(t) \leqslant c$. Из второго уравнения системы (1.3) видно, что при этом $\dot{x}_2 \leqslant c < 0$ и через конечное время x_2 станет меньше любого наперед заданного значения, откуда следует, что проекция траектории на плоскость (x_1,x_2) непременно попадет либо в область D_1, либо в D_2^+ (рис. 1), т.е. область $D_3^- \times R^1$ не может содержать целых траекторий. Аналогично доказывается, что область $D_3^+ \times R^1$ не содержит целых траекторий.

Таким образом, функция $V(x)$ удовлетворяет всем условиям теоремы Барбашина–Красовского [14] и, следовательно, начало координат является асимптотически устойчивым положением равновесия системы (1.3), (2.2) в целом. Теорема 1 доказана.

5. Стабилизация при ограниченном ресурсе управления. При постановке задачи стабилизации в разд. 1 предполагалось, что ресурс управления не ограничен. В ряде приложений такое предположение вполне допустимо. Например, в упоминавшемся в разд. 1 примере механической системы ввиду ограниченности максимальной тяги и ускорения, правая часть (1.4) ограничена и при достаточно большой мощности шагового мотора коэффициенты могут быть выбраны так, чтобы управление в представляющей интерес области пространства состояний не достигало насыщения.

В случае же, когда ограниченностью управления нельзя пренебречь, применяя функцию насыщения к правой части (2.2), получим обратную связь в виде трех вложенных сатураторов:

$$U_c(x) = -k_6 \mathrm{sat}\left(\frac{k_5}{k_6}\left(x_3 + \mathrm{sat}\left(k_3\left(x_2 + \mathrm{sat}\left(k_1 x_1\right)\right)\right)\right)\right), \tag{5.1}$$

где k_6 ресурс управления (выражение безразмерного коэффициента k_6 через размерные выводится так же, как и для остальных коэффициентов в разд. 1; см. [10]).

Необходимые условия устойчивости системы (1.3), (5.1) устанавливаются как и в случае неограниченного управления. При любых коэффициентах обратной связи найдется такая окрестность нуля, в которой все три сатуратора не насыщены и система (1.3), (5.1) линейна. Так как в окрестности нуля обе системы имеют одинаковый вид (3.8), получаем те же необходимые условия, что и для системы с неограниченным управлением: положительность всех коэффициентов и выполнение условия $k_5 > k_1$.

Исследование устойчивости системы с ограниченным управлением представляет собой гораздо более сложную задачу. Численные эксперименты показали [10], что потеря системой устойчивости сопровождается возникновением так называемых скрытых колебаний (аттракторов) [15], причем множество аттракторов может быть несчетным [10]. Как и в случае неограниченного управления, достаточные условия глобальной устойчивости могут быть получены с помощью функции Ляпунова. На настоящий момент, однако, построить функцию Ляпунова для системы (1.3), (5.1) авторам не удалось.

Очевидно, что в случае обратной связи (5.1) выполнения необходимых условий не достаточно для глобальной устойчивости. Из общих соображений очевидно, что, каковы ни были положительные коэффициенты $k_1, \ldots, k_5$, система не может быть стабилизирована при любых отклонениях от равновесия, если ресурс управления недостаточен. Кроме того, управление, гарантирующее глобальную устойчивость, может оказаться неэффективным, не обеспечивая желаемую скорость приближения к положению равновесия. С другой стороны, на практике, требование глобальной устойчивости является излишним, так как начальные отклонения от равновесия, как и действующие на систему возмущения, обычно ограничены. В таких случаях, интерес представляют нахождение областей притяжения нулевого решения для заданного набора коэффициентов обратной связи и/или построение их оценок.

Оценки областей притяжения могут быть получены с помощью известных подходов к исследованию устойчивости, основанных, например, на построении функции Лурье–Постникова [16] или погружении в класс линейных нестационарных систем с последующим применением методов абсолютной устойчивости [4, 9, 17, 18]. Работа над нахождением достаточных условий и построением оценок областей продолжается. В настоящей же работе ограничимся одним специальным случаем достаточных условий.

Представленные далее в этом разделе достаточные условия глобальной устойчивости системы с ограниченным управлением получены сведением к рассмотренному выше случаю неограниченного управления. В разд. 1 доказано, что область

$$\Pi = \{x : |x_3| \leqslant 1\} \tag{5.2}$$

является инвариантным множеством безразмерной системы (1.3), (2.2). Легко видеть, что доказательство не зависит от того, ограничено или неограничено управление, так что П является также инвариантным множеством и для системы (1.3), (5.1). Сначала докажем, что при любых начальных условиях система (1.3), (5.1) гарантировано попадет в область П, а затем найдем, при каких условиях на коэффициенты управление не достигает насыщения в этой области.

Л е м м а 2. Каковы бы ни были начальные условия, система (1.3), (5.1) за конечное время попадет в область П.

Д о к а з а т е л ь с т в о л е м м ы 2. Рассмотрим функцию $v(x) = \frac{x_3^2}{2}$, положительно определенную при всех $x_3 \neq 0$, и продифференцируем ее в силу системы (1.3), (5.1): $\dot{v}(x) = -x_3 k_6 \mathrm{sat}(k_5(x_3 + \mathrm{sat}(k_3(x_2 + \mathrm{sat}(k_1 x_1))))/k_6)$. Функция $\dot{v}(x)$ определенно отрицательна в области $|x_3| > 1$, т.е. множество $|x_3| \leqslant 1$ является притягивающим для решений системы (1.1), (2.2). Пусть для определенности $x_3 > 0$. Предположим, что утверждение леммы неверно и траектория не пересекает плоскость $x_3 = 1$. Так как $x_3 \to U_1(x_1,x_2)$ и $|U_1(x_1,x_2)| \leqslant 1$, такое возможно только, если, начиная с некоторого момента t^*, функция U_1 тождественно равна единице на траектории системы: $U_1(x_1(t),x_2(t)) \equiv 1\ \forall t > t^*$. Последнее означает, что проекция траектории на плоскость (x_1,x_2) лежит в области D_3^- (рис. 1). Однако при доказательстве теоремы 1 было установлено, что области $D_3^- \times R^1$ и $D_3^+ \times R^1$ не могут содержать целых траекторий. Следовательно, предположение не верно и найдется конечный момент времени, после которого траектория окажется в инвариантном множестве П. Лемма 2 доказана.

Нетрудно видеть, что на множестве П управление (2.2) ограничено: $|U(x)| \leqslant 2k_5$. Сравнивая правые части формул (2.2) и (5.1), находим, что если $k_6 \geqslant 2k_5$, то управление (5.1) не достигает насыщения в области П, т.е. $U_c(x) \equiv U(x) \forall x \in$ П, и, значит, устойчивость системы (1.3), (5.1) следует из устойчивости системы (1.3), (1.4). Таким образом, с учетом инвариантности множества П и леммы 2 мы доказали следующее утверждение.

Т е о р е м а 2. Пусть все коэффициенты обратной связи (5.1) положительны и удовлетворяют условиям $k_5 > k_1$ и $k_6 \geqslant 2k_5$. Тогда система (1.3), (5.1) глобально асимптотически устойчива.

З а м е ч а н и е 2. Значение теоремы 1 в том, что она устанавливает глобальную стабилизируемость цепочки трех интеграторов с помощью ограниченного управления в виде вложенных сатураторов. Другими словами, при ограниченном ресурсе управления коэффициенты обратной связи всегда могут быть выбраны так, что система стабилизируется из любого начального состояния. На практике, однако, управление с коэффициентами, удовлетворяющими условиям теоремы 1, может оказаться достаточно консервативным при малом ресурсе управления k_6 из-за малого коэффициента k_1 (так как $k_1 < \frac{k_6}{2}$ ответственного за скорость убывания отклонения в окрестности равновесия.

6. Численные примеры. В качестве иллюстрации приведены результаты численных расчетов для обратной связи (2.2) с коэффициентами $k_1 = 1$, $k_3 = 3$ и $k_5 = 9$. На рис. 2 изображена инвариантная область системы, ограниченная поверхностью уровня функции Ляпунова (4.1) $V(x) = k_5^2$. Как видно из рисунка, форма поверхности заметно отличается от эллипсоидальной формы квадратичной функции Ляпунова линейной системы. Для систем, конструкция которых не допускает нарушения фазового ограничения (как в вышеупомянутой механической системе), больший интерес представляет множество, образованное пересечением инвариантных областей $V_c = \{x : V(x) \leqslant c\}$ и П (5.2), так как “хвосты” первой, отсекаемые плоскостями $x_3 = \pm 1$, содержат только траектории с начальными условиями, не удовлетворяющими фазовому ограничению. На рис. 3 и 4 показана цилиндроподобная область, полученная пересечением множества V_c, изображенного на рис. 2 с множеством П, вместе с несколькими траекториями, начинающимися на ограничивающей область поверхности уровня.

На рис. 3 изображены траектории, начинающиеся в точках, в которых производная функции Ляпунова в силу системы равна нулю. Начальные сегменты таких траекторий (показаны сплошными линиями) лежат на поверхности инвариантной области. После того, как проекция траектории на плоскость (x_1,x_2) пересекает границу D_3 (рис. 1), производная становится отрицательной и траектория уходит внутрь области. Участки траекторий, лежащие внутри области, изображены пунктирными линиями. Траектории, начинающиеся в точках, в которых производная отрицательна, сразу уходят внутрь области (рис. 4).

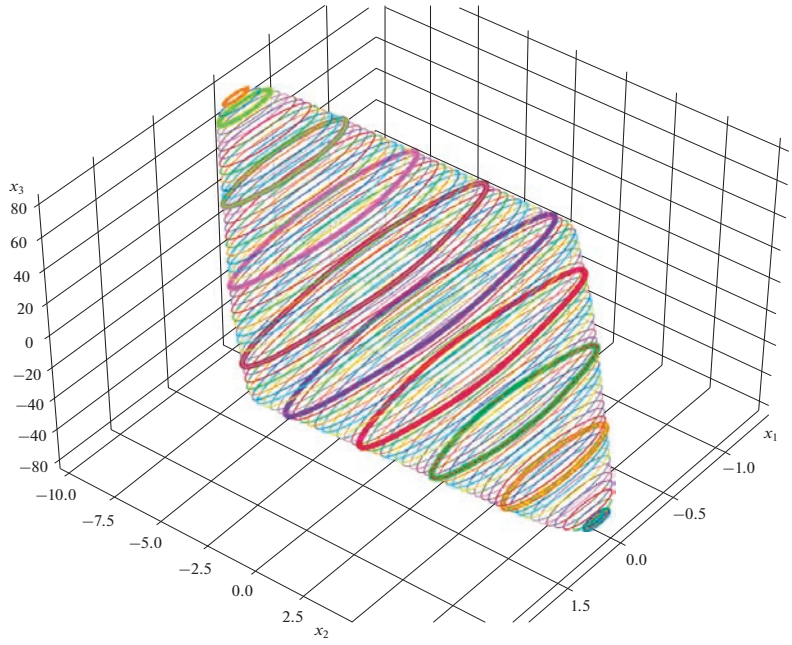

Рис. 2. Поверхность уровня функции Ляпунова (4.1)

Результаты решения задачи стабилизации системы для начальных условий $x_1(0) = -1.5$, $x_2(0) = -1.0$, $x_3(0) = 0.9$ представлены на рис. 5, демонстрирующие эффективность стабилизации. Кривые, помеченные цифрами *1–4*, показывают графики зависимостей от времени отклонения x_1, скорости x_2, ускорения x_3 и управления U соответственно. В начальный момент времени система движется в противоположном от положения равновесия направлении, величина отклонения после естественного роста на начальном этапе быстро (экспоненциально) убывает, фазовое ограничение выполняется для любого t, управление умеренно ограничено и не приводит к перерегулированию.

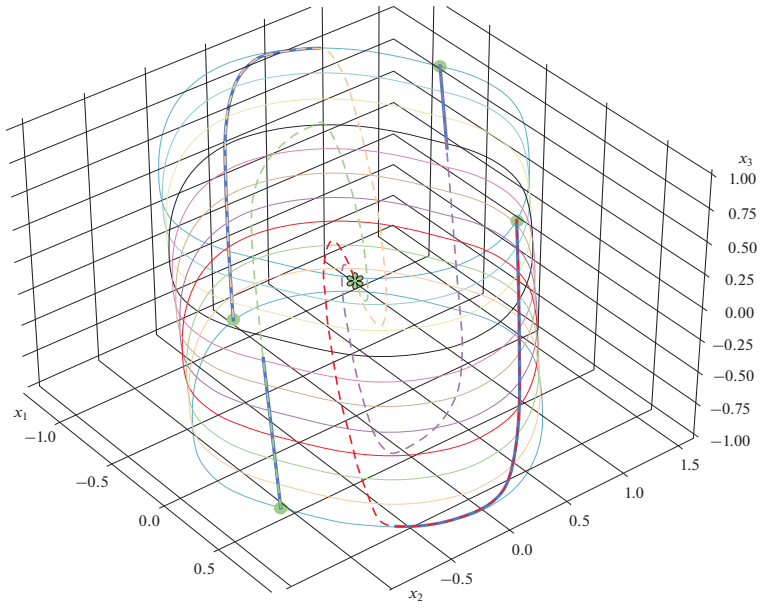

Рис. 3. Пересечение инвариантных областей $\left\{x : V(x) \le k_5^2\right\}$ и П

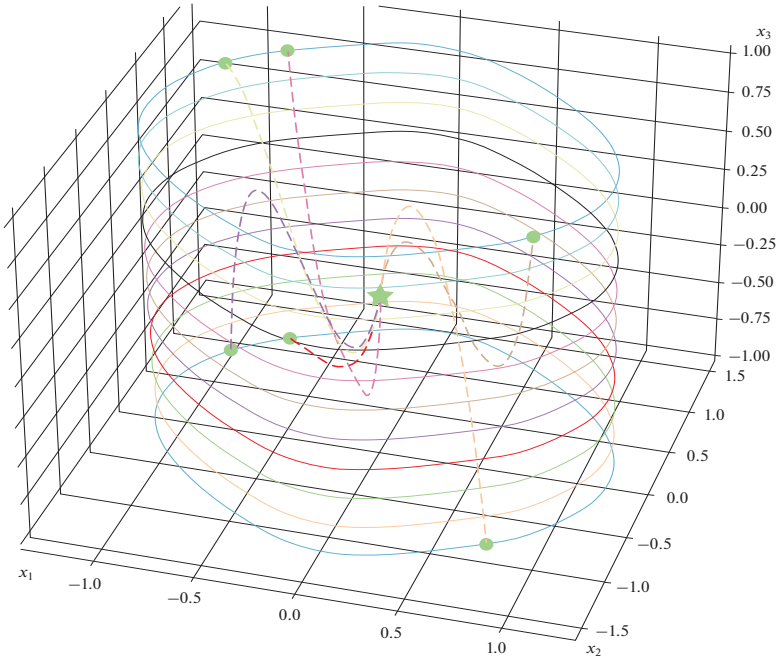

Рис. 4. Пересечение инвариантных областей $\left\{x : V(x) \le k_5^2\right\}$ и П

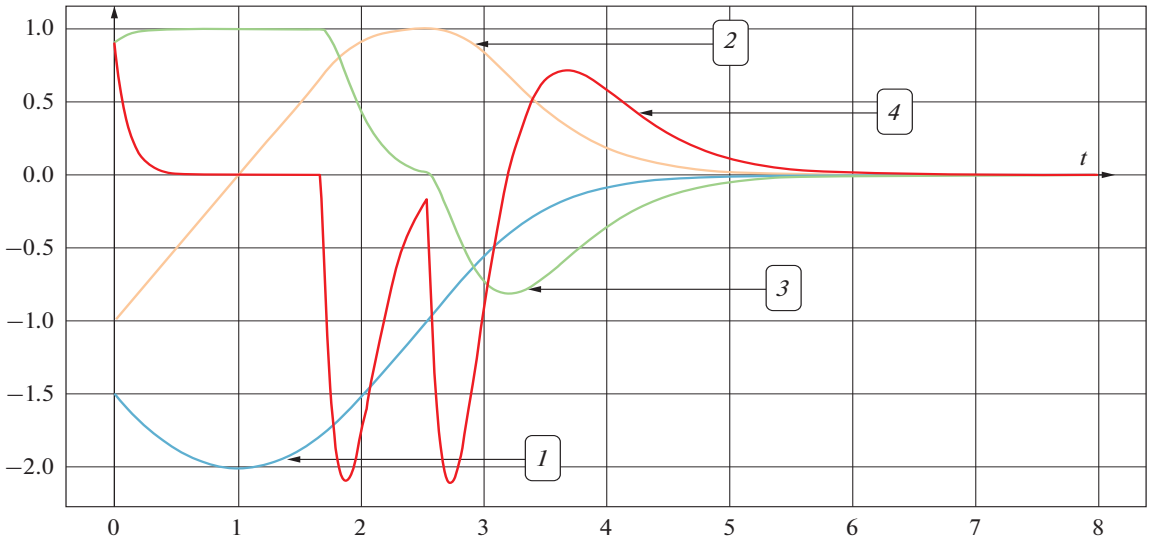

Рис. 5. Графики зависимостей от времени отклонения $x_1(t)$ (*1*), скорости $x_2(t)$ (*2*), ускорения $x_3(t)$ (*3*) и управления $U(t)$ (*4*)

Заключение. Рассмотрена задача стабилизации цепочки трех интеграторов с помощью непрерывного управления в виде вложенных сатураторов, гарантирующего выполнения фазового ограничения на третью переменную состояния. Обсуждаются преимущества предлагаемой обратной связи. С помощью перехода к безразмерным переменным исходная, зависящая от пяти коэффициентов обратной связи, задача сведена к исследованию трехпараметрической системы. Главный результат работы – построение функции Ляпунова, с помощью которой доказано, что для глобальной устойчивости замкнутой системы достаточно выполнения необходимых усло-

вий локальной устойчивости. Найдены также достаточные условия глобальной устойчивости для системы с ограниченным ресурсом управления. Приведены численные примеры, иллюстрирующие эффективность стабилизации с помощью предлагаемой обратной связи.

СПИСОК ЛИТЕРАТУРЫ

1. *Kurzhanski A.B., Varaiya P.* Solution Examples on Ellipsoidal Methods: Computation inHigh Dimensions. Cham, Switzerland: Springer, 2014.
2. *Teel A.R.* Global Stabilization and Restricted Tracking for Multiple Integrators with Bounded Controls // Sys. & Cont. Lett. 1992. V. 18. № 3. P. 165–171.
3. *Teel A.R.* A Nonlinear Small Gain Theorem for the Analysis of Control Systems with Saturation // Trans. Autom. Contr. IEEE. 1996. V. 41. № 9. P. 1256–1270.
4. *Li Y., Lin Z.* Stability and Performance of Control Systems with Actuator Saturation. Basel: Birkhauser, 2018.
5. *Olfati-Saber R.* Nonlinear Control of Underactuated Mechanical Systems with Application to Robotics and Aerospace Vehicles, Ph.D. Dissertation, Massachusetts Institute of Technology. Dept. of Electrical Engineering and Computer Science, 2001.
6. *Hua M.D., Samson C.* Time Sub-optimal Nonlinear Pi and Pid Controllers applied to Longitudinal Headway Car Control // Intern. J. Control. 2011. V. 84. P. 1717–1728.
7. *Pesterev A.V., Morozov Yu.V., Matrosov I.V.* On Optimal Selection of Coefficients of a Controller in the Point Stabilization Problem for a Robot-wheel // Communicat. Comput. Inform. Sci. (CCIS). 2020. V. 1340. P. 236–249.
8. *Pesterev A.V., Morozov Yu.V.* Optimizing Coefficients of a Controller in the Point Stabilization Problem for a Robot-wheel // Lect. Notes Comput. Sci. 2021. V. 13078. P. 191–202.
9. *Pesterev A.V., Morozov Yu.V.* The Best Ellipsoidal Estimates of Invariant Sets for a Third-Order Switched Affine System // Lect. Notes Comput. Sci. 2022. V. 13781. P. 66–78.
10. *Pesterev A.V., Morozov Yu.V.* Optimizing a Feedback in the Form of Nested Saturators to Stabilize the Chain of Three Integrators // Lect. NotesComput. Sci. 2023. V. 14395. P. 88–88.
11. *Морозов Ю.В., Пестерев А.В.* Глобальная устойчивость гибридной аффинной системы 4-го порядка // Изв. РАН. ТиСУ. 2023. № 5. С. 3–15.
12. *Пестерев А.В., Morozov Yu.V.* Глобальная стабилизация интегратора второго порядка обратной связью в виде вложенных сатураторов // АиТ. 2024. № 4. С. 55–60.
13. *Поляк Б.Т., Щербаков П.С.* Робастная устойчивость и управление. М.: Наука, 2002.
14. *Барбашин Е.А.* Введение в теорию устойчивости. Сер.: Физико-математическая библиотека инженера. М.: Наука, 1967.
15. *Кузнецов Н.В.* Теория скрытых колебаний и устойчивость систем управления // Изв. РАН. ТиСУ. 2020. № 5. С. 5–27.
16. *Лурье А.И., Постников В.Н.* К теории устойчивости регулируемых систем // ПММ. 1944. № 3. С. 246–248.
17. *Рапопорт Л.Б.* Оценка области притяжения в задаче управления колесным роботом // АиТ. 2006. № 9. С. 69–89.
18. *Generalov A., Rapoport L., Shavin M.* Attraction Domains in the Control Problem of a Wheeled Robot Following a Curvilinear Path over an Uneven Surface // Lect. NotesComput. Sci. 2021. V. 13078. P. 176–190.