СОДЕРЖАНИЕ

Том 61, номер 7, 2021 год

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

МАТЕМАТИЧЕСКАЯ ФИЗИКА

ИНФОРМАТИКА

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1059–1069

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

УДК 517.977.5

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ ЗАДАЧ ОПТИМИЗАЦИИ БОЛЬШИХ РАЗМЕРНОСТЕЙ С ИСПОЛЬЗОВАНИЕМ МОДИФИКАЦИИ МЕТОДА Б.Т. ПОЛЯКА[1)]

© 2021 г. А. Н. Андрианов[2,*], А. С. Аникин[1,**], А. Ю. Горнов[1,*]**

[1] *664033 Иркутск, ул. Лермонтова, 134, Институт динамики систем и теории управления СО РАН, Россия*

[2] *125047 Москва, Миусская пл., 4, Институт прикладной математики им. М.В. Келдыша РАН, Россия*

**e-mail: gornov@icc.ru*

***e-mail: anikin@icc.ru*

****e-mail: and@a5.kiam.ru*

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Предложена модификация специального метода выпуклой оптимизации Б.Т. Поляка. Свойства соответствующего алгоритма исследованы путем вычислительных экспериментов для задач выпуклой сепарабельной и несепарабельной оптимизации, невыпуклых задач оптимизации потенциалов атомно-молекулярных кластеров и модельной задачи оптимального управления. Реализованы последовательные и параллельные версии алгоритма, позволившие решить задачи с размерностями до ста миллиардов переменных. Библ. 10. Фиг. 4. Табл. 6.

Ключевые слова: выпуклая оптимизация, метод Б.Т. Поляка, задачи больших размерностей.

DOI: 10.31857/S0044466921070036

1. ВВЕДЕНИЕ

В ранних работах Бориса Теодоровича Поляка (см., например, [1]) был предложен (суб)градиентный метод поиска экстремума в выпуклых задачах математического программирования. Большого внимания специалистов метод не привлек, что связано, очевидно, с некоторыми его “неудобными” особенностями. Во-первых, сходимость метода гарантировалась только при минимизации выпуклых функций, нарушение выпуклости могло приводить к нарушению релаксации на итерациях. Во-вторых, для работы метода требовалось априорное знание точного значения функции в экстремальных точках. Сформулированные требования и наличие развиваемых в те годы методов сопряженных градиентов Флетчера–Ривса и Полака–Поляка–Рибьера (см. [2]) привели к незаслуженному, на наш взгляд, снижению активности по исследованию и использованию предложенного алгоритма.

Новый всплеск интереса к методу появился лишь в последние годы, что связано, очевидно, с рассмотрением больших и сверхбольших задач оптимизации. В соответствии с современной классификацией (см., например, [3], [4]) задачи выпуклой оптимизации можно разделить по размерности на: “Small” – до 100 переменных, “Medium” – от 10^3 до 10^4, “Large” – от 10^5 до 10^7 и “Huge” – более 10^8 переменных. В соответствии с доминирующей в настоящее время точкой зрения решать задачи выпуклой оптимизации размерности 10^8 и более при отсутствии разреженности (при “плотных” векторах) с использованием современной вычислительной техники малореально.

Основные недостатки метода Поляка при рассмотрении задач, в которых указанные требования не выполняются, в том числе, невыпуклых, возможно преодолевать путем введения несложных модификаций. Например, шаг метода по направлению градиента в общем случае может приводить в точку, в которой значение функции больше, чем значение в текущей точке. В этой ситуации можно, не слишком утруждаясь, организовать кратное уменьшение шага, например, в

[1)]Работа выполнена при финансовой поддержке РФФИ (код проекта № 18-07-00587).

два раза, при этом свойство релаксации будет гарантировано, если, конечно, еще не найдено решение из множества оптимальных. Однако сильные стороны метода (“дешевая” итерация, малые требования к оперативной памяти) могут оказаться очень удобными при рассмотрении задач больших и сверхбольших размерностей.

В работе рассматриваются модификации метода Поляка, ориентированные на решение больших и свербольших задач оптимизации. На простейших примерах из семейств выпуклых функций сепарабельного и квазисепарабельного типа исследуются возможности численного решения гладких задач безусловной минимизации растущих размерностей. На примерах мультиэкстремальных задач оптимизации потенциалов атомно-молекулярных кластеров Морса (см. [5]) и Китинга (см. [6]) исследуются глобализующие свойства модификации метода Поляка. Для нелинейных задач оптимального управления оцениваются возможности решения аппроксимативных задач на сетках с уменьшающимся постоянным шагом. Рассматриваемые модификации метода реализованы на языке C (компиляторы BCC, GCC, Clang и ICC) под управлением операционных систем Linux, Mac OS X и Windows с использованием технологий распараллеливания OpenMP и MPI. Приводимые в работе вычислительные эксперименты выполнялись как на обычных персональных компьютерах, так и на кластерных системах МВС-100К Межведомственного Суперкомпьютерного Центра (МСЦ) РАН и Института прикладной математики им. М.В. Келдыша (ИПМ) РАН.

2. ОБЩАЯ СТРУКТУРА МЕТОДОВ ГРАДИЕНТНОГО ТИПА

Рассматривается стандартная задача конечномерной оптимизации:

$$f(x) \to \min, \quad f(x) \in \mathbb{R}, \quad x \in \mathbb{R}^n. \tag{1}$$

Итерация классического алгоритма градиентного типа записывается в следующем традиционном виде:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

где направление спуска d_k выбирается как антиградиент минимизируемой функции $f(x)$ в точке x_k:

$$d_k = -\nabla f(x_k). \tag{3}$$

Размер шага α_k может выбираться разными способами, опирающимися на учет свойств целевой функции. Например, для дифференцируемых функций с известной константой Липшица оптимальным является выбор шага $\alpha_k = 1/L$. Однако такой способ не всегда применим для задач рассматриваемого класса в силу отсутствия информации о значении константы Липшица. Величина шага в направлении спуска может в общем случае вычисляться по формуле

$$\alpha_k = \arg\min_{\alpha} f(x_k + \alpha d_k). \tag{4}$$

Данный вариант обеспечивает наилучшую функцию релаксации на одну итерацию, но требует значительных вычислительных затрат. Для сверхбольших задач оптимизации предлагается использовать вычислительные технологии, основанные на применении методов, не требующих постоянных вычислений оптимизируемой функции и позволяющих исследовать прикладные задачи оптимизации больших и сверхбольших размерностей.

3. МЕТОД Б.Т. ПОЛЯКА И ЕГО МОДИФИКАЦИЯ

Метод был предложен Б.Т. Поляком в [1] (и, по нашим сведениям, дальнейшего развития не получил):

$$x_{i+1} = x_i - \frac{f(x_i) - f^*}{\|\nabla f(x_i)\|^2} \nabla f(x_i), \tag{5}$$

где f^* – известное минимальное значение функции.

Для одномерного случая метод совпадает с методом Ньютона для решения уравнения $f(x) = f^*$.

Поскольку в большинстве реальных постановок задач оптимизации f^* заранее не известно, предлагается ввести некоторую величину $\delta_f > 0$, определяющую прогнозируемое значение оптимизируемой функции на каждом шаге:

$$f^* = f(x_i) - \delta_f,$$
$$x_{i+1} = x_i - \frac{\delta_f}{\|\nabla f(x_i)\|^2}\nabla f(x_i). \tag{6}$$

Сходимость метода следует из теоремы, доказанной в [1]:

Теорема 1. *Пусть $f(x)$ является выпуклой, непрерывной, Q выпукло и замкнуто, существует точка минимума $x^* \in Q$, $f(x^*) = f^*$ и $\|\nabla f(x)\| \le c$ на $S = \left\{x \in Q, \left\|x - x^0\right\| \le \left\|x^* - x^0\right\|\right\}$. Тогда в методе* (5) *$f(x^n) \to f^*$, а x^n слабо сходится к некоторой точке минимума.*

Здесь функция $f(x)$ определена на некотором множестве Q. В случае неактивных границ и использования множеств простой структуры (например, параллелепипедного типа), приведенные результаты применимы к рассматриваемой постановке (1).

Алгоритм предлагаемой модификации метода Б.Т. Поляка строится следующим образом.

В качестве входных данных используются: начальная точка x_0, точность по норме градиента $\varepsilon_\nabla > 0$, точность по функции $\varepsilon_f > 0$, прогнозная оценка $\delta_f > 0$, коэффициент корректировки шага $0 < k < 1$, минимально допустимый шаг $\varepsilon_\lambda > 0$, $i = 0$ – номер итерации.

Algorithm 1. Основной цикл алгоритма

1: Вычисляется $f(x_i)$, $\nabla f(x_i)$, $\|\nabla f(x_i)\|$
2: **if** $\|\nabla f(x_i)\| < \varepsilon_\nabla$ или $(f(x_{i-1}) - f(x_i)) < \varepsilon_f$ **then**
3: переходим на шаг 13
4: **end if**
5: Находится шаг $\lambda = \dfrac{\delta_f}{\|\nabla f(x_n)\|^2}$
6: Вычисляется $x_{i+1} = x_i - \lambda \cdot \nabla f(x_i)$
7: **if** $f(x_{i+1}) < f(x_i)$, **then**
8: полагается $i = i + 1$ и выполняется переход на шаг 1
9: **end if**
10: **if** $\lambda > \varepsilon_\lambda$ **then**
11: $\lambda = \lambda \cdot k$ и выполняется переход к шагу 6
12: **end if**
13: Возвращается x_i – как найденная точка минимума

Утверждение о сходимости предлагаемого метода для выпуклых задач с известным f^* и половинным делением шага ($k = 0.5$) является тривиальным следствием теоремы 1, поскольку последовательность модифицированного метода обязательно содержит в себе последовательность исходного, сходимость которого строго доказана.

Можно выделить следующие особенности предложенной модификации.

1. При соответствующей настройке алгоритмических параметров (δ_f) предложенный метод может делать “большие” шаги, что добавляет ему глобализующие свойства.

2. Метод имеет высокий потенциал параллелизма.

3. Алгоритм имеет простую конструкцию, что имеет большое значение при реализации метода на ряде высокопроизводительных архитектур, например, GPU.

Таблица 1. Объем памяти (ОЗУ), требуемой для хранения n-мерного вектора вещественных чисел (float – одинарная точность, double – двойная)

n	Float		Double	
10^2	0.39	Kb	0.78	Kb
10^3	3.91	Kb	7.81	Kb
10^4	39.06	Kb	78.13	Kb
10^5	390.63	Kb	781.25	Kb
10^6	3.81	Mb	7.63	Mb
10^7	38.15	Mb	76.29	Mb
10^8	381.47	Mb	762.94	Mb
10^9	3.73	Gb	7.45	Gb
10^{10}	37.25	Gb	74.51	Gb
10^{11}	372.53	Gb	745.06	Gb
10^{12}	3.63	Tb	7.28	Tb

Рассмотрим далее примеры использования предложенной модификации метода при решении экстремальных задач различных классов.

4. ВЫПУКЛАЯ ОПТИМИЗАЦИЯ. СЕПАРАБЕЛЬНАЯ ТЕСТОВАЯ ЗАДАЧА

Решение задач безусловной минимизации класса “Huge-Scale” естественным образом сопряжено с рядом сложностей, одна из которых – собственно, размерность. Увеличение числа переменных очевидным образом влечет за собой повышение требований к объему памяти доступной вычислительной системы. При рассмотрении задач с предельными размерностями возникают ситуации, когда требуемый объем памяти слишком велик для одного вычислительного узла, что влечет за собой необходимость использования распределенных суперкомпьютерных мощностей. Объем памяти, требуемый для задач различной размерности, приведен в табл. 1.

Можно увидеть, что с точки зрения потребления памяти задачи с размерностями до 10^9 вполне возможно решать на современных вычислительных системах с общей памятью, задачи с большими размерностями требуют использования памяти суперкомпьютеров.

Рассмотрим сепарабельную функцию, имеющую следующий вид:

$$f(x) = \sum_{i=1}^{n} (x_i^2 + x_i^6). \tag{7}$$

Очевидно, что данная функция имеет минимум в точке, где все компоненты вектора x равны 0 и значение функции в этой точке также равно 0.

Расчеты проводились на суперкомпьютерах ИПМ РАН и МСЦ РАН, имеющих ограничение в 1 ГБ ОЗУ на процессорное ядро. Данное ограничение необходимо учитывать при выборе числа требуемых процессоров, чтобы не допустить “перерасхода” памяти на вычислительных узлах. Результаты расчетов для задач с различной размерностью приведены в табл. 2 и на фиг. 1.

Максимальная размерность тестовой сепарабельной задачи, которую удалось решить – 10^{11} переменных. На фиг. 1 хорошо видна линейная зависимость времени расчета от числа переменных, что говорит о хорошей масштабируемости предложенной модификации алгоритма.

5. ВЫПУКЛАЯ ОПТИМИЗАЦИЯ. НЕСЕПАРАБЕЛЬНАЯ ТЕСТОВАЯ ЗАДАЧА

Рассмотрим тестовую несепарабельную функцию вида

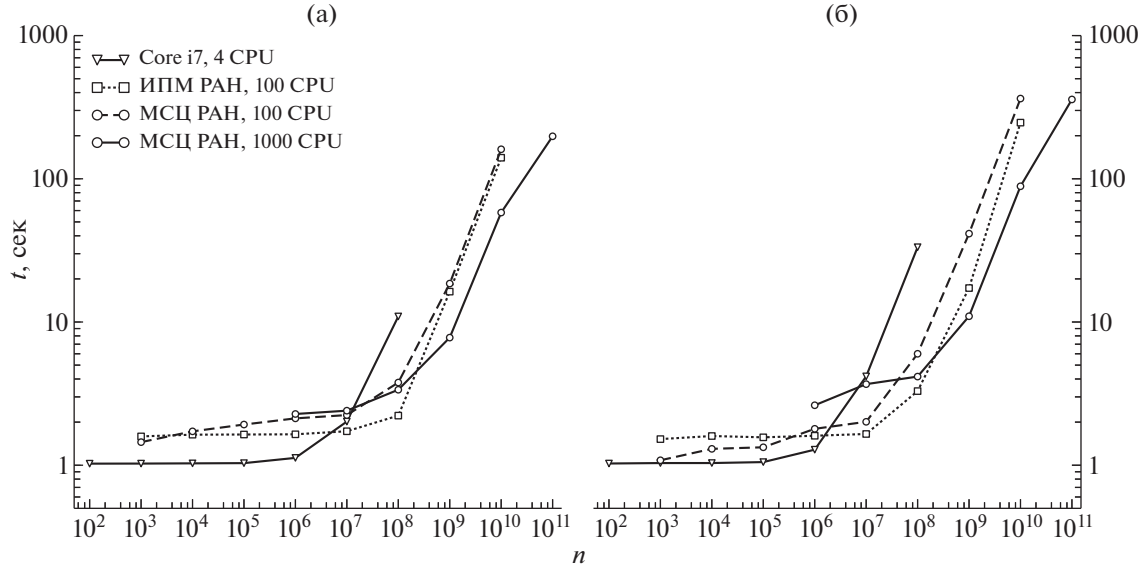
$$f(x) = \sum_{i=1}^{n} x_i^2 + \sum_{i=2}^{n} (x_i - x_{i-1})^2. \tag{8}$$

Таблица 2. Время [с], затраченное на работу алгоритма в зависимости от размерности тестовой сепарабельной функции

n	Core i7, 4 CPU	ИПМ, 100 CPU	МСЦ, 100 CPU	МСЦ, 1000 CPU
10^2	1.02559	–	–	–
10^3	1.02685	1.58932	1.44929	–
10^4	1.03031	1.63431	1.72392	–
10^5	1.03406	1.63837	1.92358	–
10^6	1.12513	1.64386	2.12392	2.27892
10^7	2.01723	1.72646	2.23966	2.40381
10^8	10.93844	2.22012	3.77897	3.36539
10^9	–	16.29881	18.55662	7.78179
10^{10}	–	140.30873	160.74543	58.09801
10^{11}	–	–	–	198.05384

При реализации варианта алгоритма для этого класса задач потребовалось применение технологии обмена сообщениями MPI для кластерных вычислительных систем. Произведенные оценки показали, что обмен данными между вычислительными узлами при нахождении градиента функции составляют незначительную величину от общего времени расчетов. Основным лимитирующим фактором для алгоритма все так же является объем доступной памяти.

Результаты расчетов для задач с различной размерностью приведены в табл. 3 и на фиг. 1. Произведенные вычислительные эксперименты показали, что для несепарабельных задач также имеет место линейная зависимость временных затрат от размерности. В частности, решение “максимальной” задачи с 10^{11} переменными с использованием 1000 процессоров потребовало всего 6 мин.



Фиг. 1. Время, затраченное на работу алгоритма в зависимости от размерности задачи: (а) – сепарабельная функция, (б) – несепарабельная функция.

Таблица 3. Время [с], затраченное на работу алгоритма в зависимости от размерности тестовой несепарабельной функции

n	Core i7, 4 CPU	ИПМ, 100 CPU	МСЦ, 100 CPU	МСЦ, 1000 CPU
10^2	1.02845	–	–	–
10^3	1.03536	1.52187	1.08275	–
10^4	1.03566	1.59923	1.30111	–
10^5	1.05166	1.56577	1.33428	–
10^6	1.28414	1.60642	1.79563	2.62236
10^7	4.17233	1.65246	2.00984	3.68216
10^8	33.29102	3.29295	5.99972	4.15825
10^9	–	17.26339	41.40902	10.97092
10^{10}	–	246.75377	363.31317	88.66335
10^{11}	–	–	–	358.11812

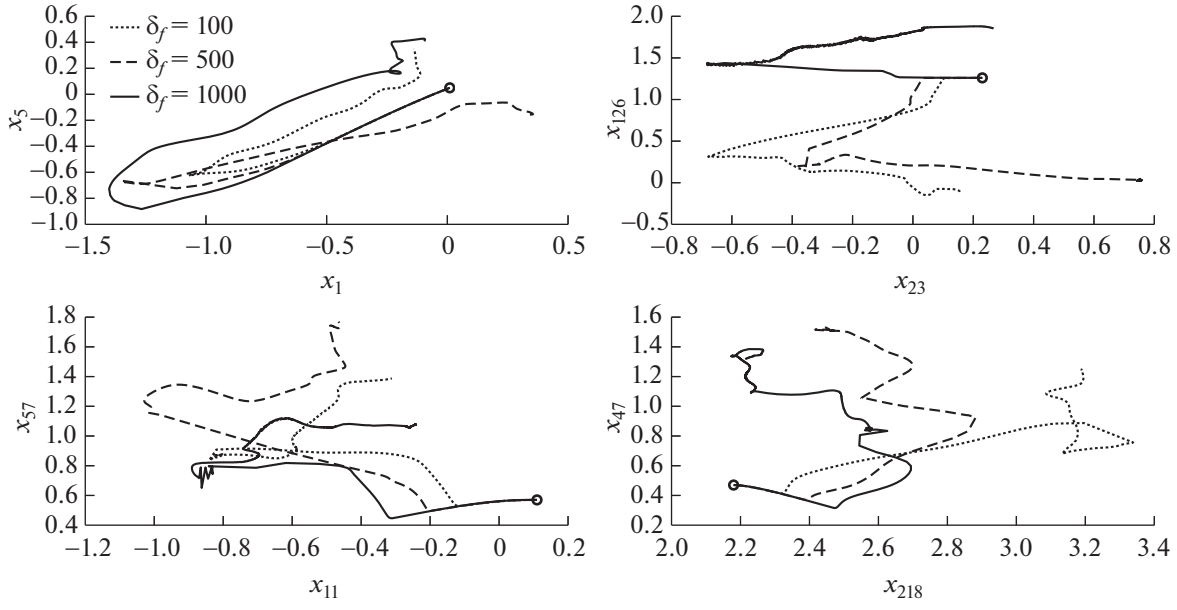
6. НЕВЫПУКЛАЯ ОПТИМИЗАЦИЯ. ПОИСК МОЛЕКУЛЯРНЫХ КЛАСТЕРОВ С МИНИМАЛЬНОЙ ЭНЕРГИЕЙ

Рассматривается потенциальная функция Морса (см. [5])

$$f(x) = V_M = \sum_{i=1}^{n}\sum_{j>i}^{n}[(e^{\rho_0(1-r_{ij})} - 1)^2 - 1]. \tag{9}$$

Задача минимизации данного потенциала является популярной “мультиэкстремальной” задачей глобальной оптимизации с астрономически растущим от размерности числом локальных экстремумов. Данная функция относится к классу несепарабельных, и оптимальное значение f^* неизвестно.

Традиционно для решения задач такого типа применяются специализированные методы невыпуклой оптимизации. Предложенная модификация алгоритма, дополненная механизмом случайного мультстарта, позволила успешно найти глобальные решения для кластеров, содер-



Фиг. 2. Проекции траекторий спусков для метода Поляка (модиф.) при различных значениях δ_f; минимизация потенциала Морса, 80 атомов (240 переменных).

Таблица 4. Минимизация потенциала Китинга, 102900 переменных (343000 атомов)

Метод	Время, с	$f_{\min}$	$\|\nabla f_{\min}\|$
Сопряженных градиентов	1087.242	4.421883e+01	3.401397e–05
Коши	2819.445	4.421883e+01	1.734354e–05
Поляка (модиф.)	19615.083	4.421883e+01	2.191246e–05

жащих до 80 атомов. При этом проявились нестандартные “глобализующие” свойства изначально локального метода, аппроксимирующего случайное распределение проб в окрестности рекордного значения. Примеры проекций траекторий спусков приведены на фиг. 2.

7. МИНИМИЗАЦИЯ ПОТЕНЦИАЛА КИТИНГА

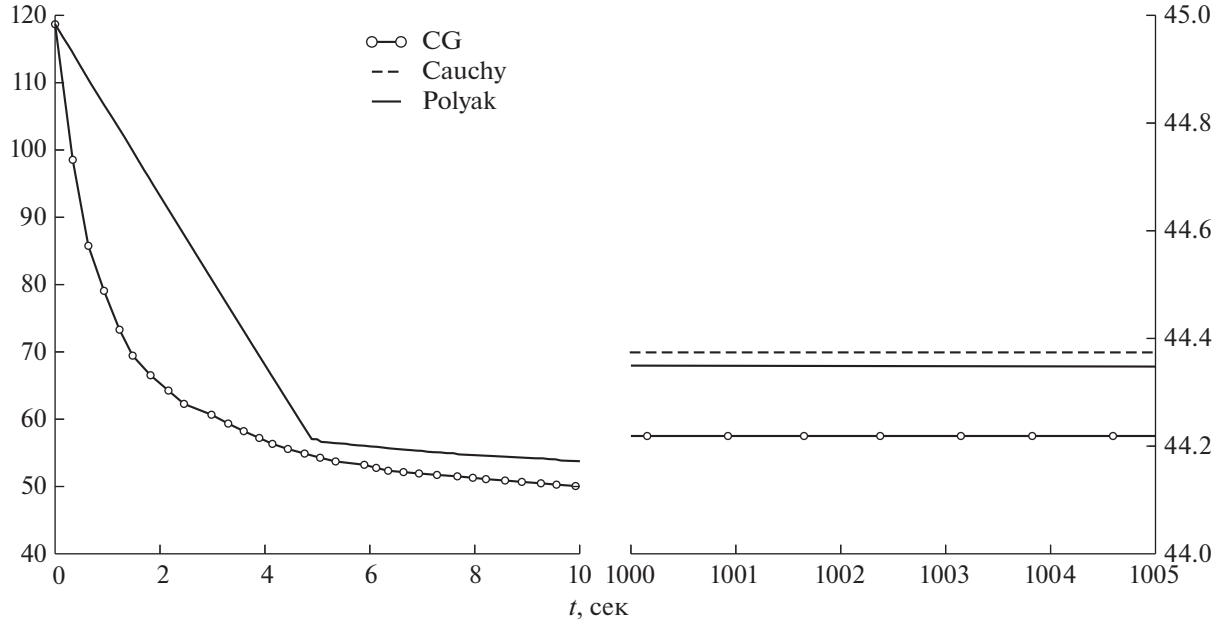
Рассмотрим задачу минимизации потенциала Китинга (см. [6]), моделирующего межатомное взаимодействие в кристаллических структурах “кремний–германий”:

$$E_k = \sum_{i=1}^{n}\left[\frac{3}{16}\sum_{j=1}^{4}\frac{\alpha_{ij}}{d_{ij}^2}\left\{\|r_i - r_j\|^2 - d_{ij}^2\right\}^2 + \frac{3}{8}\sum_{j=1}^{4}\sum_{k=j+1}^{4}\frac{\beta_{ijk}}{d_{ij}d_{ik}}\left\{\langle r_i - r_j, r_i - r_k\rangle + \frac{d_{ij}d_{ik}}{3}\right\}^2\right].$$

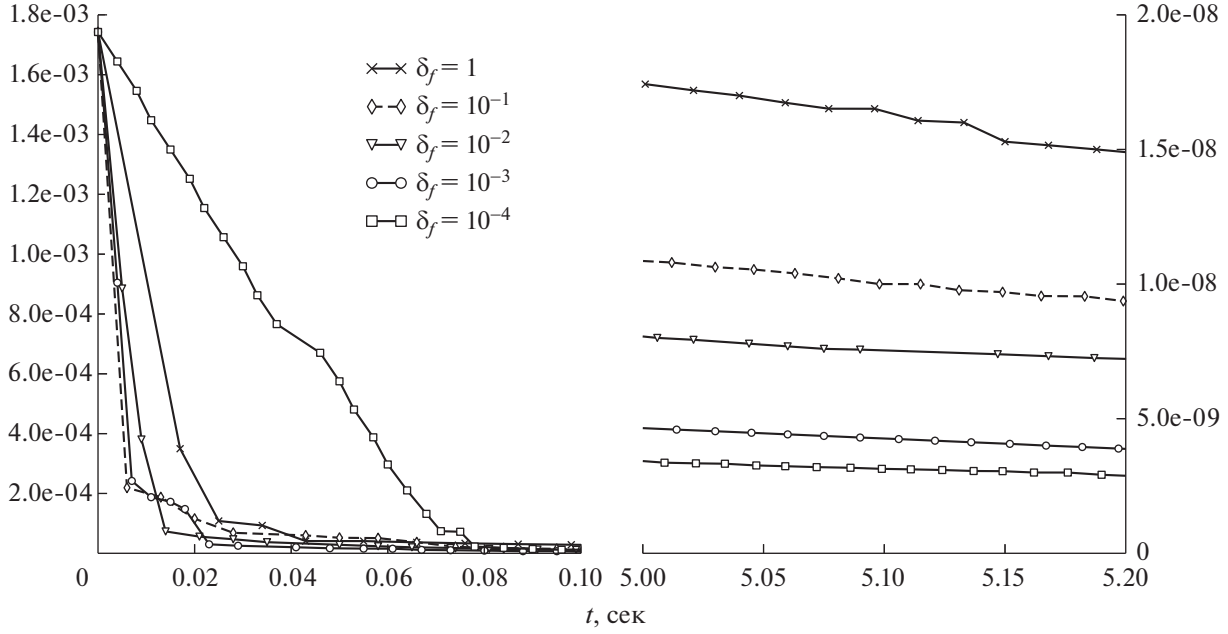
С применением предложенной модификации метода авторами производилась минимизация потенциала Китинга для кристалла с 1029000 переменными (343000 атомов в кристаллической решетке). Проводилось сравнение результатов расчетов с результатами вычислений, полученных с помощью метода сопряженных градиентов (вариант Полак–Поляк–Рибьер) и “метода Коши” (модификация метода наискорейшего спуска). Результаты расчетов приведены в табл. 4. На фиг. 3 показано поведение (сходимость) сравниваемых алгоритмов оптимизации.

Проводилось исследование поведения модифицированного алгоритма при различных значениях параметра δ_f на задаче минимизации потенциала Китинга для кристалла с 24000 переменными (8000 атомов), результаты показаны на фиг. 4.

Проведенные эксперименты показали, что и в этих задачах, в общем случае невыпуклых, с применением предложенной модификации алгоритма удалось получить верные решения за приемлемое время, хотя и немного проиграв в эффективности.



Фиг. 3. Минимизация потенциала Китинга, 1029000 переменных (343000 атомов).



Фиг. 4. Минимизация потенциала Китинга, 24000 переменных (8000 атомов).

8. ЗАДАЧА ОПТИМАЛЬНОГО УПРАВЛЕНИЯ

Традиционная постановка простейшей задачи оптимального управления (так называемая задача со свободным правым концом) заключается в поиске минимума терминального функционала

$$I(u) = \phi(x(t_1)), \tag{10}$$

определенного на траекториях управляемой системы дифференциальных уравнений (дифференциального включения)

$$\dot{x} = f(x(t), u(t), t), \tag{11}$$

заданной на интервале времени $t \in [t_0, t_1]$. Начальный фазовый вектор $x_0 = x(t_0)$ предполагается фиксированным, на управления наложены ограничения $\underline{u} \le u(t) \le \overline{u}$, где $\underline{u}$, $\overline{u}$ – фиксированные векторы. Рассматриваемая задача занимает центральное место в теории оптимального управления, к такой постановке сводится большое количество задач более сложных классов (с интегральными функционалами, с терминальными и фазовыми ограничениями, с нефиксированным временем и др.). Для задачи в такой постановке получен значительный объем глубоких теоретических результатов и предложено множество вычислительных методов. Однако практика численного решения задач такого класса, особенно прикладных, несмотря на усилия множества специалистов по теории управления, позволяет сделать вывод о доминировании методов, основанных на редукции к задачам конечномерной оптимизации. После введения сетки разбиения интервала времени, например, с равномерным шагом h, оказывается возможным сформулировать задачу конечномерной оптимизации, аппроксимирующую задачу в исходной постановке. В аппроксимативной задаче динамика системы описывается рекуррентными соотношениями $x(t+h) = f_h(x(t), u(t), t, h)$, все остальные условия сохраняются. Аппроксимативные задачи, очевидно, представляют собой специальный класс задач математического программирования с условиями типа равенств и неравенств. Если зафиксировать разностную схему, по которой выполняется дискретизация (Эйлера, Рунге–Кутты или другую), то можно воспользоваться предложенным в теории оптимального управления методом оценки градиента функционала, основанным на применении формализма сопряженных переменных. Получаемые по такой схеме алгоритмы оказываются весьма близкими к алгоритмам, которые генерируются с помощью методик быстрого автоматического дифференцирования (см., например, [7]). Вычисление градиента функционала по любой из указанных схем требует решения всего двух задач Коши. Таким образом, задача оптимального управления со свободным правым концом может быть рас-

Таблица 5. Время решения задачи оптимального управления программным комплексом OPTCON-A (N – число узлов сетки дискретизации)

N	$I(u^*)$	t, с	Задачи Коши	Норма градиента	Итерации
101	1.191521713372e+01	0	135	5.7e–3	15
201	1.190999877810e+01	0	317	2.4e–4	23
401	1.190841888965e+01	0	276	8.0e–5	24
801	1.190817286834e+01	0	182	2.2e–4	21
1601	1.190804296061e+01	0	241	1.2e–5	21
3201	1.190804923806e+01	0	167	6.8e–5	17
6401	1.190801439789e+01	1	109	1.8e–5	16
12801	1.190805357054e+01	3	206	3.5e–5	19
25601	1.190806771202e+01	8	237	4.7e–5	21
51201	1.190812331446e+01	13	203	4.4e–5	19
102401	1.190811071137e+01	22	167	3.8e–5	17
204801	1.190815091482e+01	44	166	4.0e–5	17
409601	1.190814768992e+01	88	166	2.9e–5	17
819201	1.190847626971e+01	144	131	2.9e–5	15

смотрена как задача безусловной минимизации с “дешевым”, алгоритмически оцениваемым градиентом, но вычисляемым с некоторой погрешностью, зависящей от размеров шага дискретизации h. Размер шага дискретизации непрерывной задачи задает число переменных в аппроксимативной задаче, равное, грубо говоря, $N = (t_1 - t_0)/h$.

Эффективность предложенной модификации метода Поляка попробуем оценить с помощью численных экспериментов для одной из самых популярных тестовых задач – задачи успокоения нелинейного маятника. Будем последовательно решать аппроксимативные задачи с растущим числом переменных N (и, соответственно, числом прямых ограничений на переменные), фиксируя процессорное время расчета, число итераций, решенных задач Коши и достигнутое значение функционала.

Динамика модели описывается системой (см., например, [8])

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = u - \sin x_1, \tag{12}$$

определенной на интервале времени $t \in [0, 5]$ с начальными условиями $x_1(0) = 5$, $x_2(0) = 0$. Требуется минимизировать функционал $I(u) = x_1^2(5) + x_2^2(5)$ при ограничениях на управления $|u(t)| \le 1$. В задаче имеет место два локальных экстремума, глобально оптимальное значение функционала $I(u^*) = 11.908013767$, полученное с применением метода интегрирования DOPRI8 (8-й порядок точности), достигается на релейном управлении:

$$u^*(t) = \begin{cases} +1, & t \in [0.0,\ 0.98244286] \cup [4.55036911,\ 5.0], \\ -1, & t \in [0.98244286,\ 4.55036911]. \end{cases}$$

Для решения задач Коши использовался метод интегрирования из семейства Рунге–Кутты 2-го порядка (так называемый метод “Эйлера с итерациями”). Расчет прерывался при выполнении условия $I(u^{k-1}) - I(u^k) < 10^{-6}$. Все задачи решались, начиная с одинаковых начальных приближений $u^0(t) \equiv -0.5$. В этой задаче, очевидно, известна гарантированная, но неоптимальная, нижняя оценка значения функционала, равная 0.

Эксперимент выполнялся с использованием стандартного программного комплекса (ПК) OPTCON-A (см. [9], [10]) в состав которого включена предложенная модификация алгоритма. Результаты эксперимента приведены в табл. 5.

Максимальной размерностью задач, решенных при использовании ПК OPTCON" A, оказалась размерность 819201, что связано, очевидно, с излишним объемом внутренней памяти комплекса, используемой для работы других алгоритмов. Процессорное время решения при размер-

Таблица 6. Время [с] решения задачи оптимального управления специализированной программой (N – число узлов сетки дискретизации)

N	AMD 6220 Linux	i7-2640M Linux			i7-2640M Mac OS X		i5 4260U Mac OS X	
	icc	icc	gcc	clang	gcc	clang	gcc	clang
	14.0.0	15.0.1	4.8.2	3.4.2	4.8.3	3.4.2	4.8.3	3.5SVN
101	0.006	0.001	0.006	0.009	0.001	0.001	0.002	0.002
201	0.014	0.003	0.012	0.005	0.003	0.003	0.004	0.004
401	0.018	0.007	0.014	0.014	0.006	0.006	0.009	0.008
801	0.033	0.010	0.023	0.028	0.011	0.010	0.015	0.015
1601	0.019	0.010	0.019	0.014	0.006	0.007	0.008	0.008
3201	0.085	0.031	0.065	0.070	0.044	0.034	0.045	0.046
6401	0.078	0.041	0.099	0.100	0.046	0.051	0.066	0.069
12801	0.289	0.159	0.271	0.311	0.156	0.154	0.220	0.255
25601	0.674	0.298	0.620	0.687	0.366	0.380	0.500	0.495
51201	1.030	0.521	1.060	1.234	0.627	0.613	0.870	0.942
102401	1.890	0.887	1.799	2.023	1.142	1.160	1.505	2.100
204801	3.546	1.765	4.415	4.015	2.352	2.319	2.866	3.044
409601	7.251	3.576	7.133	8.034	4.588	4.511	5.509	5.938
819201	11.662	5.714	11.608	12.883	7.447	7.315	8.752	8.682
1638401	18.197	8.929	18.203	20.443	11.588	11.381	13.547	13.313
3276801	35.353	17.975	35.947	40.622	22.972	25.862	27.110	26.504
6553601	50.279	26.991	52.610	59.321	33.680	33.090	39.626	39.432
13107201	103.237	–	–	–	–	–	–	–
26214401	205.603	–	–	–	–	–	–	–
52428801	346.577	–	–	–	–	–	–	–
104857601	692.295	–	–	–	–	–	–	–

ностях задач от 101 до 3201 переменной было менее 0.5 с и оценено в 0 с, поскольку используемый датчик времени позволяет учитывать его с точностью до 1. Начиная с размерности 401 переменных, во всех задачах получены значения функционала, с точностью до четырех знаков совпадающие с известным оптимальным значением. Быстрый рост процессорного времени с ростом размерности задач связан с неэффективностью работы процедур аппроксимации управления общего назначения, не позволяющих учитывать специфические особенности алгоритма.

Второй эксперимент проводился со специальной программной реализацией алгоритма, проверенной с использованием нескольких популярных компиляторов на нескольких различных компьютерах. В эксперименте участвовали следующие пары “процессор–компилятор”.

1. Intel Core i7-2640M; Linux; icc-15.0.1, gcc-4.8.2, gcc-4.9.1, clang-3.4.2, clang-3.5.0.

2. Intel Core i7-2640M; Mac OS X; gcc-4.8.3, clang-3.4.2.

3. Intel Core i5 4260U; Mac OS X; gcc-4.8.3, clang-600.0.56 (based on LLVM 3.5svn).

4. AMD Opteron 6220; Linux; icc-14.0.0.

Результаты выполненных расчетов приведены в табл. 6. Наилучшую производительность показала версия алгоритма, реализованная с использованием компилятора icc.

Проведенные эксперименты показали принципиальную возможность решения аппроксимативных задач оптимального управления с размерностями, превышающими 10^8 даже без применения параллельных вычислительных технологий.

9. ЗАКЛЮЧЕНИЕ

Рассмотренная в работе модификация метода Поляка позволила создать вычислительные технологии, обладающие способностью эффективно решать экстремальные задачи размерностей “Huge-Scale” из различных классов. На основе численных экспериментов выявлены ограничения по размерности решаемых “плотных” задач, которые оказались значительно большими, чем предполагалось теоретически (см. [3]). Метод Б.Т. Поляка 1969 г. с применением простейших модификаций может рассматриваться в качестве удобного и высокомасштабируемого инструмента для численного решения “больших” задач конечномерной и бесконечномерной оптимизации.

СПИСОК ЛИТЕРАТУРЫ

1. *Поляк Б.Т.* Минимизация негладких функционалов // Ж. вычисл. матем. и матем. физ. 1969. Т. 9. № 3. С. 509–521.
2. *Поляк Б.Т.* Введение в оптимизацию. М.: Наука, 1983. 384 с.
3. *Нестеров Ю.Е.* Введение в выпуклую оптимизацию. М.: МЦНМО, 2010. 280 с.
4. Yurii Nesterov Subgradient Methods for Huge-Scale Optimization Problems // Math. Program. 2014. V. 146. Iss. 1-2. P. 275–297.
5. *Marques J.M.C., Pais A.A.C.C., Abreu P.E.* On the use of big-bang method to generate low-energy structures of atomic clusters modeled with pair potentials of different ranges // J. Comput. Chemistry. 2012. V. 33. № 4. P. 442–452.
6. *Keating P.N.* Effect of Invariance Requirements on the Elastic Strain Energy of Crystals with Application to the Diamond Structure // Phys. Rev. 1966. V. 145. P. 637–645.
7. *Евтушенко Ю.Г.* Оптимизация и быстрое автоматическое дифференцирование. Препринт ВЦ им. А.А. Дородницына РАН, 2013. 144 с.
8. *Горнов А.Ю.* Вычислительные технологии решения задач оптимального управления. Новосибирск: Наука, 2009. 278 с.
9. *Gornov A.Yu., Zarodnyuk T.S., Anikin A.S.* The computational technique for nonlinear nonconvex optimal control problems based on modification gully method // DEStech Transact. Comput. Sci. Eng. 2018. P. 152–162.
10. *Gornov A.Yu., Tyatyushkin A.I., Finkelstein E.A.* Numerical methods for solving terminal optimal control problems // Comput. Math. and Math. Phys. 2016. V. 56. № 2. P. 221–234.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1070–1081

ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

УДК 517.9

КОРПОРАТИВНАЯ ДИНАМИКА В ЦЕПОЧКАХ СВЯЗАННЫХ ЛОГИСТИЧЕСКИХ УРАВНЕНИЙ С ЗАПАЗДЫВАНИЕМ[1)]

© 2021 г. С. А. Кащенко

150003 Ярославль, ул. Советская, 14, Ярославский государственный университет им. П.Г. Демидова, Россия

e-mail: kasch@uniyar.ac.ru

Поступила в редакцию 14.02.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Рассматривается локальная динамика связанных цепочек одинаковых осцилляторов. В качестве базовой модели осциллятора предложено известное логистическое уравнение с запаздыванием. Осуществлен переход к изучению пространственно распределенной модели. Рассмотрены представляющие наибольший интерес два типа связей: диффузионные и однонаправленные. В задаче об устойчивости состояния равновесия выделены критические случаи. Они, как оказывается, имеют бесконечную размеренность: бесконечно много корней характеристического уравнения стремятся к мнимой оси при стремлении к нулю малого параметра, характеризующего величину, обратную к числу элементов цепочки. В качестве основного результата построены специальные нелинейные краевые задачи, нелокальная динамика которых описывает поведение всех решений цепочки из окрестности состояния равновесия. Библ. 33.

Ключевые слова: бифуркации, устойчивость, нормальные формы, сингулярные возмущения, динамика.

DOI: 10.31857/S0044466921070085

ВВЕДЕНИЕ

В настоящее время особое внимание уделяется таким важным объектам, как цепочки взаимодействующих осцилляторов. Эти цепочки возникают при моделировании многих прикладных задач в радиофизике (см. [1], [2]), лазерной оптике (см. [3]–[5]), механике (см. [6], [7]), теории нейронных сетей (см. [8], [9]), биофизике (см. [10]), математической экологии (см. [11]–[16]) и др. В данной работе исследуются актуальные для биофизики и математической экологии цепочки связанных логистических уравнений с запаздыванием. В качестве базового объекта рассматривается хорошо известное логистическое уравнение с запаздыванием

$$\dot u = r\big[1 - u(t-T)\big]u,$$

где $u = u(t) \geq 0$ – нормированная численность (плотность) популяции в момент времени t, $r > 0$ – коэффициент мальтузианского роста, а коэффициент $T > 0$ – время запаздывания.

Отметим, что с помощью нормировки времени $t \to Tt$ можно коэффициент T принять равным 1, а получившийся при этим коэффициент rT можно опять переобозначить через r. Таким образом, далее в качестве базового рассматриваем уравнение

$$\dot u = r\big[1 - u(t-1)\big]u, \quad u \geq 0. \tag{1}$$

Напомним (см., например, [11]–[13]), что при условиях $0 < r \leq \dfrac{\pi}{2}$ состояние равновесия $u_0 \equiv 1$ этого уравнения асимптотически устойчиво, а при $r > \dfrac{\pi}{2}$ в уравнении (1) существует устойчивый цикл.

[1)]Работа выполнена в рамках реализации программы развития регионального научно-образовательного математического центра (ЯрГУ) при финансовой поддержке Министерства науки и высшего образования РФ (дополнительное соглашение 075-02-2020-1514/1 к Соглашению о предоставлении из федерального бюджета субсидии 075-02-2020-1514).

При $0 < r - \frac{\pi}{2} \ll 1$ его асимптотика имеет вид

$$u = 1 + \left(r - \frac{\pi}{2}\right)^{1/2} C\cos\left(\left(\frac{\pi}{2} + O\left(r - \frac{\pi}{2}\right)\right)t\right) + O\left(r - \frac{\pi}{2}\right), \quad C = \left(40(3\pi - 2)^{-1}\right)^{1/2}. \tag{2}$$

При $r \gg 1$ соответствующий цикл имеет ярко выраженный релаксационный характер (см. [14]), причем и амплитуда цикла, и его период неограниченно растут при $r \to \infty$.

Цепочкой связанных логистических уравнений с запаздыванием называют систему из N уравнений

$$u_j = r\left[1 - u_j(t-1)\right]u_j + d\left(\sum_{i=1, i\neq j}^{N} \alpha_{ij}u_i(t) - u_j\right), \quad j = 1, 2, \ldots, N, \tag{3}$$

в которой удобно считать, что количество популяций N является четным, а коэффициент d положителен. Будем предполагать, что эта цепочка замкнута в кольцо, т.е. $j = \pm 1, \pm 2, \ldots$ и $u_{j+N} \equiv u_j(t)$.

Относительно коэффициентов α_{ij} примем три идеологически важных ограничения. Во-первых, из биологических соображений вытекает, что эти коэффициенты неотрицательны: $\alpha_{ij} \geq 0$, иначе при некоторых начальных условиях численность хотя бы одной из популяций может стать отрицательной. Во-вторых, предполагаем, что среда однородна. Это означает, что для коэффициентов связей α_{ij} выполнены условия $\alpha_{ij} = \alpha_{j-i}$, $\alpha_{j+N} = \alpha_j$, $\alpha_0 = 0$. Тогда систему (3) можно представить в виде

$$u_j = r\left[1 - u_j(t-1)\right]u_j + d\left(\sum_{i=-\frac{1}{2}N}^{\frac{1}{2}N} \alpha_{j-i}u_i(t) - u_j(t)\right). \tag{4}$$

В-третьих, предполагаем, что выполнено условие

$$\sum_{1}^{N} \alpha_j = 1. \tag{5}$$

Это означает, что в системе (4) имеются однородные решения

$$u_j(t) \equiv u(t), \quad j = 1, 2, \ldots, N,$$

где $u(t)$ – решение уравнения (1).

Как одни из наиболее важных и часто встречающихся в приложениях, отметим два вида связей: 1) диффузионная связь, когда

$$\alpha_1 = \alpha_{-1} = \frac{1}{2}, \quad \alpha_j = 0 \quad \text{при} \quad j = \pm 2, \ldots, \pm\frac{1}{2}N, \tag{6}$$

2) однонаправленная или адвективная связь, когда

$$\alpha_1 = 1, \quad \alpha_j = 0 \quad \text{при} \quad j = -1, \pm 2, \ldots, \pm\frac{1}{2}N. \tag{7}$$

Значение $u_j(t)$ можно ассоциировать со значением плотности популяции с номером j, находящейся в точке некоторой “окружности” L с угловой координатой x_j, т.е.

$$u_j(t) = u(t, x_j).$$

Основное предположение в настоящей работе состоит в том, что значение N предполагается достаточно большим, т.е. для параметра $\varepsilon = 2\pi N^{-1}$ выполнено условие

$$0 < \varepsilon \ll 1. \tag{8}$$

При малых ε количество значений x_j на окружности L является достаточно большим, поэтому представляется естественным перейти от дискретной переменной x_j к непрерывной пространственной переменной $x \in [0, 2\pi]$, имея в виду, что для $u(t, x_{j+n})$ выполнено равенство $u(t, x + \varepsilon n)$. В более общем случае система (4) принимает вид краевой задачи

$$\frac{\partial u}{\partial t} = r\left[1 - u(t-1, x)\right]u + d\left(\int_{-\infty}^{\infty} F(s, \varepsilon)u(t, x+s)ds - u\right), \tag{9}$$

$$u(t, x + 2\pi) \equiv u(t, x). \tag{10}$$

Относительно функции $F(s,\varepsilon)$ полагаем, что

$$\int_{-\infty}^{\infty} F(s,\varepsilon)ds = 1, \quad F(s,\varepsilon) \geq 0. \tag{11}$$

Например, в случае (6) $F(s,\varepsilon) = F_\delta(s+\varepsilon) + F_\delta(s-\varepsilon)$, а в случае (7) $F(s,\varepsilon) = F_\delta(s+\varepsilon)$. Здесь $F_\delta(s)$ – δ-функция, сосредоточенная в точке $s = 0$, т.е. для любого $\delta > 0$ имеем $\int_{-\delta}^{\delta} F_\delta(s)ds = 1$.

Ниже будем рассматривать достаточно гладкие функции $F(s,\varepsilon)$. По-видимому, наиболее важным с прикладной точки зрения являются функции F, состоящие из комбинаций функций вида $c_1 \exp\left(-(s-c_2)^2\sigma^{-2}\right)$ ($\sigma > 0, c_{1,2}$ – некоторые постоянные). Так, обобщением диффузионного типа связей является функция

$$F(s,\varepsilon) = \frac{1}{2\sigma\sqrt{\pi}}\left[\exp\left(-(s+\varepsilon)^2\sigma^{-2}\right) + \exp\left(-(s-\varepsilon)^2\sigma^{-2}\right)\right], \tag{12}$$

а обобщением одностороннего типа связей – функция

$$F(s,\varepsilon) = \frac{1}{\sigma\sqrt{\pi}}\exp\left(-(s+\varepsilon)^2\sigma^{-2}\right). \tag{13}$$

В качестве фазового пространства, т.е. пространства начальных условий, фиксируем $C_{[-1,0]\times[0,2\pi]}$ пространство непрерывных на отрезке $t \in [-1, 0]$ и непрерывных и 2π-периодических по пространственной переменной функций.

Краевая задача (9), (10) имеет однородное состояние равновесия $u(t,x) \equiv u_0 = 1$. Поставим вопрос об исследовании при малых ε локальной динамики (9), (10), т.е. об исследовании поведения при $t \to \infty$ всех решений (9), (10) с начальными условиями из некоторой достаточно малой в норме $C_{[-1,0]\times[0,2\pi]}$ и не зависящей от ε окрестности состояния равновесия u_0.

Тем самым речь идет об изучении динамики распределенных цепочек логистических уравнений с запаздыванием. Отметим, что исследованию динамики в различных цепочках связанных систем были посвящены результаты многих авторов (см., например, [17]–[24]). В настоящей работе, используя специальные методы локального анализа (см. [25]–[28]), будут построены квазинормальные формы – эволюционные краевые задачи, нелокальная динамика которых определяет поведение всех решений исходной краевой задачи (7), (6) в окрестности u_0.

Приведем без доказательства два стандартных утверждения, являющихся аналогом классических теорем А.М. Ляпунова об устойчивости по первому приближению. Рассмотрим линеаризованную в окрестности u_0 краевую задачу

$$\frac{\partial v}{\partial t} = -rv(t-1,x) + d\left(\int_{-\infty}^{\infty} F(s,\varepsilon)v(t,x+s)ds - v\right), \tag{14}$$

$$v(t,x+2\pi) \equiv v(t,x). \tag{15}$$

Характеристическое уравнение для нее имеет вид

$$\lambda + r\exp(-\lambda) = d\left(\int_{-\infty}^{\infty} F(s,\varepsilon)\exp(iks)ds - 1\right), \quad k = 0, \pm 1, \pm 2, \ldots \tag{16}$$

Теорема 1. *Пусть все корни уравнения* (16) *имеют отрицательную вещественную часть и отделены от мнимой оси при* $\varepsilon \to 0$. *Тогда найдется такое* $\varepsilon_0 > 0$, *что при всех* $\varepsilon \in (0,\varepsilon_0)$ *все решения краевой задачи* (9), (10) *из некоторой достаточно малой и не зависящей от* ε *окрестности состояния равновесия* u_0 *стремятся к* u_0 *при* $t \to \infty$.

Теорема 2. *Пусть уравнение* (16) *имеет корень с положительной и отделенной от нуля при* $\varepsilon \to 0$ *вещественной частью. Тогда при малых* ε *состояние равновесия* u_0 *в* (9), (10) *неустойчиво и в некоторой достаточной малой и не зависящей от* ε *окрестности* u_0 *не существует аттрактора этой краевой задачи*.

Таким образом, в условиях теоремы 1 поставленная задача о локальной в окрестности u_0 динамике тривиальна, а в условии теоремы 2 она не может быть исследована методами локального анализа.

Ниже будем предполагать, что реализуется критический случай, т.е. уравнение (16) не имеет корней с положительной и отделенной от нуля при $\varepsilon \to 0$ вещественной частью, но существует корень, вещественная часть которого стремится к нулю при $\varepsilon \to 0$.

Ограничимся здесь рассмотрением только указанных выше двух наиболее важных и интересных случаев, когда выполнены условия (12) или (13). Отметим, что методика рассмотрения общего случая та же, что и для этих двух случаев. Важно подчеркнуть, что будет рассмотрена ситуация, когда в (12), (13) параметр σ является достаточно малым, что приведет к появлению новых динамических эффектов. В связи с этим интересно выявить роль диффузионного слагаемого, когда в (9) добавляется выражение $\varkappa \dfrac{\partial^2 u}{\partial x^2}$, причем коэффициент $\varkappa$ тоже является малым параметром. Тем самым будет рассмотрена краевая задача

$$\frac{\partial u}{\partial t} = r\left[1 - u(t-1,x)\right]u + d\left(\int\limits_{-\infty}^{\infty} F(s,\varepsilon)u(t,x+s)ds - u\right) + \varkappa\frac{\partial^2 u}{\partial x^2}, \tag{17}$$

$$u(t,x+2\pi) \equiv u(t,x). \tag{18}$$

В следующем разделе рассмотрены вопросы о динамике краевой задачи (9), (10) в случае связей диффузионного типа. Показано, что в зависимости от величины параметра σ могут реализоваться принципиально различные ситуации. Все они описаны в соответствующих подразделах. В разд. 2 исследуется динамика (9), (10) при односторонних связях.

Будут построены специальные нелинейные краевые задачи, не содержащие малых параметров, которые играют роль нормальных форм. Кроме этого, будет рассмотрена важная задача о взаимодействующих осцилляторах в цепочке со слабыми связями. Речь пойдет об изучении динамики в случае, когда параметр $\varkappa$ является достаточно малым. В заключительном разделе сформулированы выводы.

1. ДИНАМИКА ЦЕПОЧЕК В СЛУЧАЕ ДИФФУЗИОННОГО ТИПА СВЯЗЕЙ

Предполагаем, что функция $F(s,\varepsilon)$ задана равенством (12). В зависимости от параметра σ можно выделить три принципиально различные ситуации. В первой из них, самой простой, предполагается, что параметр $\sigma > 0$ как-то фиксирован и, естественно, не зависит от малого параметра ε. Этот случай рассмотрен в п. 1.1. В п. 1.2 предполагаем, что найдется такое значение $\sigma_0 > 0$, что

$$\sigma = \varepsilon\sigma_0. \tag{19}$$

При этом условии реализуется упомянутый выше критический случай бесконечной размерности. Наконец, в п. 1.3 предполагаем, что параметр σ еще более мал: $\sigma = o(\varepsilon)$. Точнее, будем для некоторого фиксированного $\sigma_0 > 0$ рассматривать соотношение

$$\sigma = \varepsilon^2\sigma_0. \tag{20}$$

Этот случай наиболее сложен и интересен. Он естественным образом обобщает случай “чисто диффузионных” связей, когда $\sigma \sim 0$.

1.1. Динамика цепочек при фиксированном значении σ

Фиксируем в формуле (12) произвольное значение $\sigma_0 > 0$. Необходимое и достаточное условие отрицательности при всех $\varepsilon > 0$ вещественных частей всех собственных значений характеристических уравнений (16) состоит в выполнении неравенств

$$0 < r < \frac{\pi}{2}. \tag{21}$$

При условии $r = \dfrac{\pi}{2}$ уравнение (16) имеет ровно два чисто мнимых корня $\lambda_{\pm} = \pm i\dfrac{\pi}{2}$, а вещественные части остальных коней отрицательны и отделены от нуля при $\varepsilon \to 0$. Тем самым выполнены условия хорошо изученной бифуркации Андронова–Хопфа. Пусть для произвольно фиксированного значения r_1 имеем

$$r = \frac{\pi}{2} + \varepsilon^2 r_1. \tag{22}$$

Тогда близкие при $\varepsilon \ll 1$ к $\lambda_\pm$ корни $\lambda_\pm(\varepsilon)$ уравнения (16) имеют вид

$$\lambda_+(\varepsilon) = \lambda_-(\varepsilon),$$
$$\lambda_+(\varepsilon) = i\frac{\pi}{2} + \varepsilon^2\lambda_{10} + O(\varepsilon^4), \tag{23}$$
$$\text{где} \quad \lambda_{10} = \left(1 + \frac{\pi^2}{4}\right)^{-1}\left(\frac{\pi}{2} + i\right)r_1.$$

При этих условиях и при достаточно малых ε краевая задача (9), (10) имеет в окрестности $u_0 = 1$ двумерное устойчивое локальное интегральное инвариантное многообразие $M(\varepsilon)$, на котором эту краевую задачу можно с точностью до $O(\varepsilon^4)$ записать в виде специального скалярного комплексного обыкновенного дифференциального уравнения

$$\frac{d\xi}{d\tau} = \lambda_{10}\xi + g\xi|\xi|^2, \tag{24}$$

в котором $\tau = \varepsilon^2 t$ – медленное время, а $\xi(\tau)$ – медленно меняющаяся амплитуда в асимптотическом представлении решений на многообразии $M(\varepsilon)$:

$$u = 1 + \varepsilon\left(\xi(\tau)\exp\left(i\frac{\pi}{2}t\right) + \overline{\xi}(\tau)\exp\left(-i\frac{\pi}{2}t\right)\right) + \varepsilon^2 u_2(t,\tau) + \varepsilon^3 u_3(t,\tau) + \ldots. \tag{25}$$

Здесь функции $u_j(t,\tau)$ – 4-периодические по t. Подставим формальное выражение (25) в (9) и будем собирать коэффициенты при одинаковых степенях ε. Сначала, приравнивая коэффициенты при ε^2, находим, что

$$u_2 = \frac{2-i}{5}\xi^2\exp(i\pi t) + \frac{2+i}{5}\xi^2\exp(-i\pi t). \tag{26}$$

На следующем шаге из условия разрешимости получающегося уравнения относительно u_3 приходим к необходимости выполнения соотношения (24), в котором

$$g = -\frac{\pi}{2}[3\pi - 2 + i(\pi + 6)]\left(10\left(1 + \frac{4}{\pi^2}\right)\right)^{-1}. \tag{27}$$

Сформулируем итоговые утверждения. Доказательства их хорошо известны (см., например, [11], [12]).

Теорема 3. *Пусть $r_1 < 0$. Тогда при всех достаточно малых ε решение краевой задачи* (9), (10) *из некоторой достаточно малой и не зависящей от ε окрестности состояния равновесия $u_0 = 1$ стремится к* 1 *при $t \to \infty$.*

Теорема 4. *Пусть $r_1 > 0$. Тогда все, кроме нулевого, решения уравнения* (24) *стремятся к орбитально устойчивому циклу*

$$\xi_0(\tau) = \left[10\frac{\pi}{2}r_1(3\pi - 2)^{-1}\right]^{1/2}\xi_0\exp(i\phi_0\tau),$$
$$\phi_0 = \operatorname{Im}\lambda_{10} + \xi_0^2\operatorname{Im}g, \tag{28}$$

а все решения ($\not\equiv 1$) *из $M(\varepsilon)$ при $t \to \infty$ стремятся к циклу*

$$u_0(t,\varepsilon) = 1 + \varepsilon\left(\xi_0\left(\varepsilon^2 t\right)\exp\left(i\frac{\pi}{2}t\right) + \overline{\xi}_0\left(\varepsilon^2 t\right)\exp\left(-i\frac{\pi}{2}t\right)\right) + O(\varepsilon^2). \tag{29}$$

Таким образом, в рассмотренной ситуации краевая задача (9), (10) может иметь в окрестности u_0 только однородный цикл, являющийся решением в условии (22) логистического уравнения (2). По-видимому, рассмотренный здесь случай интереса не представляет.

1.2. Динамика цепочек при значении σ порядка ε

Предполагаем, что выполнено условие (19). Тогда характеристическое уравнение (16) имеет совокупность корней $\lambda_m(\varepsilon)$ и $\overline{\lambda}_m(\varepsilon)$ ($m = 0, \pm 1, \pm 2, \ldots$), вещественные части которых стремятся к нулю при $\varepsilon \to \infty$. Для этих корней справедливо представление

$$\lambda_m(\varepsilon) + \left(\frac{\pi}{2} + \varepsilon^2 r_1\right)\exp(-\lambda_m(\varepsilon)) = d\left(\int\limits_{-\infty}^{\infty} F(s,\varepsilon)\exp(ims)ds - 1\right) = \\ = d\left(\cos(\varepsilon m)\exp\left(-\varepsilon^2 m^2\sigma_0^2\right) - 1\right). \tag{30}$$

Отсюда получаем, что для каждого целого m выполнено асимптотическое равенство

$$\lambda_m(\varepsilon) = i\frac{\pi}{2} + \varepsilon^2\lambda_1 + \ldots, \quad \lambda_1 = \lambda_{10} - \left(1 + i\frac{\pi}{2}\right)^{-1} d\left(\frac{1}{2} + \sigma_0^2\right)m^2.$$

Каждому такому корню отвечает решение $v_m(t,x)$ линейной краевой задачи (14), (15), для которого

$$v_m(t,x) = \exp\left(i\frac{\pi}{2}t + imx\right)\nu_m(\tau),$$
$$\text{где} \quad \nu_m(\tau) = \nu_m\exp\left(\left(-\varepsilon^2\lambda_1 + O(\varepsilon^4)\right)t\right).$$

Введем в рассмотрение формальный ряд

$$u(t,x,\varepsilon) = 1 + \varepsilon\left(\exp\left(i\frac{\pi}{2}t\right)\sum_{m=-\infty}^{\infty}\xi_m(\tau)\exp(imx) + \right. \\ \left. + \exp\left(-i\frac{\pi}{2}t\right)\sum_{m=-\infty}^{\infty}\overline{\xi}_m(\tau)\exp(-imx)\right) + \varepsilon^2 u_2(t,\tau,x) + \varepsilon^3 u_3(t,\tau,x) + \ldots. \tag{31}$$

Здесь $\tau = \varepsilon^2 t$ – медленное время, $\xi_m(\tau)$ – неизвестные медленно меняющиеся амплитуды, а функции $u_j(t,\tau,x)$ – периодичны по t и x. Отметим, что в линейном приближении, т.е. при $u_j \equiv 0$, формула (31) задает совокупность решений линейной краевой задачи (14), (15).

Выражение (31) можно существенно упростить. Для этого положим

$$\xi(\tau,x) = \sum_{m=-\infty}^{\infty}\xi_m(\tau)\exp(imx).$$

Тогда из (31) вытекает, что

$$u(t,x,\varepsilon) = 1 + \varepsilon\left(\xi(\tau,x)\exp\left(i\frac{\pi}{2}t\right) + \overline{\xi}(\tau,x)\exp\left(-i\frac{\pi}{2}t\right)\right) + \\ + \varepsilon^2 u_2(t,\tau,x) + \varepsilon^3 u_3(t,\tau,x) + \ldots. \tag{32}$$

Подставим (31) в (9), (10) и в получившемся формальном тождестве будем приравнивать коэффициенты при одинаковых степенях ε. На первом шаге при ε^1 тождество выполнено, а на втором шаге, собирая коэффициенты при ε^2, приходим к равенству (26), в котором $\xi = \xi(\tau,x)$. На следующем шаге из условия разрешимости получающегося уравнения относительно u_3 приходим к краевой задаче для определения $\xi(\tau,x)$:

$$\frac{\partial\xi}{\partial\tau} = d_0\frac{\partial^2\xi}{\partial x^2} + \lambda_{10}\xi + g\xi|\xi|^2, \quad \xi(t,x+2\pi) \equiv \xi(\tau,x), \tag{33}$$

где $d_0 = \left(1 + i\frac{\pi}{2}\right)^{-1}\left(\frac{1}{2} + \sigma_0^2\right)$, а коэффициенты λ_{10} и g те же, что и в (23), и (27).

Сформулируем основные результаты.

Теорема 5. *Пусть краевая задача* (33) *имеет ограниченное при* $\tau \to \infty$ *решение* $\xi_0(\tau, x)$. *Тогда функция*

$$u_0(t,x,\varepsilon) = 1 + \varepsilon\left(\xi_0(\tau,x)\exp\left(i\frac{\pi}{2}t\right) + \overline{\xi}_0(\tau,x)\exp\left(-i\frac{\pi}{2}t\right)\right) + \\ + \varepsilon^2\left(\frac{2-i}{5}\xi_0^2(\tau,x)\exp(i\pi t) + \frac{2+i}{5}\overline{\xi}_0^2(\tau,x)\exp(-i\pi t)\right)$$

удовлетворяет краевой задаче (9), (10) *с точностью до* $O(\varepsilon^4)$.

Вопрос о существовании и об устойчивости точного решения (9), (10), близкого при $\varepsilon \to 0$ к соответствующему решению краевой задачи (33), может быть решен, например, в случае, когда $\xi_0(\tau, x)$– периодическое решение, обладающее свойством грубости. Под грубостью будем понимать следующее. Если $\xi_0(\tau, x) \equiv \text{const} \cdot \exp(i\omega\tau + imx)$, то лишь один мультипликатор линеаризованной на $\xi_0(\tau, x)$ краевой задачи равен по модулю 1. В остальных случаях условие грубости состоит в том, что только два мультипликатора линеаризованной на $\xi_0(\tau, x)$ краевой задачи равны по модулю 1.

Теорема 6. *Пусть* $\xi_0(\tau, x)$ *– периодическое с периодом* ω_0 *решение краевой задачи* (33), *обладающее свойством грубости. Тогда при всех достаточно малых* ε *краевая задача* (9), (10) *имеет периодическое по t решение* $u_0(t, x, \varepsilon)$ *с периодом* $\omega_0 + O(\varepsilon)$ *той же, что и* $\xi_0(\tau, x)$, *устойчивости и для которого верно асимптотическое равенство*

$$u_0(t,x,\varepsilon) = 1 + \varepsilon\left(\xi_0((1+O(\varepsilon))\varepsilon^2 t, x)\exp\left(i\frac{\pi}{2}t\right) + \overline{\xi}_0\left((1+O(\varepsilon))\varepsilon^2 t, x\right)\exp\left(-i\frac{\pi}{2}t\right)\right) + \ldots . \tag{34}$$

Доказательство теоремы 5 вытекает непосредственно из приведенного построения асимптотики решения краевой задачи (9), (10). Обоснование теоремы 6 стандартно (см., например, [29], [30]), но громоздко, поэтому его опустим.

Замечание 1. При рассмотрении более общей по сравнению с (9), (10) краевой задачи (17), (18) с малым коэффициентом диффузии $\varkappa = \varepsilon^2\varkappa^0$, изменения невелики. Коэффициент d_0 в (33) дополняется еще одним слагаемым:

$$d_0 = \left(1 + i\frac{\pi}{2}\right)^{-1} d\left(\frac{1}{2} + \sigma_0^2\right) + \left(1 + i\frac{\pi}{2}\right)^{-1}\varkappa^0.$$

Отметим, что краевая задача (33) может обладать богатой динамикой, поэтому этот же вывод справедлив и для цепочки рассматриваемых осцилляторов.

1.3. Динамика цепочек при $\sigma = O(\varepsilon^2)$

Предполагаем, что выполнено условие (20). Выделим все те корни характеристического уравнения (16), вещественные части которых стремятся к нулю при $\varepsilon \to 0$. Корни $\lambda = \lambda(k)$ этого уравнения находятся из формулы

$$\lambda + \left(r_0 + \varepsilon^2 r_1\right)\exp(-\lambda) = d\left(\int\limits_{-\infty}^{\infty} F(s,\varepsilon)\exp(iks)ds - 1\right) = \\ = d\left(\cos(z)\exp\left(-\varepsilon^4\sigma_0^2 z^2\right) - 1\right), \tag{35}$$

где $z = \varepsilon k$. Условие стремления вещественных частей корней к нулю обусловлено обращением в нуль с точностью до $o(1)$(при $\varepsilon \to 0$) правой части в (36). Этому условию удовлетворяют такие номера $k = k(\varepsilon)$, для которых $\cos(z) \sim 1$. Для описания таких номеров введем обозначения. Фиксируем произвольное целое n и через $\theta_n = \theta_n(\varepsilon) \in [0, 1)$ обозначим выражение, которое дополняет до целого значение $2\pi n\varepsilon^{-1}$. Оказывается, что функцию $\theta_n(\varepsilon)$ можно считать тождественно равной нулю. Дело в том, что введенный выше параметр ε был определен как $\varepsilon = 2\pi N^{-1}$. Поэтому $2\pi n\varepsilon^{-1} = nN$, т.е. целое.

Тогда совокупность номеров $k(\varepsilon)$ корней $\lambda(k(\varepsilon))$ в рассматриваемом случае состоит из значений

$$k(\varepsilon) = 2\pi n\varepsilon^{-1} + m, \quad m, n = 0, \pm 1, \pm 2, \ldots . \tag{36}$$

Эти корни удобно обозначать через $\lambda_{m,n}(\varepsilon)$. Для них имеем асимптотическое выражение

$$\lambda_{m,n}(\varepsilon) = i\frac{\pi}{2} - \varepsilon^2\left(1 + i\frac{\pi}{2}\right)^{-1}\left(m^2 + 4\pi^2\sigma_0^2 n^2\right) + O(\varepsilon^4).$$

Следуя изложенной выше схеме исследования, введем в рассмотрение формальный ряд

$$\begin{aligned} u = 1 + \varepsilon\Bigg(&\exp\left(i\frac{\pi}{2}t\right)\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\xi_{m,n}(\tau)\exp\left(i\left(2\pi n\varepsilon^{-1} + m\right)x\right) + \\ &+ \varepsilon\left(-i\frac{\pi}{2}t\right)\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\overline{\xi}_{m,n}(\tau)\exp\left(-i\left(2\pi n\varepsilon^{-1} + m\right)x\right)\Bigg) + \varepsilon^2 u_2(t,\tau,x) + \varepsilon^3 u_3(t,\tau,x) + \ldots, \end{aligned} \tag{37}$$

где $\tau = \varepsilon^2 t$, а функции $u_j(t,\tau,x)$ периодичны по t и x.

Положим $y = 2\pi\varepsilon^{-1}x$ и

$$\xi(\tau,x,y) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\xi_{m,n}(\tau)\exp(iny + imx).$$

Тогда выражение (37) можно упростить:

$$u = 1 + \varepsilon\left(\exp\left(i\frac{\pi}{2}t\right)\xi(\tau,x,y) + \exp\left(-i\frac{\pi}{2}t\right)\overline{\xi}(\tau,x,y)\right) + \varepsilon^2 u_2 + \varepsilon^3 u_3 + \ldots. \tag{38}$$

Подставим (38) в (9) и будем приравнивать коэффициенты при одинаковых степенях ε. Сначала находим $u_2(\tau,t,x)$, а затем из условия разрешимости уравнения относительно u_3 приходим к выражению для определения $\xi(\tau,x,y)$:

$$\frac{\partial\xi}{\partial\tau} = \left(1 + i\frac{\pi}{2}\right)^{-1}\left(\frac{\partial^2\xi}{\partial x^2} + 4\pi^2\sigma_0^2\frac{\partial^2\xi}{\partial y^2}\right) + \lambda_{10}\xi + g\xi|\xi|^2, \tag{39}$$

$$\xi(\tau,x+2\pi,y) \equiv \xi(\tau,x,y+2\pi) \equiv \xi(\tau,x,y), \tag{40}$$

где коэффициенты λ_{10} и g те же, что и в (24).

Основные результаты здесь в идеологическом плане повторяют утверждение теорем 5 и 6. Приведем для примера аналог теоремы 5.

Теорема 7. *Пусть* $\xi_0(\tau,x,y)$ *— ограниченное при* $\tau\to\infty$ *решение краевой задачи* (39), (40). *Тогда функция*

$$u_0(t,x,\varepsilon) = 1 + \varepsilon\left(\exp\left(i\frac{\pi}{2}t\right)\xi_0\left(\varepsilon^2 t,x,2\pi^{-1}x\right) + \exp\left(-i\frac{\pi}{2}t\right)\overline{\xi}_0\left(\varepsilon^2 t,x,2\pi\varepsilon^{-1}x\right)\right) + \varepsilon^2 u_2 \tag{41}$$

удовлетворяет краевой задаче (9), (10) *с точностью до* $O(\varepsilon^3)$.

Краевые задачи (33) и (38), (40) численно исследовались многими авторами (см., например, [31]). Показано, что для таких краевых задач, особенно для (39), (40), характерны сложные и нерегулярные колебания. Согласно формулам (32) и (38), связывающих решения этих краевых задач и решений краевой задачи (9), (10), такой же вывод можно сформулировать и о решениях (9), (10).

2. ДИНАМИКА ЦЕПОЧЕК ПРИ ОДНОСТОРОННИХ СВЯЗЯХ

Рассмотрим краевую задачу (9), (10), в которой для функции $F(s,\varepsilon)$ выполнено равенство (13), а для параметра σ имеет место соотношение (19).

Предположим, что в отсутствие связей состояние равновесия $u_0 = 1$ логистического уравнения (1) асимптотически устойчиво. Тем самым значения параметра r удовлетворяют неравенству

$$0 < r < \frac{\pi}{2}. \tag{42}$$

Пункт 2.1 посвящен анализу линеаризованной на u_0 краевой задаче, а в п. 2.2 построена нелинейная краевая задача, которая играет роль нормальной формы.

2.1. Линейный анализ

Характеристическое уравнение для линеаризованной на u_0 краевой задачи в рассматриваемом случае приобретает вид

$$\lambda + r\exp(-\lambda) = d\left(\exp\left(iz - \sigma_0^2 z^2\right) - 1\right), \tag{43}$$

где $d > 0$, $z = \varepsilon k$, $k = 0, \pm 1, \pm 2, \ldots$. Для ответа на вопрос об устойчивости состояния равновесия в краевой задаче (14), (15) исследуем расположение корней уравнения (43).

Приведем без доказательств несколько простых утверждений о корнях (43).

Лемма 1. *При всех $d > 0$ и при всех $z \in [\pi(2n+1), \pi(2n+2)]$, $n = 0, \pm 1, \pm 2, \ldots$, уравнение* (43) *не может иметь корней с нулевой вещественной частью.*

Лемма 2. *Для каждого $z \in (2\pi n, \pi(2n+1))$ найдется такое $d > 0$, что при $d = d_z$ уравнение* (43) *имеет корень с нулевой вещественной частью.*

Введем обозначение: $d(r) = \min\limits_{-\infty < z < \infty} d_z = d_{z(r)}$.

Тогда при всех $d \in (0, d(r))$ состояние равновесия краевой задачи (14), (15) асимптотически устойчиво. При $d = d(r)$ и $z = z(r)$ уравнение (43) имеет корни $\lambda_\pm(r)$ с нулевой вещественной частью: $\lambda_\pm(r) = \pm i\omega(r)$, $\omega(r) > 0$.

Лемма 3. *Для значений $\omega(r)$ и $z(r)$ выполнены неравенства*

$$0 < z(r) < \pi, \quad \frac{\pi}{2} < \omega(r) < \frac{3\pi}{2}.$$

Рассмотрим отдельно вопросы об асимптотике выражений $d(r)$, $\omega(r)$ и $z(r)$ при $r \to 0$ и при $r \to \dfrac{\pi}{2}$.

Пусть сначала $r \to 0$. Обозначим через ω_0 корень из промежутка $\left(\dfrac{\pi}{2}, \pi\right)$ уравнения $\operatorname{tg}\omega = -\omega^{-1}$ и положим

$$c_0 = \left(1 + \sigma_0^2\right)\omega_0^2\left(\omega_0^2 + 4\right), \quad z_0 = \omega_0 c_0^{-1}.$$

Лемма 4. *При $r \to 0$ имеют место асимптотические равенства*

$$d(r) = c_0 r^{-1}(1 + o(1)), \quad \omega(r) = \omega_0 + o(1), \quad z(r) = z_0 r(1 + o(1)).$$

Пусть затем $r = \dfrac{\pi}{2} - \mu$ и $0 < \mu \ll 1$. Введем обозначения: через $z_{00} \in \left(0, \dfrac{\pi}{2}\right)$ обозначим наименьший корень уравнения

$$z = \left(\frac{\pi}{2} - 2\sigma_0^2 z\right)\left(1 + \pi\sigma_0^2 z\right)^{-1}.$$

Положим

$$c_{00} = \left(f_2 + \frac{2}{\pi} f_1\right)^{-1}, \quad w_{00} = -\frac{2}{\pi} f_1 \left(f_2 + \frac{2}{\pi} f_1\right)^{-1},$$
$$f_1 = \cos z_{00} \cdot \exp\left(-\sigma_0^2 z_{00}^2\right) - 1,$$
$$f_2 = \sin z_{00} \cdot \exp\left(-\sigma_0^2 z_{00}^2\right).$$

Лемма 5. *При всех достаточно малых μ имеют место асимптотические равенства*

$$d(r) = c_{00}\mu(1 + o(1)),$$
$$\omega(r) = \frac{\pi}{2} + \omega_{00}\mu(1 + o(1)), \quad z(r) = z_{00} + o(1).$$

Обоснования лемм 4, 5 достаточно простые, но громоздкие. Поэтому их опустим.

Фиксируем далее значение $r_0 \in \left(0, \dfrac{\pi}{2}\right)$ и произвольные величины r_1 и d_1. Положим в (9), (10)

$$r = r_0 + \varepsilon^2 r_1, \quad d = d(r_0) + \varepsilon^2 d_1. \tag{44}$$

Ниже через $\theta = \theta(\varepsilon) \in [0, 1)$ обозначаем выражение, которое дополняет до целого величину $z(r_0)\varepsilon^{-1}$. Исследуем асимптотику всех корней уравнения (43), близких к мнимой оси. Обозначим их через $\lambda_m(\varepsilon)$ и $\lambda_m(\varepsilon)$, $m = 0, \pm 1, \pm 2, \ldots$. Имеем равенства

$$\lambda_m(\varepsilon) = i\omega(r) + \varepsilon i R_1(\theta + m) + \varepsilon^2\left(R_{20} + (\theta + m)^2 R_2\right) + \ldots, \tag{45}$$

где

$$R_1 = (1 - r_0\exp(-i\omega(r_0)))^{-1} d(r_0)z(r_0)\left(1 + 2i\sigma_0^2\right)\exp\left(-\sigma_0^2 z^2(r_0) + iz_0(r_0)\right),$$

$$R_{20} = (1 - r_0\exp(-i\omega(r_0)))^{-1}\left[d_1\left(1 - \exp\left(-\sigma_0^2 z^2(r_0) + iz_0(r_0)\right) - 1\right) - r_1\exp(-i\omega(r_0))\right],$$

$$R_2 = (1 - r_0\exp(-i\omega(r_0)))^{-1}\left[\frac{1}{2}r_0\exp(-i\omega(r_0))R_1^2 + d(r_0)\left(2\sigma_0^2 z^2(r_0) - \left(\sigma_0^2 + \frac{1}{2}\right)\right)\exp\left(-\sigma_0^2 z^2(r_0) + iz(r_0)\right)\right].$$

Важно отметить, что

$$\operatorname{Im} R_1 = 0 \quad \text{и} \quad \operatorname{Re} R_2 < 0. \tag{46}$$

2.2. Построение квазинормальной формы

Введем в рассмотрение формальный ряд

$$\begin{aligned} u = 1 + \varepsilon\Bigg(&\exp\left(i\omega(r_0)t\right)\sum_{m=-\infty}^{\infty}\xi_m(\tau)\exp\left(i\left(z(r_0)\varepsilon^{-1} + \theta + m\right)x + \varepsilon i R_1(\theta + m)t\right) + \\ &+ \exp(-i\omega(r_0)t)\sum_{m=-\infty}^{\infty}\overline{\xi}_m(\tau)\exp\left(-i\left(z(r_0)\varepsilon^{-1} + \theta + m\right)x - \varepsilon i R_1(\theta + m)t\right)\Bigg) + \\ &+ \varepsilon^2 u_2(t, \tau, x, \varepsilon) + \varepsilon^3 u_3(t, \tau, x, \varepsilon) + \ldots, \quad \tau = \varepsilon^2 t. \end{aligned} \tag{47}$$

Это выражение можно существенно упростить. Положим

$$\xi(\tau, y) = \sum_{m=-\infty}^{\infty}\xi_m(\tau)\exp(imy), \quad y = x + \varepsilon R_1 t.$$

Тогда от (47) переходим к представлению

$$\begin{aligned} u = 1 + \varepsilon\Big(&\exp\left(i\left(\omega(r_0) + \varepsilon R_1\theta\right)t + i\left(z(r_0)\varepsilon^{-1} + \theta\right)x\right)\xi(\tau, y) + \\ &+ \exp\left(-i\left(\omega(r_0) + \varepsilon R_1\theta\right)t - i\left(z(r_0)\varepsilon^{-1} + \theta\right)x\right)\overline{\xi}(\tau, y)\Big) + \\ &+ \varepsilon^2 u_2(t, \tau, x, y) + \varepsilon^3 u_3(t, \tau, x, y) + \ldots. \end{aligned} \tag{48}$$

Фигурирующие здесь функции $u_j(t, \tau, x, y)$ периодичны по t, x и y.

Подставим (48) в (9). Тогда, применяя стандартные процедуры, сначала находим $u_2(t, \tau, x, y)$:

$$\begin{aligned} u_2(t, \tau, x, y) = &u_{20}|\xi(\tau, y)|^2 + \\ &+ u_{21}\xi^2(\tau, y)\exp\left(2i\left(\omega(r_0) + \varepsilon R_1\theta\right)t + 2i\left(z(r_0)\varepsilon^{-1} + \theta\right)x\right) + \\ &+ u_{21}\overline{\xi}^2(\tau, y)\exp\left(-2i\left(\omega(r_0) + \varepsilon R_1\theta\right)t - 2i\left(z(r_0)\varepsilon^{-1} + \theta\right)x\right), \end{aligned}$$

где

$$u_{20} = -2\cos\omega(r_0),$$

$$u_{21} = -2r\cos(2\omega(r_0))\left[2i\omega(r_0) + r_0\exp(-2i\omega(r_0)) - d(r_0)(\exp(-2i\omega(r_0) - 4\sigma_0^2 z^2(r_0)) - 1)\right]^{-1}.$$

На следующем этапе получим уравнение для $u_3(t, \tau, x, y)$, из условия разрешимости которого в указанном классе функций приходим к краевой задаче для определения $\xi(\tau, y)$:

$$\frac{\partial\xi}{\partial\tau} = R_2\frac{\partial^2\xi}{\partial y^2} - i\theta R_2\frac{\partial\xi}{\partial y} + \left(R_{20} + \theta^2 R^2\right)\xi + q\xi|\xi|^2, \tag{49}$$

$$\xi(\tau, y + 2\pi) \equiv \xi(\tau, y). \tag{50}$$

Для коэффициента q имеем равенство

$$q = r_0(1 - r_0\exp(-i\omega(r_0)))^{-1}\left[2\cos(\omega(r_0))(1 + \exp(-i\omega(r_0)) - u_{21}(\exp(i\omega(r_0)) + \exp(-2i\omega(r_0)))\right].$$

Для формулировки основного результата введем еще одно обозначение. Фиксируем произвольно значение $\theta_0 \in [0, 1)$ и через $\varepsilon_n(\theta_0)$ обозначим такую последовательность, что $\varepsilon_n(\theta_0) \to 0$ при $n \to \infty$ и для каждого n выполнено равенство $\theta(\varepsilon_n(\theta_0)) = \theta_0$.

Из приведенных выше построений вытекает следующее утверждение.

Теорема 8. *Пусть при некотором* $\theta = \theta_0$ *краевая задача* (49), (50) *имеет ограниченное при* $\tau \to \infty$, $y \in [0, 2\pi]$ *решение* $\xi_o(\tau, y)$. *Тогда функция*

$$u_0 = 1 + \varepsilon\left(\exp\left(i\left(\omega(r_0) + \varepsilon R_1\theta_0\right)t + i\left(z(r_0)\varepsilon^{-1} + \theta_0\right)x\right)\xi_0(\tau, y) + \right.$$
$$\left. + \exp\left(-i\left(\omega(r_0) + \varepsilon R_1\theta_0\right)t - i\left(z(r_0)\varepsilon^{-1} + \theta_0\right)x\right)\overline{\xi}_0(\tau, y)\right), \quad \tau = \varepsilon^2 t, \quad y = x + \varepsilon R_1 t,$$

при условиях (13), (19) *и при* $\varepsilon = \varepsilon_n(\theta_0)$ *удовлетворяет краевой задаче* (9), (10) *с точностью до* $O(\varepsilon^2)$.

ВЫВОДЫ

Показано, что рассмотренные критические случаи в задаче об устойчивости распределенной цепочки логистических уравнений с запаздыванием имеют бесконечную размерность. Это приводит к тому, что описание их локальной динамики сводится к исследованию нелокального поведения решений краевых задач типа Гинзбурга–Ландау. Известно (см., например., [31]), что динамика таких объектов может быть сложной, причем для них характерны нерегулярные колебания, явления мультистабильности и др. Сами динамические эффекты существенно зависят от выбора связей. Показано, что в ряде случаев решения содержат быстро и медленно осциллирующие по пространственной переменной составляющие. Основные результаты определяют структуру асимптотических по невязке решений исходных краевых задач. Вопрос о существовании, устойчивости и более сложных асимптотических разложений точных решений, близких к построенным, может быть решен в случае периодических решений нормализованных уравнений.

Остановимся отдельно на роли фигурирующего выше параметра $\theta = \theta(\varepsilon) \in [0, 1)$. Напомним, что динамические свойства исходной системы определяются квазинормальной формой (49), (50), куда входит параметр θ. При различных значениях этого параметра динамика (49), (50), а значит, и краевой задачи (9), (10), может меняться. Детально это показано в [32]. Отсюда следует, что при $\varepsilon \to 0$ может происходить бесконечный процесс прямых и обратных бифуркаций.

Сформулируем еще один вывод общего плана. Выше было показано, что квазинормальные формы, определяющие динамику исходной краевой задачи, являются уравнениями Гинзбурга–Ландау. Устойчивость простейших решений этих уравнений исследована в [33]. В частности, установлено, что свойства их устойчивости во многом определяются мнимыми составляющими коэффициентов диффузии и ляпуновской величины (коэффициенты g и q в (49) и (50)). Численный анализ соответствующего критерия позволил сформулировать вывод о неустойчивости простейших решений вида $\text{const} \cdot \exp(i\omega t + ikx)$. Таким образом, в рассмотренных цепочках синхронизация решений является достаточно редким явлением.

Отметим, что в зависимости от коэффициента σ функции $F(s, \varepsilon)$ ((12), (13)) в качестве квазинормальной формы могут выступать параболические краевые задачи как с одной, так и с двумя пространственными переменными. Кроме этого, выявлены порядки (по параметру ε) коэффициентов диффузии в исходной краевой задаче, которые делают сопоставимым вклад диффузионного слагаемого со слагаемым, обеспечивающим связь элементов цепочки.

СПИСОК ЛИТЕРАТУРЫ

1. *Maurer J., Libchaber A.* Effect of the Prandtl number on the onset of turbulence in liquid ^{4}He // J. Phys. Lett. (France) 1982. V. 41. P. 515.
2. *Кузнецов С.П., Пономаренко В.И., Селезнев Е.П.* Автономная система – генератор гиперболического хаоса. Схемотехническое моделирование и эксперимент // Изв. вузов. Прикладная нелинейная динамика. 2013. Т. 21. № 5. С. 17–30.
3. *Brun E., Derighette B., Meier D., Holzner R., Raveni M.* Observation of order and chaos in a nuclear spin-flip laser // J. Opt. Soc. Am. B. 1985. V. 2. P. 156.

4. *Dangoisse D., Glorieux P., Hennequin D.* Chaos in a CO_2 laser with modulated parameters: Experiments and numerical simulations // Phys. Rev. A. 1987. V. 36. P. 4775.
5. *Chembo Y.K., Jacquot M., Dudley J.M., Larger L.* Ikeda-like chaos on a dynamically filtered supercontinuum light source // Phys. Rev. A. 2016. V. 94. P. 023847.
6. *Thompson J.M.T., Stewart H.B.* Nonlinear Dynamics and Chaos. Chichester: Wiley, 1986.
7. *Foss J., Longtin A., Mensour B., Milton J.* Multistability and delayed recurrent loops // Phys. Rev. Lett. 1996. V. 76. P. 708.
8. *Sysoev I.V., Ponomarenko V.I., Kulminskiy D.D., Prokhorov M.D.* Recovery of couplings and parameters of elements in networks of time-delay systems from time series // Phys. Rev. E. 2016. V. 94. P. 052207.
9. *Ponomarenko V.I., Kulminskiy D.D., Prokhorov M.D.* Chimeralike states in networks of bistable time-delayed feedback oscillators coupled via the mean field // Phys. Rev. E. 2017. V. 96. P. 022209.
10. *Караваев А.С., Ишбулатов Ю.М., Киселев А.Р., Пономаренко В.И., Прохоров М.Д., Миронов С.А., Шварц В.А., Гриднев В.И., Безручко Б.П.* Модель сердечно-сосудистой системы человека с автономным контуром регуляции среднего артериального давления // Физиология человека. 2017. Т. 43. № 1. С. 70–80.
11. *Kuang Y.* Delay Differential Equations : With Applications in Population Dynamics. Boston: Academic Press, 1993. 410 p. (Mathematics in science and engineering; 191). ISBN 0124276105.
12. *Wu J.* Theory and applications of partial functional differential equations. New York: Springer Verlag, 1996. 439 p. (Applied mathematical sciences; 119). ISBN 9780387947716.
13. *Gourley S.A., Sou J.W.-H., Wu J.H.* Nonlocality of reaction-diffusion equations induced by delay: biological modeling and nonlinear dynamics // J. Math. Sci. 2004. V. 124. № 4. P. 5119–5153. https://doi.org/10.1023/B:JOTH.0000047249.39572.6d
14. *Kashchenko S.A.* Asymptotics of the solutions of the generalized hutchinson equation // Automat. Control and Comp. Sci. 2013. V. 47. № 7. P. 470–494. https://doi.org/10.3103/S0146411613070079
15. *Кащенко С.А.* Динамика логистического уравнения с двумя запаздываниями // Дифференц. ур-ния. 2016. Т. 52. № 5. С. 561–571. https://doi.org/10.1134/S0374064116050022
16. *Кащенко С.А., Логинов Д.О.* Бифуркации при варьировании граничных условий в логистическом уравнении с запаздыванием и диффузией // Матем. заметки. 2019. Т. 106. № 1. С. 138–143. https://doi.org/10.4213/mzm12438
17. *Kuramoto Y.* Chemical oscillations, waves and turbulence. Springer, 1984.
18. *Kuramoto Y., Battogtokh D.* Coexisting of coherence and incoherence in nonlocally coupled phase oscillators // Nonlinear Phenom Complex Syst. 2002. V. 5. № 4. P. 380.
19. *Haken H.* Brain dynamics: synchronization and activity patterns in pulse-coupled neural nets with delays and noise. Springer, 2002.
20. *Osipov G.V., Kurths J., Zhou Ch.* Synchronization in Oscillatory Networks. Berlin: Springer, 2007.
21. *Afraimovich V.S., Nekorkin V.I., Osipov G.V., and Shalfeev V.D.* Stability, structures and chaos in nonlinear synchronization networks. Singapore: World Scientific, 1994.
22. *Крюков А.К., Осипов, Г.В., Половинкин А.В.* Мультистабильность синхронных режимов в ансамблях неидентичных осциллятолров: Цепочка и решетка связанных элементов // Изв. вузов “ПНД”, 2009. Т. 17. № 2.
23. *Крюков А.К., Канаков О.И., Осипов Г.В.* Волны синхронизации в ансамблях слабонелинейных осцилляторов // Изв. вузов “ПНД”. 2009. Т. 17. № 1.
24. *Pikovsky A.S., Rosenblum M.G., Kurths J.* Cambridge: Cambridge Univ. Press, 2001.
25. *Kashchenko I.S., Kashchenko S.A.* Dynamics of the Kuramoto equation with spatially distributed control // Comm. Nonlin. Sci. and Numer. Simulat. 2016. V. 34. P. 123–129. https://doi.org/10.1016/j.cnsns.2015.10.011
26. *Кащенко С.А.* О квазинормальных формах для параболических уравнений с малой диффузией // Докл. АН СССР. 1988. Т. 299. № 5. С. 1049–1052.
27. *Kaschenko S.A.* Normalization in the systems with small diffusion // Inter. J. of Bifurcat. and Chaos in Appl. Sci. and Eng. 1996. V. 6. № 6. P. 1093–1109. https://doi.org/10.1142/S021812749600059X
28. *Кащенко С.А.* Простейшие критические случаи в динамике нелинейных систем с малой диффузией // Тр. Московского матем. общества. 2018. Т. 79. Вып. 1. С. 97–115. https://doi.org/10.1090/mosc/285
29. *Kashchenko S.A.* Asymptotics of periodic solutions of autonomous parabolic equations with small diffusion // Siberian Math. J. 1986. V. 27. № 6. P. 880–889.
30. *Кащенко С.А.* Бифуркации в окрестности цикла при малых возмущениях с большим запаздыванием // Ж. вычисл матем. и матем. физ. 2000. Т. 40. № 5. С. 693–702.
31. *Ахромеева Т.С., Курдюмов С.П., Малинецкий Г.Г., Самарский А.А.*Нестационарные структуры и диффузионный хаос. М.: Наука, 1992. 544 с.
32. *Kashchenko I.S., Kashchenko S.A.* Infinite process of forward and backward bifurcations in the logistic equation with two delays // Nonlin. Phenomena in Complex Syst. 2019. V. 22. № 4. P. 407–412.
33. *Kashchenko A.A.* Analysis of running waves stability in the Ginzburg–Landau equation with small diffusion // Automat. Control and Comp. Sci. 2015. V. 49. № 7. P. 514–517.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1082–1100

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 517.929

ЛОКАЛЬНО-ОДНОМЕРНАЯ СХЕМА ДЛЯ ПЕРВОЙ НАЧАЛЬНО-КРАЕВОЙ ЗАДАЧИ ДЛЯ МНОГОМЕРНОГО УРАВНЕНИЯ КОНВЕКЦИИ–ДИФФУЗИИ ДРОБНОГО ПОРЯДКА[1)]

© 2021 г. А. А. Алиханов[1,*], М. Х. Бештоков[2,], М. Х. Шхануков-Лафишев[2,***]**

[1] *355017 Ставрополь, ул. Пушкина, 1, ФГАОУ ВО "Северо-Кавказский федеральный университет", Россия*

[2] *360004 Нальчик, ул. Шортанова, 89а, ИПМатем. и автоматизации, КБНЦ РАН, Россия*

**e-mail: alikhanov-tom@yandex.ru*

***e-mail: beshtokov-murat@yandex.ru*

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

Поступила в редакцию 14.09.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Исследуется первая краевая задача для уравнения конвекции–диффузии дробного порядка. Построена локально-одномерная разностная схема. С помощью принципа максимума получена априорная оценка в равномерной метрике. Доказаны устойчивость и сходимость рассматриваемой разностной схемы. Построен алгоритм приближенного решения локально-одномерной разностной схемы. Проведены численные расчеты, иллюстрирующие полученные теоретические результаты в работе. Библ. 32. Табл. 2.

Ключевые слова: дифференциальное уравнение в частных производных, уравнение конвекции–диффузии, производная дробного порядка, дробная производная по времени в смысле Капуто, локально-одномерная разностная схема, устойчивость и сходимость разностных схем.

DOI: 10.31857/S0044466921070024

ВВЕДЕНИЕ

В последние годы возрос интерес к исследованию дифференциальных уравнений дробного порядка, в которых неизвестная функция содержится под знаком производной дробного порядка. Интегралы и производные нецелого порядка и дробные интегро-дифференциальные уравнения находят множество применений в современных исследованиях в теоретической физике, механике и прикладной математике. Для описания структуры неупорядоченных сред и протекающих в них процессов широко используется теория фракталов (см. [1]–[17]). Примерами неупорядоченных сред являются пористые тела. При этом фракталами могут быть поровое пространство, скелет породы, поверхность скелета породы и т.д. В случае, когда пространство представляет собой фрактал с размерностью Хаусдорфа–Безиковича d_f, погруженный в сплошную среду с размерностью d ($d \geq d_f$, $d = 2, 3$), для описания движения примеси в потоке однородной среды используется дифференциальное уравнение дробного порядка (см. [18]).

Перенос, описываемый операторами с дробными производными, на больших расстояниях от источника приводит к совершенно иному поведению относительно малых концентраций по сравнению с классической диффузией. Эти малые концентрации, или "далекие хвосты распределений", при дробной диффузии подчинены степенному закону убывания, и их существование может заставить пересмотреть существующие ранее представления о безопасности, базирующиеся на представлениях об экспоненциальной скорости затухания (см. [19], [20]).

В [21, с. 199] и [22] дается описание геометрии облаков, размеры которых заключены в широком диапазоне от 1 до 1.2×10^6 км^2. Выяснено, что периметр облака связан с фрактальной размер-

[1)]Работа выполнена при финансовой поддержке РФФИ и ГФЕН Китая в рамках научного проекта № 20-51-53007.

ностью облака $D = 1.35 \pm 0.05$. Заметим, что порядок дробной производной связан с размерностью фрактала (см. [3], [4], [14]).

В [23] найдена связь между порядком дробной производной и фрактальной размерностью.

В [24] рассматривается локально-одномерная схема для решения линейных и квазилинейных уравнений параболического типа с любым числом p пространственных переменных, пригодная для произвольной области G. Доказана равномерная устойчивость локально-одномерной схемы по правой части, краевым и начальным данным. Показано, что локально-одномерные схемы дают точность $O(h^2 + \tau)$.

В [25] рассмотрена локально-одномерная схема для уравнения теплопроводности с дробной по времени производной без учета движения самой среды. Построена экономичная аддитивная схема в области сложной формы. Показано, что построенная схема обладает свойством суммарной аппроксимации $\psi = O(h_\alpha^2 + \tau)$ в регулярных узлах, в нерегулярных узлах $\psi = O(1)$, где h_α и τ – шаги сетки по направлению x_α и времени t.

Построению локально-одномерных схем для численного решения различных краевых задач для уравнения параболического типа с дробной производной по времени в многомерной области посвящены работы [5], [25]–[27], в которых априорные оценки были получены лишь при условии, когда $\frac{1}{2} < \alpha < 1$.

В настоящей работе рассмотрено построение локально-одномерной (экономичной) разностной схемы для численного решения первой краевой задачи для уравнения переноса пассивных примесей дробного порядка в многомерном случае, основная идея которого состоит в сведении перехода со слоя на слой к последовательному решению ряда одномерных задач по каждому из координатных направлений. Построена локально-одномерная разностная схема. С помощью принципа максимума получена априорная оценка для решения задачи в разностной трактовке, откуда следует равномерная сходимость локально-одномерной схемы в классе достаточно гладких решений при $0 < \alpha < 1$, где α – порядок дробной производной. Построен алгоритм решения локально-одномерной разностной схемы. Проведены численные эксперименты.

1. ПОСТАНОВКА ПЕРВОЙ НАЧАЛЬНО-КРАЕВОЙ ЗАДАЧИ

В замкнутом цилиндре $\bar{Q}_T = \bar{G} \times [0 \le t \le T]$, основанием которого является p-мерный прямоугольный параллелепипед $G = \{x = (x_1, \ldots, x_p) : 0 < x_k < l_k,\ k = 1, 2, \ldots, p\}$ с границей Γ, $\bar{G} = G \cup \Gamma$, рассмотрим следующую начально-краевую задачу:

$$\partial_{0t}^{\alpha} u = Lu + f(x,t), \quad (x,t) \in Q_T, \tag{1.1}$$

$$u|_{\Gamma} = \mu(x,t), \quad 0 \le t \le T, \tag{1.2}$$

$$u(x,0) = u_0(x), \quad x \in \bar{G}, \quad \bar{G} = G \cup \Gamma, \tag{1.3}$$

где

$$\partial_{0t}^{\alpha} u = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{\partial u(x,\eta)}{\partial \eta} \frac{d\eta}{(t-\eta)^{\alpha}}, \quad 0 < \alpha < 1,$$

есть дробная производная Капуто порядка α,

$$L = \sum_{k=1}^{p} L_k, \quad L_k u = \frac{\partial}{\partial x_k}\left(\Theta_k(x,t) \frac{\partial u}{\partial x_k}\right) + r_k(x,t) \frac{\partial u}{\partial x_k} - q_k(x,t)u, \quad k = 1, 2, \ldots, p,$$

$u(x,t)$ – концентрация примеси в точке x в момент времени t,

$\Theta_k(x,t)$ – коэффициент турбулентной диффузии по направлениям x_k,

$r_k(x,t)$ – компоненты вектора скорости воздушных потоков по направлениям x_k,

$$0 < c_0 \le \Theta_k(x,t), \quad q_k(x,t) \le c_1, \quad |r_k(x,t)| \le c_2,$$

c_0, c_1, c_2 – положительные постоянные, $Q_T = G \times (0 < t \le T]$, $k = 1, 2, \ldots, p$,

$$x = (x_1, \ldots, x_p), \quad x' = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_p).$$

В дальнейшем будем предполагать, что коэффициенты уравнения и граничных условий (1.1)–(1.3) удовлетворяют необходимым по ходу изложения условиям, обеспечивающим нужную гладкость решения $u(x,t)$ в цилиндре Q_T.

Локально-одномерные разностные схемы для уравнения диффузии дробного порядка в многомерной области для случая, когда оператор $Lu = \sum_{k=1}^{p} L_k u$, $L_k u = \dfrac{\partial u^2}{\partial x_k^2}$, рассмотрены в [5], а для случая, когда оператор $L_k u = \dfrac{\partial}{\partial x_k}\left(k_k \dfrac{\partial u}{\partial x_k}\right)$ с краевыми условиями III рода, рассмотрены в [27].

В той же области вместо задачи (1.1)–(1.3) рассмотрим следующую задачу с малым параметром ε:

$$\varepsilon u_t^{\varepsilon} + \partial_{0t}^{\alpha} u^{\varepsilon} = Lu^{\varepsilon} + f(x,t), \quad (x,t) \in Q_T, \tag{1.4}$$

$$u^{\varepsilon}\big|_{\Gamma} = \mu(x,t), \quad 0 \le t \le T, \tag{1.5}$$

$$u^{\varepsilon}(x,0) = u_0(x), \quad x \in \bar{G}, \tag{1.6}$$

где $\varepsilon = \text{const} > 0$.

Так как при $t = 0$ начальные условия для уравнения (1.1) и (1.4) совпадают, то в окрестности $t = 0$ у производной u_t^{ε} не возникает особенности типа пограничного слоя (см. [28], [29, с. 10]).

Покажем, что $u^{\varepsilon} \to u$ в некоторой норме при $\varepsilon \to 0$. Обозначим $\tilde{z} = u^{\varepsilon} - u$ и подставим $u^{\varepsilon} = \tilde{z} + u$ в задачу (1.4)–(1.6). Тогда получим

$$\varepsilon \tilde{z}_t + \partial_{0t}^{\alpha} \tilde{z} = L\tilde{z} + \tilde{f}(x,t), \quad (x,t) \in Q_T, \tag{1.7}$$

$$\tilde{z}\big|_{\Gamma} = 0, \quad 0 \le t \le T, \tag{1.8}$$

$$\tilde{z}(x,0) = 0, \quad x \in \bar{G}, \quad \bar{G} = G + \Gamma, \tag{1.9}$$

где $\tilde{f}(x,t) = -\varepsilon \dfrac{\partial u}{\partial t}$.

Для получения априорной оценки воспользуемся методом энергетических неравенств. Умножим уравнение (1.7) скалярно на $\tilde{\mathbf{z}}$ и получим энергетическое тождество

$$\left(\varepsilon \frac{\partial \tilde{z}}{\partial t}, \tilde{z}\right) + \left(\partial_{0t}^{\alpha} \tilde{z}, \tilde{z}\right) = \left(\sum_{k=1}^{p} \frac{\partial}{\partial x_k}\left(\Theta_k(x,t) \frac{\partial \tilde{z}}{\partial x_k}\right), \tilde{z}\right) + \left(\sum_{k=1}^{p} r_k(x,t) \frac{\partial \tilde{z}}{\partial x_k}, \tilde{z}\right) - \left(\sum_{k=1}^{p} q_k(x,t)\tilde{z}, \tilde{z}\right) + (\tilde{f}(x,t), \tilde{z}). \tag{1.10}$$

Будем пользоваться скалярным произведением и нормой

$$(u,v) = \int_G uv\,dx, \quad (u,u) = \|u\|_0^2, \quad \|u\|_{L_2(0,l_k)}^2 = \int_0^{l_k} u^2(x,t)\,dx_k.$$

Далее через M_i, $i = 1, 2, \ldots$, обозначаются положительные постоянные, зависящие только от входных данных рассматриваемой задачи.

Используя лемму 1 из [30], преобразуем интегралы, входящие в тождество (1.10):

$$\left(\varepsilon \frac{\partial \tilde{z}}{\partial t}, \tilde{z}\right) = \frac{\varepsilon}{2} \frac{\partial}{\partial t} \|\tilde{z}\|_0^2, \tag{1.11}$$

$$\begin{aligned} \left(\partial_{0t}^{\alpha} \tilde{z}, \tilde{z}\right) &= \left(\frac{1}{p} \sum_{k=1}^{p} \partial_{0t}^{\alpha} \tilde{z}, \tilde{z}\right) = \frac{1}{p} \int_G \sum_{k=1}^{p} \tilde{z} \partial_{0t}^{\alpha} \tilde{z}\,dx = \frac{1}{p} \sum_{k=1}^{p} \int_G \tilde{z} \partial_{0t}^{\alpha} \tilde{z}\,dx = \frac{1}{p} \sum_{k=1}^{p} \int_{G'} \left(\int_0^{l_k} \tilde{z} \partial_{0t}^{\alpha} \tilde{z}\,dx_k \right) dx' \ge \\ &\ge \frac{1}{2p} \sum_{k=1}^{p} \int_{G'} \left(\int_0^{l_k} \partial_{0t}^{\alpha} \tilde{z}^2\,dx_k \right) dx' = \frac{1}{2p} \sum_{k=1}^{p} \int_{G'} \partial_{0t}^{\alpha} \|\tilde{z}\|_{L_2(0,l_k)}^2\,dx' = \frac{1}{2p} \sum_{k=1}^{p} \partial_{0t}^{\alpha} \|\tilde{z}\|_0^2 = \frac{1}{2} \partial_{0t}^{\alpha} \|\tilde{z}\|_0^2, \end{aligned} \tag{1.12}$$

$$\left(\sum_{k=1}^{p}\frac{\partial}{\partial x_k}\left(\Theta_k(x,t)\frac{\partial\tilde z}{\partial x_k}\right),\tilde z\right)=\sum_{k=1}^{p}\int_{G'}\Theta_k(x,t)\tilde z\frac{\partial\tilde z}{\partial x_k}\Big|_0^{l_k}dx'-\sum_{k=1}^{p}\int_G\Theta_k(x,t)\left(\frac{\partial\tilde z}{\partial x_k}\right)^2dx=$$
$$=-\sum_{k=1}^{p}\int_G\Theta_k(x,t)\left(\frac{\partial\tilde z}{\partial x_k}\right)^2dx\le -c_0\|\tilde z_x\|_0^2, \tag{1.13}$$

где $u_x^2=\sum_{k=1}^{p}u_{x_k}^2$, $G'=\{x'=(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_p):0<x_k<l_k\}$, $dx'=dx_1\cdots dx_{k-1}dx_{k+1}\cdots dx_p$.

Далее, для оценки слагаемых в правой части применим ε-неравенство Коши

$$\left(\sum_{k=1}^{p}r_k(x,t)\frac{\partial\tilde z}{\partial x_k},\tilde z\right)=\int_G\sum_{k=1}^{p}r_k(x,t)\frac{\partial\tilde z}{\partial x_k}\tilde z\,dx=\sum_{k=1}^{p}\int_G r_k(x,t)\frac{\partial\tilde z}{\partial x_k}\tilde z\,dx\le$$
$$\le\varepsilon_1\sum_{k=1}^{p}\int_G\left(\frac{\partial\tilde z}{\partial x_k}\right)^2dx+M_1^{\varepsilon_1}\sum_{k=1}^{p}\int_G\tilde z^2dx, \tag{1.14}$$

$$\left(\tilde f(x,t),\tilde z\right)\le\frac12\|\tilde f\|_0^2+\frac12\|\tilde z\|_0^2. \tag{1.15}$$

Учитывая преобразования (1.11)–(1.15), из (1.10) получаем неравенство

$$\frac{\varepsilon}{2}\frac{\partial}{\partial t}\|\tilde z\|_0^2+\frac12\partial_{0t}^{\alpha}\|\tilde z\|_0^2+c_0\|\tilde z_x\|_0^2+c_0\|\tilde z\|_0^2\le\varepsilon_1\|\tilde z_x\|_0^2+M_2^{\varepsilon_1}\|\tilde z\|_0^2+\frac12\|\tilde f\|_0^2. \tag{1.16}$$

Выбирая $\varepsilon_1=\frac{c_0}{2}$, из неравенства (1.16) находим

$$\varepsilon\frac{\partial}{\partial t}\|\tilde z\|_0^2+\partial_{0t}^{\alpha}\|\tilde z\|_0^2+\|\tilde z_x\|_0^2+\|\tilde z\|_0^2\le M_3\|\tilde z\|_0^2+M_4\|\tilde f\|_0^2. \tag{1.17}$$

Проинтегрируем (1.17) по τ от 0 до t, тогда получим

$$\varepsilon\|\tilde z\|_0^2+D_{0t}^{\alpha-1}\|\tilde z\|_0^2+\int_0^t\left(\|\tilde z\|_0^2+\|\tilde z_x\|_0^2\right)d\tau\le M_5\int_0^t\|\tilde z\|_0^2d\tau+M_6\int_0^t\|\tilde f\|_0^2d\tau, \tag{1.18}$$

где $D_{0t}^{\alpha-1}u=\frac{1}{\Gamma(1-\alpha)}\int_0^t\frac{u\,d\tau}{(t-\tau)^{\alpha}}$ – дробный интеграл Римана–Лиувилля порядка $1-\alpha$, $0<\alpha<1$.

В (1.18) покажем, что $\int_0^t\|\tilde z\|_0^2d\tau=D_{0t}^{-\alpha}\left(D_{0t}^{\alpha-1}\|\tilde z\|_0^2\right)$:

$$D_{0t}^{-\alpha}\left(D_{0t}^{\alpha-1}\|\tilde z\|_0^2\right)=\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\int_0^t\frac{d\tau}{(t-\tau)^{1-\alpha}}\int_0^{\tau}\frac{\|\tilde z\|_0^2\,ds}{(\tau-s)^{\alpha}}=\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\int_0^t\|\tilde z\|_0^2ds\int_s^{\tau}\frac{d\tau}{(t-\tau)^{1-\alpha}(\tau-s)^{\alpha}}=$$
$$=\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\int_0^t\|\tilde z\|_0^2ds\int_0^{\infty}\frac{dv}{(1+v)v^{\alpha}}=\frac{B(1-\alpha,\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)}\int_0^t\|\tilde z\|_0^2d\tau=\int_0^t\|\tilde z\|_0^2d\tau, \tag{1.19}$$

где $D_{0t}^{-\alpha}u=\frac{1}{\Gamma(\alpha)}\int_0^t\frac{u\,d\tau}{(t-\tau)^{1-\alpha}}$ – дробный интеграл Римана–Лиувилля порядка α, $0<\alpha<1$.

Учитывая (1.19), из (1.18) получаем

$$\varepsilon\|\tilde z\|_0^2+D_{0t}^{\alpha-1}\|\tilde z\|_0^2+\int_0^t\left(\|\tilde z\|_0^2+\|\tilde z_x\|_0^2\right)d\tau\le M_5D_{0t}^{-\alpha}\left(D_{0t}^{\alpha-1}\|\tilde z\|_0^2\right)+M_6\int_0^t\|\tilde f\|_0^2\,d\tau. \tag{1.20}$$

С помощью леммы 2 (см. [30]) из (1.20) получаем неравенство

$$\varepsilon\|\tilde z\|_0^2+D_{0t}^{\alpha-1}\|\tilde z\|_0^2+\|\tilde z\|_{2,Q_t}^2+\|\tilde z_x\|_{2,Q_t}^2\le M_7\int_0^t\|\tilde f\|_0^2d\tau=\varepsilon^2M_7\int_0^t\|u_\tau\|_0^2d\tau=O(\varepsilon^2), \tag{1.21}$$

где M – зависит только от входных данных задач (1.1)–(1.3), $\|\tilde z_x\|_{2,Q_t}^2=\int_0^t\|\tilde z_x\|_0^2\,d\tau$.

Из априорной оценки (1.21) следует сходимость u^ε к u при $\varepsilon \to 0$ в норме $\|\tilde{z}\|_1^2 = \varepsilon\|\tilde{z}\|_0^2 + D_{0t}^{\alpha-1}\|\tilde{z}\|_0^2 + \|\tilde{z}\|_{2,Q_t}^2 + \|\tilde{z}_x\|_{2,Q_t}^2$. Поэтому при малом ε решение задачи (1.4)–(1.6) будем принимать за приближенное решение первой краевой задачи для уравнения конвекции–диффузии дробного порядка (1.1)–(1.3).

2. ПОСТРОЕНИЕ ЛОКАЛЬНО-ОДНОМЕРНОЙ СХЕМЫ (ЛОС)

Пространственную сетку выберем равномерной по каждому направлению Ox_k с шагом $h_k = \dfrac{l_k}{N_k}$, $k = 1, 2, \ldots, p$:

$$\overline{\omega}_{h_k} = \left\{ x_k^{(i_k)} = i_k h_k : i_k = 0, 1, \ldots, N_k, h_k = \frac{l_k}{N_k}, k = 1, 2, \ldots, p \right\}, \quad \overline{\omega} = \prod_{k=1}^{p} \overline{\omega}_{h_k}.$$

На отрезке $0 \le t \le T$ введем равномерную сетку

$$\overline{\omega}_\tau' = \left\{ 0, t_{j+\frac{k}{p}} = \left(j + \frac{k}{p} \right)\tau,\ j = 0, 1, \ldots, j_0 - 1,\ \tau = \frac{T}{j_0},\ k = 1, 2, \ldots, p \right\},$$

содержащую, наряду с узлами $t_j = j\tau$, фиктивные узлы $t_{j+\frac{k}{p}}$, $k = 1, 2, \ldots, p-1$. Будем обозначать через ω_τ' множество узлов сетки $\overline{\omega}_\tau'$, для которых $t > 0$.

На равномерной сетке $\overline{\omega}_{h\tau}$ по аналогии с [31] уравнению (1.4) поставим в соответствие цепочку “одномерных” уравнений, для этого перепишем уравнение (1.4) в виде

$$\pounds^\varepsilon = \varepsilon u_t^\varepsilon + \partial_{0t}^\alpha u^\varepsilon - Lu^\varepsilon - f = 0$$

или

$$\sum_{k=1}^{p} \pounds_k u^\varepsilon = 0, \quad J_k u^\varepsilon = \frac{\varepsilon}{p} u_t^\varepsilon + \frac{1}{p}\partial_{0t}^\alpha u^\varepsilon - L_k u^\varepsilon - f_k,$$

где $f_k(x,t)$, $k = 1, 2, \ldots, p$, – произвольные функции, обладающие той же гладкостью, что и $f(x,t)$, и удовлетворяющие условию $\sum_{k=1}^{p} f_k = f$.

На каждом полуинтервале $\Delta_k = \left(t_{j+\frac{k-1}{p}}, t_{j+\frac{k}{p}} \right]$, $k = 1, 2, \ldots, p$, будем последовательно решать задачи

$$\begin{gathered} \pounds_k \vartheta_{(k)} = 0, \quad x \in G, \quad t \in \Delta_k, \quad k = 1, 2, \ldots, p, \\ \vartheta_{(k)} = \mu(x,t) \quad \text{при} \quad x \in \Gamma_k, \end{gathered} \tag{2.1}$$

полагая при этом

$$\begin{gathered} \vartheta_{(1)}(x,0) = u_0(x), \quad \vartheta_{(1)}(x,t_j) = \vartheta_{(p)}(x,t_j), \quad j = 1, 2, \ldots, j_0 - 1, \\ \vartheta_{(k)}\left(x, t_{j+\frac{k-1}{p}} \right) = \vartheta_{(k-1)}\left(x, t_{j+\frac{k-1}{p}} \right), \quad k = 2, 3, \ldots, p, \end{gathered} \tag{2.2}$$

где Γ_k – множество граничных точек по направлению x_k.

Аналогично [31, с. 401], получим для уравнения (2.1) номера k монотонную схему второго порядка аппроксимации по h_k, для которой справедлив принцип максимума при любых τ и h_k, $k = 1, 2, \ldots, p$. Для этого рассмотрим уравнение (2.1) при фиксированном k с возмущенным оператором $\tilde{L}_k$:

$$\frac{\varepsilon}{p}\vartheta_t + \frac{1}{p}\partial_{0t}^\alpha \vartheta_{(k)} = \tilde{L}_k \vartheta_{(k)} + f_k, \quad t \in \Delta_k, \quad k = 1, 2, \ldots, p, \tag{2.3}$$

где

$$\tilde{L}_k\vartheta_{(k)} = \chi_k\frac{\partial}{\partial x_k}\left(\Theta_k(x,t)\frac{\partial\vartheta_{(k)}}{\partial x_k}\right) + r_k(x,t)\frac{\partial\vartheta_{(k)}}{\partial x_k} - q_k(x,t)\vartheta_{(k)},$$

$\chi_k = \dfrac{1}{1+R_k}$, $R_k = 0.5h_k\dfrac{|r_k|}{\Theta_k}$ – разностное число Рейнольдса.

Каждое из уравнений (2.3) заменим разностной схемой

$$\begin{gathered}\frac{\varepsilon}{p}y_{\bar{t}}^{\,j+\frac{k}{p}} + \frac{1}{p}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{pj+k}\left(t_{j+\frac{k-s+1}{p}}^{1-\alpha} - t_{j+\frac{k-s}{p}}^{1-\alpha}\right)y_t^{\frac{s}{p}} = \tilde{\Lambda}_k\left(\sigma_k y^{\,j+\frac{k}{p}} + (1-\sigma_k)y^{\,j+\frac{k-1}{p}}\right) + \varphi_k^{\,j+\frac{k}{p}},\\ x\in\omega_h,\quad k=1,2,\ldots,p,\\ y^{\,j+\frac{k}{p}}\Big|_{\gamma_{h,k}} = \mu^{\,j+\frac{k}{p}},\quad j=0,1,\ldots,j_0-1,\\ y(x,0)=u_0(x),\quad x\in\bar{G},\end{gathered}\tag{2.4}$$

где

$$\frac{1}{\Gamma(1-\alpha)}\int_0^{t_{j+\frac{k}{p}}}\frac{\partial u(x,\eta)}{\partial\eta}\frac{d\eta}{\left(t_{j+\frac{k}{p}}-\eta\right)^{\alpha}} = \frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{pj+k}\left(t_{j+\frac{k-s+1}{p}}^{1-\alpha} - t_{j+\frac{k-s}{p}}^{1-\alpha}\right)u_t^{\frac{s}{p}} + O\left(\frac{\tau}{p}\right),$$

$$y_t^{\frac{s}{p}} = \frac{y^{\frac{s}{p}} - y^{\frac{s-1}{p}}}{\frac{\tau}{p}},\quad \mu^{\,j+\frac{k}{p}} = \mu\left(x,t_{j+\frac{k}{p}}\right),\quad \varphi_k^{\,j+\frac{k}{p}} = f_k\left(x,t_{j+\frac{k}{p}}\right),\quad k=1,2,\ldots,p,$$

σ_k – произвольные параметры, $\gamma_{h,k}$ – множество граничных по направлению x_k узлов,

$$x\in\bar{\omega}_h = \left\{x_i = (i_1h_1,\ldots,i_ph_p)\in\bar{G},\ i_k = 0,1,\ldots N_k, h_k = \frac{l_k}{N_k}\right\},$$

$$\tilde{\Lambda}_k y^{\,j+\frac{k}{p}} = \chi_k\left(a_k y_{\bar{x}}^{\,j+\frac{k}{p}}\right)_{x_k} + b_k^+a_k^{(+1)}y_{x_k}^{\,j+\frac{k}{p}} + b_k^-a_k y_{\bar{x}_k}^{\,j+\frac{k}{p}} - d_k y^{\,j+\frac{k}{p}},$$

$$d_k^{\,j+\frac{k}{p}} = q\left(x_i,t^{\,j+\frac{k}{p}}\right),\quad a^{(+1)} = a_{i+1},\quad a_i^j = \Theta\left(x_{i-1/2},\bar{t}\right),\quad \bar{t} = t_{j+\frac{1}{2}},$$

$$r_k^+ = 0.5(r_k+|r_k|)\ge 0,\quad r_k^- = 0.5(r_k-|r_k|)\le 0,\quad b_k^+ = \frac{r_k^+}{\Theta_k},\quad b_k^- = \frac{r_k^-}{\Theta_k},\quad r_k = r_k^+ + r_k^-,$$

3. ПОГРЕШНОСТЬ АППРОКСИМАЦИИ ЛОС

Перейдем к изучению погрешности аппроксимации (невязки) локально-одномерной схемы и убедимся в том, что каждое в отдельности уравнение (2.4) номера k не аппроксимирует уравнение (1.4), но сумма погрешностей аппроксимации:

$$\psi = \psi_1 + \ldots + \psi_p,$$

стремится к нулю при τ и $|h|$, стремящимся к нулю.

Будем считать $\sigma_k = 1$, $k = 1, 2, \ldots, p$. Пусть $u = u(x,t)$ – решение задачи (1.4)–(1.6), а $y^{\,j+\frac{k}{p}}$ – решение разностной задачи (2.4). Характеристикой точности локально-одномерной схемы являет-

ся разность $y^{j+1} - u^{j+1} = z^{j+1}$. Промежуточные значения $y^{j+\frac{k}{p}}$ будем сравнивать с $u^{j+\frac{k}{p}} = u\left(x, t_{j+\frac{k}{p}}\right)$, полагая $z^{j+\frac{k}{p}} = y^{j+\frac{k}{p}} - u^{j+\frac{k}{p}}$. Подставляя $y^{j+\frac{k}{p}} = z^{j+\frac{k}{p}} + u^{j+\frac{k}{p}}$ в разностное уравнение (2.4), получаем

$$\frac{\varepsilon}{p} z_{\bar{t}}^{j+\frac{k}{p}} + \frac{1}{p}\frac{1}{\Gamma(2-\alpha)} \sum_{s=1}^{pj+k} \left(t^{1-\alpha}_{j+\frac{k-s+1}{p}} - t^{1-\alpha}_{j+\frac{k-s}{p}} \right) z_{\bar{t}}^{\frac{s}{p}} = \tilde{\Lambda}_k z^{j+\frac{k}{p}} + \psi_k^{j+\frac{k}{p}}, \tag{3.1}$$

$$z^{j+\frac{k}{p}}\Big|_{\gamma_{h,k}} = 0, \quad z(x,0) = 0, \tag{3.2}$$

где

$$\psi_k^{j+\frac{k}{p}} = \tilde{\Lambda}_k u^{j+\frac{k}{p}} + \varphi_k^{j+\frac{k}{p}} - \frac{1}{p}\frac{1}{\Gamma(2-\alpha)} \sum_{s=1}^{pj+k} \left(t^{1-\alpha}_{j+\frac{k-s+1}{p}} - t^{1-\alpha}_{j+\frac{k-s}{p}} \right) u_{\bar{t}}^{\frac{s}{p}} - \frac{\varepsilon}{p} u_t^{j+\frac{k}{p}}.$$

Обозначив через

$$\mathring{\psi}_k = \left(L_k u + f_k - \frac{\varepsilon}{p} u_t - \frac{1}{p} \partial_{0t}^{\alpha} u \right)^{j+\frac{1}{2}} \tag{3.3}$$

и, замечая, что

$$\sum_{k=1}^{p} \mathring{\psi}_k = 0,$$

если

$$\sum_{k=1}^{p} f_k = f,$$

представим $\psi_k = \psi_k^{j+\frac{k}{p}}$ в виде

$$\psi_k = \mathring{\psi}_k + \overset{*}{\psi}_k,$$

где

$$\psi_k^{j+\frac{k}{p}} = \left(\tilde{\Lambda}_k u^{j+\frac{k}{p}} - L_k u^{j+\frac{1}{2}} \right) + \left(\varphi_k^{j+\frac{k}{p}} - f_k^{j+\frac{1}{2}} \right) - \left(\frac{1}{p} \Delta^{\alpha}_{0t_{j+\frac{k}{p}}} u^{j+\frac{k}{p}} - \frac{1}{p} (\partial_{0t}^{\alpha} u)^{j+\frac{1}{2}} \right) -$$

$$- \left(\frac{\varepsilon}{p} u_{\bar{t}}^{j+\frac{k}{p}} - \frac{\varepsilon}{p} u_t^{j+\frac{1}{2}} \right) + \mathring{\psi}_k = \mathring{\psi}_k + \overset{*}{\psi}_k,$$

$$\overset{*}{\psi}_k = \left(\tilde{\Lambda}_k u^{j+\frac{k}{p}} - L_k u^{j+\frac{1}{2}} \right) + \left(\varphi_k^{j+\frac{k}{p}} - f_k^{j+\frac{1}{2}} \right) - \left(\frac{1}{p} \Delta^{\alpha}_{0t_{j+\frac{k}{p}}} u^{j+\frac{k}{p}} - \frac{1}{p} (\partial_{0t}^{\alpha} u)^{j+\frac{1}{2}} \right) - \left(\frac{\varepsilon}{p} u_{\bar{t}}^{j+\frac{k}{p}} - \frac{\varepsilon}{p} u_t^{j+\frac{1}{2}} \right).$$

Ясно, что $\overset{*}{\psi}_k = O(h_k^2 + \tau)$, так как каждая из схем (2.4) номера k аппроксимирует в обычном смысле соответствующее уравнение (2.3), т.е. $\|\overset{*}{\psi}_k\|$ стремится к нулю (в некоторой норме) при $|h| \to 0$, $\tau \to 0$. Таким образом, ЛОС (2.4) обладает суммарной аппроксимацией

$$\overset{*}{\psi}_k = O(h_k^2 + \tau), \quad \mathring{\psi}_k = O(1), \quad \sum_{k=1}^{p} \mathring{\psi}_k = 0,$$

$$\psi = \sum_{k=1}^{p} \psi_k = \sum_{k=1}^{p} \left(\mathring{\psi}_k + \overset{*}{\psi}_k \right) = \sum_{k=1}^{p} \overset{*}{\psi}_k = O\left(|h|^2 + \tau \right).$$

4. УСТОЙЧИВОСТЬ ЛОС

Получим априорную оценку в сеточной норме C для решения разностной задачи (2.4), выражающую устойчивость локально-одномерной схемы по начальным данным и правой части. Исследование устойчивости разностной схемы (2.4) будем проводить с помощью принципа максимума (см. [31, с. 226]), для чего решение задачи (2.4) представим в виде суммы

$$y = \overline{y} + v,$$

где $\overline{y}$ – решение однородных уравнений (2.4) с неоднородными краевыми и начальными условиями

$$\overline{y}^{\,j+\frac{k}{p}}\Big|_{\gamma_{h,k}} = \mu^{\,j+\frac{k}{p}},$$

$$\overline{y}(x,0) = u_0(x),$$

v – решение неоднородных уравнений (2.4) с однородными краевыми и начальными условиями

$$\begin{gathered}\frac{\varepsilon}{p}\overline{y}_{\bar{t}}^{\,j+\frac{k}{p}} + \frac{1}{p}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{pj+k}\left(t_{j+\frac{k-s+1}{p}}^{1-\alpha} - t_{j+\frac{k-s}{p}}^{1-\alpha}\right)\overline{y}_t^{\frac{s}{p}} = \tilde{\Lambda}_k\overline{y}^{\,j+\frac{k}{p}},\\ \overline{y}^{\,j+\frac{k}{p}}\Big|_{\gamma_{h,k}} = \mu^{\,j+\frac{k}{p}},\\ \overline{y}(x,0) = u_0(x),\end{gathered} \tag{4.1}$$

$$\begin{gathered}\frac{\varepsilon}{p}v_{\bar{t}}^{\,j+\frac{k}{p}} + \frac{1}{p}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{pj+k}\left(t_{j+\frac{k-s+1}{p}}^{1-\alpha} - t_{j+\frac{k-s}{p}}^{1-\alpha}\right)v_t^{\frac{s}{p}} = \tilde{\Lambda}_k v^{\,j+\frac{k}{p}} + \varphi_k^{\,j+\frac{k}{p}},\\ v^{\,j+\frac{k}{p}}\Big|_{\gamma_{h,k}} = 0,\\ v(x,0) = 0.\end{gathered} \tag{4.2}$$

Получим оценку для $\overline{y}$, записав уравнение (4.1) в канонической форме. В точке $P = P\left(x_{i_k}, t_{j+\frac{k}{p}}\right)$ имеем

$$\begin{aligned}&\left[\frac{\varepsilon}{\tau} + \frac{\gamma}{\tau^\alpha} + \frac{\chi_k a_{k,i_k+1}}{h_k^2} + \frac{\chi_k a_{k,i_k}}{h_k^2} + \frac{b_k^+ a_{k,i_k+1}}{h_k} - \frac{b_k^- a_{k,i_k}}{h_k} + d_k\right]\overline{y}_{i_k}^{\,j+\frac{k}{p}} = \frac{\chi_k a_{k,i_k+1}}{h_k^2}\overline{y}_{i_k+1}^{\,j+\frac{k}{p}} + \frac{\chi_k a_{k,i_k}}{h_k^2}\overline{y}_{i_k-1}^{\,j+\frac{k}{p}} + \frac{b_k^+ a_{k,i_k+1}}{h_k}\overline{y}_{i_k+1}^{\,j+\frac{k}{p}} -\\ &- \frac{b_k^- a_{k,i_k}}{h_k}\overline{y}_{i_{k-1}}^{\,j+\frac{k}{p}} + \left[\frac{\varepsilon}{\tau} + \frac{\gamma}{\tau^\alpha}(2-2^{1-\alpha})\right]\overline{y}_{i_k}^{\,j+\frac{k-1}{p}} + \frac{1}{\tau}\frac{1}{\Gamma(2-\alpha)}\left[\left(t_{j+\frac{k}{p}}^{1-\alpha} - t_{j+\frac{k-1}{p}}^{1-\alpha}\right)\overline{y}_{i_k}^{0} +\right.\\ &\left. + \left(-t_{j+\frac{k}{p}}^{1-\alpha} + 2t_{j+\frac{k-1}{p}}^{1-\alpha} - t_{j+\frac{k-2}{p}}^{1-\alpha}\right)\overline{y}_{i_k}^{\frac{1}{p}} + \ldots + \left(-t_{\frac{3}{p}}^{1-\alpha} + 2t_{\frac{2}{p}}^{1-\alpha} - t_{\frac{1}{p}}^{1-\alpha}\right)\overline{y}_{i_k}^{\,j+\frac{k-2}{p}}\right],\end{aligned} \tag{4.3}$$

где $\gamma = \dfrac{1}{p^{1-\alpha}\Gamma(2-\alpha)}$.

Справедлива следующая (см. [5])

Лемма. *Пусть $l = pj + k - 1 \geq 1$, тогда имеет место неравенство*

$$-t_{j+\frac{k}{p}}^{1-\alpha} + 2t_{j+\frac{k-1}{p}}^{1-\alpha} - t_{j+\frac{k-2}{p}}^{1-\alpha} > 0, \quad j = 0, 1, \ldots, j_0 - 1, \quad k = 2, 3, \ldots, p. \tag{4.4}$$

В [31] доказан принцип максимума и получены априорные оценки для решения сеточного уравнения общего вида

$$A(P)y(P) = \sum_{Q\in Ш'(P)} B(P,Q)y(Q) + F(P), \quad P \in \Omega,$$

$$y(P) = \mu(P) \quad \text{при} \quad P \in S,$$

где P, Q – узлы сетки $\Omega + S$, Ш'(P) – окрестность узла P, не содержащего самого узла P. Коэффициенты $A(P)$, $B(P,Q)$ удовлетворяют условиям

$$A(P) > 0, \quad B(P,Q) > 0, \quad D(P) = A(P) - \sum_{Q \in \text{Ш}'(P)} B(P,Q) \geq 0. \tag{4.5}$$

Обозначим через $P(x,t')$, где $x \in \omega_h$, $t' \in \omega'_\tau$, узел $(p+1)$-мерной сетки $\mathfrak{Q} = \omega_h \times \omega'_\tau$, через S – границу Ω, состоящую из узлов $P(x,0)$ при $x \in \bar{\omega}_h$ и узлов $P\left(x, t_{j+\frac{k}{p}}\right)$ при $t_{j+\frac{k}{p}} \in \omega'_\tau$ и $x \in \gamma_{h,k}$ для всех $k = 1, 2, \ldots, p$ и $j = 0, 1, \ldots, j_0$.

Справедливы следующие теоремы (см. [32]).

Теорема 1 (см. [32, с. 344]). *Пусть коэффициенты уравнения*

$$A(P)y(P) = \sum_{Q \in \text{Ш}'(P)} B(P,Q)y(Q) + F(P), \quad P \in \Omega, \tag{*}$$

удовлетворяют условиям

$$A(P) > 0, \quad B(P,Q) \geq 0, \quad D(P) > 0, \quad P \in \overset{*}{\omega},$$

$$A(P) > 0, \quad B(P,Q) > 0, \quad D(P) = F(P) = 0, \quad P \in \mathring{\omega},$$

где $\mathring{\omega}$ *– некоторое связное подмножество множества* ω, *а* $\overset{*}{\omega}$ *– дополнение* $\mathring{\omega}$ *до* ω.

Тогда для решения задачи (*) *справедлива оценка*

$$\|y\|_C \leq \left\| \frac{F(P)}{D(P)} \right\|_{C^*},$$

где

$$\|f\|_C = \max_{P \in \omega} |f(P)|, \quad \|f\|_{C^*} = \max_{P \in \omega^*} |f(P)|.$$

Теорема 2 (см. [32, с. 347]). *Если выполнены условия*

$$D'\left(P_{(n+1)}\right) > 0 \quad \textit{для всех} \quad P_{(n+1)} \in \omega, \quad A\left(P_{(n+1)}\right) > 0, \quad B\left(P_{(n+1)}, Q\right) \geq 0$$

для всех $Q \in \text{Ш}''_n$, $Q \in \text{Ш}'_{n+1}$,

$$\sum_{Q \in \text{Ш}''_n} B\left(P_{(n+1)}, Q\right) > 0, \quad \frac{1}{D'\left(P_{(n+1)}\right)} \sum_{Q \in \text{Ш}^*_n} B\left(P_{(n+1)}, Q\right) \leq 1 + c_1\tau,$$

где $c_1 = \text{const} > 0$ *не зависит от* τ, h.

Тогда для решения задачи

$$A\left(P_{(n+1)}\right)y\left(P_{(n+1)}\right) = \sum_{Q \in \text{Ш}'_{n+1}} B(P_{(n+1)}, Q)y(Q) + \Phi\left(P_{(n+1)}\right),$$

где

$$P_{(n+1)} = P\left(x, t_{n+1}\right),$$

$$\Phi(P_{(n+1)}) = \sum_{Q \in \text{Ш}'_n} B(P_{(n+1)}, Q)y(Q) + F\left(P_{(n+1)}\right),$$

$$D'(P_{(n+1)}) = A\left(P_{(n+1)}\right) - \sum_{Q \in \text{Ш}''_m} B\left(P_{(n+1)}, Q\right),$$

справедлива оценка

$$\|y_{n+1}\|_{C_h} \leq e^{c_1 t_n} \left(\|y_0\|_{C_h} + \sum_{k=1}^{n+1} \tau \|\tilde{F}_k\|_{C_h} \right).$$

Проверим выполнимость условий теоремы 1, опираясь на лемму. Тогда, учитывая положительность выражений, стоящих в круглых скобках, имеем, что коэффициенты уравнения (4.3) в точке $P = P\left(x_{i_k}, t_{j+\frac{k}{p}}\right)$ удовлетворяют условиям (4.5) и $D(P) = 0$.

Из теоремы 2 следует, что для решения задачи (4.1) верна оценка

$$\left\|\bar{y}^{j}\right\|_{C} \le \|u_0\|_{C} + \max_{0<t'\le j\tau}\|\mu(x,t')\|_{C_\gamma}, \tag{4.6}$$

где

$$\|y\|_{C} = \max_{x\in\bar{\omega}_h}|y|, \quad \|y\|_{C_\gamma} = \max_{x\in\gamma_h}|y|.$$

Переходим к оценке функции v. Уравнение (4.2) перепишем в виде

$$\begin{gathered}\left(\frac{\varepsilon}{p} + \frac{1}{p}\frac{1}{\Gamma(2-\alpha)}\left(\frac{\tau}{p}\right)^{1-\alpha}\right)v_{\bar{t}}^{\,j+\frac{k}{p}} = \tilde{\Lambda}_k v^{\,j+\frac{k}{p}} + \tilde{\varphi}_k^{\,j+\frac{k}{p}},\\ v^{\,j+\frac{k}{p}}\Big|_{\gamma_{h,k}} = 0,\\ v(x,0) = 0,\end{gathered} \tag{4.7}$$

где

$$\tilde{\varphi}_k^{\,j+\frac{k}{p}} = \varphi^{\,j+\frac{k}{p}} - \frac{1}{p}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{pj+k-1}\left(t_{j+\frac{k-s+1}{p}}^{1-\alpha} - t_{j+\frac{k-s}{p}}^{1-\alpha}\right)v_{\bar{t}}^{\frac{s}{p}}.$$

Уравнение (4.7) приведем к каноническому виду

$$\begin{gathered}\left[\frac{\varepsilon}{\tau} + \frac{\gamma}{\tau^\alpha} + \frac{\chi_{i_k}a_{k,i_k+1}}{h_k^2} + \frac{\chi_{i_k}a_{k,i_k}}{h_k^2} + \frac{b_k^+a_{k,i_k+1}}{h_k} - \frac{b_k^-a_{k,i_k}}{h_k} + d_k\right]v_{i_k}^{\,j+\frac{k}{p}} =\\ = \frac{1}{h_k^2}\left[\chi_{i_k}a_{k,i_k+1}\nu_{i_k+1}^{\,j+\frac{k}{p}} + \chi_{i_k}a_{k,i_k}\nu_{i_k-1}^{\,j+\frac{k}{p}}\right] + \frac{b_k^+a_{k,i_k+1}}{h_k}\nu_{i_k+1}^{\,j+\frac{k}{p}} - \frac{b_k^-a_{k,i_k}}{h_k}\nu_{i_k-1}^{\,j+\frac{k}{p}} + \Phi\left(P_{j+\frac{k}{p}}\right),\end{gathered}$$

где

$$\Phi\left(P_{j+\frac{k}{p}}\right) = \left[\frac{\varepsilon}{\tau} + \frac{\gamma}{\tau^\alpha}\left(2-2^{1-\alpha}\right)\right]v_{i_k}^{\,j+\frac{k-1}{p}} + \bar{\varphi}_k^{\,j+\frac{k}{p}},$$

$$\bar{\varphi}_k^{\,j+\frac{k}{p}} = \varphi_k^{\,j+\frac{k}{p}} + \frac{1}{\Gamma(2-\alpha)}\frac{1}{\tau}\left(t_{\frac{2}{p}}^{1-\alpha} - t_{\frac{1}{p}}^{1-\alpha}\right)v_{i_k}^{\,j+\frac{k-2}{p}} - \frac{1}{\tau}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{pj+k-2}\left(t_{j+\frac{k-s+1}{p}}^{1-\alpha} - t_{j+\frac{k-s}{p}}^{1-\alpha}\right)\left(v_{i_k}^{\frac{s}{p}} - v_{i_k}^{\frac{s-1}{p}}\right).$$

Проверим выполнимость условий теоремы 2

$$\begin{gathered}D'(P_{(k)}) = A(P_{(k)}) - \sum_{Q\in Ш_k'(P)}B\left(P_{(k)},Q\right) = \frac{\varepsilon}{\tau} + \frac{\gamma}{\tau^\alpha} + d_k \ge \frac{\varepsilon}{\tau} + \frac{\gamma}{\tau^\alpha} > 0,\\ P_{(k)} = P\left(x, t_{j+\frac{k}{p}}\right), \quad A(P_{(k)}) > 0, \quad B\left(P_{(k)},Q\right) > 0,\end{gathered} \tag{4.8}$$

для всех $Q \in Ш_{k-1}''$, $Q \in Ш_k'$,

$$\sum_{Q\in Ш_{k-1}''}B\left(P_{(k)},Q\right) = \frac{\varepsilon}{\tau} + \frac{\gamma}{\tau^\alpha}\left(2-2^{1-\alpha}\right) > 0,$$

$$\frac{1}{D'(P_{(k)})}\sum_{Q\in Ш''_{k-1}} B(P_{(k)},Q)=\frac{\frac{\varepsilon}{\tau}+\frac{\gamma(2-2^{1-\alpha})}{\tau^{\alpha}}}{\frac{\varepsilon}{\tau}+\frac{\gamma}{\tau^{\alpha}}}\le 1,$$

где $Ш'\left(P\left(x,t_{j+\frac{k}{p}}\right)\right)=Ш'_k+Ш'_{k-1}$, $Ш'_k$ – множество узлов $Q=Q(\xi,t_k)\in Ш'(P(x,t_k))$, $Ш'_{k-1}$ – множество узлов $Q=Q(\xi,t_{k-1})\in Ш'(P(x,t_{k-1}))$.

На основании теоремы 2, в силу (4.8), получаем оценку для v

$$\left\|v^{j+\frac{k}{p}}\right\|_C\le\frac{1}{\frac{\varepsilon}{\tau}+\frac{\gamma}{\tau^{\alpha}}}\left\|\overline{\varphi}_k^{j+\frac{k}{p}}\right\|_C+\frac{\frac{\varepsilon}{\tau}+\frac{\gamma(2-2^{1-\alpha})}{\tau^{\alpha}}}{\frac{\varepsilon}{\tau}+\frac{\gamma}{\tau^{\alpha}}}\left\|v^{j+\frac{k-1}{p}}\right\|_C. \tag{4.9}$$

Оценим $\left\|\overline{\varphi}_k^{j+\frac{k}{p}}\right\|_C$, где

$$\begin{aligned}\overline{\varphi}_k^{j+\frac{k}{p}}&=\varphi_k^{j+\frac{k}{p}}+\frac{1}{\Gamma(2-\alpha)}\frac{1}{\tau}\left(t_{\frac{2}{p}}^{1-\alpha}-t_{\frac{1}{p}}^{1-\alpha}\right)v_{i_k}^{j+\frac{k-2}{p}}-\frac{1}{p}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{pj+k-2}\left(t_{j+\frac{k-s+1}{p}}^{1-\alpha}-t_{j+\frac{k-s}{p}}^{1-\alpha}\right)v_t^{\frac{s}{p}}=\\&=\varphi_k^{j+\frac{k}{p}}+\frac{1}{\tau}\frac{1}{\Gamma(2-\alpha)}\left[\left(t_{j+\frac{k}{p}}^{1-\alpha}-t_{j+\frac{k-1}{p}}^{1-\alpha}\right)v_{i_k}^0+\left(-t_{j+\frac{k}{p}}^{1-\alpha}+2t_{j+\frac{k-1}{p}}^{1-\alpha}-t_{j+\frac{k-2}{p}}^{1-\alpha}\right)v_{i_k}^{\frac{1}{p}}+\ldots+\right.\\&\left.+\left(-t_{\frac{3}{p}}^{1-\alpha}+2t_{\frac{2}{p}}^{1-\alpha}-t_{\frac{1}{p}}^{1-\alpha}\right)v_{i_k}^{j+\frac{k-2}{p}}\right].\end{aligned} \tag{4.10}$$

Так как, в силу леммы, выражения, стоящие в круглых скобках положительны, то из (4.10) получаем оценку

$$\left\|\overline{\varphi}_k^{j+\frac{k}{p}}\right\|_C\le\left\|\varphi_k^{j+\frac{k}{p}}\right\|_C+\frac{\gamma(2^{1-\alpha}-1)}{\tau^{\alpha}}\max_{0\le s\le k-2}\left\|v^{j+\frac{s}{p}}\right\|_C. \tag{4.11}$$

С помощью (4.11) из (4.9) находим

$$\max_{0\le s\le k}\left\|v^{j+\frac{s}{p}}\right\|_C\le\max_{0\le s\le k-1}\left\|v^{j+\frac{s}{p}}\right\|_C+\frac{\tau}{\varepsilon+\gamma\tau^{1-\alpha}}\max_{0\le s\le k}\left\|\varphi_k^{j+\frac{s}{p}}\right\|_C. \tag{4.12}$$

Просуммировав (4.12) сначала по $k=1,2,\ldots,p$, затем по $j'=0,1,\ldots,j$, получим оценку

$$\left\|v^{j+1}\right\|_C\le\left\|v^0\right\|_C+\sum_{j'=0}^{j}\frac{\tau}{\varepsilon+\gamma\tau^{1-\alpha}}\sum_{k=1}^{p}\max_{0\le s\le k}\left\|\varphi_k^{j'+\frac{s}{p}}\right\|_C. \tag{4.13}$$

Из (4.6) и (4.13) следует окончательная оценка

$$\left\|y^{j+1}\right\|_C\le\left\|y^0\right\|_C+\max_{0<t'\le t_{j+1}}\left\|\mu(x,t')\right\|_{C_\gamma}+\sum_{j'=0}^{j}\frac{\tau}{\varepsilon+\gamma\tau^{1-\alpha}}\sum_{k=1}^{p}\max_{0\le s\le k}\left\|\varphi_k^{j'+\frac{s}{p}}\right\|_C. \tag{4.14}$$

Таким образом, справедлива

Теорема 3. *Локально-одномерная схема* (2.4) *устойчива по начальным данным и правой части, так что для решения задачи* (2.4) *справедлива оценка* (4.14).

5. РАВНОМЕРНАЯ СХОДИМОСТЬ ЛОС

Чтобы использовать свойство $\sum_{k=1}^{p} \mathring{\psi}_k = 0$, $\mathring{\psi} = O(1)$, представим по аналогии с [31] решение задачи для погрешности (3.1), (3.2) в виде суммы

$$z_{(k)} = v_{(k)} + \eta_{(k)}, \quad z_{(k)} = z^{j+\frac{k}{p}}, \tag{5.1}$$

где $\eta_{(k)}$ определяется условиями

$$\frac{\varepsilon}{p}\eta_{\bar{t}}^{j+\frac{k}{p}} + \frac{1}{p}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{pj+k}\left(t_{j+\frac{k-s+1}{p}}^{1-\alpha} - t_{j+\frac{k-s}{p}}^{1-\alpha}\right)\eta_t^{\frac{s}{p}} = \mathring{\psi}_k, \quad x \in \omega_h + \gamma_{h,k}, \quad k = 1, 2, \ldots, p, \tag{5.2}$$
$$\eta(x,0) = 0.$$

Функция $v_{(k)}$ определяется условиями

$$\frac{\varepsilon}{p}v_{\bar{t}}^{j+\frac{k}{p}} + \frac{1}{p}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{pj+k}\left(t_{j+\frac{k-s+1}{p}}^{1-\alpha} - t_{j+\frac{k-s}{p}}^{1-\alpha}\right)v_t^{\frac{s}{p}} = \tilde{\Lambda}_k v_{(k)} + \tilde{\psi}_k, \tag{5.3}$$
$$v_{(k)}\big|_{\gamma_{h,k}} = -\eta_{(k)}, \quad v(x,0) = 0,$$

где

$$\tilde{\psi}_k = \overset{*}{\psi}_k + \tilde{\Lambda}_k\eta_{(k)}, \quad \overset{*}{\psi}_k = O(h_k^2 + \tau).$$

Покажем, что

$$\eta^{j+\frac{k}{p}} = O\left(\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\right), \quad k = 1, 2, \ldots, p, \quad j = 0, 1, \ldots, j_0 - 1.$$

Ради простоты рассмотрим двумерный случай ($p = 2$). Сначала положим $j = 0$, т.е. рассмотрим первый слой $(0, t_1]$. Тогда задача (5.2) примет вид

$$\frac{\varepsilon}{2}\eta_{\bar{t}}^{\frac{k}{2}} + \frac{1}{2}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{k}\left(t_{\frac{k-s+1}{2}}^{1-\alpha} - t_{\frac{k-s}{2}}^{1-\alpha}\right)\eta_t^{\frac{s}{2}} = \mathring{\psi}_k, \quad k = 1, 2.$$

Пусть $k = 1$, тогда получим

$$\frac{\varepsilon}{2}\eta_{\bar{t}}^{\frac{1}{2}} + \frac{1}{2}\frac{1}{\Gamma(2-\alpha)}t_{\frac{1}{2}}^{1-\alpha}\eta_t^{\frac{1}{2}} = \mathring{\psi}_1. \tag{5.4}$$

При $k = 2$ получаем

$$\frac{\varepsilon}{2}\eta_{\bar{t}}^{1} + \frac{1}{2}\frac{1}{\Gamma(2-\alpha)}\left[\left(t_1^{1-\alpha} - t_{\frac{1}{2}}^{1-\alpha}\right)\eta_t^{\frac{1}{2}} + t_{\frac{1}{2}}^{1-\alpha}\eta_{\bar{t}}^{1}\right] = \mathring{\psi}_2. \tag{5.5}$$

Складывая выражения (5.4) и (5.5), получаем

$$\frac{\varepsilon}{2}\eta_{\bar{t}}^{\frac{1}{2}} + \frac{\varepsilon}{2}\eta_{\bar{t}}^{1} + \frac{1}{2}\frac{1}{\Gamma(2-\alpha)}\frac{1}{\tau^{\alpha}}\left[\left(1 - \frac{1}{2^{1-\alpha}}\right)\eta^{\frac{1}{2}} + \frac{1}{2^{1-\alpha}}\eta^{1}\right] = 0. \tag{5.6}$$

Из (5.4) находим

$$\eta^{\frac{1}{2}} = \frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\mathring{\psi}_1 = -\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\mathring{\psi}_2, \tag{5.7}$$

где $\gamma = \dfrac{1}{2^{1-\alpha}\Gamma(2-\alpha)}$.

Выражая η^1 из (5.6) и учитывая (5.7), получаем

$$\eta^{\frac{1}{2}}, \eta^{1} = O\left(\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\right). \tag{5.8}$$

Допустим, что при $j = n$ выполнено условие

$$\eta^{\frac{1}{2}}, \eta^{1}, \eta^{1+\frac{1}{2}}, \ldots, \eta^{n+1} = O\left(\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\right). \tag{5.9}$$

Опираясь на допущение (5.9), покажем, что аналогичное условие выполнено и при $j = n+1$. Для чего запишем уравнение (5.2) при $j = n+1$, $p = 2$:

$$\frac{\varepsilon}{2}\eta_{\bar{t}}^{n+1+\frac{k}{2}} + \frac{1}{2}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{2(n+1)+k}\left(t^{1-\alpha}_{n+1+\frac{k-s+1}{2}} - t^{1-\alpha}_{n+1+\frac{k-s}{2}}\right)\eta_{\bar{t}}^{\frac{s}{2}} = \mathring{\psi}_k, \quad k = 1, 2. \tag{5.10}$$

Полагая в (5.10) $k = 1$, находим

$$\begin{aligned}
&\tau^{1-\alpha}\left[\left(n+\frac{3}{2}\right)^{1-\alpha} - 2(n+1)^{1-\alpha} + \left(n+\frac{1}{2}\right)^{1-\alpha}\right]\eta^{\frac{1}{2}} + \\
&+ \tau^{1-\alpha}\left[(n+1)^{1-\alpha} - 2\left(n+\frac{1}{2}\right)^{1-\alpha} + n^{1-\alpha}\right]\eta^{1} + \cdots - \Gamma(2-\alpha)\left(\varepsilon - \frac{\tau^{1-\alpha}}{\Gamma(2-\alpha)}(1-2^{\alpha})\right)\eta^{n+1} + \\
&+ \Gamma(2-\alpha)(\varepsilon + \gamma\tau^{1-\alpha})\eta^{n+\frac{3}{2}} = 2\Gamma(2-\alpha)\tau\mathring{\psi}_1.
\end{aligned} \tag{5.11}$$

Откуда с учетом (5.9) и достаточной ограниченности коэффициентов при $\eta^{\frac{1}{2}}, \eta^{1}, \ldots, \eta^{n+\frac{3}{2}}$ находим $\eta^{n+\frac{3}{2}} = O\left(\dfrac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\right)$.

Положим теперь в (5.10) $k = 2$, затем сложим полученное таким образом выражение с выражением (5.11) с учетом равенства

$$\mathring{\psi}_1 + \mathring{\psi}_2 = 0.$$

Тогда получим

$$\eta^{\frac{1}{2}}, \eta^{1}, \ldots, \eta^{n+1}, \eta^{n+\frac{3}{2}}, \eta^{n+2} = O\left(\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\right). \tag{5.12}$$

Итак, равенство (5.12) выполнено при любом значении j. Аналогично можно показать, что

$$\eta^{j+\frac{k}{p}} = O\left(\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\right), \quad k = 1, 2, \ldots, p, \quad j = 0, 1, \ldots, j_0 - 1.$$

Для оценки решения задачи (5.3) воспользуемся теоремой 3:

$$\left\|v^{j+1}\right\|_C \le \max_{0<j'+\frac{k}{p}\le j+1}\left\|\eta^{j'+\frac{k}{p}}\right\|_{C_\gamma} + \sum_{j'=0}^{j}\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\sum_{k=1}^{p}\max_{0\le s\le k}\left\|\tilde{\psi}_k^{j'+\frac{s}{p}}\right\|_C, \tag{5.13}$$

где $\tilde{\psi}_k = \overset{*}{\psi}_k + \tilde{\Lambda}_k\eta_{(k)}$.

Если существуют непрерывные в замкнутой области $\bar{Q}_T$ производные $\dfrac{\partial^4 u}{\partial x_k^2 \partial x_\nu^2}$, $k \ne \nu$, то

$$\tilde{\Lambda}_k\eta_{(k)} = -\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}a_k\tilde{\Lambda}_k\left(\mathring{\psi}_{k+1} + \ldots + \mathring{\psi}_p\right) = O\left(\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\right),$$

a_k – известные постоянные.

Тогда из оценки (5.13) находим, что

$$\left\|v^{j+1}\right\|_C \le M\left(\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}} + p\frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\sum_{j'=0}^{j}\left(h^2 + \frac{\tau}{\varepsilon + \gamma\tau^{1-\alpha}}\right)\right) \le M\left(\frac{h^2}{\varepsilon + \tau^{1-\alpha}} + \frac{\tau}{(\varepsilon + \tau^{1-\alpha})^2}\right), \quad h = \max_{1\le k\le p}h_k.$$

Откуда получаем

$$\left\|z^{j+1}\right\|_C \le \left\|\eta^{j+1}\right\|_C + \left\|v^{j+1}\right\|_C \le O\left(\frac{h^2}{\varepsilon+\tau^{1-\alpha}} + \frac{\tau}{(\varepsilon+\tau^{1-\alpha})^2}\right).$$

Итак, справедлива теорема

Теорема 4. *Пусть задача* (1.4)–(1.6) *имеет единственное непрерывное решение* $u(x,t)$ *в* $\overline{Q}_T$ *при всех значениях* ε *и существуют непрерывные в* $\overline{Q}_T$ *производные*

$$\frac{\partial^2 u}{\partial t^2}, \frac{\partial^4 u}{\partial x_k^2 \partial x_\nu^2}, \frac{\partial^3 u}{\partial x_k^2 \partial t}, \frac{\partial^{2+\alpha} u}{\partial x_k^2 \partial t^\alpha}, \frac{\partial^2 f}{\partial x_k^2}, \quad 1 \le k, \quad \nu \le p, \quad k \ne \nu, \quad 0 < \alpha < 1,$$

тогда решение разностной задачи (2.4) *равномерно сходится к решению дифференциальной задачи* (1.1)–(1.3) *со скоростью*

$$O\left(\frac{h^2}{\varepsilon+\tau^{1-\alpha}} + \frac{\tau}{(\varepsilon+\tau^{1-\alpha})^2} + \varepsilon\right), \quad h^2 = o(\varepsilon+\tau^{1-\alpha}), \quad \tau = o((\varepsilon+\tau^{1-\alpha})^2),$$

где ε *– малый параметр.*

Очевидно, что скорость сходимости будет определяться наилучшим образом, если

$$\frac{h^2}{\varepsilon+\tau^{1-\alpha}} + \frac{\tau}{(\varepsilon+\tau^{1-\alpha})^2} = \varepsilon.$$

Пусть $\varepsilon = \tau^\gamma$, тогда из последнего получаем

$$h^2(\tau^\gamma + \tau^{1-\alpha}) + \tau = \tau^\gamma(\tau^\gamma + \tau^{1-\alpha})^2$$

или

$$\tau \le \tau^\gamma(\tau^\gamma + \tau^{1-\alpha})^2.$$

Следовательно,

$$\min\{\gamma, 1-\alpha\} = \frac{1-\gamma}{2},$$

откуда получаем, что

$$\varepsilon = \begin{cases} \tau^{\frac{1}{3}}, & 0 < \alpha \le \frac{2}{3}, \\ \tau^{2\alpha-1}, & \frac{2}{3} < \alpha < 1. \end{cases} \tag{5.14}$$

Тогда справедливо

Следствие. Если ε определяется из условия (5.14), тогда решение разностной задачи (2.4) равномерно сходится к решению дифференциальной задачи (1.1)–(1.3) со скоростью

$$O\left(\frac{h^2}{\tau^{\frac{1}{3}}} + \tau^{\frac{1}{3}}\right), \quad \text{если} \quad 0 < \alpha \le \frac{2}{3},$$

и

$$O\left(\frac{h^2}{\tau^{1-\alpha}} + \tau^{2\alpha-1}\right), \quad \text{если} \quad \frac{2}{3} < \alpha < 1.$$

При $\alpha \to 1$ получаем, что решение разностной задачи (2.4) равномерно сходится к решению дифференциальной задачи (1.1)–(1.3) со скоростью $O(h^2+\tau)$.

6. АЛГОРИТМ ЧИСЛЕННОГО РЕШЕНИЯ

Для численного решения поставленной задачи (1.1)–(1.3) выпишем расчетные формулы ($0 \le x_k \le l_k$, $k = 1, 2$, $p = 2$):

$$\varepsilon\frac{\partial u}{\partial t} + \partial_{0t}^{\alpha}u = \frac{\partial}{\partial x_1}\left(\Theta_1(x_1,x_2,t)\frac{\partial u}{\partial x_1}\right) + \frac{\partial}{\partial x_2}\left(\Theta_2(x_1,x_2,t)\frac{\partial u}{\partial x_2}\right) + r_1(x_1,x_2,t)\frac{\partial u}{\partial x_1} + \\ + r_2(x_1,x_2,t)\frac{\partial u}{\partial x_2} - q_1(x_1,x_2,t)u(x_1,x_2,t) - q_2(x_1,x_2,t)u(x_1,x_2,t) + f(x_1,x_2,t), \tag{6.1}$$

$$\begin{cases} u(0,x_2,t) = \mu_{11}(x_2,t), \quad u(l_1,x_2,t) = \mu_{12}(x_2,t), \\ u(x_1,0,t) = \mu_{21}(x_1,t), \quad u(x_1,l_2,t) = \mu_{22}(x_1,t), \end{cases} \tag{6.2}$$

$$u(x_1,x_2,0) = u_0(x_1,x_2). \tag{6.3}$$

Рассмотрим сетку $x_k^{(i_k)} = i_k h_k$, $k = 1, 2$, $t_j = j\tau$, где $i_k = 0, 1, \ldots, N_k$, $h_k = l_k/N_k$, $j = 0, 1, \ldots, m$, $\tau = T/m$. Вводим один дробный шаг $t_{j+\frac{1}{2}} = t_j + 0.5\tau$. Обозначим через $y_{i_1,i_2}^{j+\frac{k}{p}} = y^{j+\frac{k}{p}} = y(i_1h_1, i_2h_2, (j+0.5k)\tau)$, $k = 1, 2$, сеточную функцию.

Напишем локально-одномерную схему

$$\varepsilon\frac{y^{j+\frac{1}{2}} - y^{j}}{\tau} + \frac{1}{2}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{2j+1}\left(t_{j+\frac{2-s}{2}}^{1-\alpha} - t_{j+\frac{1-s}{2}}^{1-\alpha}\right)y_{\bar{t}}^{\frac{s}{2}} = \tilde{\Lambda}_1 y^{j+\frac{1}{2}} + \varphi_1,$$

$$\varepsilon\frac{y^{j+1} - y^{j+\frac{1}{2}}}{\tau} + \frac{1}{2}\frac{1}{\Gamma(2-\alpha)}\sum_{s=1}^{2j+2}\left(t_{j+\frac{3-s}{2}}^{1-\alpha} - t_{j+\frac{2-s}{2}}^{1-\alpha}\right)y_{\bar{t}}^{\frac{s}{2}} = \tilde{\Lambda}_2 y^{j+1} + \varphi_2, \tag{6.4}$$

$$y_{0,i_2}^{j+\frac{1}{2}} = \mu_{11}\left(i_2h_2, t_{j+\frac{1}{2}}\right), \quad y_{N_1,i_2}^{j+\frac{1}{2}} = \mu_{12}\left(i_2h_2, t_{j+\frac{1}{2}}\right),$$

$$y_{i_1,0}^{j+1} = \mu_{21}\left(i_1h_1, t_{j+1}\right), \quad y_{i_1,N_2}^{j+1} = \mu_{22}\left(i_1h_1, t_{j+1}\right), \tag{6.5}$$

$$y_{i_1,i_2}^{0} = u_0\left(i_1h_1, i_2, h_2\right),$$

$$\tilde{\Lambda}_k y^{j+\frac{k}{p}} = \varkappa_k\left(a_k y_{\bar{x}_k}^{j+\frac{k}{p}}\right)_{x_k} + b_k^{+}a_k^{(+1)}y_{x_k}^{j+\frac{k}{p}} + b_k^{-}a_k y_{\bar{x}_k}^{j+\frac{k}{p}} - d_k y^{j+\frac{k}{p}}, \quad k = 1, 2, \tag{6.6}$$

$$\varphi_k = \frac{1}{2}f(x_1,x_2,t_{j+0.5k}) \quad \text{или} \quad \varphi_1 = 0, \quad \varphi_2 = f(x_1,x_2,t_{j+1}).$$

Приведем расчетные формулы для решения задачи (6.4)–(6.6).

На первом этапе находим решение $y_{i_1,i_2}^{j+\frac{1}{2}}$. Для этого при каждом значении $i_2 = \overline{1, N_2 - 1}$ решается следующая задача:

$$A_{1(i_1,i_2)}y_{i_1-1,i_2}^{j+\frac{1}{2}} - C_{1(i_1,i_2)}y_{i_1,i_2}^{j+\frac{1}{2}} + B_{1(i_1,i_2)}y_{i_1+1,i_2}^{j+\frac{1}{2}} = -F_{1(i_1,i_2)}^{j+\frac{1}{2}}, \quad 0 < i_1 < N_1,$$

$$y_{0,i_2}^{j+\frac{1}{2}} = \mu_{11}\left(i_2h_2, t_{j+\frac{1}{2}}\right), \quad y_{N_1,i_2}^{j+\frac{1}{2}} = \mu_{12}\left(i_2h_2, t_{j+\frac{1}{2}}\right), \tag{6.7}$$

где

$$A_{1(i_1,i_2)} = \frac{(\varkappa_1)_{i_1,i_2}(a_1)_{i_1,i_2}}{h_1^2} - \frac{(b_1^{-})_{i_1,i_2}(a_1)_{i_1,i_2}}{h_1},$$

$$B_{1(i_1,i_2)} = \frac{(\varkappa_1)_{i_1,i_2}(a_1)_{i_1+1,i_2}}{h_1^2} + \frac{(b_1^{+})_{i_1,i_2}(a_1)_{i_1+1,i_2}}{h_1},$$

$$C_{1(i_1,i_2)} = A_{1(i_1,i_2)} + B_{1(i_1,i_2)} + \frac{\varepsilon}{\tau} + \frac{1}{2\tau^{\alpha}\Gamma(2-\alpha)} + \frac{1}{p}d_{1(i_1,i_2)},$$

Таблица 1. Результаты численных экпериментов при $0 < \alpha \le \frac{2}{3}$

α	h	$\|z\|_{C(\overline{w}_{h\tau})}$	ПС в $\|\cdot\|_{C(\overline{w}_{h\tau})}$	$O(\tau^{1/3})$
0.1	1/2	0.029143444		0.33
	1/4	0.025357432	0.1004	
	1/8	0.021421778	0.1217	
	1/16	0.018023784	0.1246	
	1/32	0.014033112	0.1805	
	1/64	0.010004234	0.2441	
	1/128	0.006583291	0.3019	
0.2	1/2	0.029013789		0.33
	1/4	0.025301206	0.0988	
	1/8	0.021371432	0.1218	
	1/16	0.017972508	0.1249	
	1/32	0.013990212	0.1807	
	1/64	0.009976202	0.2439	
	1/128	0.006567871	0.3015	
0.3	1/2	0.028951506		0.33
	1/4	0.025219101	0.0996	
	1/8	0.021285824	0.1223	
	1/16	0.017871468	0.1261	
	1/32	0.013892420	0.1817	
	1/64	0.009902294	0.2442	
	1/128	0.006521075	0.3013	
0.4	1/2	0.028874976		0.33
	1/4	0.025100031	0.1011	
	1/8	0.021141306	0.1238	
	1/16	0.017674549	0.1292	
	1/32	0.013673165	0.1852	
	1/64	0.009711621	0.2468	
	1/128	0.006381999	0.3029	
0.5	1/2	0.028782052		0.33
	1/4	0.024929275	0.1037	
	1/8	0.020901055	0.1271	
	1/16	0.017300340	0.1364	
	1/32	0.013200657	0.1951	
	1/64	0.009253975	0.2562	
	1/128	0.005996327	0.3130	
0.6	1/2	0.028670815		0.33
	1/4	0.024688137	0.1079	
	1/8	0.020511074	0.1337	
	1/16	0.016619461	0.1518	
	1/32	0.012273815	0.2186	
	1/64	0.008259210	0.2858	
	1/128	0.005092653	0.3488	

Таблица 2. Результаты численных экпериментов при $\frac{2}{3} < \alpha < 1$

α	h	$\|z\|_{C(\overline{w}_{h\tau})}$	ПС в $\|\cdot\|_{C(\overline{w}_{h\tau})}$	$O(\tau^{2\alpha-1})$
0.7	1/2	0.028539879		0.4
	1/4	0.024131209	0.1210	
	1/8	0.021062646	0.0981	
	1/16	0.016421549	0.1795	
	1/32	0.011107562	0.2820	
	1/64	0.006532123	0.3830	
0.8	1/2	0.028445640		0.6
	1/4	0.024093465	0.1198	
	1/8	0.020990772	0.0994	
	1/16	0.015942124	0.1985	
	1/32	0.009999178	0.3365	
	1/64	0.005163160	0.4768	
0.9	1/2	0.028288873		0.8
	1/4	0.023793274	0.1248	
	1/8	0.019898778	0.1289	
	1/16	0.013266177	0.2925	
	1/32	0.006775084	0.4847	
	1/64	0.002731669	0.6552	
0.99	1/2	0.028105342		0.98
	1/4	0.023231860	0.1374	
	1/8	0.017992042	0.1844	
	1/16	0.009925780	0.4291	
	1/32	0.004029305	0.6503	
	1/64	0.001270309	0.8327	

$$F^{j+\frac{1}{2}}_{1(i_1,i_2)} = \frac{\varepsilon}{\tau} y^{j}_{i_1,i_2} + \frac{1}{2\tau^{\alpha}\Gamma(2-\alpha)} y^{j}_{i_1,i_2} + \frac{1}{\tau\Gamma(2-\alpha)} \sum_{s=1}^{2j} \left(t^{1-\alpha}_{j+\frac{2-s}{2}} - t^{1-\alpha}_{j+\frac{1-s}{2}} \right) y^{\frac{s}{2}}_{t} + \varphi_{1(i_1,i_2)}.$$

На втором этапе находим решение $y^{j+1}_{i_1,i_2}$. Для этого, как и в первом случае, при каждом значении $i_1 = \overline{1, N_1 - 1}$ решается задача

$$\begin{gathered}
A_{2(i_1,i_2)} y^{j+1}_{i_1,i_2-1} - C_{2(i_1,i_2)} y^{j+1}_{i_1,i_2} + B_{2(i_1,i_2)} y^{j+1}_{i_1,i_2+1} = -F^{j+1}_{2(i_1,i_2)}, \quad 0 < i_2 < N_2, \\
y^{j+1}_{i_1,0} = \mu_{21}(i_1 h_1, t_{j+1}), \quad y^{j+1}_{i_1,N_2} = \mu_{22}(i_1 h_1, t_{j+1}), \\
A_{2(i_1,i_2)} = \frac{(\varkappa_2)_{i_1,i_2}(a_2)_{i_1,i_2}}{h_2^2} - \frac{(b_2^-)_{i_1,i_2}(a_2)_{i_1,i_2}}{h_2}, \\
B_{2(i_1,i_2)} = \frac{(\varkappa_2)_{i_1,i_2}(a_2)_{i_1,i_2+1}}{h_2^2} + \frac{(b_2^+)_{i_1,i_2}(a_2)_{i_1,i_2+1}}{h_2}, \\
C_{2(i_1,i_2)} = A_{2(i_1,i_2)} + B_{2(i_1,i_2)} + \frac{\varepsilon}{\tau} + \frac{1}{2\tau^{\alpha}\Gamma(2-\alpha)} + \frac{1}{p} d_{2(i_1,i_2)}, \\
F^{j+1}_{2(i_1,i_2)} = \frac{\varepsilon}{\tau} y^{j+\frac{1}{2}}_{i_1,i_2} + \frac{1}{2\tau^{\alpha}\Gamma(2-\alpha)} y^{j+\frac{1}{2}}_{i_1,i_2} + \frac{1}{\tau\Gamma(2-\alpha)} \sum_{s=1}^{2j+1} \left(t^{1-\alpha}_{j+\frac{3-s}{2}} - t^{1-\alpha}_{j+\frac{2-s}{2}} \right) y^{\frac{s}{2}}_{t} + \varphi_{2(i_1,i_2)}.
\end{gathered} \tag{6.8}$$

Каждая из задач (6.7), (6.8) решается методом прогонки (см. [32]).

7. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

Коэффициенты уравнения и граничных условий задачи (1.1)–(1.3) подбираются таким образом, чтобы точным решением задачи была функция $u(x,t) = t^3(x_1^4 - l_1x_1^3)(x_2^4 - l_2x_2^3)$.

Ниже в табл. 1 и 2 при уменьшении размера сетки приведены максимальное значение погрешности ($z = y - u$) и порядок сходимости в норме $\|\cdot\|_{C(\overline{w}_{h\tau})}$, где $\|y\|_{C(\overline{w}_{h\tau})} = \max\limits_{(x_i,t_j)\in\overline{w}_{h\tau}} |y|$ при $0 < \alpha < 1$, когда $h^2 = \tau$. Погрешность уменьшается в соответствии с порядком аппроксимации

$$O\left(\frac{h^2}{\tau^{\frac{1}{3}}} + \tau^{\frac{1}{3}}\right), \quad \text{если} \quad 0 < \alpha \le \frac{2}{3},$$

и

$$O\left(\frac{h^2}{\tau^{1-\alpha}} + \tau^{2\alpha-1}\right), \quad \text{если} \quad \frac{2}{3} < \alpha < 1,$$

где $h = \max\limits_{1\le\alpha\le p} h_\alpha$,

$$\varepsilon = \begin{cases} \tau^{\frac{1}{3}}, & 0 < \alpha \le \frac{2}{3}, \\ \tau^{2\alpha-1}, & \frac{2}{3} < \alpha < 1. \end{cases}$$

Таким образом, проведены численные расчеты тестовых примеров на ЭВМ, иллюстрирующие полученные в работе теоретические выкладки.

СПИСОК ЛИТЕРАТУРЫ

1. *Динариев О.Ю.* Фильтрация в трещиноватой среде с фрактальной геометрией трещин // Изв. РАН. Механ. жидкости и газа. 1990. № 5. С. 66–70.
2. *Кобелев В.Л., Кобелев Я.Л., Романов Е.П.* Недебаевская релаксация и диффузия во фрактальном пространстве // Докл. РАН. 1998. Т. 361. № 6. С. 755–758.
3. *Кобелев В.Л., Кобелев Я.Л., Романов Е.П.* Автоволновые процессы при нелинейной фрактальной диффузии // Докл. РАН. 1999. Т. 369. № 3. С. 332–333.
4. *Кочубей А.Ю.* Диффузия дробного порядка // Дифференц. ур-ния. 1990. Т. 26. С. 660–670.
5. *Лафишева М.М., Шхануков-Лафишев М.Х.* Локально-одномерная разностная схема для уравнения диффузии дробного порядка // Ж. вычисл. матем. и матем. физ. 2008. Т. 48. № 10. С. 1878–1887.
6. *Бештоков М.Х.* К краевым задачам для вырождающихся псевдопараболических уравнений с дробной производной Герасимова–Капуто // Изв. вузов. Математика. 2018. № 10. С. 3–16.
7. *Бештоков М.Х.* Численное исследование начально-краевых задач для уравнения соболевского типа с дробной по времени производной // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 2. С. 185–202.
8. *Бештоков М.Х.* Краевые задачи для псевдопараболического уравнения с дробной производной Капуто // Дифференц. уравнения. Т. 55. № 7. 2019. С. 919–928.
9. *Бештоков М.Х., Водахова В.А.* Нелокальные краевые задачи для уравнения конвекции–диффузии дробного порядка // Вестн. Удмуртского ун-та. Математика. Механика. Компьютерные науки. 2019. Т. 29. Вып. 4. С. 459–482.
10. *Бештоков М.Х., Эржибова Ф.А.* К краевым задачам для интегро-дифференциальных уравнений дробного порядка // Матем. труды. 2020. Т. 23 № 1. С. 16–36.
11. *Мальшаков А.В.* Уравнения гидродинамики для пористых сред со структурой порового пространства, обладающей фрактальной геометрией // ИФЖ. 1992. Т. 62. № 3. С. 405–410.
12. *Нахушев А.М.* Дробное исчисление и его применение. М.: Физматгиз, 2003, 272 с.
13. *Учайкин В.В.* Метод дробных производных. Ульяновск: “Артишок”, 2008. 512 с.
14. *Шогенов В.Х., Кумыкова С.К., Шхануков-Лафишев М.Х.* Обобщенное уравнение переноса и дробные производные // Докл. АМАН. 1996. Т. 2. № 1. С. 43–45.
15. *Oldham K.B., Spanier J.* The Fractional Calculus. New York–London: Acad. Press, 1974. 234 p.
16. *Podlubny I.* Fractional Differential Equations. San-Diego: Acad. Press, San Diego-Boston-New York-London-Sydney-Tokyo-Toronto, 1999. 368 p.
17. *Самко С.Г., Килбас А.А., Маричев О.И.* Интегралы и производные дробного порядка и некоторые их приложения. Минск: Наука и техника, 1987. 688 с.

18. *Nigmatulin R.R.* The realization of generalized transfer equation in a medium with fractal geometry // Phys. Status Solidi. B. 1986. V. 133. P. 425–430.
19. *Головизнин В.М., Кисилев В.П., Короткин И.А., Юрков Ю.П.* Некоторые особенности вычислительных алгоритмов для уравнений дробной диффузии // Препринт IBRAE-2002-01. М: ИБРАЭ РАН, 2002.
20. *Головизнин В.М., Кисилев В.П., Короткин И.А.* Численные методы решения уравнения диффузии с дробной производной в одномерном случае // Препринт IBRAE-2002-01. М: ИБРАЭ РАН, 2002.
21. *Федер Е.* Фракталы. М.: Мир, 1991. 260 с.
22. *Lovejoy S.* Area-perimeter relation for rain and cloud areas // Science. 1982. V. 216. P. 185–187.
23. *Шогенов В.Х., Ахубеков А.А., Ахубеков Р.А.* Метод дробного дифференцирования в теории броуновского движения // Изв. вузов. Северо-Кавказский регион. 2004. № 1. С. 46–50.
24. *Самарский А.А.* Об одном экономичном разностном методе решения многомерного параболического уравнения в произвольной области // Ж. вычисл. матем. и матем. физ. Т. 2. № 5. 1962. С. 787–811.
25. *Баззаев А.К., Шхануков-Лафишев М.Х.* Локально-одномерные схемы для уравнения диффузии с дробной производной по времени в области произвольной формы // Ж. вычисл. матем. и матем. физ. 2016. Т. 56. № 1. С. 113–123.
26. *Ашабоков Б.А., Бештокова З.В., Шхануков-Лафишев М.Х.* Локально-одномерная разностная схема для уравнения переноса примесей дробного порядка// Ж. вычисл. матем. и матем. физ. 2017. Т. 57. № 9. С. 1517–1529.
27. *Баззаев А.К., Шхануков-Лафишев М.Х.* Локально-одномерная схема для уравнения диффузии дробного порядка с краевыми условиями III рода // Ж. вычисл. матем. и матем. физ. 2010. Т. 50. № 7. С. 1200–1208.
28. *Вишик М.И., Люстерник Л.А.* Регулярное вырождение и пограничный слой для линейных дифференциальных уравнений с малым параметром // Успехи матем. наук. 1967. Т. 12. № 5. С. 3–122.
29. *Годунов С.К., Рябенький В.С.* Разностные схемы. М.: Наука, 1977. 439 с.
30. *Alikhanov A.A.* Boundary value problems for the diffusion equation of the variable order in differential and difference settings // Appl. Math. 2012. V. 219. P. 3938–3946.
31. *Самарский А.А.* Теория разностных схем. М.: Наука, 1983. 616 с.
32. *Самарский А.А., Гулин А.В.* Устойчивость разностных схем. М.: Наука, 1973. 415 с.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1101–1112

МАТЕМАТИЧЕСКАЯ ФИЗИКА

УДК 51-71

МЕТОД МОДЕЛИРОВАНИЯ ПАРАМЕТРОВ ИОНОСФЕРЫ И ОБНАРУЖЕНИЯ ИОНОСФЕРНЫХ ВОЗМУЩЕНИЙ[1)]

© 2021 г. О. В. Мандрикова[1,*], Ю. А. Полозов[1,], Н. В. Фетисова[1,***]**

[1] *684034 Камчатский край, п. Паратунка, ул. Мирная, 7, Институт космофизических исследований и распространения радиоволн ДВО РАН, Россия*

**e-mail: oksanam1@mail.ru*

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

**** e-mail: nv.glushkova@yandex.ru*

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Предложен автоматизированный метод анализа параметров ионосферы и обнаружения ионосферных аномалий. Основу метода составляет разработанная авторами обобщенная многокомпонентная модель параметров ионосферы. Идентификация модели основана на комплексном подходе, объединяющем методы вейвлет-преобразования с моделями авторегрессии – проинтегрированного скользящего среднего. Приведены оценки эффективности метода, описаны операции обнаружения ионосферных аномалий и оценки их параметров. На примере обработки параметров ионосферы (критической частоты ионосферы foF2) района Камчатки показана возможность применения метода в режиме оперативного анализа данных (по мере поступления данных в систему обработки). На основе метода обнаружены короткопериодные аномальные изменения, предшествующие магнитным бурями и характеризующие возникновение колебательных процессов в ионосфере на фоне повышенной солнечной активности. Метод реализован в системе комплексного анализа геофизических данных Aurora. Библ. 27. Фиг. 4.

Ключевые слова: модель временного ряда, вейвлет-преобразование, авторегрессионные модели, параметры ионосферы, аномалии.

DOI: 10.31857/S0044466921070139

1. ВВЕДЕНИЕ

Работа направлена на создание автоматизированных методов анализа временных рядов параметров ионосферы и выделения ионосферных аномалий. Атмосфера Земли на высотах примерно от 60 до 1000 км называется ионосферой. Данная область сильно влияет на радиоволны и условия их распространения (см. [1]). На структуру ионосферы, а также на регулярные суточные и сезонные вариации ее параметров существенное влияние оказывают солнечная и геомагнитная активности, географические координаты (выделяют полярные и авроральные, среднеширотные и экваториальные области) (см. [1]–[4]). Вариации различных факторов в околоземном космическом пространстве (солнечные вспышечные события, колебания параметров солнечного ветра, магнитные бури и суббури) оказывают влияние на ионосферу (см. [1]–[7]). Реакция ионосферы проявляется в виде формирования ионосферных неоднородностей, причиной которых являются энергетические частицы и излучение, возникающее во время солнечных вспышек и (или) магнитосферных возмущений (см. [1]–[4]). Неоднородности (возмущения) в ионосфере представляются в виде областей с электронной концентрацией, отличной от некоторого характерного (спокойного) уровня (см. [1]–[4]). Резкое существенное изменение (повышение/понижение) электронной концентрации в возмущенные периоды называется ионосферной бурей (см. [1]–[4]). Динамика параметров ионосферы в периоды ионосферных бурь, как правило, может характеризоваться различными фазами – отрицательной (понижение электронной концентрации) и положительной (повышение электронной концентрации) (см. [1]–[3]). Простран-

[1)]Работа выполнена в рамках ГЗ по теме “Динамика физических процессов в активных зонах ближнего космоса и геосфер” (2018-2020) № гос. регистрации АААА-А17-117080110043-4.

ственно-временное распределение ионосферных бурь является сложным и зависит от многих факторов – географического положения, уровня солнечной активности, процессов в магнитосфере и др. (см. [1]–[4]). В настоящее время механизмы возникновения и протекания ионосферных бурь достаточно изучены (см., например, [1], [2]), но их своевременное прогнозирование не реализовано (см. [1]–[9]).

Исследование ионосферы и оценка ее состояния основаны на анализе регистрируемых ионосферных параметров (критическая частота F2-слоя ионосферы (foF2), полное электронное содержание и другие параметры). Предметами исследования в работе являются анализ вариаций критической частоты ионосферы (foF2) и обнаружение аномальных изменений. Критическая частота foF2 – параметр ионосферы, характеризующий электронную концентрацию F2-слоя ионосферы, и является максимальной частотой радиоволны, отражающейся от слоя ионосферы при вертикальном падении (см. [4]). На основе метода вертикального радиозондирования выполняется регистрация данных foF2 с использованием цифрового ионозонда. После предварительной обработки данные foF2 представляются в виде временных рядов. Полученные временные ряды foF2 имеют сложную нестационарную структуру, которая формируется при воздействии солнечной активности и других факторов (координаты станций регистрации, геомагнитная активность и др.). Возникновение ионосферных бурь сопровождается аномальными изменениями различной длительности, амплитуды и вида (всплески, пики) в получаемых временных рядах foF2. Данные изменения несут информацию о динамике ионосферной бури и ее интенсивности. Аномальные изменения в ионосфере оказывают негативное влияние на работу сложных технических систем (телекоммуникационные, навигационные и радиолокационные системы (см. [8])). Ионосферные аномалии могут вызывать отказы и сбои оборудования, поэтому их своевременное обнаружение является важной прикладной задачей.

Задачи анализа ионосферных данных решаются разными научными коллективами (см. [1], [4], [10]–[19]). В работах используются подходы, основанные на нейронных сетях (см. [1], [11]–[14], [19]), традиционный метод скользящей медианы (см. [4], [15]), эмпирические модели (см. [10], [18]), физические модели (см. [16], [17]) и др. Широко используемые классические методы, основанные на процедуре сглаживания, являются недостаточно эффективными и могут вести к потере части важной информации (см. [10], [15], [20], [21]). Качество работы современных методов (например, подходы на основе нейронных сетей (см. [1], [11]–[14])), их точность и эффективность зависят от используемых оперативных данных, получаемых на основе аппаратного мониторинга (например, временные ряды солнечной и магнитной активности, межпланетного магнитного поля и др.). Качество данных, получаемых в возмущенные периоды, существенно снижается из-за повышения уровня помех и наличия пропусков в измерениях. Область применения физических моделей ограничена отсутствием информации о динамике ионосферных процессов (см. [16], [17]). Разработанный авторами гибридный подход, применяемый в работе, базируется на совместном использовании аппарата вейвлет-преобразования (см. [22]) и методов авторегрессии – проинтегрированного скользящего среднего (АРПСС, см. [23]). Впервые данный подход был предложен в [21], [24] и применялся для анализа ионосферных данных. В основе подхода лежит обобщенная многокомпонентная модель параметров ионосферы (ОМКМ) (см. [20], [25], [26]). Важным преимуществом предлагаемого метода является возможность его полной автоматизации. В работе представлено описание процедур выделения ионосферных аномалий и оценки их параметров, а также приведены оценки эффективности метода. Предлагаемый метод был реализован в системе комплексного анализа геофизических данных Aurora (http://lsaoperanalysis.ikir.ru/lsaoperanalysis.html).

2. ОПИСАНИЕ МЕТОДА

2.1. Обобщенная многокомпонентная модель параметров ионосферы

В соответствии с [20], [21] представим временной ряд параметров ионосферы в виде

$$f(t) = A^{\text{РЕГ}}(t) + U(t) + e(t) = \sum_{\mu=\overline{1,T}} \sum_{j} G_j^{\mu} \alpha_j^{\mu}(t) + \sum_{i,\eta} \beta_{i,\eta}^{\text{возм}}(t) + e(t), \tag{1}$$

где $A^{\text{РЕГ}}(t) = \sum_{\mu=\overline{1,T}} \sum_j G_j^{\mu} \alpha_j^{\mu}(t)$ ($\mu = \overline{1,T}$ – номер компоненты) – регулярная компонента модели; составляющие $\alpha_j^{\mu}(t)$ имеют разномасштабную структуру (определяются локальными факторами

и включают сезонные вариации параметров, суточные колебания и др.); $U(t) = \sum_{i,\eta} \beta_{i,\eta}^{\text{возм}}(t)$ – аномальная компонента модели, описывающая динамику ионосферных параметров в возмущенные периоды (колебательные процессы в периоды повышенной солнечной активности, магнитосферных возмущений и др.), в периоды спокойной ионосферы предполагается, что компонента $U(t) = 0$; $e(t)$ – шумовая составляющая.

На основе совмещения операций кратномасштабного анализа (КМА) и методов АРПСС получим параметрическое представление компоненты $A^{\text{РЕГ}}(t)$ модели (1) (см. [20]):

$$A^{\text{РЕГ}}(t) = \sum_{\mu=\overline{1,T}} \sum_{k=\overline{1,N_{j^{\text{per}}}^{\mu}}} s_{j^{\text{рег}},k}^{\mu} b_{j^{\text{рег}},k}^{\mu}(t) + e(t), \tag{2}$$

где $s_{j^{\text{per}},k}^{\mu} = \sum_{l=1}^{p_{j^{\text{per}}}^{\mu}} \gamma_{j^{\text{per}},l}^{\mu} \omega_{j^{\text{per}},k-l}^{\mu} - \sum_{n=1}^{h_{j^{\text{per}}}^{\mu}} \theta_{j^{\text{per}},n}^{\mu} a_{j^{\text{per}},k-n}^{\mu}$ – оценочное значение регулярной μ-й составляющей; $p_{j^{\text{per}}}^{\mu}$, $\gamma_{j^{\text{per}},l}^{\mu}$ – порядок и параметры авторегрессии μ-й составляющей; $\omega_{j^{\text{per}},k}^{\mu} = \nabla^{\nu^{\mu}} \delta_{j^{\text{per}},k}^{\mu}$; ν^{μ} – порядок разности μ-й составляющей; $\delta_{-m^{\text{per}},k}^{1} = c_{-m^{\text{per}},k}$, $\delta_{j^{\text{per}},k}^{\mu} = d_{j^{\text{per}},k}$, $\mu = \overline{2,T}$; T – количество моделируемых составляющих; $c_{-m^{\text{per}},k} = \left\langle f, \phi_{-m^{\text{per}},k} \right\rangle$ – вейвлет-коэффициенты сглаженной составляющей КМА масштаба m^{per}; $d_{j^{\text{per}},k} = \left\langle f, \Psi_{j^{\text{per}},k} \right\rangle$ – вейвлет-коэффициенты детализирующих составляющих КМА масштабов j^{per}; $\phi_{-m^{\text{per}},k}(t)$ – масштабирующая функция; $\Psi_{j^{\text{per}},k}(t)$ – вейвлет-функция; $h_{j^{\text{per}}}^{\mu}$, $\theta_{j^{\text{per}},n}^{\mu}$ – порядок и параметры скользящего среднего μ-й составляющей; $a_{j^{\text{per}},k}^{\mu}$ – остаточные ошибки модели μ-й составляющей; $b_{-m^{\text{per}},k}^{1} = \phi_{-m^{\text{per}},k}$; $b_{j^{\text{per}},k}^{\mu} = \Psi_{j^{\text{per}},k}$, $\mu = \overline{2,T}$; $N_{j^{\text{per}}}^{\mu}$ – длина μ-й составляющей.

Аномальная компонента $U(t)$ модели (1) включает составляющие $\beta_{i,\eta}^{\text{возм}}(t)$, которые определяют нестационарные процессы в ионосфере, и в вейвлет-пространстве могут быть представлены в виде

$$\sum_{i,\eta} \beta_{i,\eta}^{\text{возм}}(t) = \sum_{\eta,n} P_{1,\eta}(d_{\eta,n}) \Psi_{\eta,n}(t) + \sum_{\eta,n} P_{2,\eta}(d_{\eta,n}) \Psi_{\eta,n}(t) + \sum_{\eta,n} P_{3,\eta}(d_{\eta,n}) \Psi_{\eta,n}(t),$$

$$P_{1,\eta}(x) = \begin{cases} 0, & \text{если} \quad |x| \le T_{1,\eta} \quad \text{или} \quad |x| > T_{2,\eta}, \\ x, & \text{если} \quad T_{1,\eta} < |x| \le T_{2,\eta}, \end{cases}$$

$$P_{2,\eta}(x) = \begin{cases} 0, & \text{если} \quad |x| \le T_{2,\eta} \quad \text{или} \quad |x| > T_{3,\eta}, \\ x, & \text{если} \quad T_{2,\eta} < |x| \le T_{3,\eta}, \end{cases} \tag{3}$$

$$P_{3,\eta}(x) = \begin{cases} 0, & \text{если} \quad |x| \le T_{3,\eta}, \\ x, & \text{если} \quad |x| > T_{3,\eta}, \end{cases}$$

где $d_{\eta,n} = \left\langle U, \Psi_{\eta,n} \right\rangle$ – вейвлет-коэффициенты на масштабе η; $\{\Psi_{\eta,n}\}_{\eta,n \in Z}$ – вейвлет-базис. Амплитуда вейвлет-коэффициентов $|d_{\eta,n}|$, следуя [20], [25], определена в качестве меры интенсивности для ионосферных аномалий на масштабе η (соотношение (3)). Таким образом, для определения *аномалии малой интенсивности* (класс 1) используются пороги $T_{1,\eta}$, для определения *аномалии умеренной интенсивности* (класс 2) – пороги $T_{2,\eta}$, для определения *аномалии высокой интенсивности* (класс 3) – пороги $T_{3,\eta}$.

На основе представлений (2) и (3) получаем *обобщенную многокомпонентную модель временного ряда параметров ионосферы*:

$$f(t) = A^{\text{РЕГ}}(t) + U(t) + e(t) = \sum_{\mu=\overline{1,T}} \sum_{k=\overline{1,N_{j^{\text{per}}}^{\mu}}} s_{j^{\text{per}},k}^{\mu} b_{j^{\text{per}},k}^{\mu}(t) + \sum_{i=\overline{1,3}} \sum_{\eta,n} P_{i,\eta}(d_{\eta,n}) \Psi_{\eta,n}(t) + e(t).$$

2.2. Операции обнаружения и оценки параметров ионосферных аномалий

1. Очевидно, что вследствие аномальных изменений во временном ходе параметров ионосферы возрастут остаточные ошибки компоненты $A^{\text{РЕГ}}(t)$ модели (см. соотношение (2)). Следовательно, обнаружение аномалий может быть основано на проверке условия (см. [20], [21])

$$\varepsilon^{\mu}_{j^{\text{per}}} = \sum_{q=1}^{Q_{\mu}} |a^{\mu}_{j^{\text{per}},k+q}| > H_{\mu,j^{\text{per}}}, \tag{4}$$

где $q \geq 1$ – шаг упреждения данных, Q_{μ} – длина упреждения данных на основе μ-й составляющей модели, $a^{\mu}_{j^{\text{per}},k+q} = s^{\mu,\text{факт}}_{j^{\text{per}},k+q} - s^{\mu,\text{модель}}_{j^{\text{per}},k+q}$, $s^{\mu,\text{модель}}_{j^{\text{per}},k} = \sum_{l=1}^{p^{\mu}_{j^{\text{per}}}} \gamma^{\mu}_{j^{\text{per}},l}\omega^{\mu}_{j^{\text{per}},k+q-l} - \sum_{n=1}^{h^{\mu}_{j^{\text{per}}}} \theta^{\mu}_{j^{\text{per}},n}a^{\mu}_{j^{\text{per}},k+q-n}$, $H_{\mu,j^{\text{per}}}$ – пороговое значение μ-й составляющей. Пороговое значение $H_{\mu,j^{\text{per}}}$, следуя [23], определено на основе оценки дисперсии остаточных ошибок модели с учетом вероятностных пределов (см. [20]):

$$H_{\mu,j^{\text{per}}}(Q_{\mu}) = u_{\xi/2}\left\{1 + \sum_{q=1}^{Q_{\mu}-1}(\psi^{\mu}_{j^{\text{per}},q})^2\right\}^{1/2}\sigma_{a^{\mu}_{j^{\text{per}}}}, \tag{5}$$

где $u_{\xi/2}$ – квантиль уровня $(1-\xi/2)$ стандартного нормального распределения; $\sigma^2_{a^{\mu}_{j^{\text{per}}}}$ – дисперсия остаточных ошибок модели μ-й составляющей; $\psi^{\mu}_{j^{\text{per}},q}$ – весовые коэффициенты модели μ-й составляющей, которые определяются из соотношения

$$\left(1 - \varphi^{\mu}_{j^{\text{per}},1}B - \varphi^{\mu}_{j^{\text{per}},2}B^2 - \ldots - \varphi^{\mu}_{j^{\text{per}},p^{\mu}_{j^{\text{per}}}+\nu^{\mu}}B^{p^{\mu}_{j^{\text{per}}}+\nu^{\mu}}\right)(1 + \psi^{\mu}_{j^{\text{per}},1}B + \psi^{\mu}_{j^{\text{per}},2}B^2 + \ldots) =$$
$$= \left(1 - \theta^{\mu}_{j^{\text{per}},1}B - \theta^{\mu}_{j^{\text{per}},2}B^2 - \ldots - \theta^{\mu}_{j^{\text{per}},h^{\mu}_{j^{\text{per}}}}B^{h^{\mu}_{j^{\text{per}}}}\right),$$

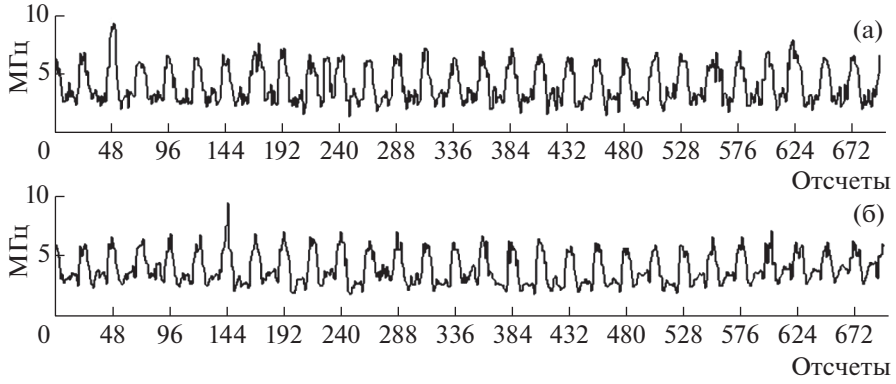
где $\varphi^{\mu}_{j^{\text{per}},p^{\mu}_{j^{\text{per}}}+\nu^{\mu}}$ – обобщенный оператор авторегрессии: $\varphi^{\mu}_{j^{\text{per}}} = \gamma^{\mu}_{j^{\text{per}}}(B)(1-B)^{\nu^{\mu}}$, B – оператор сдвига назад: $B^l\omega^{\mu}_{j^{\text{per}},k}(t) = \omega^{\mu}_{j^{\text{per}},k-l}(t)$, $\psi^{\mu}_{j^{\text{per}},0} = 1$.

Поскольку амплитуда остаточной ошибки $\left|a^{\mu}_{j^{\text{per}},k}\right|$ характеризует величину отклонения текущего значения функции от ее характерного уровня, интенсивность аномалии на масштабе j^{per} может быть оценена как

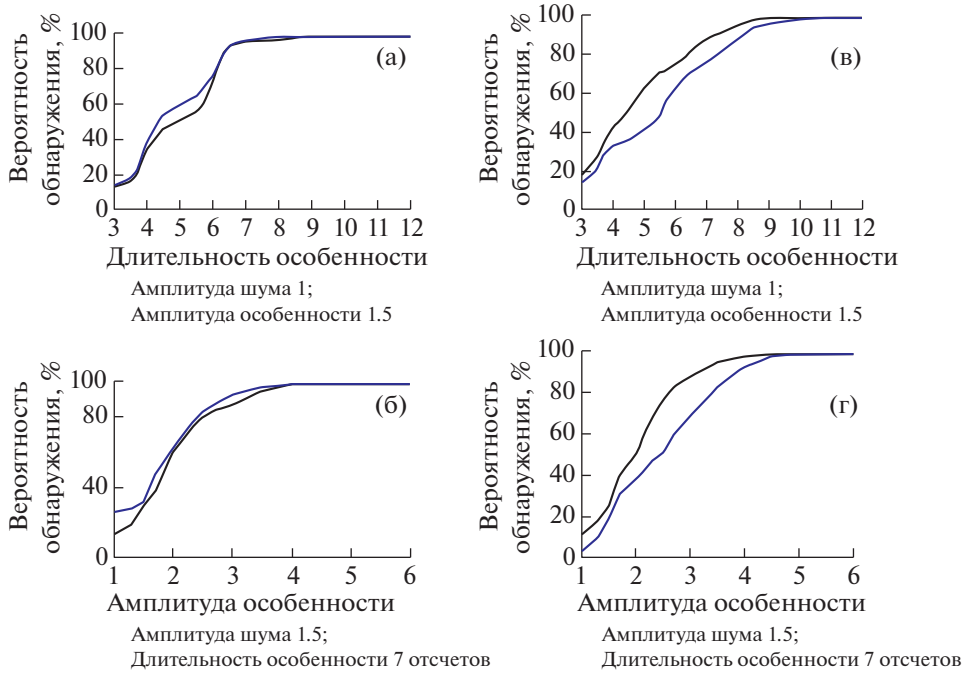
$$Y^{\mu}_{j^{\text{per}},\overline{k+1,k+L_{\mu}}} = \frac{\sqrt{\dfrac{1}{L_{\mu}}\sum_{l=1}^{L_{\mu}}(a^{\mu}_{j^{\text{per}},k+l})^2}}{H_{\mu,j^{\text{per}}}}, \tag{6}$$

где L_{μ} – длина временного окна.

Эффективность операций (4), (5) оценивалась на основе статистического моделирования. Структура модельных временных рядов соответствовала регистрируемым временным рядам параметров ионосферы и включала следующие компоненты: (1) – медианные значения данных foF2 (использовались данные, регистрируемые в периоды спокойной ионосферы при отсутствии геомагнитных возмущений и землетрясений на Камчатке с энергетическим классом $Ks > 12$); (2) – локальные особенности различного вида, длительности и амплитуды; (3) – белый шум. Использовались локальные особенности различного вида: синусоида, моделированная функцией Гаусса, прямоугольный импульс, треугольный импульс. Длительность локальных особенностей изменялась от 3 до 17 отсчетов, их амплитуда варьировалась в диапазоне 1.5–6, амплитуда белого шума – в диапазоне 0.5–3. В качестве примера на фиг. 1 представлен модельный временной ряд с шумом и локальными особенностями в виде треугольного импульса (фиг. 1а), и регистрируемые временные ряды критической частоты ионосферы (foF2) станции “Паратунка”. Оценки зависимости вероятности обнаружения локальных особенностей вида “треугольный импульс” с учетом их длительности и амплитуды представлены на фиг. 2. В соответствии со структурой временных рядов ионосферных параметров для различных сезонов и уровней активности Солнца (СА)



Фиг. 1. (а) – Модельный временной ряд; (б) – временной ряд foF2, регистрируемый на станции “Паратунка” в период с 01.01.2011 по 30.01.2011.



Фиг. 2. Графики вероятностей обнаружения локальных особенностей в зависимости от их амплитуды и длительности (низкая СА показана черным цветом, высокая СА – синим): (а), (б) – зимний сезон; (в), (г) – летний сезон.

оценки выполнялись раздельно. Анализ фиг. 2 показывает, что метод позволяет выделять локальные особенности, имеющие длительность не менее семи отсчетов для зимнего сезона (сигнал/шум = 2) и от девяти отсчетов и более для летнего сезона с вероятностью от 93%. Данный результат показывает высокую эффективность метода для выделения аномальных изменений в ионосфере, связанных с возникновением ионосферных бурь.

2. Составляющие $\beta_{i,\eta}^{\text{возм}}(t)$ аномальной компоненты $U(t)$ (см. соотношение (3)) являются случайными функциями, поэтому для их идентификации логично применить адаптивные пороги $P_{i,\eta}^{\text{ад}}, i = \overline{1,3}$, и коэффициенты $d_{\eta,n}$ в соотношении (3) принять равными

$$d_{\eta,n} = \begin{cases} d_{\eta,n}^{1+}, & \text{если} \quad P_{1,\eta}^{\text{ад}} < (d_{\eta,n} - d_{\eta,n}^{\text{med}}) \le P_{2,\eta}^{\text{ад}}, \\ 0, & \text{если} \quad \left|d_{\eta,n} - d_{\eta,n}^{\text{med}}\right| < P_{1,\eta}^{\text{ад}} \quad \text{или} \quad \left|d_{\eta,n} - d_{\eta,n}^{\text{med}}\right| > P_{2,\eta}^{\text{ад}}, \\ d_{\eta,n}^{1-}, & \text{если} \quad -P_{2,\eta}^{\text{ад}} \le (d_{\eta,n} - d_{\eta,n}^{\text{med}}) < -P_{1,\eta}^{\text{ад}}, \end{cases}$$

$$d_{\eta,n} = \begin{cases} d_{\eta,n}^{2+}, & \text{если} \quad P_{2,\eta}^{\text{ад}} < (d_{\eta,n} - d_{\eta,n}^{\text{med}}) \le P_{3,\eta}^{\text{ад}}, \\ 0, & \text{если} \quad \left|d_{\eta,n} - d_{\eta,n}^{\text{med}}\right| < P_{2,\eta}^{\text{ад}} \quad \text{или} \quad \left|d_{\eta,n} - d_{\eta,n}^{\text{med}}\right| > P_{3,\eta}^{\text{ад}}, \\ d_{\eta,n}^{2-}, & \text{если} \quad -P_{3,\eta}^{\text{ад}} \le (d_{\eta,n} - d_{\eta,n}^{\text{med}}) < -P_{2,\eta}^{\text{ад}}, \end{cases} \tag{7}$$

$$d_{\eta,n} = \begin{cases} d_{\eta,n}^{3+}, & \text{если} \quad (d_{\eta,n} - d_{\eta,n}^{\text{med}}) > P_{3,\eta}^{\text{ад}}, \\ 0, & \text{если} \quad \left|d_{\eta,n} - d_{\eta,n}^{\text{med}}\right| < P_{3,\eta}^{\text{ад}}, \\ d_{\eta,n}^{3-}, & \text{если} \quad (d_{\eta,n} - d_{\eta,n}^{\text{med}}) < -P_{3,\eta}^{\text{ад}}, \end{cases}$$

где $P_{i,\eta}^{\text{ад}} = V_i St_\eta$, V_i – пороговый коэффициент, величина

$$St_\eta = \sqrt{\frac{1}{\Phi - 1}\sum\nolimits_{n=1}^{\Phi}(d_{\eta,n} - \overline{d_{\eta,n}})^2},$$

$\overline{d_{\eta,n}}$ и $d_{\eta,n}^{\text{med}}$ – среднее значение и медианное значение соответственно. Значения с учетом суточного хода ионосферных данных вычисляются на основе скользящего временного окна, имеющего длительность Ф. Положительные аномалии класса i определяются на основе вейвлет-коэффициентов $d_{\eta,n}^{i+}$, отрицательные аномалии класса i определяются на основе вейвлет-коэффициентов $d_{\eta,n}^{i-}$.

Для различных классов i оценку интенсивности положительных ($J^{i+}(n)$) и отрицательных ($J^{i-}(n)$) аномалий в момент времени $t = n$ можно определить по формуле

$$J^{i+(-)}(n) = \sum_{\eta}\left|d_{\eta,n}^{i+(-)}\right|, \tag{8}$$

а общую интенсивность положительных ($J^{+}(n)$) и отрицательных ($J^{-}(n)$) аномалий по формуле

$$J^{+(-)}(n) = \sum_{\eta}\left|d_{\eta,n}^{+(-)}\right|. \tag{9}$$

Для оценки адаптивных порогов $P_{i,\eta}^{ad}, i = \overline{1,3}$, минимизировался апостериорный риск (см. [27]). Пороги разбивают пространство значений X анализируемой функции на четыре непересекающиеся области $\mathrm{X}_i, i = \overline{0,3}$. В таком случае *правило выбора решения* устанавливает соответствие между решениями о наличии/отсутствии аномалии класса i и областями. В соответствии с [27], используя *правило выбора решения* для заданного состояния h_j^i (характеризует наличие/отсутствие аномалии класса i), определим среднюю величину потерь в виде

$$R_j^i(x) = \sum_{l=0}^{3}\Pi_{il}P\{x \in X_l/h_j^i\},$$

где Π_{il} – функция потерь, $P\{x \in X_l/h_j^i\}$ – условная вероятность попадания выборки в область X_l, если в действительности имеет место состояние $h_j^i, i \ne l$, где i, l – индексы состояний (знак "/" означает условную вероятность).

Путем усреднения условной функции риска по всем состояниям h_j^i найдем *средний риск* в виде

$$R = \sum_{i=0}^{3} p_i R_j^i,$$

где p_i – априорная вероятность состояния h_j^i.

Наилучшим правилом будет такое, для которого средний риск будет наименьшим (байесовский риск, см. [27]). Учитывая, что априорное распределение состояний p_i нам не известно, для выбора *наилучшего правила* будем использовать *апостериорный риск*. Апостериорные вероятности $P\{h_j^i/x\}$ в этом случае позволяют получить наиболее полную характеристику состояний h_j^i при

располагаемых априорных данных. Для простой функции потерь $\Pi_{il} = \begin{cases} 1, & i \neq l \\ 0, & i = l \end{cases}$ апостериорный риск $R_l(x)$ равен

$$R_l(x) = \sum_{i \neq l} P\{h_j^i / x \in X_l\}.$$

В этом случае *критерием качества выбора решения* является критерий наименьшей частоты ошибок. Пороги $P_{i,\eta}^{\text{ад}}, i = \overline{1,3}$, определяются *наилучшим правилом выбора решения*, обеспечивающим наименьшее значение апостериорного риска $R_l(x)$. Оценки выполнялись отдельно для периодов высокой и низкой активности Солнца. При формировании классов также учитывались следующие особенности ионосферных данных.

1. *Аномалии малой интенсивности* (класс 1) характеризуют возникновение короткопериодных аномальных особенностей малой амплитуды ($J^{+(-)}(n)$ в диапазоне 400–600 для района Камчатки). Для формирования класса и оценки порогового коэффициента V_1 использовались данные foF2 района Камчатки, регистрируемые в периоды спокойного геомагнитного поля (K-индекс < 3).

2. *Аномалии умеренной интенсивности* (класс 2) характеризуют возникновение короткопериодных аномальных особенностей средней амплитуды (для района Камчатки $J^{+(-)}(n) > 600$ и $J^{+(-)}(n) \leq 1100$). Для формирования класса и оценки порогового коэффициента V_2 использовались данные foF2 района Камчатки, регистрируемые в периоды слабовозмущенного геомагнитного поля (K-индекс имел значения в диапазоне 3–4).

3. *Аномалии высокой интенсивности* (класс 3) характеризуют возникновение короткопериодных аномальных особенностей большой амплитуды ($J^{+(-)}(n) > 1100$ для района Камчатки). Для формирования класса и оценки порогового коэффициента V_3 использовались данные foF2 района Камчатки, регистрируемые в периоды возмущенного геомагнитного поля (K-индекс имел значения в диапазоне 5–8).

2.3. Моделирование и анализ ионосферных данных в периоды магнитных бурь

В работе при моделировании использовались 15-мин и часовые данные критической частоты ионосферы foF2, полученные на станции регистрации “Паратунка” (53.0° с.ш.; 158.7° в.д., п-ов Камчатка, ИКИР ДВО РАН). Выбирались данные в периоды спокойных геомагнитных условий (суммарное суточное значение K-индекса ≤20, максимальные значения K-индекса ≤4) и не содержащие сильных землетрясений (энергетический класс $Ks \geq 12$, http://sdis.emsd.ru/main.php). Учитывая, что ионосферные временные ряды содержат пропуски, возникающие в силу физических (например, возникновение спорадического E-слоя) и технических причин, в оценках использовались данные за периоды с наименьшим количеством пропусков – от 1 до 6%. Пропуски заполнялись на основе медианных значений, полученных с учетом суточного хода ионосферных параметров. Для операций кратномасштабного анализа (КМА) использовался вейвлет-базис Добеши порядка 3, который был определен с помощью процедуры минимизации погрешности при выполнении аппроксимации функции (см. [20]). Использовались часовые данные foF2 за период с 1968 по 2013 г. и 15-мин данные foF2 за период 2015–2018 гг. Для каждого сезона выполнялась отдельная оценка параметров моделей (см. [20]).

Для часовых данных foF2 получена модель вида

$$f(t) = \sum_{\mu=1,2} \sum_{k=1, N_{-3}^{\mu}} s_{-3,k}^{\mu} b_{-3,k}^{\mu}(t) + \sum_{i,\eta} \beta_{i,\eta}^{\text{возм}}(t) + e(t),$$

где для зимнего сезона:

$$s_{-3,k}^{1} = -0.6\omega_{-3,k-1}^{1} - 0.6\omega_{-3,k-2}^{1} + 0.4\omega_{-3,k-3}^{1} + a_{-3,k}^{1},$$

$$s_{-3,k}^{2} = -0.97\omega_{-3,k-1}^{2} - 0.93\omega_{-3,k-2}^{2} + a_{-3,k}^{2},$$

для летнего сезона высокой СА:

$$s^1_{-3,k} = -0.5\omega^1_{-3,k-1} - 0.6\omega^1_{-3,k-2} + a^1_{-3,k},$$

$$s^2_{-3,k} = -0.9\omega^2_{-3,k-1} - 0.8\omega^2_{-3,k-2} + a^2_{-3,k},$$

для летнего сезона низкой СА:

$$s^1_{-3,k} = -0.8\omega^1_{-3,k-1} - 0.7\omega^1_{-3,k-2} + a^1_{-3,k},$$

$$s^2_{-3,k} = -0.9\omega^2_{-3,k-1} - 0.9\omega^2_{-3,k-2} + a^2_{-3,k}.$$

При оценке порогов $H_{\mu,-3}$, определяющих периоды возникновения ионосферных аномалий (см. соотношения (4), (5)), использовались шаг упреждения $q = 1$ и доверительная вероятность 70%, $H_{\mu,-3}$ приняты равными:

– для зимнего сезона высокой (низкой) СА $H_{1,-3} = 1.4$ ($H_{1,-3} = 1.2$) – для сглаженной компоненты $f_{-3}(t)$; $H_{2,-3} = 0.97$ ($H_{2,-3} = 0.7$) – для детализирующей компоненты $g_{-3}(t)$;

– для летнего сезона высокой (низкой) СА $H_{1,-3} = 1.6$ ($H_{1,-3} = 1.3$) – для сглаженной компоненты $f_{-3}(t)$; $H_{2,-3} = 0.9$ ($H_{2,-3} = 0.8$) – для детализирующей компоненты $g_{-3}(t)$.

Для 15-мин данных foF2 получена модель вида

$$f(t) = \sum_{\mu=1,2} \sum_{k=1,N^\mu_{-5}} s^\mu_{-5,k} b^\mu_{-5,k}(t) + \sum_{i,\eta} \beta^{\text{возм}}_{i,\eta}(t) + e(t),$$

где для летнего сезона низкой СА:

$$s^1_{-5,k} = -0.9\omega^1_{-5,k-1} - 0.9\omega^1_{-5,k-2} + a^1_{-5,k},$$

$$s^2_{-5,k} = 0.4 - 0.3\omega^2_{-5,k-1} - 0.3\omega^2_{-5,k-2} + 0.5\omega^2_{-5,k-3} + a^2_{-5,k},$$

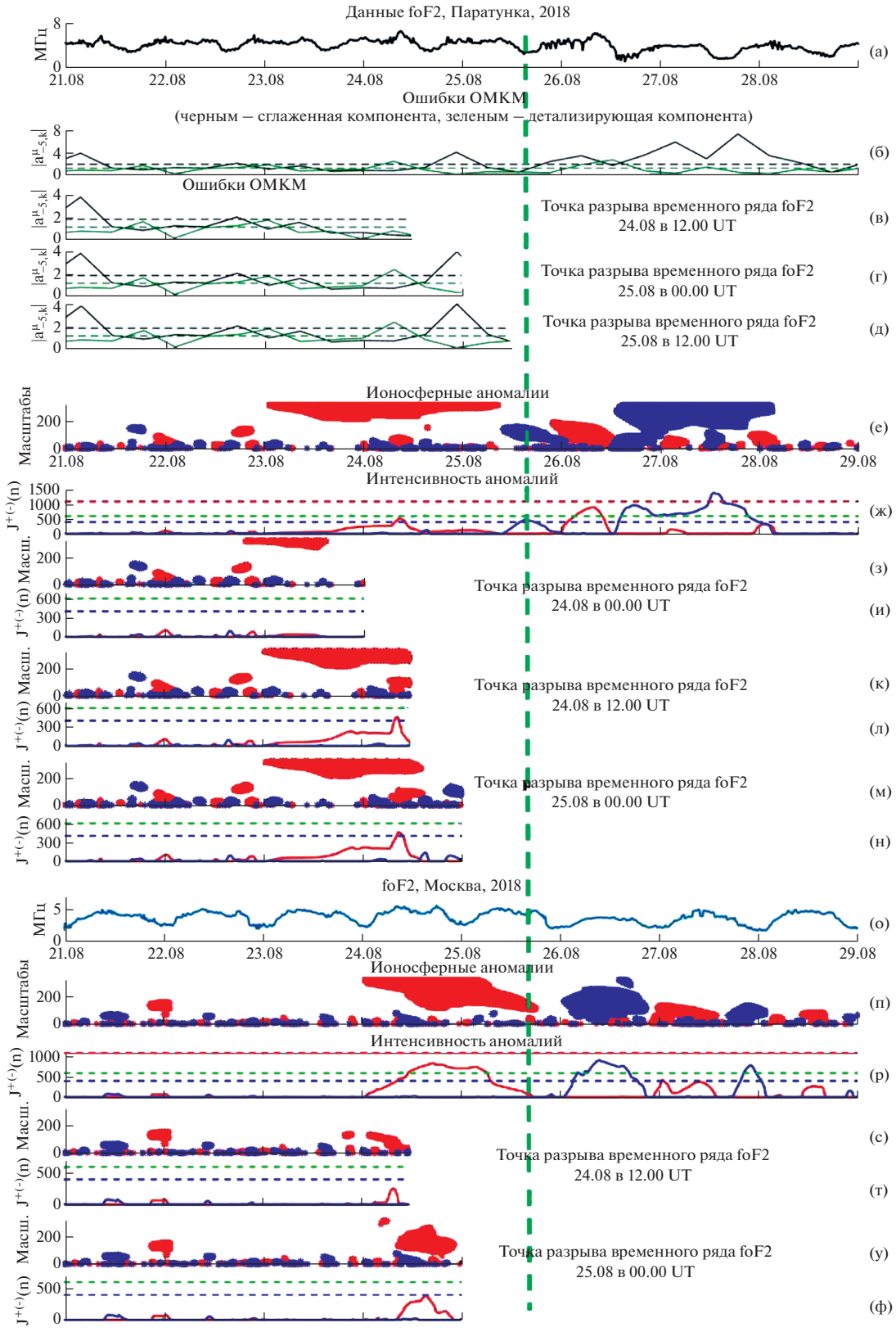
для зимнего сезона низкой СА:

$$s^1_{-5,k} = -0.6\omega^1_{-5,k-1} - 0.6\omega^1_{-5,k-2} + 0.4\omega^1_{-5,k-3} + a^1_{-5,k},$$

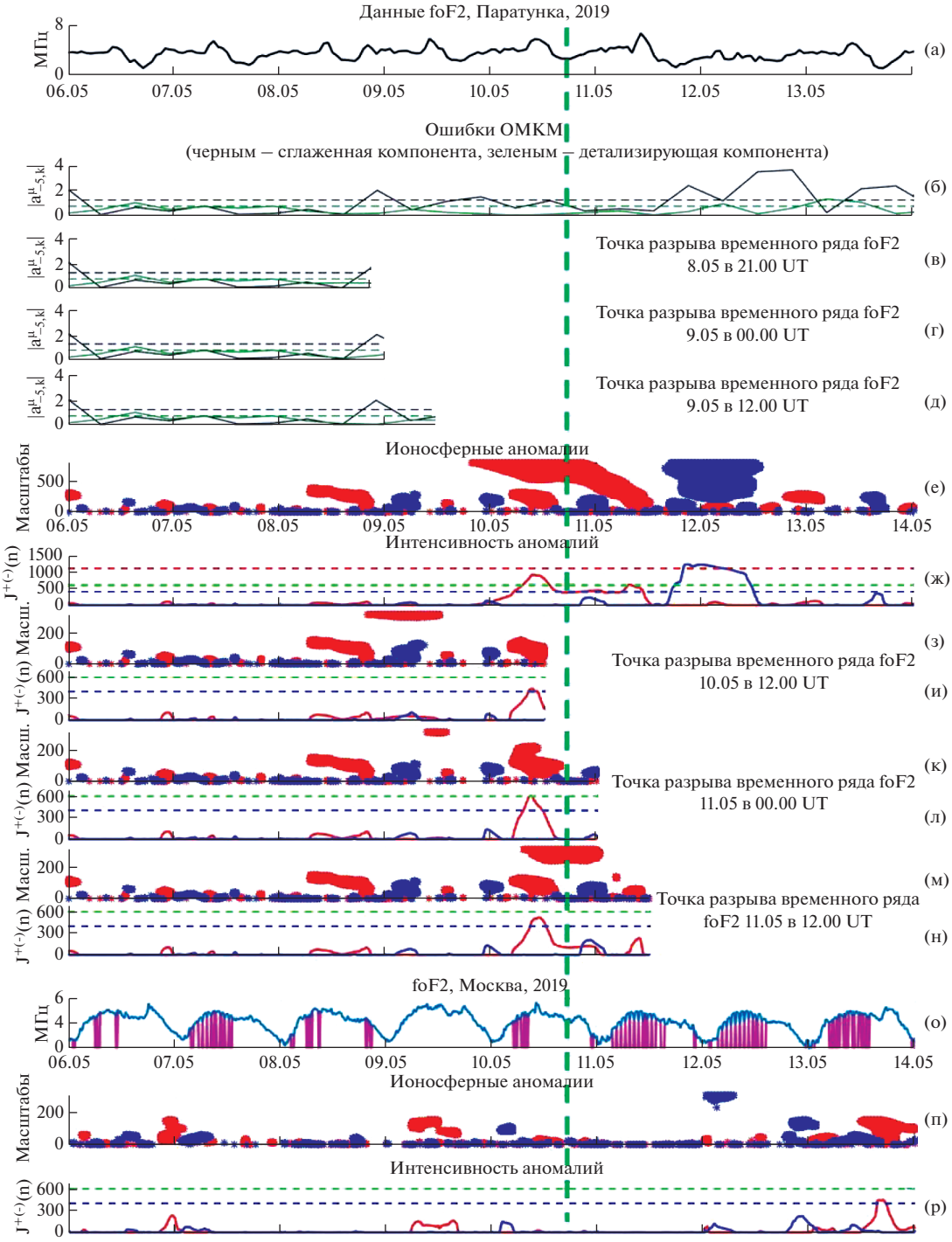
$$s^2_{-5,k} = -0.5\omega^2_{-5,k-1} - 0.4\omega^2_{-5,k-2} + 0.5\omega^2_{-5,k-3} + a^2_{-5,k}.$$

При оценке порогов $H_{\mu,-5}$ (соотношения (4), (5)) использовались шаг упреждения $q = 1$ и доверительная вероятность 70%. Для летнего сезона низкой СА пороги приняты равными $H_{1,-5} = 1.8$ – для сглаженной компоненты $f_{-5}(t)$; $H_{2,-5} = 1.1$ – для детализирующей компоненты $g_{-5}(t)$.

Результаты применения метода в период сильной магнитной бури 25–26 августа 2018 г. (фиг. 3) показывают возникновение длительных положительных ионосферных аномалий слабой интенсивности (класс 1) с максимумом на станции “Паратунка” около 10.00 UT 24 августа и на станции “Москва” – около 16.00 UT. Во время бури в ионосфере наблюдаются колебательные процессы, ошибки моделирования превысили 70% доверительный интервал (фиг. 3б). Далее 26 августа на восстановительной фазе магнитной бури ионосферные аномалии достигли наибольшей интенсивности (класс 3, фиг. 3е, ж, п, р), длительность отрицательной ионосферной бури на станции “Паратунка” составляла около 35 ч и на станции “Москва” – около 18 ч. В этот период результаты моделирования (фиг. 3б) показывают возникновение длительной ионосферной бури, о чем свидетельствуют существенные изменения во временном ходе foF2 – превышение ошибок сглаженной компоненты $f_{-3}(t)$ модели составило 3 СКО. По результатам исследования (см. [25]) в период анализируемого события наблюдается характерная динамика параметров ионосферы в возмущенный период – накануне магнитной бури на станциях возникли положительные ионосферные аномалии слабой интенсивности (незначительное повышение электронной концентрации ионосферы относительно фонового уровня), в период начальной фазы магнитной бури электронная концентрация ионосферы оставалась повышенной, в период восстановительной фазы электронная концентрация существенно понизилась и возникли интенсивные ионосферные бури. Анализ результатов последовательной обработки ионосферных данных (фиг. 3в–д, з–н, с–ф) подтверждает эффективность применения предлагаемого метода для обнаружения ионосферных аномалий в оперативном режиме. Положительная аномалия, обнаруженная накануне магнитной бури, превысила по интенсивности пороговое значение на стан-



Фиг. 3. Результаты анализа 15-мин данных foF2 станций “Паратунка” (Камчатский край) и “Москва” в период магнитной бури 25–26 августа 2018 г. (начало события отмечено вертикальной штриховой линией). Горизонтальными штриховыми линиями на (б)–(д) показаны пороги: черным – $H_{1,-5}$, зеленым – $H_{2,-5}$; на (ж), (и), (л), (н), (р), (т), (ф) показаны границы интенсивности $J^{+(-)}(n)$: красным – высокая, зеленым – умеренная, синим – малая.



Фиг. 4. Результаты анализа данных foF2 станций “Паратунка” (Камчатский край) и “Москва” в период магнитной бури 10–11 мая 2019 г. (начало события отмечено вертикальной штриховой линией). Горизонтальными штриховыми линиями на (б)–(д) показаны пороги: черным – $H_{1,-5}$, зеленым – $H_{2,-5}$; на (ж), (и), (л), (н), (р) – границы интенсивности $J^{+(-)}(n)$: красным – высокая, зеленым – умеренная, синим – малая.

ции “Паратунка” 24 августа в 09.00 UT, а на станции “Москва” приблизилась к пороговому значению на 7 ч позже.

Результаты применения метода в период магнитной бури 10–11 мая 2019 г. представлены на фиг. 4. Результаты моделирования (фиг. 4б) показывают, что 9 мая в период увеличения ССВ на-

блюдаются аномальные изменения во временном ходе foF2 – превышение ошибок сглаженной компоненты $f_{-3}(t)$ модели составило 1.6 СКО. Анализ ионосферных данных станции "Паратунка" (фиг. 4е,ж) показывает возникновение за несколько часов до начала события положительной аномалии умеренной интенсивности (класс 2) с максимумом около 11.00 UT 10 мая и длительностью около 28 ч. Результаты последовательной обработки данных (фиг. 4з–н) показывают, что интенсивность положительной аномалии на станции "Паратунка" превысила пороговое значение за 7 ч до начала магнитной бури. На восстановительной фазе магнитной бури электронная концентрация в ионосфере существенно понизилась, что привело к возникновению отрицательной ионосферной бури высокой интенсивности (класс 3) длительностью около 18 ч. По результатам моделирования (фиг. 4б) этот период сопровождался существенными изменениями временного хода данных foF2 – превышение ошибок составило более 3 СКО для компоненты $f_{-3}(t)$ модели и 1.7 СКО для компоненты $g_{-3}(t)$ модели. На станции "Москва", вследствие наличия длительных пропусков в данных (пропуски отмечены вертикальными линиями на фиг. 4о, которые могут возникать при автоматическом распознавании ионограмм), наблюдаемых как накануне, так и в период анализируемого события, достоверность результатов низкая и не подлежала анализу.

3. ЗАКЛЮЧЕНИЕ

Разработанный метод моделирования и анализа параметров ионосферы позволяет определять регулярные составляющие временного ряда и обнаруживать аномальные изменения разной амплитуды и длительности. Оценки показали высокую чувствительность метода и возможность его применения в режиме оперативного анализа данных (по мере поступления данных в систему обработки). На примере обработки критической частоты ионосферы foF2 для районов Камчатки и Москвы показана эффективность метода в задачах обнаружения ионосферных аномалий, предшествующих и сопутствующих периодам магнитных бурь. Уровень ионосферных возмущений оценивался на основе введенных классов (аномалии малой, умеренной и высокой интенсивностей). На основе метода подтверждена возможность возникновения короткопериодных аномалий слабой и умеренной интенсивности, предшествующих сильным магнитным бурям и характеризующим возникновение колебательных процессов в ионосфере на фоне повышенной солнечной активности. Метод имеет прикладную значимость в задачах прогноза космической погоды и предсказания магнитных бурь. Реализация метода выполнена в системе комплексного анализа геофизических данных Aurora (http://lsaoperanalysis.ikir.ru/lsaoperanalysis.html).

Авторы выражают благодарность институтам, выполняющим регистрацию ионосферных данных, которые использовались в работе.

СПИСОК ЛИТЕРАТУРЫ

1. *Nakamura M., Maruyama T., Shidama Y.* Using a neural network to make operational forecasts of ionospheric variations and storms at Kokubunji, Japan // J. Nation. Inst. Inform. and Comm. Technology. 2009. V. 56. № 3. P. 391–406.
2. *Danilov A.D.* Ionospheric F-region response to geomagnetic disturbances // Adv. Space Res. 2013. V. 52. № 3. P. 343–366.
3. *Danilov A.D.* F-2 region response to geomagnetic disturbances// J. Atmospheric and Solar-Terrestrial Phys. 2001. V. 63. № 2. P. 441–449.
4. *Афраймович Э.Л., Перевалова Н.П.* GPS-мониторинг верхней атмосферы Земли. Иркутск: ГУ НУ РВХ ВСНЦ СО РАМН, 2006. 480 с.
5. *Liu L., Wan W., Zhang M.L., Zhao B.* Case study on total electron content enhancements at low latitudes during low geomagnetic activities before the storms // Annales Geophysicae. 2008. V. 26. № 4. P. 893–903. https://doi.org/10.1134/S1990793115050206
6. *Liu L., Wan W., Zhang M.L., Zhao B., Ning B.* Prestorm enhancements in NmF2 and total electron content at low latitudes // J. Geophys. Res. 2008. V. 113(A02311). P. 1–12.
7. *Saranya P.L., Venkatesh K., Prasad D.S.V.V.D., Rama Rao P.V.S., Niranjan K.* Pre-storm behaviour of NmF2 and TEC (GPS) over equatorial and low latitude stations in the Indian sector // Adv. Space Res. 2011. V. 48. № 2. P. 207–217.
8. *Chernogor L.F., Rozumenko V.T.* Earth–Atmosphere–Geospace as an Open Nonlinear Dynamical System // Radio Phys. and Radio Astron. 2008. V. 13. № 2. P. 120–137.
9. *Dmitriev A.V., Suvorova A.V., Klimenko M.V., Klimenko V.V., Ratovsky K.G., Rakhmatulin R.A., Parkhomov V.* Predictable and unpredictable ionospheric disturbances during St. Patrick's Day magnetic storms of 2013 and 2015

and on 8–9 March 2008: Prediction of ionospheric disturbances // J. Geophys. Res.: Space Phys. 2017. V. 122. № 2. P. 2398–2423.

10. *Bilitza D., Reinisch B.W.* International Reference Ionosphere 2007: Improvements and new parameters // Adv. Space Res. 2008. V. 42. P. 599–609.
11. *Watthanasangmechai K., Supnithi P., Lerkvaranyu S., Tsugawa T., Nagatsuma T., Maruyama T.* TEC prediction with neural network for equatorial latitude station in Thailand // Earth, Planets and Space. 2012. V. 64. № 6. P. 473–483.
12. *Zhao X., Ning B., Liu L., Song G.* A prediction model of short-term ionospheric foF2 based on AdaBoost // Adv. Space Res. V. 53. № 3. P. 387–394. https://doi.org/10.1016/j.asr.2013.12.001.2014
13. *Sai Gowtam V., Tulasi Ram S.* An artificial neural network-based ionospheric model to predict NmF2 and hmF2 using long-term data set of FORMOSAT-3/COSMIC Radio Occultation. Observations: preliminary results // J. Geophys. Res.: Space Phys. 2017. V. 122. https://doi.org/10.1002/2017JA024795
14. *Tebabal A., Radicella S.M., Nigussie M., Damtie B., Nava B., Yizengaw E.* Local TEC modelling and forecasting using neural networks // J. Atmospheric and Solar-Terrestrial Phys. 2018. V. 172. P. 143–151. https://doi.org/10.1016/j.jastp.2018.03.004
15. *Mikhailov A., Morena B., Miro G., Marin D.* A method for foF2 monitoring over Spain using the El Arenosillo digisonde current observations // Annals of Geophys. 1999. V. 42. № 4. https://doi.org/10.4401/ag-3748
16. *Solomentsev D.V., Titov A.A., Khattatov B.V.* Three-dimensional assimilation model of the ionosphere for the European region // Geomagnetism and Aeronomy. 2013. V. 53. № 1. P. 73–84. https://doi.org/10.1134/S0016793212060114
17. *Knyazeva M.A., Namgaladze A.A., Beloushko K.E.* Field-aligned currents influence on the ionospheric electric fields: modification of the upper atmosphere model // Russ. J. Phys. Chemistry. 2015. V. 9. № 5. P. 758–763. https://doi.org/10.1134/S1990793115050206
18. *Shubin V.N., Karpachev A.T., Telegin V.A., Tchybulya K.G.* Global model SMF2 of the F2-layer maximum height // Geomagnetism and Aeronomy. 2015. V. 55. № 5. P. 609–622. https://doi.org/10.1134/S001679321505014X
19. *Song R., Zhang X., Zhou Ch., Liu J., He J.* Prediction TEC in China based on the neural networks optimized by genetic algorithm // Adv. Space Res. 2018. V. 62. № 4. https://doi.org/10.1016/j.asr.2018.03.043
20. *Mandrikova O.V., Fetisova N.V., Polozov Y.A., Solovev I.S., Kupriyanov M.S.* Method for modeling of the components of ionospheric parameter time variations and detection of anomalies in the ionosphere coupling of the high and mid latitude ionosphere and its relation to geospace dynamics // Earth, Planets and Space. 2015. V. 67. № 1. P. 131–146. https://doi.org/10.1186/s40623-015-0301-4
21. *Mandrikova O.V., Fetisova (Glushkova) N.V., Al-Kasasbeh R.T., Klionskiy D.M., Geppener V.V., Ilyash M.Y.* Ionospheric parameter modeling and anomaly discovery by combining the wavelet transform with autoregressive models // Annals of Geophys. 2015. V. 58. № 5. https://doi.org/10.4401/ag-6729
22. *Mallat S.* A wavelet tour of signal processing. 3rd Ed. London: Acad. Press, 2008. 832 p.
23. *Box G., Jenkins G.* Time series analysis: Forecasting and control. San Francisco: Holden Day, 1970. 553 p.
24. *Mandrikova O.V., Glushkova N.V., Zhivet'ev I.V.* Modeling and analysis of ionospheric parameters by a combination of wavelet transform and autoregressive models // Geomagnetism and Aeronomy. 2014. V. 54. № 5. P. 593–600. https://doi.org/10.1134/S0016793214050107
25. *Mandrikova O., Polozov Yu., Fetisova N., Zalyaev T.* Analysis of the dynamics of ionospheric parameters during periods of increased solar activity and magnetic storms // J. Atmospheric and Solar-Terrestrial Phys. 2018. V. 181. P. 116–126. https://doi.org/10.1016/j.jastp.2018.10.019
26. *Mandrikova O., Fetisova N., Polozov Yu.* Method of ionospheric parameter analysis in the problems of real-time data processing // J. Phys.: Conf. Ser. by IOP Publ. 2018. V. 1096(012091). https://doi.org/10.1088/1742-6596/1096/1/012091
27. *Левин Б.Р.* Теоретические основы статистической радиотехники. Изд. 2-е. М.: Советское радио, 1975. 392 с.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1113–1124

ИНФОРМАТИКА

УДК 519.72

МОРФОЛОГИЧЕСКИЕ И ДРУГИЕ МЕТОДЫ ИССЛЕДОВАНИЯ ПОЧТИ ЦИКЛИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ НА ПРИМЕРЕ РЯДОВ КОНЦЕНТРАЦИИ CO_2[1)]

© 2021 г. В. К. Авилов[1], В. С. Алешновский[2], А. В. Безрукова[2], В. А. Газарян[2,4], Н. А. Зюзина[2], Ю. А. Курбатова[1], Д. А. Тарбаев[2], А. И. Чуличков[2,3,*], Н. Е. Шапкина[2,5]

[1] 119071 Москва, Ленинский пр-т, 33, ИПЭЭ им. А.Н. Северцова РАН, Россия

[2] 119991 Москва, Ленинские горы, 1, стр. 2, МГУ им. М.В. Ломоносова, физический факультет, Россия

[3] 119017 Москва, Пыжевский пер., 3, ИФА им. А.М. Обухова РАН, Россия

[4] 125993 Москва, Ленинградский пр-т, 49, Финансовый университет при правительстве РФ, Россия

[5] 125412 Москва, ул. Ижорская, 13, ИТПЭ РАН, Россия

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

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

На основе методов морфологического анализа, развитых под руководством Ю.П. Пытьева, предложен метод фильтрации временных рядов, позволяющих выделить почти циклическую составляющую, длительность цикла которой не является постоянной, а значения элементов ряда внутри циклов изменчивы. Эффективность подхода иллюстрирована на примере декомпозиции временного ряда концентрации CO_2 в атмосфере Земли. Остаток ряда после фильтрации составляющей ряда, моделирующей суточную изменчивость, становится стационарным, что позволяет для его дальнейшего исследования использовать методы математической статистики и фурье-анализа. Верификация полученных результатов проводилась путем сравнения с результатами фурье-анализа. Для исследования цикличности с периодом, большим суток, используются фурье-разложение и вейвлет-анализ исходного ряда. Библ. 16. Фиг. 8.

Ключевые слова: цифровая обработка сигналов, квазипериодические сигналы, декомпозиция, форма сигнала, фурье-анализ, вейвлет-анализ.

DOI: 10.31857/S0044466921070048

1. ВВЕДЕНИЕ

В настоящее время обработка и анализ результатов длительных измерений, проводимых через заданные промежутки времени, вызывают интерес в различных сферах, от экономики до метеорологии. Для этого используется достаточно обширный арсенал методов как математически обоснованных (например, метод Фурье для стационарных рядов, вейвлет-преобразования для нестационарных), так и эвристических [1]. В этой ситуации интерес представляет разработка методов анализа временных рядов, основанных на достаточно общих математических моделях, дающих оптимальные решения на широком классе задач.

В данной работе для исследования цикличности временных рядов в первую очередь используется морфологический подход, развиваемый в школе Ю.П. Пытьева [2]–[4]. Предлагается способ морфологической фильтрации, позволяющий выделить в сигнале составляющую, которая представляет собой последовательность фрагментов, схожих по форме, но меняющихся по длительности и амплитуде в некоторых заданных пределах. Такая фильтрация может быть полезна для декомпозиции исходного временного ряда на “почти циклическую” составляющую, связанную с сезонностью, и остаток, который может быть подвергнут дальнейшему исследованию с целью выделения дополнительных циклических составляющих.

Эффективность предложенного подхода исследована на примере анализа данных измерения концентрации CO_2, представленных в виде временных рядов. В результате морфологической

[1)]Работа выполнена при частичной финансовой поддержке РФФИ (код проекта № 19-29-09044).

фильтрации ряда получен стационарный остаток, его дальнейшее исследование проведено методами фурье-анализа и математической статистики. В результате этого анализа были выделены короткопериодные составляющие ряда. Для верификации полученных результатов эти же данные исследовались методами фурье-анализа и вейвлет-анализа, было проведено сравнение результатов.

2. МОРФОЛОГИЧЕСКИЙ ПОДХОД К АНАЛИЗУ СИГНАЛОВ

На практике, в частности, при дистанционном зондировании атмосферы, при изучении временных рядов метеорологических параметров и др., анализируемый сигнал имеет вид суммы почти повторяющихся фрагментов, схожих по форме с сигналами иной формы – со стационарным или нестационарным сигналами, с широкополосной шумовой составляющей и др. Задача состоит в разделении суммы этих сигналов на составляющие и анализе параметров каждой из них.

Широко распространенный метод фурье-анализа не всегда удобен, так как позволяет представить исследуемый сигнал в виде суммы ортогональных слагаемых синусоидальной формы, в то время как при нестационарном характере сигналов каждый из участков сигнала может иметь свои характерные особенности. В этом случае более удобным является вейвлет-анализ, позволяющий исследовать динамическую структуру сигнала [5]. Вейвлет-анализ также дает возможность представить участок сигнала в виде линейной комбинации масштабируемых базисных функций заданной формы, однако, желательно, чтобы форма слагаемых была адаптирована к анализируемому сигналу.

Более гибкую модель сигнала заданной формы дает морфологический подход [2]–[4]. Морфологические методы анализа сигналов, развиваемые в школе Ю.П. Пытьева, предназначены для анализа структуры сигнала, сохраняющейся при его преобразовании, принадлежащем заданному классу. В частности, если каждый из двух сигналов представляет собой набор “пиков” и “провалов” (т.е. локальных максимумов и минимумов), то считать их сигналами одной формы разумно, если положения “пиков” и “провалов” совпадают, а упорядоченность локальных экстремумов по амплитуде сохраняется (самый высокий “пик” у обоих сигналов расположен в одной и той же точке и т.д.). Это свойство сигналов сохранится при любом монотонном преобразовании их амплитуд.

Похожий подход к анализу сигналов и изображений предложен в работах А.Н. Каркищенко и А.В. Гончарова [6] и [7], используемое в них описание изображений и сигналов получило название “знакового описания”. Изучению свойств такого описания посвящена монография [8], исследования продолжены в работе [9]. Методы, в которых используется свойство неотрицательности производных сигнала, развиваются также в работе [10].

В методах морфологического анализа [3] форма сигнала $f(t)$, $t \in T$, понимается как множество V_f сигналов, полученных из $f(t)$ всевозможными монотонными преобразованиями его амплитуды; очевидно, при таком понимании формы указанное выше характерное свойство сигнала сохраняется. Математическую модель сигнала и множество преобразований, сохраняющих форму, обычно выбирают так, чтобы для любого предъявленного для анализа сигнала $g(t)$, $t \in T$, была разрешима задача его наилучшего приближения элементами множества V_f, решение которой называют проекцией g на V_f. Операцию проецирования, в ряде случаев определенную конструктивно, называют формой сигнала (изображения) g [3].

В настоящей работе методы морфологического анализа, описанные в [2]–[4], адаптируются для анализа временного ряда концентрации CO_2, полученного на основе данных круглогодичных измерений на эколого-климатической станции “AsiaFlux” во Вьетнаме с 2011 по 2018 г., предоставленных ИПЭЭ им. А.Н. Северцова РАН [11]. Концентрация CO_2 в атмосфере регистрируется ежесекундно с последующим усреднением до получасовых значений и выражена в миллионных долях (ppm). Таким образом, рассматриваемый временной ряд концентрации CO_2 представляет собой упорядоченную последовательность усредненных получасовых значений концентрации CO_2 на определенной высоте над поверхностью Земли. Во временном ряду имеется составляющая, отражающая циклические процессы с периодом в одни сутки, амплитуда этой составляющей существенно больше других составляющих временного ряда. При этом от суток к суткам имеется существенная вариация поведения концентрации CO_2, что затрудняет применение не только фурье-, но и вейвлет-анализа. В данной работе предлагается математическая модель формы сигнала, отражающая суточную динамику концентрации CO_2, в виде функции, направление выпуклости которой меняется на противоположное в точках перегиба.

Эти точки являются параметрами формы фрагмента сигнала, моделирующего суточный ход концентрации CO_2. Считается, что два фрагмента сигнала имеют одинаковую форму, если интервалы их выпуклостей вверх и выпуклостей вниз совпадают.

Для выделения суточной циклической составляющей из исходного временного ряда выбирается составляющая, моделирующая суточную динамику концентрации CO_2 путем решения задачи наилучшего приближения участков ряда сигналами заданной формы. Далее исследуется остаток ряда, представляющий собой разность исходного ряда и его аппроксимации. Остаток удовлетворяет критерию стационарности.

3. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ВРЕМЕННОГО РЯДА

Временной ряд будем рассматривать как конечный набор значений ξ_i, $i = 1, \ldots, N$. Здесь каждое значение ξ_i есть результат наблюдения некоторого параметра в i-й момент времени. Зависимость значения параметра от времени является суммой почти циклической составляющей f_i и широкополосной случайной помехи ν_i, $i = 1, \ldots, N$.

Зададим математическую модель почти циклической составляющей f_i, $i = 1, \ldots, N$, считая, что последовательность моментов {1, 2, ..., N} времени измерений можно разбить на участки $\{i_1, \ldots, i_2 - 1\}, \{i_2, \ldots, i_3 - 1\}, \ldots, \{i_{2m-1}, \ldots, i_{2m} - 1\}$, длительность которых может изменяться от участка к участку в заданных пределах, и на каждом нечетном участке почти периодическая составляющая имеет выпуклость вверх, а на четном – вниз. В точках перегиба при переходе от выпуклости вниз к выпуклости вверх функция не убывает, а при смене выпуклости вверх на выпуклость вниз – не возрастает. Множество $V_f(i_1, \ldots, i_{2m})$ всех таких функций назовем формой почти периодического сигнала, а границы участков $1 \le i_1 \le \ldots \le i_{2m} \le N + 1$ определяют параметры его формы. Итак, форма $V_f(i_1, \ldots, i_{2m})$ почти периодического сигнала есть набор значений f_i, $i = i_1, \ldots, i_{2m} - 1$, для которых выполнены неравенства

$$\begin{gathered} 2f_j \ge f_{j-1} + f_{j+1}, \quad \text{если} \quad i_{2k-1} + 1 \le j \le i_{2k} - 2, \quad k = 1, \ldots, m, \\ 2f_j \le f_{j-1} + f_{j+1}, \quad \text{если} \quad i_{2k} + 1 \le j \le i_{2k+1} - 2, \quad k = 1, \ldots, m, \\ f_{j-1} \ge f_j, \quad \text{если} \quad j = i_{2k}, \quad k = 1, \ldots, m-1, \\ f_{j-1} \le f_j, \quad \text{если} \quad j = i_{2k+1} \quad k = 1, \ldots, m-1. \end{gathered} \tag{1}$$

Если рассматривать значения ξ_i, $i = i_1, \ldots, i_{2m} - 1$, временного ряда как координаты вектора конечномерного евклидова пространства, то множество $V_f(i_1, \ldots, i_{2m})$ является выпуклым замкнутым множеством, и для любого вектора $\boldsymbol{\xi}$ с координатами ξ_i, $i = i_1, \ldots, i_{2m} - 1$, однозначно разрешима задача наилучшего приближения

$$P(i_1, \ldots, i_{2m})\boldsymbol{\xi} = \arg\inf\{\|\boldsymbol{\xi} - \mathbf{f}\|^2 \,|\, \mathbf{f} \in V_f(i_1, \ldots, i_{2m})\} \tag{2}$$

вектора $\boldsymbol{\xi}$ векторами из $V_f(i_1, \ldots, i_{2m})$ при фиксированных параметрах формы. Ее решение $P(i_1, \ldots, i_{2m})\boldsymbol{\xi}$ называется проекцией $\boldsymbol{\xi}$ на $V_f(i_1, \ldots, i_{2m})$, а оператор проецирования $P(i_1, \ldots, i_{2m})$ взаимно однозначно определяется множеством $V_f(i_1, \ldots, i_{2m})$, и поэтому тоже называется формой сигнала $\mathbf{f}$ [3].

Сформулируем лемму, на основе которой удается найти решение задачи (2).

Лемма. *Пусть в евклидовых пространствах* R^m, R^k, $R^{m+k} = R^m \otimes R^k$, *где* $\mathbf{g} = (\mathbf{t}, \mathbf{s})$, $\mathbf{g} \in R^{m+k}$, $\mathbf{t} \in R^m$, $\mathbf{s} \in R^k$ *и* $(g_1, \ldots, g_m, g_{m+1}, \ldots, g_{m+k}) = (t_1, \ldots, t_m, s_1, \ldots, s_k)$, *выпуклые замкнутые конусы* V_{m+k}, V_m *и* V_k *заданы системами линейных неравенств*: $V_m = \{\mathbf{t} \in R^m : A\mathbf{t} \ge \mathbf{c}\}$, $V_k = \{\mathbf{s} \in R^k : B\mathbf{s} \ge \mathbf{d}\}$,

$$V_{m+k} = \{(\mathbf{t}, \mathbf{s}) \in R^{m+k} : A\mathbf{t} \ge \mathbf{c},\ B\mathbf{s} \ge \mathbf{d},\ C(\mathbf{t}, \mathbf{s}) \ge \mathbf{g}\} \tag{3}$$

для некоторых матриц A, B, C *и векторов* $\mathbf{c}$, $\mathbf{d}$, $\mathbf{g}$; P_{m+k}, P_m *и* P_k *– проекторы на конусы* V_{m+k}, V_m *и* V_k *соответственно. Тогда для некоторого вектора* $\tilde{\mathbf{g}} = (\tilde{\mathbf{t}}, \tilde{\mathbf{s}}) \in R^{m+k}$ *равенство* $P_{m+k}(\tilde{\mathbf{t}}, \tilde{\mathbf{s}}) = (P_m\tilde{\mathbf{t}}, P_k\tilde{\mathbf{s}})$ *выполняется тогда и только тогда, когда выполнено неравенство* $C(P_m\tilde{\mathbf{t}}, P_k\tilde{\mathbf{s}}) \ge \mathbf{g}$.

Эта лемма обобщает утверждения, приведенные в работах [3] и [4], и доказывается аналогично.

Лемма позволяет последовательно строить проекции фрагментов временного ряда на конус, определяемый неравенствами (1). На первом шаге строится проекция фрагмента ряда, состоящего из трех первых последовательных значений, на конус размерности 3. Далее, проекции фрагментов из 4, 5, ... значений элементов ряда на соответствующие конусы размерности 4, 5 и т.д. строятся путем последовательного присоединения к проецируемому фрагменту по одному значению.

Действительно, пусть для определенности $i_1 = 1$, и построена проекция $P_j(f_1, \ldots, f_j)$ на конус V_j, определяемый неравенствами (1), в которые входят только координаты $(f_1, \ldots, f_j)$. Тогда проекция $P_{j+1}(f_1, \ldots, f_{j+1})$ на конус V_{j+1} строится последовательным построением пары проекций $P_{j-l+1}(f_1, \ldots, f_{j-l+1})$, и $P_l(f_{j-l+2}, \ldots, f_{j+1})$ вектора $(f_1, \ldots, f_{j+1})$ на конусы V_{j-l-1} и V_{l+2} ($l = 1, 2, \ldots$) соответственно. Здесь первый конус V_{j-l-1} определяется неравенствами (1), в которые входят только координаты $(f_1, \ldots, f_{j-l+1})$, а второй V_{l+2} неравенствами (1), в которые входят только координаты $(f_{j-l}, \ldots, f_{j+1})$, $l = 1, 2, \ldots$, до тех пор, пока не будут выполнены неравенства (1), в которые входит хотя бы одна координата как из первого набора, так и из второго. Заметим, что как в первый, так и во второй наборы координат могут входить одна, две и более координат, в первых двух случаях могут отсутствовать неравенства, связывающие только координаты из соответствующего набора; в этой ситуации ограничения на эти координаты отсутствуют, а конус превращается в линейное подпространство.

Рассмотрим подробнее, как строится проекция $P_{j+1}(f_1, \ldots, f_{j+1})$ при известной $P_j(f_1, \ldots, f_j)$. Пусть имеется два набора координат: $(f_1, \ldots, f_j)$ и f_{j+1}. Если выполнены все неравенства из (1), в которые входят f_{j+1} и хотя бы одна координата из (f_{j-1}, f_j), то искомая проекция построена и равна $P_{j+1}(f_1, \ldots, f_{j+1}) = (P_j(f_1, \ldots, f_j), f_{j+1})$, так как выполнены условия леммы при $B = 0$, $\mathbf{d} = 0$.

Если эти неравенства не выполняются, следует построить две проекции: $P_{j-1}(f_1, \ldots, f_{j-1})$ (она уже вычислена на предыдущих шагах) вектора $(f_1, \ldots, f_{j-1})$ на конус V_{j-l} и $P_2(f_j, f_{j+1})$ вектора (f_j, f_{j+1}) на множество, определяемое неравенствами из (1), в которые входят только координаты (f_j, f_{j+1}); если таких неравенств нет, то $P_2(f_j, f_{j+1}) = (f_j, f_{j+1})$. Далее для координат построенных проекций проверяются выполнения неравенств (1), в которые входят как координаты из набора (f_j, f_{j+1}), так и из набора (f_{j-2}, f_{j-1}). Если эти неравенства выполняются, то искомая проекция построена, иначе следует продолжить процесс.

Поясним, как построить проекции на конусы на начальных шагах. Пусть, например, как и прежде, $i_1 = 1$, а $i_2 > 4$.

Шаг 1. Выберем первые три координаты f_1, f_2, f_3, и построим проекцию $P_3(f_1, f_2, f_3) = (q_1, q_2, q_3)$ трехмерного вектора (f_1, f_2, f_3) на трехмерный конус

$$V_3 = \{(t_1, t_2, t_3) : 2t_2 \geq t_1 + t_3\}, \tag{4}$$

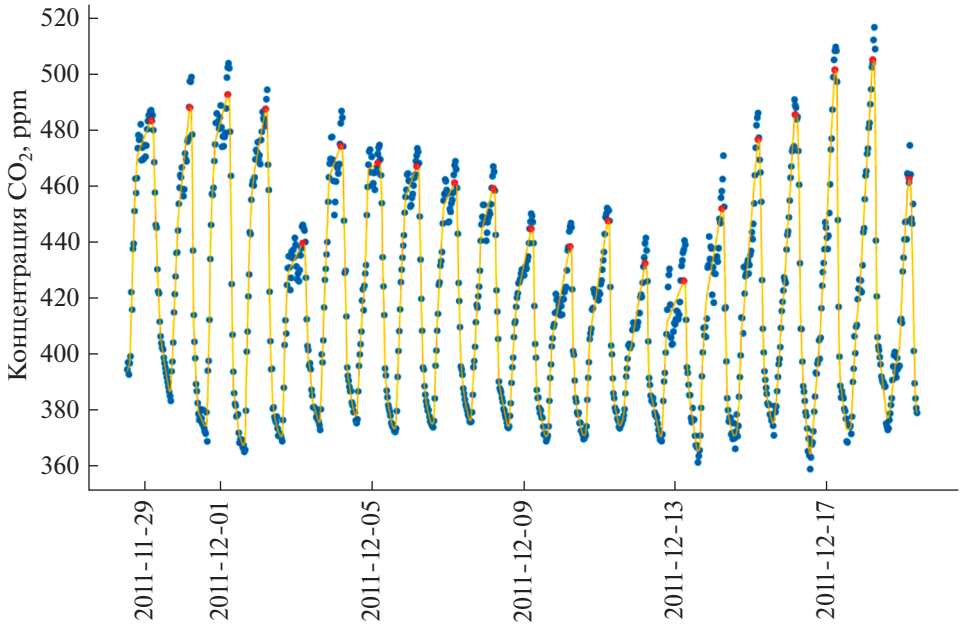
решая задачу выпуклого программирования

$$\inf_{\mathbf{t}} \left\{ \sum_{i=1}^{3} (f_i - t_i)^2 \,|\, 2t_2 \geq t_1 + t_3 \right\}.$$

Эта проекция есть либо сам вектор (f_1, f_2, f_3), если $2f_2 \geq f_1 + f_3$, либо наилучшая среднеквадратичная аппроксимация вектора (f_1, f_2, f_3) вектором с координатами $(q_1, (q_1, +q_3)/2, q_3)$, т.е. линейно зависящими от времени. Заметим, что эта ситуация соответствует утверждению леммы при $B = 0$, $C = 0$, $\mathbf{d} = 0$, $\mathbf{g} = 0$.

Шаг 2. Добавим к набору координат (f_1, f_2, f_3) вектора $\mathbf{f}$ четвертую координату f_4. Если выполнены неравенства $2q_3 \geq q_2 + f_4$ для координат проекции $P_3(f_1, f_2, f_3) = (q_1, q_2, q_3)$, вычисленной на первом шаге, и f_4, то согласно лемме проекция $P_{3+1}(f_1, f_2, f_3, f_4)$ на конус $V_{3+1} = \{(t_1, t_2, t_3, s_1) : 2t_2 \geq t_1 + t_3,\ 2t_3 \geq t_2 + s_1\}$ в четырехмерном пространстве равна $P_{3+1}(f_1, f_2, f_3, f_4) = (P_3(f_1, f_2, f_3), f_4)$, что соответствует условиям леммы при $\mathbf{d} = 0$. Если это неравенство не выполнено, то дальнейшие действия зависят от результата первого шага.

1. Если $P_3(f_1, f_2, f_3) = (f_1, f_2, f_3)$, то следует построить проекцию $P_3(f_2, f_3, f_4)$ трехмерного вектора (f_2, f_3, f_4) на конус V_3, определенный в (4), и проверить включение $(f_1, P_3(f_2, f_3, f_4)) \in V_{3+1}$. Если



Фиг. 1. Временной ряд концентрации CO_2; точки – результат регистрации, сплошная линия – результат морфологической фильтрации. Шаг по времени 0.5 ч.

это включение выполнено, то $(f_1, P_3(f_2, f_3, f_4))$ – искомая проекция, так как выполнены условия леммы при $A = 0$, $\mathbf{c} = 0$ для $V_{1+3} = V_{3+1}$.

2. Если $P_3(f_1, f_2, f_3) \neq (f_1, f_2, f_3)$, то проекция $P_{3+1}(f_1, f_2, f_3, f_4)$ есть результат аппроксимации фрагмента (f_1, f_2, f_3, f_4) временного ряда фрагментом, значения которого есть линейная функция времени, в соответствии с условиями леммы при $B = 0$, $C = 0$, $\mathbf{d} = 0$, $\mathbf{g} = 0$.

Итак, если заданы параметры формы $V_f(i_1,\ldots,i_{2m})$ временного ряда, то построить наилучшее приближение любого временного ряда рядами заданной формы можно, основываясь на приведенных здесь результатах. Однако на практике интервалы выпуклости вверх и вниз могут быть непостоянными и меняться в известных пределах: $(i_1,\ldots,i_{2m}) \in I$, $I = \{(i_1,\ldots,i_{2m}) : \underline{i}_k \le i_k \le \overline{i}_k,\ k = 1,\ldots,2m\}$. Интерес представляет выбор значений этих параметров, при которых анализируемый временной ряд наиболее близок к соответствующей форме $V_f(i_1,\ldots,i_{2m})$.

Задачу наилучшего приближения временного ряда сигналами формы $V_f(i_1,\ldots,i_{2m})$ выбором параметров формы назовем задачей морфологической фильтрации временного ряда. Формально это есть задача на поиск минимума

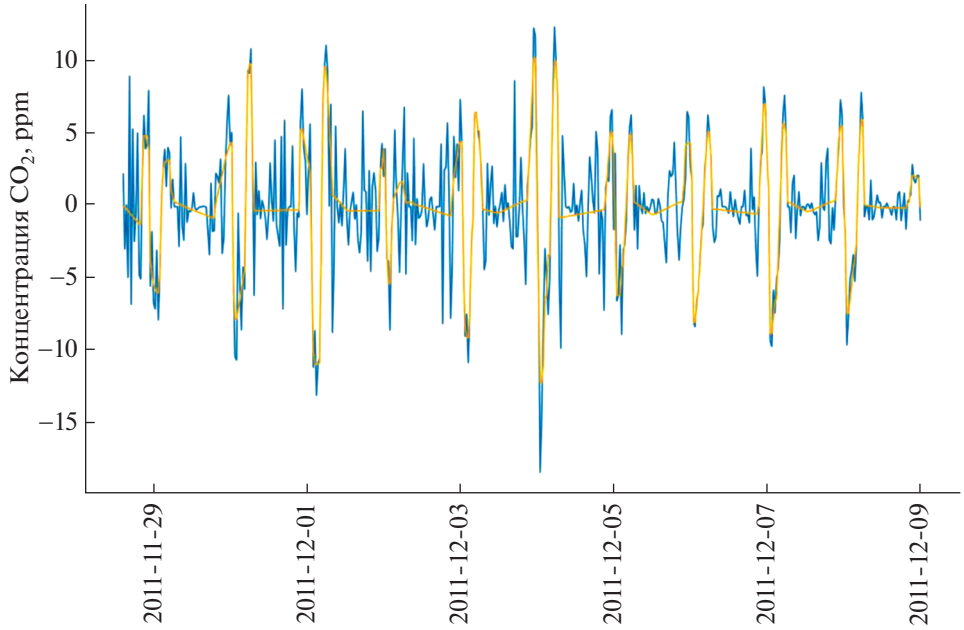
$$\inf_{\mathbf{i}\in I} \inf\left\{\|\boldsymbol{\xi} - \mathbf{f}\|^2 \,|\, \mathbf{f} \in V_f\,(i_1,\ldots,i_{2m})\right\},$$

где I – множество допустимых значений параметров формы $V_f(i_1,\ldots,i_{2m})$.

4. ЗАДАЧА ФИЛЬТРАЦИИ СИГНАЛА ЗАДАННОЙ ФОРМЫ

Рассмотрим пример анализа временного ряда, при котором используется морфологическая фильтрация, описанная в разд. 2 настоящей статьи. На фиг. 1 точками отображена временная зависимость концентрации CO_2, измеренная с шагом 0.5 ч. По вертикали отложена концентрация в миллионных долях (ppm), по горизонтали – время. Видно, что фрагменты временного ряда длительностью примерно 24 ч схожи по форме, однако имеются вариации как в амплитуде, так и в длительности фрагментов. В целом, 24-часовые фрагменты хорошо описываются сигналами, математическая модель формы которых описана в предыдущем разделе настоящей статьи, однако имеются и отличия, которые тоже могут представлять интерес для исследователя.

Для описания суточного хода концентрации CO_2 была применена морфологическая фильтрация сигнала. Морфологическая фильтрация осуществляется следующим образом. Задаются



Фиг. 2. Остаток временного ряда концентрации CO_2 после вычитания результата морфологической фильтрации и результат фильтрации остатка для выделения парных пиков.

ограничения на параметры формы, затем для каждого возможного значения параметров i_1 и i_2 строится проекция фрагмента временного ряда на множество фрагментов, выпуклых вверх, и выбираются такие их значения, которые минимизируют отличие анализируемого фрагмента от заданной формы. Затем для найденного значения i_2 и для каждого возможного значения i_3 строится проекция фрагмента временного ряда на множество фрагментов, выпуклых вниз. Далее проверяется условие сшивания фрагментов, выпуклых вниз и вверх. Условие сшивания задается неравенствами (3). Если они не выполнены, корректируются значение найденной проекции и значение i_2 положения точки перегиба, далее эта процедура повторяется для каждого участка заданного типа выпуклости.

На фиг. 1 приведен результат морфологической фильтрации (сплошная линия). Хорошо видны повторяющиеся участки длительностью примерно 24 ч, отвечающие суточной изменчивости концентрации CO_2. Средний период, вычисленный по исследованному фрагменту временного ряда, составил 24.0 ч.

В результате применения алгоритма морфологической фильтрации к временному ряду концентрации CO_2 наряду с выделением суточной циклической составляющей из исходного временного ряда проведено исследование остаточной составляющей, представляющей собой разность исходного ряда и его аппроксимации, полученной в результате морфологической фильтрации. График значений остатков показан на фиг. 2; на нем явно угадываются повторяющиеся парные пики. Они были выделены из остаточной составляющей повторной морфологической фильтрацией с периодом, меньшим, чем 24 ч. Результат повторной морфологической фильтрации приведен на фиг. 2 в виде кривой оранжевого цвета.

Проверка остаточной составляющей на стационарность с помощью расширенного теста Дики–Фуллера [12] показала, что, в отличие от исходного ряда, ряд остатков удовлетворяет условию стационарности, следовательно, для дальнейшего выделения циклических компонент ряда с периодами в пределах суток возможно применение фурье-анализа.

Приближенное представление ряда остатков X_t в виде частичной суммы ряда Фурье

$$X_t \approx a_0 + \sum_{i=1}^{I}\left[a_i \cos\left(\frac{2\pi t}{T_i}\right) + b_i \sin\left(\frac{2\pi t}{T_i}\right)\right], \tag{5}$$

где $t = 1, 2, \ldots, N$, N – число элементов ряда, a_i, b_i – коэффициенты Фурье, T_i – период i-й гармоники, $i = 1, \ldots, I$, I – число гармоник, позволяет обнаружить колебания ряда X_t остаточной

составляющей с периодами в пределах суток, вносящие наибольший вклад во временной ряд X_t, и оценить вклад i-й гармоники с помощью периодограммы – значений $\mathrm{Per}_i = \frac{N}{2}(a_i^2 + b_i^2)$, $i = 1, \ldots, I$, а ее относительный вклад в суммарную периодограмму ряда $(X_t - a_0)$ оценить значением

$$\frac{\mathrm{Per}_j}{\sum_{i=1}^{I} \mathrm{Per}_i} = \frac{\frac{N}{2}(a_j^2 + b_j^2)}{\frac{N}{2}\sum_{i=1}^{I}(a_i^2 + b_i^2)}, \quad i = 1,\ldots,I.$$

В результате применения фурье-анализа к ряду остаточной составляющей показано, что наиболее существенный вклад вносят гармонические компоненты разложения Фурье (5) с периодами 11.9, 7.95 и 6 ч.

Алгоритм морфологической фильтрации требует достаточно больших вычислительных затрат по сравнению с линейными разложениями по заданной системе функций, поэтому для исследования рядов большой длительности воспользуемся более быстрыми алгоритмами фурье-анализа и вейвлет-анализа.

5. СТАТИСТИЧЕСКИЙ АНАЛИЗ ВРЕМЕННОГО РЯДА КОНЦЕНТРАЦИИ УГЛЕКИСЛОГО ГАЗА

Статистический анализ временных рядов концентрации углекислого газа также проводится на основе данных эколого-климатической станции “AsiaFlux” во Вьетнаме с 2011 по 2018 г. На фиг. 3 графически представлен временной ряд концентрации CO_2 за период с 2011 по 2018 г. На основе визуального анализа графика концентрации CO_2 выдвинем предположение о том, что представленный ряд содержит регулярную и случайную компоненты. Такое предположение подтверждается также заключениями метеорологов о механизмах формирования уровня концентрации CO_2 в атмосфере под воздействием различных природных и антропогенных факторов. При этом регулярная компонента включает основную тенденцию ряда динамики показателя (тренд), а также его сезонные и циклические составляющие, характеризующиеся различными периодами колебаний концентрации CO_2.

Визуальный анализ рядов динамики концентрации CO_2 позволяет также высказать предположение о том, что значительных изменений амплитуды колебаний концентрации CO_2 с 2011 г. по 2018 г. как в сторону заметного увеличения, так и в сторону уменьшения, не наблюдается, поэтому в качестве модели декомпозиции изучаемых рядов динамики предлагается аддитивная модель [13]:

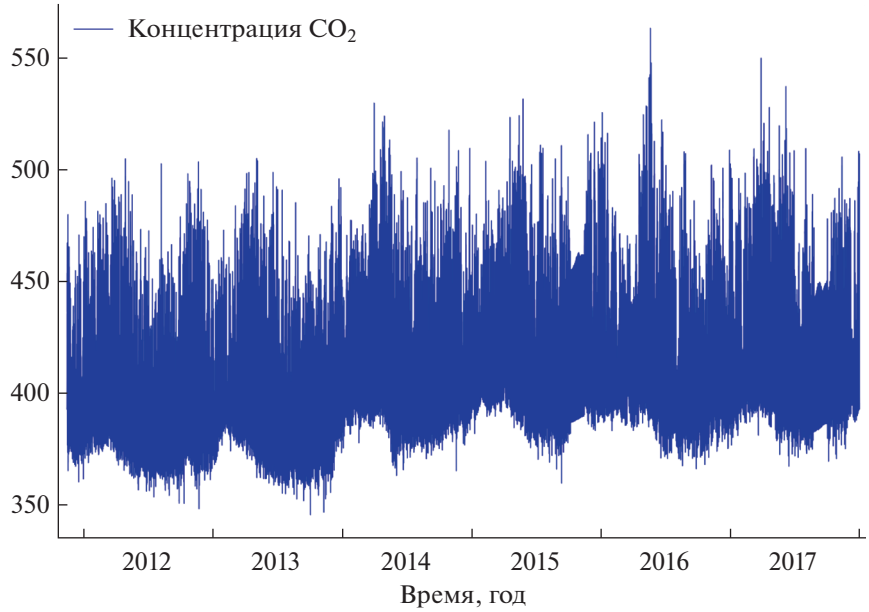
$$X_t = S_t + C_t + u_t + \varepsilon_t, \tag{6}$$

где S_t – постоянная сезонная составляющая, C_t – циклическая составляющая, u_t – тренд, определяющий основную тенденцию временного ряда, ε_t – нерегулярная составляющая, $t = 1, 2, \ldots, N$, где N – число элементов ряда.

При изучении основной тенденции изменчивости концентрации CO_2 за рассматриваемый временной период целесообразно на начальном этапе, предшествующем декомпозиции, провести процедуру сглаживания исходного ряда с целью повышения его устойчивости по отношению к выбросам различной природы. В результате сглаживания методом простой скользящей средней исходного ряда динамики концентрации CO_2 X_t, $t = 1, \ldots, N$ (6), получаем ряд скользящих средних $\hat{X}_i$:

$$\hat{X}_i = \frac{\sum_{t=i-p}^{i+p} X_t}{K}, \quad p = \frac{K-1}{2}, \quad t = 1,\ldots,N, \quad i = p+1,\ldots,N-p, \tag{7}$$

где K – интервал сглаживания, равный периоду колебаний концентрации CO_2, который будем задавать различным образом, учитывая периоды колебаний, вносящие наибольший вклад в ряд динамики концентрации CO_2. Если провести усреднение ежесекундных значений концентрации CO_2 не за полчаса, а за сутки, т.е. в качестве X_t, $t = 1, \ldots, N$, рассмотреть среднесуточные зна-



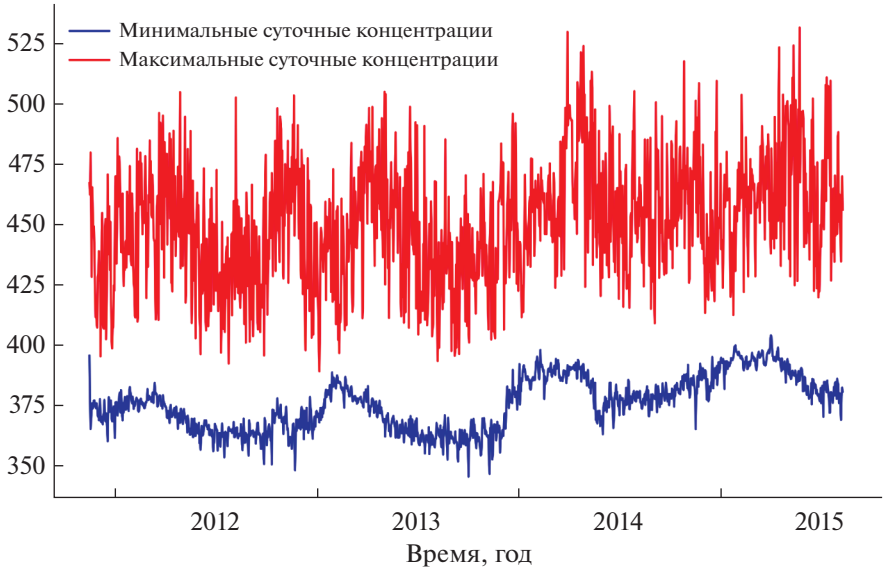
Фиг. 3. Графическое представление временного ряда концентрации CO_2 (2011–2018 гг.)

чения уровней ряда, а интервал сглаживания задать равным 365 дней, в результате сглаживания получится ряд $\hat{X}_t$, $t = 1, \ldots, N$, не содержащий влияния сезонных изменений. Тогда для определения постоянной сезонной компоненты S_t следует усреднить разность значений X_t и $\hat{X}_i$ для каждого дня $t = k + Kj$, $t = 1, 2, \ldots, N$, на протяжении восьмилетнего периода наблюдений за концентрацией CO_2, где k – день года, j – год, $k = 1, \ldots, K$, $j = 0, \ldots, J-1$:

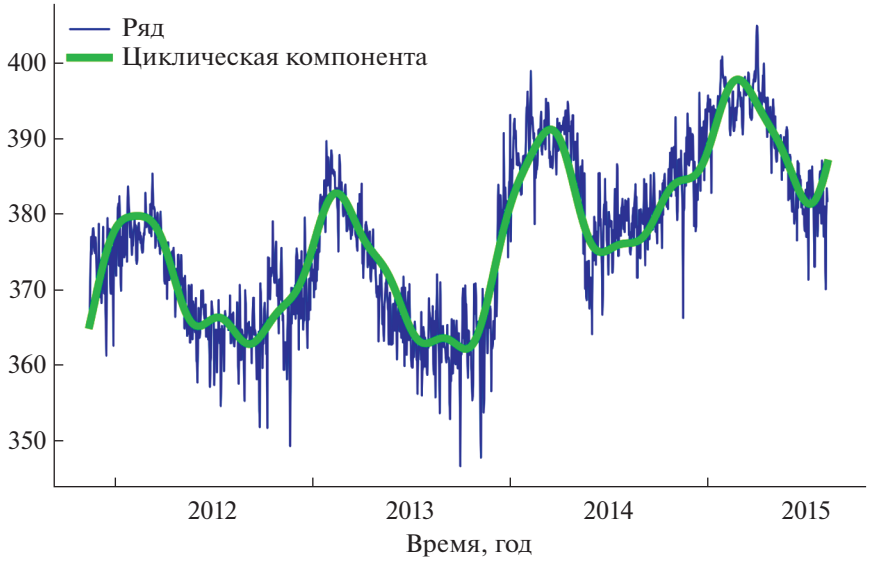
$$S_t = \frac{\sum_{j=0}^{J-1} [X_{k+Kj} - \hat{X}_{k+Kj}]}{J},$$

где $J = 8$ лет, $K = 365$ дней. Для дальнейшего изучения основной тенденции ряда (6), не зависящей от сезонных колебаний, следует вычесть из исходного ряда X_t, $t = 1, \ldots, N$, ряд сезонной компоненты S_t, $t = 1, \ldots, N$, и снова провести усреднение ряда $X_t - S_t$, $t = 1, \ldots, N$, методом скользящих средних (7). Поскольку в этом случае характер основной тенденции может зависеть от выбора периода усреднения, проведем анализ циклических колебаний ряда динамики концентрации CO_2.

Наряду с временными рядами концентрации CO_2 на различных высотах над поверхностью Земли, важное значение для анализа климатических изменений имеют ряды динамики экстремальных (минимальных и максимальных) суточных значений концентрации CO_2 (фиг. 4). Проверка временных рядов концентрации CO_2 на стационарность с помощью расширенного теста Дики–Фуллера [12] показала, что в то время как ряды динамики нельзя считать стационарными, временные ряды экстремальных суточных значений концентрации CO_2 являются стационарными, поэтому для анализа их циклических колебаний применен фурье-анализ. В результате применения фурье-анализа к ряду минимальных суточных значений концентрации CO_2 выделены следующие колебания концентрации CO_2, вносящие наибольший вклад во временной ряд X_t: 11.37 мес, 3.73 г, 15.1, 6.5, 9, 7.6 и 3.5 мес (перечислены в порядке убывания значений периодограммы). Графическое представление циклической составляющей C_t временного ряда (6) показано на фиг. 5. При анализе максимальной суточной концентрации были выделены совпадающие по значениям, но отличающиеся по величине периодограммы периоды, приведенные далее также в порядке убывания значений периодограммы: 6.5, 15.1, 11.37, 7.6, 5.7 мес. Проверка остаточных составляющих ε_t (5) рядов максимальных и минимальных суточных концентраций с помощью критерия Колмогорова–Смирнова показала, что ряды остатков подчиняются нормальному закону распределения.



Фиг. 4. Графики минимальных и максимальных суточных значений концентрации CO_2.



Фиг. 5. Зависимость минимальных суточных значений концентрации CO_2 и его циклической компоненты от времени.

Таким образом, с помощью морфологического анализа исходного временного ряда концентрации CO_2 и остаточной составляющей, представляющей собой разность исходного ряда и его аппроксимации, полученной в результате морфологической фильтрации, были выявлены суточные колебания концентрации CO_2, а также циклические составляющие с периодами в пределах суток: 12, 8 и 6 ч. Для определения циклических компонент, имеющих более длительные периоды, был применен фурье-анализ стационарных рядов минимальных и максимальных суточных значений концентрации CO_2. В результате фурье-анализа показано, что наибольший вклад в исследуемые за период с 2011 по 2018 г. временные ряды вносят гармоники около 1 г., 3.7 года, 6.5 и 9 мес.

6. ВЕЙВЛЕТ-АНАЛИЗ ВРЕМЕННОГО РЯДА КОНЦЕНТРАЦИИ УГЛЕКИСЛОГО ГАЗА

В случае, когда ряд не является стационарным и преобразовать его в стационарный не удается, оправдано применение вейвлет-анализа. Вейвлеты – это семейство функций, которые полу-

чаются из одной функции посредством ее сдвигов и растяжений по оси времени. С помощью вейвлет-преобразования функция рассматривается в виде разложения на колебания, локализованные по времени и частоте. Таким образом, вейвлет-анализ применяется для обработки нестационарных сигналов, так как отвечает специфике временных рядов, которые демонстрируют эволюцию во времени своих основных характеристик – среднего значения, дисперсии, периодов, амплитуду и т.д. [14] и [15].

Исследуемый временной ряд показателей концентрации CO_2 за полтора года, усредненных с периодом в 30 мин, в ходе работы был проверен на стационарность с помощью расширенного теста Дикки–Фуллера. Результат оказался предсказуем – ряды нестационарны, тем самым обосновано применение вейвлет-анализа. Временные ряды минимумов и максимумов на различных высотах, напротив, оказались стационарными, что, вообще говоря, не исключает возможность использования вейвлет-преобразования в качестве одной из методик исследования, в связи с этим вышеупомянутые ряды динамики также использовались в качестве анализируемых временных рядов.

Сущность непрерывного вейвлет-преобразования заключается в следующем. С помощью подходящего материнского вейвлета $\Psi(t)$ вычисляются вейвлет-функции $\Psi_{a,b}(t)$:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right),$$

где параметр a, называемый параметром масштаба вейвлет-преобразования, принимает строго положительные значения и отвечает за ширину вейвлета; величина b есть параметр сдвига, который определяет положение вейвлета на оси t.

Далее определяются вейвлет-коэффициенты $W(a,b)$:

$$W(a,b) = \frac{1}{\sqrt{a}}\int\limits_{-\infty}^{\infty} f(t)\Psi^*_{a,b}(t)dt,$$

где $f(t)$ – анализируемый ряд динамики, * обозначает комплексное сопряжение [5].

После этого проводится качественный анализ картины вейвлет-коэффициентов и построение интегрального спектра:

$$S(a) = \int\limits_{a_1}^{a_2} |W(a,b)|^2\, db,$$

который показывает наличие циклов в исходном временном ряде. Масштаб a связан с координатами оси времени t как $t = \frac{a}{F_c}$, где F_c – центральная частота вейвлета.

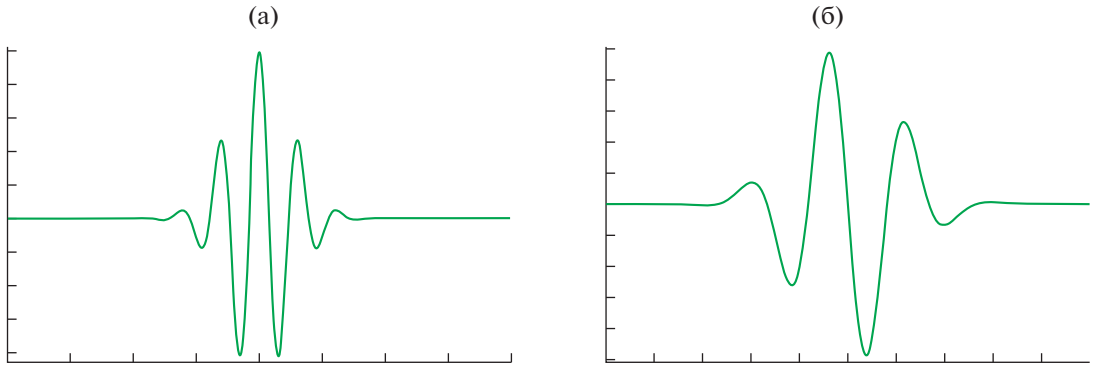
Для исследуемых временных рядов в качестве материнских вейвлетов были выбраны вейвлет Морле и вейвлет Гаусса 7, дающие наиболее информативные и наглядные результаты [16]. Центральная частота данных материнских вейвлетов – 0.8125 и 0.6 соответственно (фиг. 6). Применение иных материнских вейвлетов давали результаты, принципиально не отличающиеся от приведенных в настоящей статье, но с менее выраженными частотно-временными особенностями.

Картины вейвлет-коэффициентов позволяют подробно исследовать структуру периодичностей, содержащихся в исследуемом ряде динамики (в отличие, например, от метода фурье-анализа). Масштаб, коэффициенты которого максимальны (т.е. вносят максимальный вклад во временной ряд) характеризует определенный период цикличности. Иными словами, светлые горизонтальные линии спектра и соответствуют искомым циклам (фиг. 7а и 8а). Картина вейвлет-коэффициентов позволяет качественно оценить величину цикличностей и время, когда они наблюдаются.

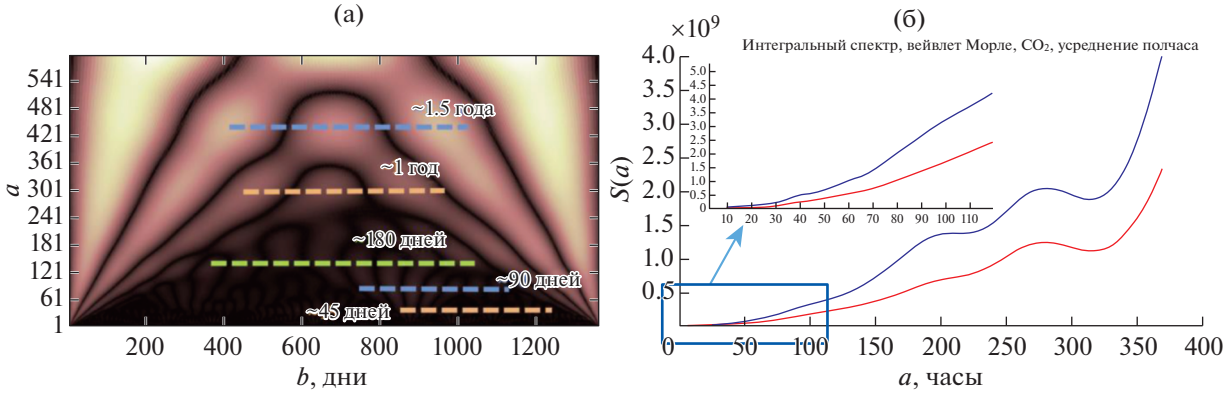
Интегральный спектр, в отличие от картины вейвлет-коэффициентов, позволяет рассчитать величину периода цикла (фиг. 7б и 8б). Локальные экстремумы графика соответствуют максимальным коэффициентам на данном масштабе, показывая значения периода циклов.

Также были получены картины вейвлет-коэффициентов и интегральные спектры для различных показателей концентрации углекислого газа в атмосфере.

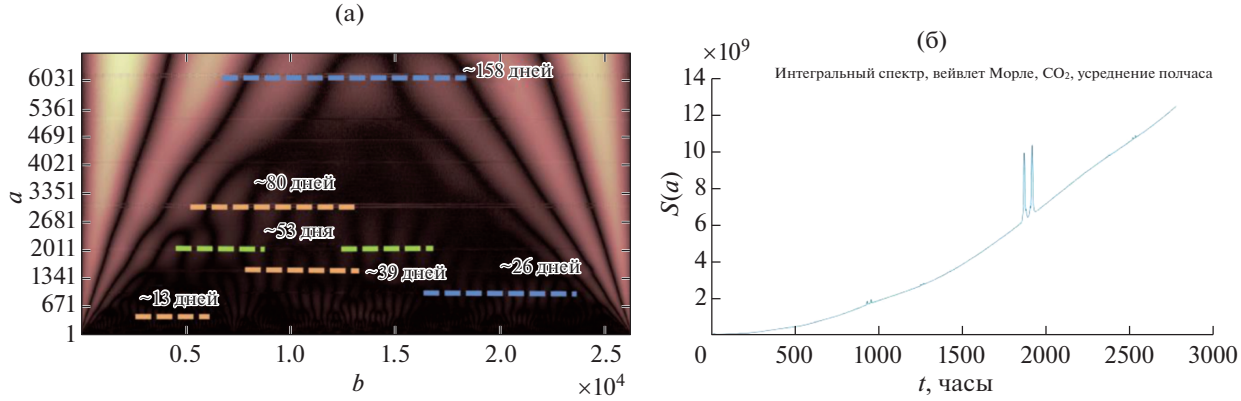
На фиг. 7а представлена картина вейвлет-коэффициентов минимумов и максимумов концентраций углекислого газа, на которой выделены цикличности, что позволяет качественно наблюдать нестационарную структуру временного ряда, показывающего наличие множественных непродолжительных цикличностей.



Фиг. 6. Вейвлет Морле (а) и Гаусса 7 (б).



Фиг. 7. Коэффициенты вейвлет-разложения временного ряда минимальных суточных концентраций CO_2 с базовым вейвлетом Морле (а), интегральный спектр (б).



Фиг. 8. Коэффициенты вейвлет-разложения временного ряда концентраций CO_2, усредненного по получасовым интервалам, с базовым вейвлетом Морле (а), интегральный спектр (б).

Соответствующий интегральный спектр (фиг. 7б) также является достаточно размытым. Можно наблюдать пики в 23, 45, около 80 дней для минимума концентрации, а также общие и для показателей минимумов и максимумов пики в 1.04 и 1.5 года.

Данные временные ряды были также проанализированы с помощью вейвлета Гаусса. Картины вейвлет-коэффициентов для данного материнского вейвлета аналогичны, а интегральные спектры размыты еще сильнее, но некоторые слабые пики совпадают с результатами, полученными с помощью вейвлета Морле.

При анализе данных концентрации углекислого газа за полтора года с усреднением в полчаса также выявлено множество непродолжительных (вносящих вклад в ряд динамики на протяжении только части области исследования) цикличностей (фиг. 8а). В связи с чем для получения более информативных результатов необходимо анализировать отдельные участки временного ряда, которые следует выделить в согласовании с климатическими условиями местности.

7. ЗАКЛЮЧЕНИЕ

Рассмотрены математические методы анализа рядов динамики с целью изучения их цикличности – метод морфологической фильтрации, предложенный в данной статье, метод, основанный на анализе фурье-разложения временного ряда, и метод вейвлет-анализа. Метод морфологической фильтрации позволил выделить суточную циклическую составляющую и анализировать остаток ряда для оценивания составляющих с периодом, меньшим, чем 24 ч. Метод, основанный на фурье-преобразовании, позволил выделить циклические составляющие анализируемого ряда с периодом, как большим, так и меньшим, чем 24 ч. Вейвлет-анализ позволил выделить и локализовать во времени составляющие с периодом, большим, чем 24 ч.

Метод морфологической фильтрации дает возможность преобразовать нестационарный ряд в стационарный, что позволяет провести дальнейшее исследование ряда методами фурье-анализа и статистики. В случае нестационарных рядов выделить циклы также позволяет применение вейвлет-анализа. В нашем случае – это 24, 12, 8 и 6 ч для периодов, не превосходящих суточные. Метод вейвлет-анализа позволил локализовать циклические участки с периодом в несколько дней: 23, 45, около 80 дней для минимума концентрации, а также общие и для показателей минимумов и максимумов на всех высотах пики в 1.04 и 1.5 года.

Периодичность в 1.04 года соответствует в целом годовому циклу изменения концентрации CO_2 и обусловлен сезонной динамикой осадков и фенологическими изменениями в развитии растительности, а в 24 ч соответствует суточному циклу. Объяснение природы циклов с другими периодами требует дальнейших исследований.

Таким образом, предложен набор методов, позволяющих выявить цикличность рядов в широком временном диапазоне.

СПИСОК ЛИТЕРАТУРЫ

1. *Кричевский М.Л.* Временные ряды в менеджменте. Ч. 1, 2. М.: РУСФИНС, 2016.
2. *Пытьев Ю.П.* Морфологические понятия в задачах анализа изображений // Докл. АН СССР. 1975. Т. 224. № 6. С. 1283–1286.
3. *Пытьев Ю.П., Чуличков А.И.* Методы морфологического анализа изображений // М.: Физматлит, 2010. 336 с.
4. *Demin D.S., Chulichkov A.I.* Filtering of monotonic convex noise-distorted signals and estimates of positions of special points // J. of Math. Sci. 2011. V. 172. № 6. P. 770–781.
5. *Astafyeva N.M.* Wavelet analysis: basic theory and some applications // Phys. Usp. 1996. V. 39. P. 1085–1108.
6. *Гончаров А.В.* Исследование свойств знакового представления изображений в задачах распознавания образов // Известия ЮФУ. Технические науки. 2009. Тематический выпуск. С. 178–188.
7. *Каркищенко А.Н.* Исследование устойчивости знакового представления изображений // Автоматика и телемехан. 2010. Т. 9. С. 57–69.
8. *Броневич А.Г., Каркищенко А.Н., Лепский А.Е.* Анализ неопределенности выделения информативных признаков и представлений изображений. М.: Физматлит, 2013. 320 с.
9. *Мясников В.В.* Локальное порядковое преобразование цифровых изображений // Компьютерная оптика. 2015. Т. 39. № 3. С. 397–405.
10. *Terentiev E.N., Farshakova I.I., Prikhodko I.N.* Problems of accurate localization objects in images // AIP Conference Proc. 2019. V. 2171. P. 110009-1–110009-4.
11. *Дещеревская О.А., Авилов В.К., Ба Зуй Динь, Конг Хуан Чан, Курбатова Ю.А.* Современный климат национального парка Кат Тьен (Южный Вьетнам): использование климатических данных для экологических исследований // Геофизические процессы и биосфера. 2013. Т. 12. № 2. С. 5–33.
12. *Dickey D.A., Fuller W.A.* Distribution of the Estimators for Autoregressive Time Series with a Unit Root // J. of the American Statistical Association. 1979. V. 74. P. 427–431.
13. *Вуколов Э.А.* Основы статистического анализа. М.: Форум, 2008. 464 с.
14. *Sallie Baliunas, Peter Frick, Dmitry Sokoloff, Willie Soon.* Time scales and trends in the Central England Temperature data (1659–1990): A wavelet analysis // Geophysical Research Letters. 1997. V. 24. № 11. P. 1351–1354.
15. *Витязев В.В.* Вейвлет-анализ временных рядов. СПб.: Изд-во СПбГУ, 2001.
16. *Асташов Р.А., Голубинский А.Н.* Обоснование выбора материнского вейвлета непрерывного вейвлет-преобразования для анализа речевых сигналов // Вестник Воронежского института МВД России. 2014. № 1. С. 7–15.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1125–1136

ИНФОРМАТИКА

УДК 519.72

ДИНАМИЧЕСКИЕ БАЙЕСОВСКИЕ СЕТИ КАК ИНСТРУМЕНТ ТЕСТИРОВАНИЯ ВЕБ-ПРИЛОЖЕНИЙ МЕТОДОМ ФАЗЗИНГА

© 2021 г. Т. В. Азарнова[1,*], П. В. Полухин[1]

[1] *394018 Воронеж, Университетская пл., 1, ВГУ, Россия*

**e-mail: ivdas92@mail.ru*

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

В работе рассмотрены вопросы моделирования процессов тестирования веб-приложений методом фаззинга с помощью динамических байесовских сетей. Сформулированы основные принципы оптимизации структуры анализируемых динамических байесовских сетей и предложены гибридные алгоритмы обучения и вероятностного вывода с использованием квазиньютоновских алгоритмов и элементов теории достаточных статистик. Библ. 12. Фиг. 3. Табл. 3.

Ключевые слова: динамические байесовские сети, марковский процесс, критерий Шварца, вероятностный вывод, многочастичный фильтр, критерий условной независимости, теорема Рао–Блекквелла–Колмогорова, алгоритм Левенберга–Марквардта, метод Бройдена.

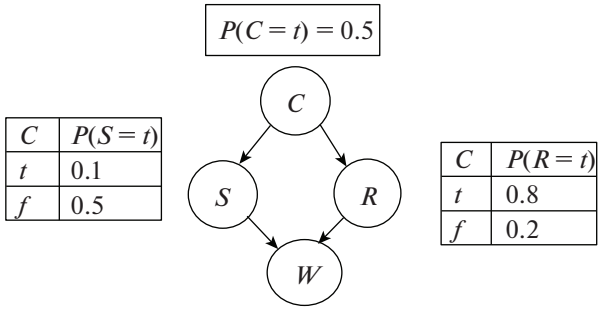
DOI: 10.31857/S004446692107005X

1. ВВЕДЕНИЕ

Предметная область, связанная с разработкой веб-приложений, является достаточно обширной и включает в себя широкий спектр направлений, касающихся проектирования, тестирования и сопровождения веб-приложений. В настоящее время веб-приложения реализуют огромное количество научных, инженерных задач, и успешно используются в процессе организации различных бизнес-процессов. Большинство средств виртуализации, облачных платформ и средств хранения больших данных строятся по архитектуре клиент-сервер и осуществляют механизмы удаленного доступа и управления с помощью сетевых служб и протоколов. Веб-приложения открывают значительные преимущества как для целевых пользователей, так и для создающих их компаний, позволяя систематизировать и самостоятельно производить поддержку и обновление различных функциональных элементов. Несмотря на целый ряд преимуществ использования веб-приложений, они имеют и проблемные зоны. Процесс разработки и функционирования веб-приложений и сервисов сопряжен с возникновением ошибок различного уровня критичности. Многопользовательская эксплуатация веб-приложений может приводить к неконтролируемым потокам входной информации, которые способны вызывать непредсказуемые последствия для устойчивости и безопасности функционирования веб-приложений, в частности, создавать угрозы раскрытия и утечки конфиденциальной информации. Аспекты, связанные с вопросами информационной безопасности, находят широкое отражение в работах российских и зарубежных исследователей, создаются специализированные компании, работающие в области обеспечения безопасности веб-приложений. Особую роль играют открытые проекты, в частности OWASP, специализирующиеся на классификации ошибок, выработке рекомендаций и проектировании механизмов блокирования угроз безопасности веб-приложений. Среди наиболее критичных программных ошибок выделяют: инъекции, межсайтовый скриптинг (XSS) и обход систем аутентификации и управления сессиями. Инъекции классифицируются по типу используемой программной реализации: SQL, инъекции команд и кода [1]. Межсайтовый скриптинг классифицируется по способу внедрения JavaScript кода в веб-страницу – хранимые, отраженные и dom. Механизмы обхода системы аутентификации и управления сессиями реализуют двухэтапные подходы, комбинируя другие методы обнаружения уязвимостей, в особенности, инъекции и межсайтовый скриптинг.

Среди методов обнаружения рассмотренных классов уязвимостей веб-приложений широкое распространение получили методы на основе сканирования, заключающиеся в формировании

специальных тестовых шаблонов. Однако данные методы обладают рядом существенных недостатков, а именно они не позволяют находить аномальные ошибки. Тестирование методом фаззинга было предложено Б. Миллером в качестве альтернативного метода, позволяющего устранить ряд недостатков метода сканирования. Сущность фаззинга [2] заключается в формировании случайного набора тестовых выборок с целью вызова в целевом приложении события сбоя или ошибки и организации мониторинга за реакцией приложения для определения места возникновения анализируемой ошибки. Фаззинг достаточно часто сопоставляют с методом анализа граничных значений, задающим некоторую область допустимых для приложений параметров, и способным отслеживать значения, выходящие за пределы допустимой области. Структурно выделяют несколько групп фаззинга в зависимости от методов тестирования, а также формирования тестовых выборок. С точки зрения формирования выборок, выделяют порождающий и мутационный фаззинг. Основное отличие заключается в том, что мутационный фаззинг осуществляет более осмысленное формирование выборок на базе определенных знаний о структуре или функциях приложения. Среди методов фаззинга выделяют процедуры белого, черного и серого ящиков. Метод белого ящика наиболее применим для обнаружения ошибок, связанных с логикой выполнения программ, обработкой параметров, может использоваться как составной элемент системы анализа покрытия кода тестами. Данный метод широко применяется в процессе решения задач статического анализа кода, используется в системах непрерывной разработки в качестве основного элемента тестирования и поиска программных ошибок. Основной подход, используемый в методе черного ящика, позволяет производить анализ приложения без наличия информации о его внутренних механизмах функционирования. В процессе тестирования методом черного ящика используются случайные генерации наборов входных данных. При этом от тестировщика необходимо лишь иметь общее представление о механизмах функционирования приложения, а также определить набор входных параметров. Существует несколько основных разновидностей тестирования методом черного ящика: эквивалентное разбиение, анализ граничных значений, отладка переходных состояний, функциональные диаграммы, тестирование всех пар значений. Метод серого ящика представляет собой синтез двух описанных выше методов тестирования. Применение данного метода позволяет производить тестирование приложений с частичной информацией о логике обработки параметров, передаваемых между отдельными модулями или подпрограммами. В качестве таких подпрограмм выступают компоненты, предоставляющие механизмы работы с данными, а также реализующие процедуру взаимодействия между процессами и сервисами. С точки зрения моделирования процесса тестирования, тестирование методом серого ящика может быть представлено в виде модели черного ящика с элементами обратной связи. В данном случае обратная связь позволяет формировать управляющие воздействия и своевременно обновлять поток тестовых данных, передаваемый в качестве входных данных рассматриваемой модели, используя ретроспективный анализ. Такой подход позволяет оптимизировать процедуру формирования только тех сценариев, которые позволяют обнаруживать определенные классы ошибок с максимальной вероятностью. Процесс фаззинга серого ящика зачастую носит разделенный во времени характер с поэтапным накоплением статистической информации. Процедуру фаззинга веб-приложений можно представить в виде дискретного или непрерывного стохастического процесса, разделенного на несколько временных срезов. Временные срезы соответствуют периодичности накопления обучающей выборки, используемой для формирования тестовых последовательностей. Применение стохастических процессов для моделирования тестирования методом фаззинга позволяет осуществлять прогнозирование возникновения определенных ошибок при подаче на вход различных типов тестовых данных, а также производить адаптацию тестов для анализа определенного класса программных ошибок за счет применения алгоритмов сглаживания. Эффективным и хорошо апробированным инструментом моделирования стохастических процессов являются динамические байесовские сети (ДБС). Аппарат ДБС включает целый комплекс процедур структуризации, математических методов и алгоритмов, позволяющих моделировать процедуры тестирования методом фаззинга в виде комплексного и адаптивного стохастического процесса. В данной работе рассматриваются различные аспекты применения ДБС для моделирования процессов тестирования веб-приложений методом фаззинга: проектирование ДБС для тестирования определенных уязвимостей, обучения их структуры и параметров, построения процедур вероятностного вывода для достижения целей тестирования. Для реализации процедуры обучения ДБС на нескольких временных срезах используются свидетельства, накопленные до текущего временного среза, формирование транзитивных связей производится на основе проверки гипотез об условной независимости.



C	$P(S=t)$
t	0.1
f	0.5

C	$P(R=t)$
t	0.8
f	0.2

C	R	$P(W=t)$
t	t	0.99
t	f	0.9
f	t	0.9
f	f	0

Фиг. 1. Структура БС с ТУВ (t – истина, f – ложь).

2. ГИБРИДНЫЕ АЛГОРИТМЫ ОБУЧЕНИЯ ДИНАМИЧЕСКИХ БАЙЕСОВСКИХ СЕТЕЙ

Динамические байесовские сети представляют собой разновидность временных моделей и могут быть представлены в виде совокупности статических байесовских сетей (БС), моделирующих исследуемый процесс на определенном временном отрезке. Процедура представления ДБС в виде набора БС является развертыванием сети. БС представляет собой ориентированный ациклический граф с множеством вершин X, представляющих собой дискретные или непрерывные случайные величины, и с множеством дуг G, отражающих отношение родитель–потомок [3]. В качестве множества родительских вершин для множества вершин Y рассматривается множество

$$\text{Parents}(Y) = \{x : \exists y \in Y, \exists(x, y) \in G\}.$$

Множество детей для множества вершин Y определяется в виде:

$$\text{Children}(Y) = \{x : \exists y \in Y, \exists(y, x) \in G\}.$$

В теории статических байесовских сетей принято использовать топологическую нумерацию, при которой родительские вершины получают меньшие номера, чем дети. Существование топологической нумерации для любой байесовской сети доказано. Каждая вершина БС характеризуется таблицей условных вероятностей (ТУВ). Пример простейшей БС приведен на фиг. 1.

При исследовании байесовских сетей важнейшую роль играют гипотезы об условной независимости, факторизации, разделении. Гипотеза об условной независимости представляет собой предположение, что каждый узел y при известных значениях родителей $\text{Parents}(y)$ не зависит от любого множества X, такого, что $x \notin Y$ и $X \subsetneq Y$. Формальное представление гипотезы об условной независимости для байесовской сети имеет следующий вид:

$$P(x_i, y|\text{Parents}(x_i)) = P(x_i|\text{Parents}(x_i))P(y|\text{Parents}(x_i)).$$

Гипотеза о факторизации – это предположение о том, что совместная вероятность есть произведение условных вероятностей каждого узла при известных значениях родителей:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\text{Parents}(X_i)).$$

Гипотеза о разделении представляет собой предположение о том, что для множеств вершин X, Y, Z, $X \cap Y = \oslash$, $X \cap Z = \oslash$, $Z \cap Y = \oslash$ справедливо утверждение о том, что если Z разделяет множества X, Y, то X, Y являются условно независимыми при известных значениях Z. Говорят, что Z разделяет множества X, Y, если оно разделяет все пары вершин, входящие в X и Y, блокируя все маршруты между соответствующими переменными.

Для определения факта условной независимости определенной вершины от всех остальных узлов сети необходимо определить ее марковское покрытие (МП). Под марковским покрытием понимается некоторое множество, в которое входит сама вершина, ее родительские, дочерние вершины, а также множество родителей дочерних вершин. Марковское покрытие для переменной $y \in X$ принято обозначать M^y.

Для динамической байесовской сети обозначим через X_t множество вершин слоя t, а через E_t – множество свидетельств (наблюдаемых переменных) слоя t. Полное совместное распределение вероятностей для ДБС [4] с учетом связей между слоями и свидетельствами на каждом слое можно записать в виде:

$$P(X_0, X_1, \ldots, X_t, E_1, \ldots, E_t) = P(X_0)\prod_{i=1}^{t} P(X_i \mid X_{i-1})P(E_i \mid X_i), \tag{1}$$

где $P(X_0)$ – начальное распределение вероятностей, $P(X_i \mid X_{i-1})$ и $P(E_i \mid X_i)$ – соответственно модели перехода и восприятия.

Выражение (1) выполняется в рамках ограничений, связанных с правилом формирования вероятностных связей при построении модели перехода. Предполагается, что рассматривается марковский процесс I рода и, как следствие, учитываются связи лишь в двух смежных временных срезах.

Рассмотрим используемые в работе гибридные алгоритмы обучения структуры и параметров ДБС. Обучение структуры и параметров ДБС тесно связано с понятиями условной независимости и МП. Обучение структуры применяется для формирования оптимальной топологии сети, учитывающей специфику процессов тестирования веб-приложений.

Для построения структуры статических байесовских сетей широко используется алгоритм минимаксного восхождения [5], который позволяет оптимизировать процедуру построения структуры байесовской сети за счет разделения логики построения базовой топологии сети и определения ее направленности. В качестве оценочной функции используется функция правдоподобия. Проблемные позиции данного алгоритма связаны с тем, что в результате его применения даже для статических байесовских сетей может быть получен локальный оптимум, а также необходимо расширение его функциональности для нахождения структуры динамических байесовских сетей. При применении алгоритма минимаксного восхождения к динамическим байесовским сетям его нужно адаптировать к построению первичной структуры байесовской сети для начального момента времени, и структуры связей для моделей перехода и восприятия. В рамках данного исследования предполагается разработка гибридного алгоритма, позволяющего сочетать статистические подходы к оценке топологии байесовских сетей и алгоритмы определения направленности связей между отдельными узлами байесовских сетей. Для определения топологии предполагается использование алгоритма на основе вычисления марковского покрытия, формирующего модель начального состояния и модель перехода между состояниями.

Методика предлагаемого алгоритма основывается на эвристическом анализе возможных связей, а именно определение множества родительских и дочерних вершин для некоторой вершины z на основе вычисления марковского покрытия M'. В основе вычисления марковского покрытия лежит процесс выполнения тестов на условную независимость. Использование данных тестов обусловлено необходимостью оценки устойчивости связей между дочерними и родительскими узлами. В основе вычисления тестов на условную независимость вершин x_i и x_j при наличии x_k в предложенном алгоритме используется G^2 критерий:

$$G^2 = 2\sum_{a,b,c} N_{i,j,k}^{a,b,c} \ln \frac{N_{i,j,k}^{a,b,c}}{E_{i,j,k}^{a,b,c}} = 2\sum_{a,b,c} N_{i,j,k}^{a,b,c} \ln \frac{N_{i,j,k}^{a,b,c} N_3^c}{N_{i,k}^{a,c} N_{j,k}^{b,c}}, \quad E_{i,j,k}^{a,b,c} = \frac{N_{i,k}^{a,c} N_{j,k}^{b,c}}{N_k^c}, \tag{2}$$

с числом степеней свободы:

$$df = (|D_m(x_i)| - 1)(|D_m(x_j)| - 1)\prod_{x_k} |D_m(x_k)|, \tag{3}$$

где $E_{i,j,k}^{a,b,c}$ и $N_{i,j,k}^{a,b,c}$ – количество всех ожидаемых при выполнении гипотезы об условной независимости и наблюдаемых в обучающей выборке D частот событий, заключающихся в том, что

$x_i = a$, $x_j = b$, $x_k = c$; $D_m(x_k)$ – множество возможных значений (домен значений), которые может принимать узел ДБС x_k.

G-критерий в рамках реализации процедуры обучения структуры ДБС выступает в качестве оценки связи между узлами, позволяя сформировать ненаправленную структуру ДБС. Для определения направленности связей в ДБС могут использоваться различные оптимизационные алгоритмы. Процедура поиска представляет собой рекурсивную операцию по добавлению, удалению или изменению направленности ребер графа, что в свою очередь приводит к изменению параметров графа G. В качестве целевой функции используются статистические критерии на основе логарифма правдоподобия, в частности, критерии Шварца и Акаике. Обобщенное математическое представление данных критериев имеет следующий вид:

$$\begin{gathered}\Phi(M) = L(G, X, D) - mF(N),\\ L(G, X, D) = \log P(D|X) = \sum_{i=1}^{n} \log P(D_i|X),\end{gathered} \tag{4}$$

где m – общее число параметров ДБС, D – обучающая выборка, G – статическая БС, X – переменные, входящие в состав ДБС.

Из обобщенного равенства (4) можно получить выражения для критерия Шварца и Акаике, определяя значение $F(N)$ соответственно равным $F(N) = \log N/2$ и $F(N) = 1$. Задачи локального поиска решаются по отношению к каждому узлу ДБС с целью определения максимума(минимума) оценочной функции, формируемой на основе данных критериев. Одним из достаточно эффективных алгоритмов поиска экстремумов является алгоритм Левенберга–Марквардта (ЛМ, являющегося разновидностью класса регуляризованных алгоритмов на основе метода Гаусса–Ньютона (ГН)). Основным отличием подхода на основе ЛМ от классического ГН является вычисление приближенного значения матрицы Гессе H' с учетом информации относительно вторых производных. Входными данными алгоритма ЛМ является выборка $D = \{(Z_k, Y_k)\}_{k=1}^{n}$, $y_k \in \mathbb{R}^m$ – вектор ожидаемых значений, а также регрессионная модель, задаваемая в виде функции $f(Q, Z_k)$, $Q = (q_1, \ldots, q_m)$ – вектор параметров модели. В результате целевую функцию можно записать в виде

$$E_D(Q) = \frac{1}{2}\sum_{k=1}^{n} [e_k(Q)]^2, \tag{5}$$

где $e_k(Q) = y_k(Q) - f(Q, Z_k)$.

Приближенное значение матрицы Гессе для рассматриваемой задачи будет иметь следующий вид:

$$\begin{gathered}H'(Q) = [J(Q)]^m J(Q) + R(Q),\\ J(Q) = \begin{pmatrix} \dfrac{de_1}{dq_1} & \cdots & \dfrac{de_1}{dq_m} \\ \cdots & \cdots & \cdots \\ \dfrac{de_n}{dq_1} & \cdots & \dfrac{de_n}{dq_m} \end{pmatrix},\end{gathered} \tag{6}$$

где $R(Q)$ – компоненты вторых производных, $J(Q)$ – матрица Якоби, $e(Q) = [e_1, e_2, \ldots, e_n]$ – функция отображения (функция невязки).

В основе алгоритма ЛМ лежит решение системы уравнений относительно приращения градиента ΔQ путем введения дополнительного коэффициента регуляризации λ и аппроксимации компоненты $R(Q)$

$$([J(Q)]^{\mathrm{T}} J(Q) + \lambda I(Q))\Delta Q = -[J(Q)]^{\mathrm{T}} E(Q), \tag{7}$$

где $E(Q) = y - f(Q)$ есть функция отображения (невязки).

В большинстве случаев целесообразно осуществить замену единичной матрицы $I(Q)$ на диагональную приближенную матрицу Гессе $\mathrm{diag}[H'(Q)]$. Это обусловлено тем, что снижение ско-

рости аппроксимации алгоритма ЛМ приводит к увеличению параметра λ. Как следствие, слагаемое $\lambda I(Q)$ теряет свой математический смысл:

$$([J(Q)]^{\mathrm{T}} J(Q) + \lambda \operatorname{diag}[H'(Q)])\Delta Q = -[J(Q)]^{\mathrm{T}} E(Q). \tag{8}$$

Применение алгоритма ЛМ накладывает ряд ресурсных и временных ограничений, связанных с необходимостью перерасчета матрицы Якоби $J(Q)$ на каждом шаге алгоритма. Основное время выполнения алгоритма ЛМ занимает расчет матрицы Якоби для каждой из выполняемых итераций. Для этого в рамках реализации гибридного алгоритма обучения ДБС наиболее оптимальным решением является комбинирование алгоритма ЛМ с методов секущих Бройдена. Метод Бройдена позволяет получить приближенную матрицу Якоби J_{k+1}. Определим соотношение секущих для получения выражения Бройдена

$$J_{k+1}(x_{k+1} - x_k) = F(x_{k+1}) - F(x_k). \tag{9}$$

Введем следующие обозначения: $\alpha = x_{k+1} - x_k$ и $\beta = F(x_{k+1}) - F(x_k)$. Далее определим формулу Бройдена

$$J_{k+1} = J_k + \frac{(\beta - J_k)}{\alpha\alpha^{\mathrm{T}}}\alpha^m. \tag{10}$$

Из выражения (10) следует, что необходимо получить точное значение якобиана J_0 на начальном этапе алгоритма, на последующих этапах формируется лишь приближенное значение якобиана J_{k+1}.

Для оценки транзитивных связей между временными состояниями динамической байесовской сети используется теория марковских цепей. Предполагается, что процесс перехода между данными состояниями представляет собой марковский процесс I рода. Структура выполнения алгоритма разделяется на два шага. Первый шаг характеризуется вычислением множества узлов-кандидатов, которые предположительно могут входить в марковское покрытие M' для текущей переменной z. На следующем шаге происходит удаление вершин, ошибочно добавленных в M'. Это достигается за счет повторного выполнения тестов на условную независимость для каждой переменной z при наличии всех возможных подмножеств M множества M'. Марковское покрытие для переменной z динамической байесовской сети определяется в виде

$$M_{t:t+1} = M_t \cup C_{t+1},$$

где C_{t+1} – множество дочерних вершин переменной z из временного среза $t+1$.

Общая структура алгоритма построения ненаправленной динамической байесовской сети состоит из следующих шагов.

Шаг 1. Задаются начальные значения для выполнения алгоритма: текущая переменная z, множество обучающих данных D, множество кандидатов $M'_{t:t+1} = \varnothing$.

Шаг 2. Выполняются итерации среди всех узлов сети и поиск вершин f с максимальной значимостью устойчивости связи между целевой переменной z и текущей переменной x_i, при анализе всех возможных подмножеств $M^* \subset M'_{t:t+1}$. После чего f добавляется в множество $M'_{t:t+1}$.

Шаг 3. Происходит формирование результирующего марковского покрытия для каждой из вершин путем удаления узлов ошибочно добавленных в $M'_{t:t+1}$, которые определяются за счет проведения G^2 тестов на условную независимость при наличии подмножеств $M^* \subset M'_{t:t+1}$. Если соблюдается гипотеза о разделении, то текущая вершина удаляется из $M'_{t:t+1}$.

В результате выполнения алгоритма мы получаем марковское покрытие для каждого узла динамической байесовской сети, на основе которого строится временная модель для каждого из узлов ДБС. Такой подход позволяет учесть наличие связей между срезами ДБС, а также определить, какие именно вершины имеют связи в соседних временных срезах. Построение ненаправленной структуры динамической байесовской сети не позволяет определить направление связей внутри сети, следовательно, возникает необходимость использования алгоритмов локального поиска. Применительно к семантике динамической байесовской сети в качестве функции оценки используется критерий Шварца или Акаике, напрямую зависящих от добавления, удаления и изменения направленности связей между узлами байесовской сети.

В качестве алгоритма локального поиска, предназначенного для определения направленности связей между узлами сети, предлагается использовать модификацию градиентного вывода, в частности алгоритм Левенберга–Марквардта. Данный алгоритма сочетает в себе алгоритм градиентного поиска и метод Гаусса–Ньютона. Основным отличием подхода на основе Левенберга–Марквардта от классического Гаусса–Ньютона является возможность вычисления приближенного значения матрицы Гессе. Применение алгоритма Левенберга–Марквардта (ЛМ) накладывает ряд ресурсных и временных ограничений, связанных с необходимостью перерасчета матрицы Якоби на каждом шаге алгоритма. Основное время выполнения алгоритма ЛМ занимает расчет матрицы Якоби для каждой из выполняемых итераций. Для преодоления ограничений классического алгоритма ЛМ в рамках реализации гибридного алгоритма обучения ДБС наиболее оптимальным решением является комбинирование алгоритма Левенберга–Марквардта и метода секущих Бройдена. Метод Бройдена позволяет получить приближенную матрицу Якоби.

Можно определить два основных этапа гибридного алгоритма обучения структуры ДБС с использованием алгоритма Левенберга–Марквардта и метода Бройдена. На первом этапе происходит формирование множества узлов-кандидатов в состав МП для каждой из вершин путем выполнения тестов на условную независимость с использованием G^2-критерия. В результате получаем вершины и соответствующие им множества родительских и дочерних вершин для каждого из n временных срезов. На следующем этапе определяем направленности связей между узлами ДБС, применяя алгоритм Левенберга–Марквардта с методом Бройдена. Стоит отметить, что направленность связей между узлами, участвующими в формировании транзитивных связей, имеет смысл только в прямом направлении от среза $t + k$ к срезу $t + k + 1$. Если в результате выполнения алгоритма ЛМ возникают обратные по направлению связи, они должны быть исключены, так как переходный процесс между временными срезами является Марковским, исключающий формирование обратных связей.

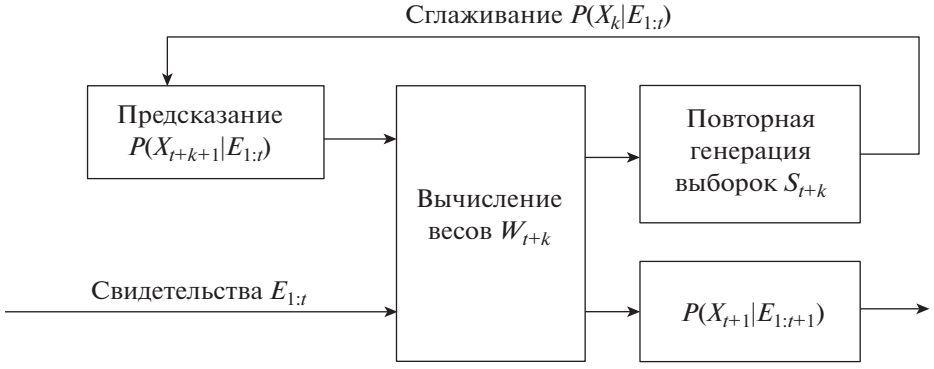
3. ГИБРИДНЫЕ АЛГОРИТМЫ ВЕРОЯТНОСТНОГО ВЫВОДА ДИНАМИЧЕСКИХ БАЙЕСОВСКИХ СЕТЕЙ

Вероятностный вывод является целевым инструментом любой стохастической модели. Применение вероятностного вывода в первую очередь направлено на получение апостериорного распределения при поступлении непрерывного потока свидетельств вплоть до текущего состояния системы. Основной реализацией вероятностного вывода во временных моделях является решение трех основных задач: фильтрация, предсказания и сглаживания. Наиболее эффективными алгоритмами вероятного вывода являются алгоритмы на основе метода Монте-Карло с применением цепей Маркова (МКМЦ), в частности, алгоритмы взвешивания с учетом правдоподобия (ВСП) и многочастичный фильтр (МЧФ).

В алгоритме ВСП осуществляется генерация только тех выборок S для переменных $X = (x_1, x_2, \ldots, x_n)$, которые в полной мере согласуются со свидетельствами E. Такое условие достигается за счет того, что в процессе выполнения вероятностного вывода происходит определение и фиксирование переменных свидетельств E, а формирование выборок осуществляется исключительно для всех оставшихся переменных, запроса Z и состояния X ДБС. В процессе выполнения алгоритма происходит формирование выборок для каждой из переменных состояния развернутой динамической байесовской сети, которые взвешиваются с учетом правдоподобия в соответствии с наблюдаемыми свидетельствами. Исходя из того, что любая переменная z зависит лишь только от родительских вершин, то распределение вероятностей для выборок может быть записано в виде следующего выражения:

$$S_w(Z, E) = \sum_{i=1}^{m} P(z_i \,|\, \mathrm{Parents}(z_i)), \quad E \subset \mathrm{Parents}(z_i). \tag{11}$$

Из формулы (11) следует, что в отличие от первичного распределения вероятностей $P(X)$ по всем переменным X, свидетельство E может вносить вклад в формирования вероятностного распределения выборок S_{ws}, так как E может входить в состав родительских вершин $\mathrm{Parents}(z_i)$. Весовая величина правдоподобия $w(z, E)$ определяется как разница между полученными и ожидаемыми распределениями вероятностей, сформированных для каждой переменной z выборок. Вес $W_{ws}(Z \,|\, E)$ вычисляется для каждой выборки Z и представляет произведение показателей



Фиг. 2. Обобщенная схема МЧФ фильтра.

правдоподобия свидетельств $E_i \in E$, если определено множество родительских вершин для каждой переменной свидетельства [6]:

$$W_{ws}(Z|E) = \sum_{i=1}^{l} P(E_i|\text{Parents}(E_i)). \tag{12}$$

Перемножая выражения (11) и (12), можно получить искомое соотношение для алгоритма взвешивания с учетом правдоподобия

$$S_{ws}(Z,E)W_{ws}(Z|E) = \sum_{i=1}^{m} P(Z_i|\text{Parents}(Z_i))\sum_{i=1}^{l} P(E_i|\text{Parents}(E_i)) = P(Z|E). \tag{13}$$

Классический алгоритм ВСП позволяет получить апостериорное распределение $P(X|E)$ за счет вычисления весов правдоподобия, однако с ростом общего числа переменных запроса и свидетельств, входящих в ДБС, наблюдаем увеличение доли выборок с низкими весами. Применение методов МКМЦ позволяет избежать данных проблем. Каждая новая генерация выборки S_{k+1} формируется на основе внесения случайного изменения в выборку S_k, полученную на предыдущем этапе выполнения метода МКМЦ. Одним из наиболее распространенных методов стохастического вероятностного вывода на основе метода МКМЦ является МЧФ фильтр [7]. Основным преимуществом МЧФ относительно других методов на основе МКМЦ является возможность использования различных подходов для оценки весов выборок, в частности рассмотренного нами метода ВСП. Такой подход позволяет комбинировать случайную генерацию методом Монте-Карло и оценку весов на основе подхода ВСП. В общем виде структура МЧФ фильтра приведена на фиг. 2.

Применительно к ДБС, выполнение МЧФ фильтрации осуществляется с учетом стохастических связей между узлами сети. Для этого поэтапно используются: начальное распределение $P(X_0)$, модель перехода $P(X_{t+1}|X_t)$ и модель восприятия $P(E_{t+1}|X_{t+1})$. Формирование начальной выборки S_t при наличии свидетельств $E_{1:t}$, полученных до текущего состояния, выполняется на основе распределения $P(X_t|E_{1:t})$. Формирование выборок S_{t+1} осуществляется за счет тиражирования свидетельств до момента времени $t+1$. Распределение вероятностей по всем выборкам для моментов времени t и $t+1$ определяется следующим образом [8]:

$$\begin{aligned} N'(X_t|E_{1:t}) &= N \times P(X_t|E_{1:t}), \\ N'(X_{t+1}|E_{1:t}) &= \sum_{X_t} P(X_{t+1}|X_t)N'(X_t|E_{1:t}). \end{aligned} \tag{14}$$

Процедура получения весов каждой из выборок основывается на применении алгоритма ВСП:

$$W(X_{t+1}|E_{1:t+1}) = P(E_{t+1}|X_{t+1})N'(X_{t+1}|E_{1:t}). \tag{15}$$

В результате получаем искомое распределение по всем N выборкам для момента времени $t+1$

$$N'(X_{t+1}|E_{1:t+1}) = N \times P(E_{t+1}|X_{t+1})N'(X_{t+1}|E_{1:t}) = N \times P(E_{t+1}|X_{t+1})\sum_{X_t} P(X_{t+1}|X_t)N'(X_t|E_{1:t}) =$$
$$= N \times P(E_{t+1}|X_{t+1})\sum_{X_t} P(X_{t+1}|X_t)P(X_t|E_{1:t}) = N \times P(X_{t+1}|E_{1:t+1}). \quad (16)$$

Распределение $N'(X_{t+1}|E_{1:t+1})$, представленное в формуле (14), зависит от общего числа выборок, используемых в процессе выполнения алгоритма МЧФ. В идеальном случае для получения требуемой точности алгоритма значение N должно выбираться достаточно большим $N \to \infty$, что накладывает ресурсные и временные ограничения на МЧФ алгоритм. Для оптимизации алгоритма МЧФ и снижения общего числа выборок, необходимых для достижения заданного уровня точности, предлагается использовать теорему Рао–Блеквелла–Колмогорова (РБК).

Сформулируем основные понятия достаточных статистик и теорему РБК. Под достаточной статистикой $T(X)$ будем понимать статистику относительно параметра θ, для которой условное распределение выборки $P(X|T(X))$ не будет зависеть от θ [9].

Если $T(X)$ является достаточной статистикой выборки X, а $T_1(X)$ – некоторая оценка параметра θ, тогда можно определить оценку $T_2(X) = \mathbb{E}_\theta(T_1(X)|T(X))$, для которой справедлива теорема РБК [10]

$$\mathbb{E}_\theta(T_1(X) - T_2(X))^2 \geq 0. \quad (17)$$

Исходя из формулы (11), можно установить соотношение дисперсий для оценок $T_1(X)$ и $T_2(X)$:

$$\mathbb{D}(T_1(X)) = \mathbb{E}(\mathbb{D}(T_1(X)|T(X))) + \mathbb{D}(\mathbb{E}(T_1(X)|T(X))) = \mathbb{E}(\mathbb{D}(T_1(X)|T(X))) + \mathbb{D}(T_2(X)),$$
$$\mathbb{D}(T_2(X)) \leq \mathbb{D}(T_1(X)). \quad (18)$$

Применение теоремы РБК для оптимизации МЧФ заключается в разделении множества переменных запроса X_t на подмножества $X_t' \subset X_t$ и $X_t'' \subset X_t$. В таком случае модель перехода будет иметь следующее представление:

$$P(X_{t+1}|X_t) = P(X_{t+1}'|X_{t:t+1}'', X_t')P(X_{t+1}''|X_t''). \quad (19)$$

В работах Дуста и Рассела [11] предполагается, что компонента $P(X_{t+1}'|E_{1:t-1}, X_{t+1}'')$ может быть определена аналитически еще до начала выполнения алгоритма МЧФ. Неизвестным остается лишь модель $P(X_{t+1}'|E_{1:t+1})$, которую и необходимо вычислить в процессе выполнения этапов МЧФ фильтра. В работе исследовано, что в большинстве случаев, применительно к семантике ДБС, компоненту $P(X_{t+1}'|E_{1:t-1}, X_{t+1}'')$ можно не задавать аналитически, а вычислить также в процессе выполнения фильтрации с помощью МЧФ. В таком случае апостериорное распределение вероятностей для следующего момента времени $t+1$, соответствующее переменным $X_t' \subset X_t$ и $X_t'' \subset X_t$, можно определить на основе цепного правила:

$$P(X_{t+1}', X_{t+1}''|E_{1:t+1}) = P(X_{t+1}''|E_{1:t+1})P(X_{t+1}'|X_{t:t+1}'', X_t')P(X_{t+1}''|X_t''). \quad (20)$$

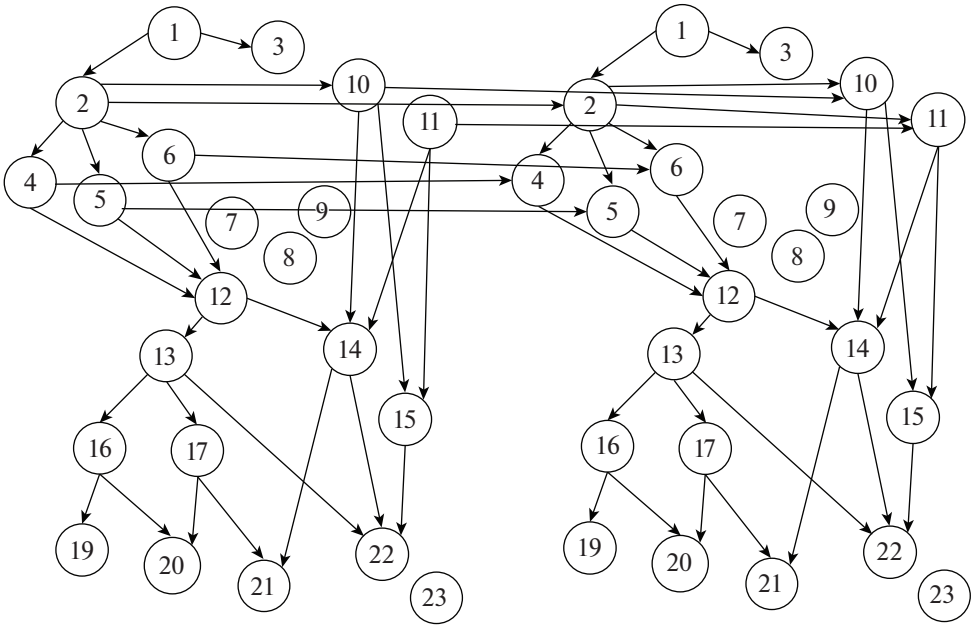
Применительно к алгоритмам ВСП, используемым в процессе определения весов выборок в МЧФ, можно определить оценку этих весов в терминах введенных обозначений:

$$W(X_{t+1}', X_{t+1}''|E_{1:t+1}) = P(E_{t+1}|X_{t+1}'', X_{t+1}')N'(X_{t+1}'', X_{t+1}'|E_{1:t}). \quad (21)$$

С учетом выражения (18) можно определить условие существования выборок для распределений $P(X_{t+1}'|E_{1:t+1})$ и $P(X_{t+1}', X_{t+1}''|E_{1:t+1})$ на основе использования теоремы РБК

$$\mathbb{D}(W(X_{t+1}'|E_{1:t+1})) \leq \mathbb{D}(W(X_{t+1}', X_{t+1}''|E_{1:t+1})). \quad (22)$$

Из неравенства (15) следует, что из апостериорного распределения вероятностей будут исключаться выборки с низким значением среднеквадратического отклонения, это приводит к тому, что вклад будут вносить лишь согласованные выборки. Для получения искомого апостериорного распределения $P(X_{t+1}'|E_{1:t+1})$ для момента времени $t+1$ введем ограничение, связанное с определением транзитивных связей между смежными временными срезами. Транзитивные свя-



Фиг. 3. Обученная структура ДБС фаззинга программных ошибок типа “инъекция”.

зи имеют смысл только между одними и теми же узлами, разнесенными между срезами. Определим распределение $P(X'_{t+1}|E_{1:t+1})$ за счет преобразования выражения (14):

$$P(X'_{t+1}|E_{1:t+1}) = N'(X''_{t+1}, X'_{t+1}|E_{1:t+1})/N. \quad (23)$$

4. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ ПО ПОСТРОЕНИЮ ДИНАМИЧЕСКОЙ БАЙЕСОВСКОЙ СЕТИ ФАЗЗИНГА ИНЪЕКЦИЙ, РЕШЕНИЕ ЗАДАЧ ВЕРОЯТНОСТНОГО ВЫВОДА

В рамках проведения вычислительного эксперимента рассмотрим тестирование систем управления сайтами (СУС). Система управления сайта представляет собой разновидность веб-приложения, предоставляющая расширенный набор по управлению, размещению и хранению различного содержимого. СУС используется как фреймворк для создания интерактивных веб-приложений с целым набором функциональных возможностей, включая механизмы обеспечения безопасности и валидации данных. Среди СУС лидирующую роль занимает WordPress, в рамках исследования остановимся на данном решении. Для обучения ДБС и получения обучающей выборки в качестве целевых приложений используется набор виртуальных машин с развернутыми СУС типа WordPress, начиная с версии 4.3 до 5.2. Система фаззинга строится в виде блочной архитектуры, где происходит автоматическая настройка среды тестирования определенного типа веб-приложений за счет классификации типовых структур приложений тестирования методом черного ящика и формирования списка входных компонентов (параметров), для которых необходимо провести тестирование. Формирование тестовых выборок происходит по результатам выполнения каждого из блоков фаззинга, являющихся узлами ДБС.

Полученная структура ДБС фаззинга “инъекций”, сформированная по результатам тестирования СУС WordPress и обучения на основе алгоритма ЛМ и метода Бройдена, приведена на фиг. 3. Описание узлов ДБС, являющихся блоками фаззинга, представлено в табл. 1.

Фигура 3 показывает, что на основе построенной ДБС формируется направленная система фаззинга. Направления указывают не только на условные зависимости вершин, но и устанавливают последовательность выполнения различных блоков фаззинга, а также взаимосвязь данных блоков. Узлы, не имеющие связей в процессе тестирования, будут отброшены. В построенной ДБС полная условная независимость допустима лишь для корневого узла сети. Необходимо отметить, что если условные связи имеют нулевое значение вероятности в ТУВ, то на следующих этапах тестирования такие блоки фаззинга будут пропущены ввиду их неэффективности к обнаружению ошибок типа “инъекции” для данного типа веб-приложений.

Таблица 1. Характеристика узлов динамической байесовской сети фаззинга инъекций веб-приложений

Узел	Характеристика
1	Определение типа инъекции: SQL, команд, кода
2, 3	Механизмы кодирования обхода межсетевых экранов веб-приложений (WAF)
4, 5, 6, 7, 8, 9	Типы инъекций: Time Based blind, Boolean Based Blind, Error Based Blind, Out of Band, Union injection, Stacked Time
10, 11	Инъекции команд и кода
12	Определение типа и версии СУБД, установленной на сервере
13	Получение структуры таблиц и баз данных СУБД
14	Исполнение команд операционной системы через SQL инъекции
15	Получение доступа к компонентам сети из командного интерфейса СУБД
15	Получения данных, хранящихся в таблицах базы данных
17	Возможность удаленной загрузки файлов, через функции СУБД
19, 20, 21, 22, 23	Нарушение механизмов аутентификации, авторизации, целостности, конфиденциальности и доступности

Таблица 2. Результаты сравнения времени выполнения различных алгоритмов обучения вывода в ДБС “инъекции”

№ п/п	Размер обучающей выборки D	Алгоритм ВС	Алгоритм АМП	Алгоритм ММВ	Алгоритм ЛМ Бройден
1	2000	0.63215 с	0.54231 с	0.49314 с	0.38201 с
2	50000	2.94313 с	2.16543 с	1.85324 с	1.45467 с
3	600000	10.57213 с	8.57732 с	7.02256 с	6.55224 с
4	1000000	18.18432 с	12.25452 с	11.05311 с	8.98432 с
5	10000000	40.09432 с	32.54146 с	28.19356 с	13.95421 с

В рамках проведения процедур обучения и вероятностного вывода используются распределенная вычислительная платформа Apache Hadoop YARN и распределенная файловая система HDFS. Данная платформа имеет в своем составе 6 вычислительных узлов, представленных серверами со следующей аппаратной конфигурацией: 2 процессора Intel Xeon-Platinum 2.5 GHz 16 ядер, 128 GB ОЗУ, жесткий диск 10 ТВ. Размер распределенной файловой системы HDFS 59.5 ТВ, канал связи между узлами обеспечивают скорость до 16 Gb в секунду. Распределенная файловая система Hadoop HDFS используется для хранения обучающих выборок, а также промежуточных данных: таблицы условных вероятностей, весовые распределения выборок, полученные методом МКМЦ. При этом часть данных, используемых в процессе выполнения гибридных алгоритмов обучения и вероятностного вывода, хранится непосредственно в оперативной в виде распределенного множества данных, представленного программной реализацией Apache Hadoop YARN. YARN представляет собой разновидность MapReduce с встроенным планировщиком нагрузки [12], распределения ресурсов и модулем отказоустойчивости, построенным по архитектуре клиент–сервер, с выраженным управляющим узлов (мастер–узел) и клиентскими узлами (рабочий узел). Из 6 узлов вычислительной системы один узел используется нами одновременно в качестве рабочего и мастер-узла. Это дает возможность задействовать все ресурсы 6 узлов в процессе решения задач обучения и вероятностного вывода. Далее в табл. 2 и 3 приведем результаты вычислительных экспериментов по оценке времени выполнения общеизвестных алгоритмов обучения (алгоритм возрастания-сокращения (ВС), инкрементных ассоциаций марковского покрытия (ИАМП), минимаксного восхождения (ММВ)) и вероятностного вывода (Метрополиса–Гастингса (МГ), выборки по значимости (ВЗ), МЧФ) и разработанных гибридных алгоритмов.

Подводя итоги вычислительного эксперимента, отметим, что разработанные алгоритмы обучения и вероятностного ДБС являются ресурсно-эффективными и достаточно легко масштабируются для выполнения на любой из параллельных систем. Отметим, что рассмотренные в ка-

Таблица 3. Результаты сравнения времени выполнения различных алгоритмов вероятностного вывода в ДБС “инъекции”

№ п/п	Размер выборки S	Алгоритм МГ	Алгоритм ВЗ	Алгоритм ВСП	Алгоритм МЧФ	Алгоритм МЧФ РБК
1	2000	0.12311 с	0.13456 с	0.23564 с	0.17654 с	0.10231 с
2	50000	4.33784 с	3.26546 с	3.31231 с	2.26766 с	2.05322 с
3	600000	8.65432 с	7.76532 с	7.81111 с	6.54355 с	5.44355 с
4	1000000	25.23121 с	26.42982 с	22.12941 с	20.31234 с	15.87431 с
5	10000000	56.26332 с	48.21942 с	36.98765 с	31.54328 с	21.12453 с

честве сравнения существующие алгоритмы обучения структуры ВС, ИАМП и ММВ имеют ряд существенных недостатков. Первый из них связан с тем, что они адаптированы для обучения лишь статических БС. Второй недостаток связан с использованием метода восхождения для определения направленностей связи между узлами ДБС. Недостаток заключается в том, что алгоритмы обладают достаточно большой вероятностью попадания оценочных функций в локальный оптимум. Из этого следует, что направленность между узлами ДБС будет задана некорректно, а полученная структура ДБС не может быть в полной мере использована для решения задач вероятностного вывода. Алгоритм ЛМ в сочетании с методом Бройдена позволяет повысить точность расчета экстремумов оценочных функций на основе логарифма правдоподобия, метрик Шварца и Акаике, а также исключить возможность получения некорректной структуры ДБС.

5. ЗАКЛЮЧЕНИЕ

Разработанные в работе алгоритмические и программные решения для оптимизации процедур фаззинга веб-приложений позволяют осуществлять тестирование, обучение и накопление статистических данных. Модели ДБС служат для представления внутренних процессов фаззинга и построения функциональных связей между отдельными блоками фаззинга. Применение такого подхода позволяет осуществлять настройку средств тестирования под специфику анализа определенных групп ошибок. При этом число срезов будет пропорционально количеству настроек или изменений, вносимых в веб-приложения в рамках расширения их функциональных возможностей или совершенствования механизмов защиты. Рассмотрена возможность применения теоремы РБК в рамках многочастичного фильтра, что позволяет оптимизировать МЧФ и гарантировать заданную точность, но при меньшем числе выборок.

СПИСОК ЛИТЕРАТУРЫ

1. *Zalewski M.* The Tangled Web. A Guide to Securing Modern Web Applications. San Francisco: No starch Press, 2012. P. 477.
2. *Саттон М., Амини П.* Fuzzing: исследование уязвимостей методом грубой силы / Пер. с англ. М.: Вильямс, 2009. С. 560.
3. *Тулупьев А.Л., Сироткин А.В., Николенко С.И.* Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. СПб.: Изд-во Санкт-Петербурского ун-та, 2009. С. 400.
4. *Korb A., Nicholson A.* Bayesian Artificial Intelligence. London: Chapman & Hal, CRC Press UK, 2004. P. 244.
5. *Tsamardinos I., Brown L.E., Aliferis C.E.* The max-min hill-climbing Bayesian network structure learning algorithm // Machine Learning. 2006. P. 31–78.
6. *Торопова А.В.* Подходы к диагностике согласованности данных в байесовских сетях доверия // Труды СПИИРАН. 2015. № 43. С. 156–178.
7. *Azarnova T.V., Polukhin P.V.* Advanced hybrid stochastic dynamic Bayesian network inference algorithm development in the context of the web applications test execution // IOP Conf. Ser. 2019. P. 052028–052035.
8. *Russel S., Norvig P.* Artificial Intelligence A Modern Approach. Boston: Prentice Hall, 2009. P. 1095.
9. *Кельберт М.Я., Сухов Ю.В.* Вероятность и статистика в примерах и задачах / Пер. с англ. Основные понятия теории вероятностей и математической статистики. М.: МЦНМО, 2007. С. 241–285.
10. *Колмогоров А.Н.* Теория вероятностей и математическая статистика. М.: Наука, 2005. С. 581.
11. *Deucet A., Freitas N., Murphy K.P., Russel S.* Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks // Proc. of 16th Conf. Uncertainty in AI. 2000. P. 176–183.
12. *Zaharia M., Chowdhury M., Das T., Dave A., McCauley M., Franklin M., Shenker S., Stoica I.* Resilent Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing // NSDI. 2012. P. 1–15.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1137–1148

ИНФОРМАТИКА

УДК 519.72

АВТОМАТИЗИРОВАННЫЙ МЕТОД АНАЛИЗА ДАННЫХ КОСМИЧЕСКИХ ЛУЧЕЙ И ВЫДЕЛЕНИЯ СПОРАДИЧЕСКИХ ЭФФЕКТОВ[1)]

© 2021 г. В. В. Геппенер[1,*], Б. С. Мандрикова[2,]**

[1] *197022 Санкт-Петербург, ул. Профессора Попова, 5, СПбГЭТУ "ЛЭТИ", Россия*

[2] *684034 Камчатский край, Паратунка, Мирная ул., 7, ИКИР ДВО РАН, Россия*

**e-mails: geppener@mail.ru*

***e-mail: 555bs5@mail.ru*

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Предложен автоматизированный метод обнаружения разномасштабных спорадических эффектов по данным наземных станций нейтронных мониторов. Метод включает использование конструкций кратномасштабного анализа и кластерных нейронных сетей типа Learning vector quantization. Обоснован выбор вейвлетов семейств Добеши и Койфлеты на этапе предобработки данных. Предложен алгоритм выбора "наилучшего" аппроксимирующего вейвлет-базиса в классе ортогональных функций. Эмпирическим путем подтверждена эффективность предлагаемого метода для обнаружения мелкомасштабных спорадических эффектов. Показана возможность численной реализации предлагаемого метода для применения в оперативном режиме. Библ. 34. Фиг. 5. Табл. 2.

Ключевые слова: метод анализа данных, нейронные сети LVQ, вейвлет-преобразование, космические лучи, спорадические эффекты.

DOI: 10.31857/S0044466921070061

1. ВВЕДЕНИЕ

Анализ космических лучей проводят при исследовании проблем солнечно-земной физики, а также в решении многих практических задач, связанных с космической погодой. Аномальные события, возникающие на Солнце, в околоземном космическом пространстве (ОКП) находят негативное отражение в работе техносферных систем, а также могут оказывать губительное воздействие на здоровье и жизнь людей [1]. В действительности до сих пор не существует решения, позволяющего получить оперативный и точный прогноз космической погоды [1]. Ключевым моментом в данной области исследования является создание автоматизированных методов анализа данных и своевременного обнаружения аномальных процессов в ОКП. В динамике потока космических лучей различают два типа возмущений – рекуррентные (характерные) вариации и спорадические (аномальные) изменения. Рекуррентные вариации определяются высокоскоростными потоками плазмы из корональных дыр на Солнце. Корональные выбросы (CMEs – coronal mass ejections) служат причиной возникновения спорадических событий в космических лучах. В потоке солнечного ветра корональные выбросы превращаются в межпланетные облака ICMEs. Спорадические эффекты, являющиеся предметом исследования, проявляются в виде форбуш-эффектов [2] и сильных протонных возрастаний Ground Level Enhancement (GLE-событий). Характеристики и параметры форбуш-эффектов определяются многими факторами, и их исследования важны для изучения процессов в ОКП и в задачах, связанных с прогнозом космической погоды. В современном представлении форбуш-эффект является гелиосферным явлением, включающим в себя аномальные понижения интенсивности потока космических лучей, восстановление характерной динамики, а также мелкомасштабные изменения плотности и ани-

[1)]Работа выполнена в рамках Государственного задания по теме "Физические процессы в системе ближнего космоса и геосфер в условиях солнечного и литосферного воздействий" (2021–2023 гг.), № гос. регистрации АААА-А21-121011290003-0.

зотропии космических лучей, возникающие перед началом крупного форбуш-понижения и служащие их предвестниками. Аномальные изменения в межпланетном пространстве и в потоке космических лучей способны вызвать отклик в магнитосфере и ионосфере Земли. GLE-события опасны сильным радиационным излучением и подвергают риску здоровье и жизнь людей [1]. Потоки космических лучей изотропно со всех направлений космического пространства достигают Земли. Попадая в атмосферу, они вступают в реакцию с содержащимися в воздухе атомами азота и кислорода [3]. Результатом реакции становятся расщепление ядер атомов и появление нестабильных элементарных частиц. Данные частицы, регистрируемые в атмосфере, называются вторичными космическими лучами.

Регистрацию и изучение временного ряда данных космических лучей проводят по измерениям сети нейтронных мониторов [4]. Сигнал космических лучей имеет сложную нестационарную структуру, и включает множество различных аномальных эффектов. Регистрируемые вторичные космические лучи содержат высокий уровень шума и зависят от метеорологических факторов и атмосферных явлений, географических координат станции, а также электромагнитной обстановки в Солнечной системе и физических условий в Галактике [5]. Неполные знания о процессах в ОКП существенно затрудняют этап построения методов и моделей анализа данных нейтронных мониторов. Традиционные классические подходы и методы (спектральные методы [7], [8], сглаживающие и регрессионные методы [6], [9]) направлены на изучение устойчивых свойств и характеристик данных, но не являются эффективными для исследования нестационарных аномальных изменений в динамике потока космических лучей. Поскольку мелкомасштабные форбуш-эффекты имеют малый носитель и небольшую амплитуду, их детектирование в шуме является весьма сложной задачей [10]. Современный метод кольца станций дает возможность вычислять характеристики и параметры временного ряда космических лучей с достаточной точностью. Но точность данного метода сильно зависит от плотности станций регистрации, а также ввиду сложных математических расчетов его автоматизация до настоящего момента не реализована. В работе, с целью преодоления описанных проблем и с учетом не изученности явлений, протекающих в ОКП, предложено использовать аппарат искусственных нейронных сетей (ИНС). Аппарат ИНС нашел широкое применение в задачах экстраполяции сложных функций и подтвердил свою эффективность в случае отсутствия полных априорных знаний об исследуемых процессах и их взаимодействиях [11], [13]. Также известным преимуществом ИНС является численная реализация нейросетевых парадигм, обеспечивающая их автоматизацию [14], [15]. Ввиду указанных преимуществ аппарата нейронных сетей, он широко используется для решения геофизических задач [12], [13]. Предлагаемый в статье подход впервые рассмотрен в работе [17], он основан на использовании конструкции ортогонального кратномасштабного анализа (КМА [18], [19]) и кластерных нейронных сетей Learning vector quantization (LVQ [29]). В работах [20]–[23] показано, что объединение аппаратов вейвлет-преобразования и нейронных сетей повышает успешность распознавания образов и аппроксимации сложных функций [23], [24]. Для использования комбинации указанных методов в работе определены и обоснованы применяемые семейства ортогональных вейвлетов. Предложен алгоритм выбора “наилучшего” аппроксимирующего вейвлет-базиса в классе ортогональных функций. Эмпирически подтверждена эффективность предлагаемого метода для обнаружения мелкомасштабных спорадических эффектов.

2. ОПИСАНИЕ МЕТОДА

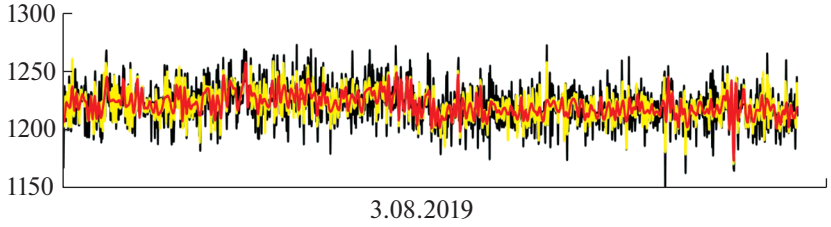
2.1. Применение кратномасштабных ортогональных вейвлет-разложений

Будем рассматривать пространство, порождаемое сдвигами и растяжениями скейлинг-функции ϕ [18], [19]:

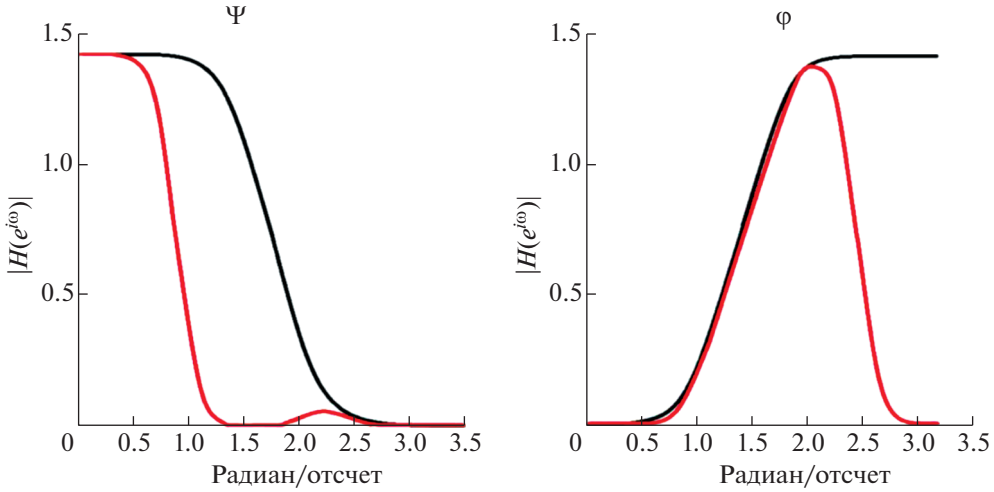
$$V_j = \operatorname{clos}^2_{L(R)}(\phi(2^j t - n)),$$

где $n \in Z$, $L^2(R)$ – пространство Лебега. Тогда, на основе отображения функции f_j в подпространства V_{j-1} и W_{j-1} (пространство W_{j-1} порождается сдвигами и растяжениями вейвлета $\Psi_{j-1,n} = 2^{\frac{j-1}{2}}\Psi(2^{j-1}t - n)$) получим ее представление в виде

$$f^j(t) = g^{j-1}(t) + f^{j-1}(t) = \sum_n d_{j-1,n}\Psi_{j-1,n}(t) + \sum_n c_{j-1,n}\phi_{j-1,n}(t). \quad (2.1)$$



Фиг. 1. Данные НМ ст. “Новосибирск” за 3 августа 2019 г.: черная кривая – первичные данные НМ; желтая кривая – данные с предобработкой (ф. Койфлет 3, разложение до уровня $m = 1$), красная кривая – данные с предобработкой (ф. Койфлет 3, разложение до уровня $m = 2$).



Фиг. 2. АЧХ функций вейвлет Ψ и масштабирующей φ: черный график – на первом уровне разложения; красный график – на втором уровне разложения (использовался вейвлет Койфлет 3).

Применяя рекурсивно операцию (2.1) m раз, получаем

$$f^{j}(t) = g^{j-1}(t) + g^{j-2}(t) + \ldots + g^{j-m}(t) + f^{j-m}(t) = \sum_{k=j-1}^{j-m}\sum_{n} d_{k,n}\Psi_{k,n}(t) + \sum_{n} c_{j-m,n}\phi_{j-m,n}(t). \qquad (2.2)$$

Составляющие в представлении (2.2) определяются как $c_j = \{c_{j,n}\}_{n\in Z} \in V_j$ и $d_j = \{d_{j,n}\}_{n\in Z} \in W_j$, вейвлет-коэффициенты $c_{j,n} = \langle f, \phi_{j,n}\rangle$, $d_{j,n} = \langle f, \Psi_{j,n}\rangle$. Составляющие g^{j-1} являются детализирующими (высокочастотными), а составляющая f^{j-m} сглаженной [18], [26].

На фиг. 1 показан результат представления временного ряда данных нейтронного монитора (НМ) станции Новосибирск в виде (2.2) для уровней вейвлет-разложения $m = 1, 2$. Анализ результатов показывает наличие во временном ряде данных НМ высокого уровня шума и подтверждает эффективность применения операций КМА для его подавления. В работе, следуя результатам [17], использовалось кратномасштабное представление данных НМ для $m = 1$. Оценки в работе [17] показали возрастание погрешности работы сети начиная с уровня вейвлет-разложения $m = 2$. Данный результат свидетельствует о наличии полезной информации в составляющей 2-го уровня вейвлет-разложения. Анализ амплитудно-частотных характеристик (АЧХ) вейвлета и масштабирующей функции (см. фиг. 2) показывает, что на 2-м уровне вейвлет-разложения детектируются колебания в диапазоне 3–14 мин.

2.2. Определение вейвлет-функции и построение “наилучшего” аппроксимирующего базиса

В работе рассматривались библиотеки ортогональных вейвлет-функций, которые позволяют получить численной устойчивости разложения вида (2.2). Определение вейвлет-функции базировалось на следующих критериях, впервые рассмотренных в работе [25].

1. Вейвлет должен иметь большое число нулевых моментов. Число нулевых моментов вейвлета, т.е.

$$\int_{-\infty}^{+\infty} t^k \Psi(t)dt = 0, \quad k = \overline{0, s-1},$$

определяет его способность детектировать особенности функции вида $\alpha \le s$, где α – порядок гладкости.

2. Вейвлет должен иметь малый носитель. Применение операций (2.1), (2.2) порождает возникновение “краевых эффектов” [19]. Величина краевого эффекта определяется по формуле $h_j = 2^j q$, где q – размер носителя вейвлета.

3. Вейвлет должен иметь высокий порядок гладкости α. Данное свойство вейвлета определяет его способность детектировать особенности высокого порядка – вида $\alpha \le s$.

Опираясь на критерии 1–3, необходимо учитывать, что возрастание числа нулевых моментов неизбежно приводит к возрастанию величины носителя функции [19]. Также важным моментом является возможность получить наилучшее приближение функции в составляющей f^{j-m} (см. (2.2)), которое обеспечивается масштабирующей функцией ϕ с большим числом нулевых моментов. Учитывая данные аспекты и опираясь на критерии 1–3, в работе определены семейства вейвлетов Добеши и Койфлеты. Вейвлеты Добеши являются единственным семейством ортогональных вейвлетов, которые имеют наименьший носитель при заданном числе нулевых моментов [27]. Койфлеты – это единственное семейство ортогональных вейвлетов, имеющих наименьший носитель при достаточном числе нулевых моментов в масштабирующей функции ϕ [26]. Для Койфлетов также выполняется следующее важное свойство (см. [26]).

Для любой $f \in C^r$ (C^r пространство r раз непрерывно дифференцируемых функций) в окрестности $2^{-m}n$ при $r \le s$ выполняется условие

$$2^{-m/2}\langle f, \phi_{-m,n}\rangle \approx f(2^{-m}n) + O(2^{-m(r+1)}).$$

Порядок приближения возрастает с ростом s, Койфлет при этом имеет носитель $3s-1$.

Построение “наилучшего” аппроксимирующего базиса в работе основывалось на минимаксном подходе [28], следуя которому погрешность получаемой оценки $\tilde{F}$ определяется в виде

$$(D, F) = \mathbf{E}\{\|F - Df\|^2\},$$

где F – оцениваемый сигнал, f – зашумленные данные, D – оператор решения, $\mathbf{E}$ – математическое ожидание $\|\cdot\|$ – норма. Минимаксный риск – это нижняя граница, вычисленная по всем операторам D [28]:

$$r(\Theta) = \inf_D \sup_{y\in\Theta} \mathbf{E}\{\|F - Df\|^2\}.$$

Рассматривая в качестве оператора D преобразования (2.1), (2.2) и следуя работе [19], оценка $\tilde{F}^\lambda$ в базисе β^λ может быть получена в виде

$$\tilde{F}^\lambda = \sum_n P_T\left(\langle f, \beta_n^\lambda\rangle\right)\beta_n^\lambda,$$

где P_T – пороговая функция. Тогда наилучший базис β^λ – есть базис, который минимизирует погрешность

$$\mathbf{E}\left\{\|F - \tilde{F}^\alpha\|^2\right\} = \min_{\lambda\in\Lambda} \mathbf{E}\left\{\|F - \tilde{F}^\lambda\|^2\right\}.$$

В этом случае определение “наилучшего” базиса может быть основано на выполнении следующих операций.

Шаг 1. На основе отображения (2.2) выполняем преобразование функции f:

$$f^j(t) = g^{j-1}(t) + g^{j-2}(t) + \ldots + g^{j-m}(t) + f^{j-m}(t) = \sum_{k=j-1}^{j-m}\sum_n d_{k,n}\Psi_{k,n}(t) + \sum_n c_{j-m,n}\phi_{j-m,n}(t).$$

Шаг 2. Путем применения пороговых функций P_{T_j} (см. (2)) получаем оценку

$$\tilde{F}^m = \sum_{k=j-1,j-m} \sum_n P_{T_j}(d_{k,n})\Psi_{k,n} + \sum_n P_{T_{j-m}}(c_{j-m,n})\varphi_{j-m,n},$$

где $T_j = \mathrm{Med}\left(\left\langle f, \beta^\lambda \right\rangle\right)$, $\beta^\lambda = \{\Psi_{k,n}, \phi_{j-m,n}\}_{k=j-1,\ldots,j-m}$, Med – медиана.

Шаг 3. Оцениваем величину

$$Q_m^\lambda = \sum_{n \in I_M^j} |c_{j-m,n}|^2 + \sum_{k=j-1,j-m} \sum_{n \in I^j} |d_{k,n}|^2,$$

где множество индексов I^j: $n \in I^j$, если $\left|\left\langle f, \beta^\lambda \right\rangle\right| \geq T_j$, и определяем “наилучший” базис β_m^α: $Q_m^\alpha = \max_{\lambda \in \Lambda} \mathbf{E}\{Q_m^\lambda\}$.

2.3. Принцип работы кластерной нейронной сети типа Learning vector quantization и схема решения задачи

Предлагаемый метод основан на применении кластерных нейронных сетей типа Learning Vector Quantization (LVQ) [29], [30]. Построение LVQ-сети включает определение числа кластеров l (количество нейронов в первом слое) и числа классов k (количество нейронов во втором слое), а также определение принадлежности каждого кластера классу [27]:

$$F_l = \sum_k w_{kl} y_k,$$

где w_{kl} – веса нейрона l второго слоя сети, связанного с нейроном k первого слоя, y_k – выходное значение нейрона k первого слоя сети.

В соответствии с решаемой задачей логично определить следующие $L = 3$ класса нейронной сети.

1. “Спокойный” класс – отсутствие спорадических эффектов. Признаками класса являются: (1) отсутствием активных пятен и вспышек на Солнце; (2) отсутствием потока солнечного ветра с видимой стороны по линии с Землей; (3) спокойная геомагнитная обстановка.

2. “Слабовозмущенный” класс – наличие мелкомасштабных спорадических эффектов. Признаками класса являются: (1) незначительные вспышки на Солнце, направленные на Землю; (2) слабые возмущения в магнитосфере.

3. “Возмущенный” класс – наличие крупномасштабных спорадических эффектов и/или GLE-событий. Признаками класса являются: (1) проникновением в окрестности Земли высокоскоростных потоков солнечного ветра и/или связанной с ним ударной волны; (2) сильные возмущения в магнитосфере.

Число кластеров сети определялось путем минимизации апостериорного риска и принято равным 20.

Кластеризация входных данных LVQ-сети базируется на применении метода наименьших квадратов:

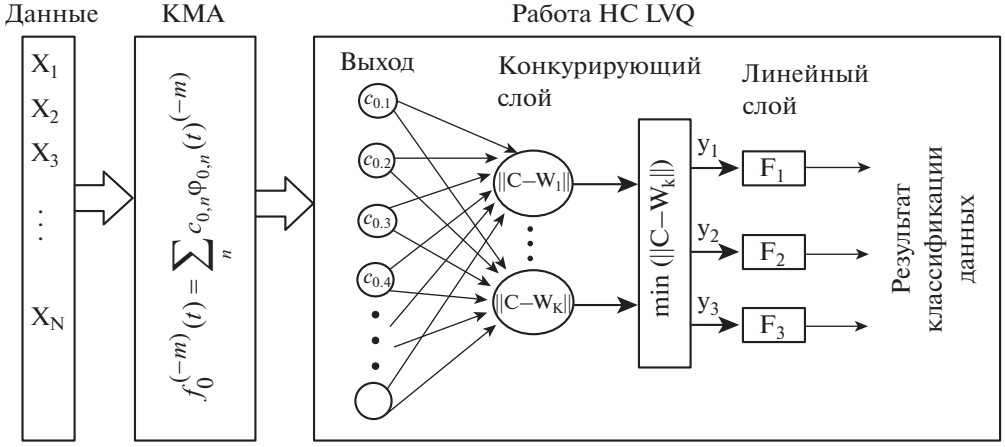
$$d_k = d(C, W_k) = \|C - W_k\| = \sqrt{\sum_{i=1}^{I} (c_i - w_{ik})^2},$$

где C – входной вектор; W_k – вектор весов нейрона k первого слоя сети, l – размерность входного вектора сети. В процессе работы сети в первом ее слое путем оценки расстояния d_k определяется нейрон-победитель p:

$$D = d_{\min}(C, W_k) = \min_k \|C - W_k\|.$$

В процессе функционирования сети один элемент выходного вектора равен 1, остальные – нулю. Таким образом, сеть решает задачу классификации.

В соответствии с предлагаемым подходом, решение задачи классификации данных нейтронных мониторов может быть представлено в виде схемы, показанной на фиг. 3. Для восстановления исходного разрешения функции выполняется операция вейвлет-восстановления:



Фиг. 3. Общая схема решения задачи.

$f_0^{(-m)}(t) = \sum_n c_{0,n}\varphi_{0,n}(t)^{(-m)}$, (верхний индекс $(-m)$ соответствует разрешению функции до выполнения операции вейвлет-восстановления).

Для оценки метода использовались минутные данные наземных станций нейтронных мониторов [32]. Определение классов нейронных сетей основывалось на анализе геомагнитных индексов – *A*, *K* и *Dst* индексы [33]. “Спокойный” класс формировался из данных за периоды, в которые *A*-индекс был менее 7, *K*-индекс был менее 3, *Dst*-индекс находился в пределах ±4. “Слабовозмущенный” класс (наличие мелкомасштабных спорадических эффектов) формировался из данных за периоды, в которые *A*-индекс был менее 18, *K*-индекс был менее 5, *Dst*-индекс находился в пределах ±8. “Возмущенный” класс (наличие крупномасштабных спорадических эффектов и/или GLE-событий) включал периоды, в которые *A*-индекс был менее 18, *K*-индекс был больше 4, *Dst*-индекс превышал ±8. Для периодов высокой и низкой солнечной активности (солнечная активность определялась по значениям индексов f10.7 [33]) сети обучались отдельно. В разложениях использовались вейвлет-функции семейств Добеши и Койфлеты (выбор семейств обоснован в п. 2.2). Входные векторы сети, следуя работе [17], имели длительность, равную трем суткам. Разложения (2.2) выполнялись для $m = 1$ [17], [34]. Перед подачей на вход сети выполнялось восстановление исходного разрешения функций на основе операции обратного вейвлет-преобразования. С целью уменьшения краевого эффекта выполнялось зеркальное дополнение функций. Оценки показали, что наименьшую погрешность сети позволяют получить вейвлеты Добеши 3-го порядка (db3) и Койфлеты 3-го порядка (coif3). Результаты оценок (см. табл. 1) подтвердили эффективность метода, включающего совместное использование кластерных нейронных сетей LVQ и процедуры ортогональных кратномасштабных вейвлет-разложений (представление (2.2)). Анализ результатов показал зависимость динамики космических

Таблица 1. Результаты работы построенных сетей

Входной сигнал сети		Первичные данные НМ	db3_1	coif3_1
Работа сети LVQ1 (данные за период высокой солнечной активности)	“Спокойный” класс	100%	100%	100%
	“Слабовозмущенный” класс	80%	87%	93%
	“Возмущенный” класс	93%	93%	93%
Работа сети LVQ2 (данные за период низкой солнечной активности)	“Спокойный” класс	73%	–	80%
	“Слабовозмущенный” класс	75%	–	75%
	“Возмущенный” класс	67%	–	83%

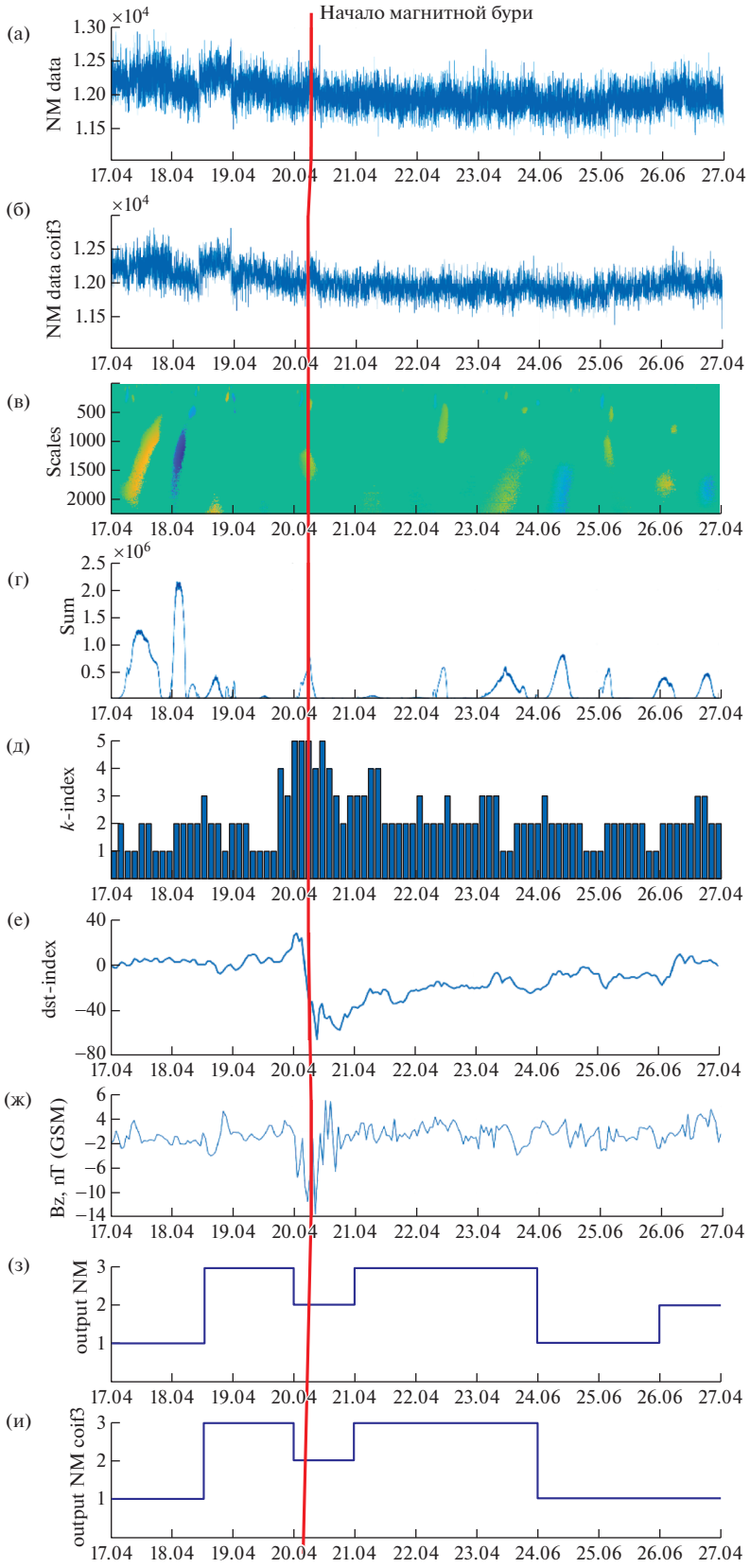
лучей от уровня солнечной активности. В период высокой солнечной активности погрешность метода не превышала 7%. В период низкой солнечной активности погрешность возросла до 21%. Поскольку мерой возмущенности потока ГКЛ является величина отклонения вариаций от характерного уровня [16], очевидно, в периоды низкой солнечной активности шкалы амплитуд вариаций имеют меньший размах. Для повышения эффективности метода в период низкой солнечной активности, вероятно, требуется увеличить размер обучающей выборки. Также повысить эффективность метода, возможно, позволит увеличение числа анализируемых станций.

3. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ МЕТОДА

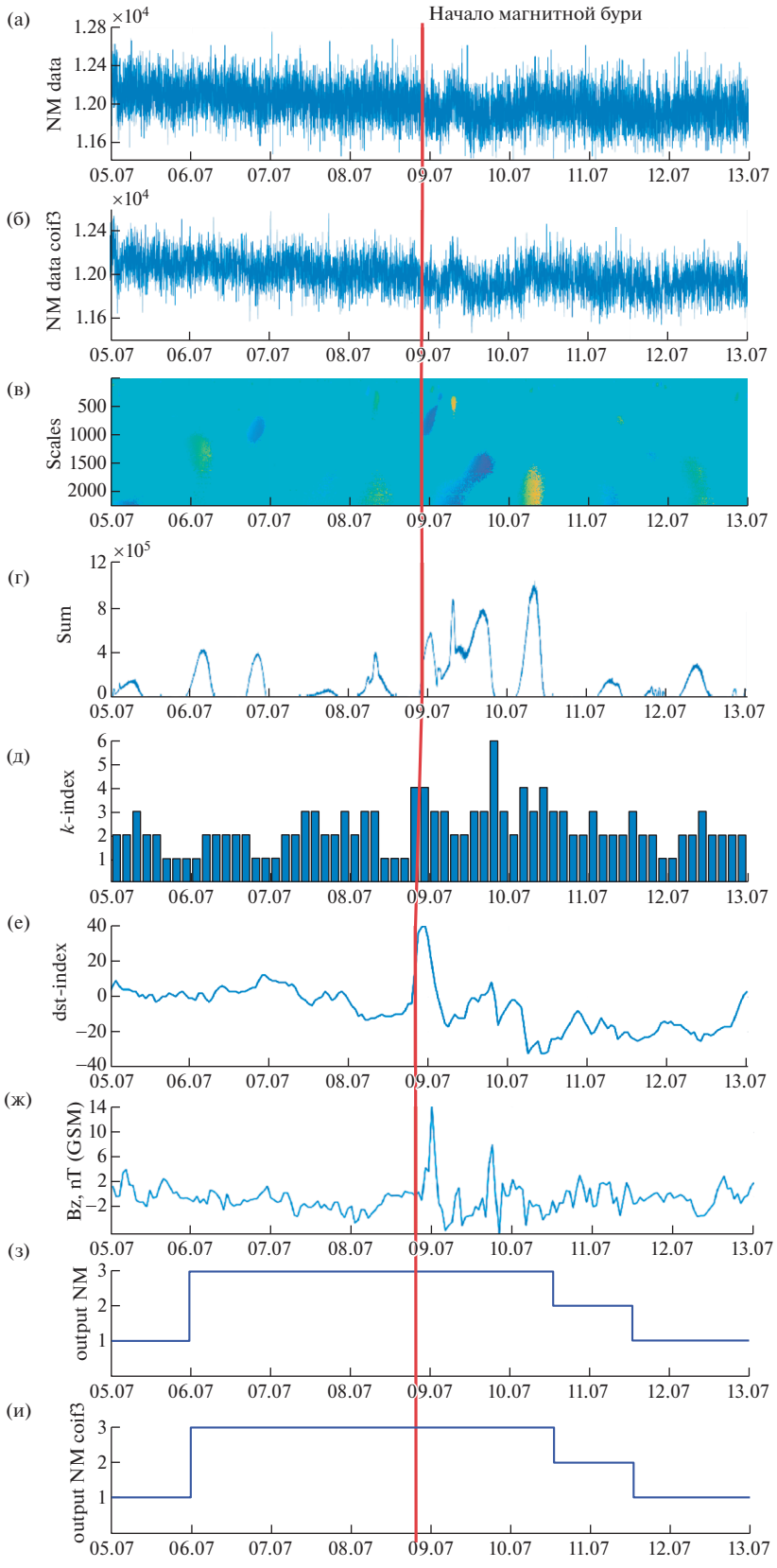
Результаты применения метода показаны на фиг. 4 и 5. Магнитная буря 20 апреля 2018 г. (фиг. 4) вызвана неоднородным скоростным потоком из корональной дыры [31]. Накануне события 17–19 апреля скорость солнечного ветра (ССВ) была в окрестности 300 км/с, флуктуации компоненты межпланетного магнитного поля (ММП) составляли $Bz = \pm 5$ nT (фиг. 4ж, [31]). Неоднородный ускоренный поток от корональной дыры (CIR) пришел в конце суток 19 апреля, 20 апреля флуктуации компоненты ММП усилились до $Bz = \pm 19$ nT, ССВ возросла до 650 км/с и оставалась в этих пределах до конца суток 21 апреля. Далее, с 22 апреля в связи с ослаблением влияния корональной дыры, ССВ уменьшилась до 350 км/с, флуктуации ММП уменьшились до $Bz = \pm 5$ nT. Результаты обработки данных показывают накануне события 18 апреля возникновение крупномасштабных аномальных изменений в динамике космических лучей (см. фиг. 4з, и). Отметим, что момент возникновения форбуш-эффекта совпадает с моментом увеличения размаха флуктуаций ММП, и в этот период наблюдается возрастание K-индекса. Аномальные изменения в данных нейтронного монитора 18 апреля также показывают результаты применения пороговых функций (фиг. 4в, г, алгоритм описан в Приложении). За несколько часов до события амплитуда форбуш-эффекта значительно увеличилась, форбуш-понижение наблюдалось в начальную и основную фазы бури. Плавное восстановление уровня космических лучей произошло по данным обработки к началу суток 26 апреля (“слабовозмущенный” класс). Заметим, что в период мелкомасштабных аномальных изменений в динамике КЛ 21–26 апреля (результаты порогового алгоритма, см. фиг. 4в, г) происходили повышения K-индекса (фиг. 4д). Сопоставление результатов работы сети с данными ОКП указывает на достоверность решений нейронной сети и подтверждает эффективность метода.

Результаты применения метода в период слабой магнитной бури 9 июля 2019 г. показаны на фиг. 5. По данным космической погоды [32] накануне события 05 июля ССВ возросла до 457 км/с. Около 18.30 UT 08 июля пришел неоднородный ускоренный поток от корональной дыры (CIR), ССВ к концу суток 08 июля возросла до 390 км/с, флуктуации компоненты ММП усилились до $Bz = \pm 8$ nT. Во время магнитной ССВ достигла значения 695 км/с, размах флуктуаций ММП увеличился до $Bz = \pm 11$ nT. В конце периода 11–13 июля ССВ находилась в пределах $v = 400$–500 км/с, размах флуктуаций компоненты ММП составлял от $Bz = \pm 4$ nT до $Bz = \pm 6$ nT. Результаты работы сети (фиг. 5з, и) показывают изменение состояния потока космических лучей с начала суток 06 июля (“возмущенный” класс) до 12:00 UT 10 июля. Результаты порогового алгоритма (фиг. 5в, г) соответствуют результатам работы сети и показывают возникновение разномасштабных аномальных изменений в динамике космических лучей, которые достигают максимальных значений в период события. С 12:00 UT 10 июля по 12:00 UT 11 июля сеть классифицировала как “Слабовозмущенный” (класс 2), в конце периода динамика потока космических лучей восстановилась (“Спокойный” класс). Сопоставление результатов нейронной сети с данными межпланетного пространства и результатами порогового алгоритма подтверждает их достоверность.

Анализ результатов применения метода в возмущенные периоды показывает высокую частоту возникновения спорадических эффектов в космических лучах в преддверии магнитных бурь (см. табл. 2). Для основных фаз магнитных бурь характерно возникновение форбуш-понижений, длительность которых по данным табл. 2 может составлять от нескольких часов до нескольких суток. В работах [17], [34] детально рассмотрены результаты метода в периоды высокой солнечной активности, которые также подтверждают его эффективность для обнаружения спорадических эффектов в динамике космических лучей.



Фиг. 4. (а) – сигнал НМ ст. Москва 2018 г., (б) – сигнал НМ с предобработкой ф. Койфлет 3, разложение до уровня $m = 1$, (в) – применение порогового алгоритма (см. Приложение), положительные аномалии изображены желтым, отрицательные – синим, (г) –интенсивность аномальных изменений (см. Приложение), (д) – *k*-индекс, (е) – *Dst*-индекс, (ж) – *Bz* компонента ММП, (з) – работа НС LVQ, (и) – работа НС LVQ2_ coif3_1.



Фиг. 5. (а) – сигнал НМ ст. Новосибирск 2019 г., (б) – сигнал НМ с предобработкой ф. Койфлет 3, разложение до уровня $m = 1$, (в) – применение порогового алгоритма (см. Приложение), положительные аномалии изображены желтым, отрицательные – синим, (г) – интенсивность аномальных изменений (см. Приложение), (д) – k-индекс, (е) – Dst-индекс, (ж) – Bz компонента ММП, (з) – работа НС LVQ, (и) – работа НС LVQ2_ coif3_1.

Таблица 2. Результаты применения метода в возмущенные периоды

Анализируемые аномальные события (периоды, станция)	Выявленные аномалии накануне события (класс/время до бури)		Основной период события (класс)	Период восстановления (класс)
10.07.13–16.07.13 Кингстон	2/24 ч	3/12 ч	3	2
15.03.15–20.03.15 Кингстон	2/48 ч	3/12 ч	3	2
16.01.16–22.01.16 Кингстон	2/24 ч		3	2
21.08.18–28.08.18 Москва	2/18 ч		3	2
12.03.18–19.03.18 Москва	2/48 ч	3/24 ч	3	2
17.04.18–26.04.18 Москва	3/24 ч		3	1
17.04.18–26.04.18 Новосибирск	3/24 ч		3	1
4.10.18–11.10.18 Москва	2/68 ч	3/9 ч	3	2
5.07.19–12.07.19 Москва	2/68 ч	3/12 ч	3	1
5.07.19–12.07.19 Новосибирск	3/68 ч		3	1
4.06.19–11.06.19 Москва	2/9 ч		2	1
11.04.14–16.04.14 Инувик	3/12 ч		3	2
11.04.14–16.04.14 Thul	3/32 ч		3	1
12.09.14–16.09.14 Thul	3/24 ч		3	2
12.09.14–16.09.14 Инувик	2/60 ч		3	2
11.04.14–16.04.14 Москва	2/48 ч		3	1
12.09.14–16.09.14 Южный полюс	2/24 ч		3	1
5.09.14–7.09.14 Инувик	2/24 ч		3	2
10.09.14–13.09.14 Инувик	2/24 ч		3	2
29.08.17–01.09.17 Моусон	2/27 ч		2	2
5.09.17–09.09.17 Моусон	3/48 ч		3	3

4. ВЫВОДЫ

Предложенный в работе метод анализа данных космических лучей подтвердил свою эффективность в задачах обнаружения разномасштабных спорадических эффектов в динамике космических лучей. Эмпирически доказана результативность совместного применения конструкции ортогонального кратномасштабного анализа с кластерными нейронными сетями векторного квантования. Предложен алгоритм определения “наилучшего” аппроксимирующего вейвлет-базиса в классе ортогональных функций, основанный на минимаксном подходе. Подтверждена применимость метода для детектирования мелкомасштабных спорадических эффектов.

Результаты оценки метода показали его высокую результативность в период высокой солнечной активности – погрешность метода составила 7%. В период низкой солнечной активности флуктуации космических лучей имеют меньший размах, что усложняет задачу детектирования аномальных особенностей и, как следствие, погрешность метода возрастает (по результатам исследования до 21%). На примере магнитных бурь 2018–2019 гг. по измерениям данных разных станций показана возможность применения метода в оперативном режиме.

В будущем авторы планируют продолжить исследование в направлении расширения спектра анализируемых станций регистрации данных космических лучей и увеличения статистического материала.

ПРИЛОЖЕНИЕ

Алгоритм выделения аномалий в динамике космических лучей и оценки их интенсивности [16], [22].

Шаг 1. Выполнение непрерывного вейвлет-преобразования:

$$(W_\Psi f_{b,s}) := |s|^{-1/2}\int\limits_{-\infty}^{+\infty} f(t)\Psi\left(\frac{t-b}{s}\right)dt, \quad f \in L^2(R), \quad s,b \in R, \quad s \neq 0.$$

Шаг 2. Применение пороговой функции P_{T_s}:

$$P_{T_s}(W_\Psi f_{b,s}) = \begin{cases} W_\Psi f_{b,s}, & \text{если} \quad (W_\Psi f_{b,s} - W_\Psi f_{b,s}^{\text{med},l}) \geq T_s^l, \\ 0, & \text{если} \quad \left|W_\Psi f_{b,s} - W_\Psi f_{b,s}^{\text{med},l}\right| < T_s^l, \\ -W_\Psi f_{b,s}, & \text{если} \quad (W_\Psi f_{b,s} - W_\Psi f_{b,s}^{\text{med},l}) < -T_s^l, \end{cases}$$

где $W_\Psi f_{b,s}^{\text{med},l}$ – медианное значение, рассчитанное в скользящем временном окне длины l, $T_s^l = U\sigma_s^l$ – порог,

$$\sigma_s^l = \sqrt{\left(\frac{1}{l} - 1\sum_{k=1}^{l}(W_\Psi f_{b,s} - \overline{W_\Psi f_{b,s}})\right)^2}$$

есть стандартное отклонение, рассчитанное в скользящем временном окне длины l, $W_\Psi f_{b,s}$ – среднее значение, U – пороговый коэффициент.

Шаг 3. Оценка интенсивности аномалий: $\text{sum}(t) = \sum\limits_s P_{T_s}(W_\Psi f_{b,s})$, которая в случае локального повышения КЛ будет положительной, а в случае локального понижения – отрицательной.

Авторы выражают благодарность институтам, выполняющим поддержку станций нейронных мониторов, которые использовались в работе.

СПИСОК ЛИТЕРАТУРЫ

1. *Eroshenko E.A., Belov A.V., Kryakunova O.N., Kurt V.G., Yanke V.G.* The alert signal of GLE of cosmic rays // In proceedings of the 31st ICRC. 2009.
2. *Forbush S.E.* On cosmic ray effects associated with magnetic storms // Eos, Trans Am Geophys Union. 1938. V. 19. P. 193–193. https://doi.org/10.1029/TR019i001p00193-1
3. *Топтыгин И.Н.* Космические лучи в межпланетных магнитных полях М.: Наука, 1938. 304 с.
4. Real time data base for the measurements of high-resolution Neutron Monitor. [Электронный ресурс] – Режим доступа: www.nmdb.eu (01.11.2019).
5. *Баренбаум А.А.* Галактика, Солнечная система, Земля. Соподчиненные процессы и эволюция. М.: ГЕОС, 2002. 394 с.
6. *Mishev A., Usoskin I.* Application of a full chain analysis using neutron monitor data for space weather studies // 25th European Cosmic Ray Symposium (ECRS 2016), Turin, Italy, September 04–09, 2016.
7. *Vipindas V., Gopinath S., Girish T.E.* Periodicity analysis of galactic cosmic rays using Fourier, Hilbert, and higher-order spectral methods // Astrophys Space Science. 2016. V. 361. 18 p. https://doi.org/10.1007/s10509-016-2719-y
8. *Livada M., Mavromichalaki H., Plainaki C.* Galactic cosmic ray spectral index: the case of Forbush decreases of March 2012 // Astrophys. Space Science. 2018. V. 363. P. 8. https://doi.org/10.1007/s10509-017-3230-9
9. *Ni Sulan, Gu B.H., Zhiyi.* Interplanetary coronal mass ejection induced forbush decrease event: simulation study with one-dimensional stochastic differential method. 2017. V. 63. P. 1–8. https://doi.org/10.7498/aps.66.139601
10. *Kota J., Jokipii J.R.* The role of corotating interaction regions in cosmic-ray modulation // Geophys. Res. Lett. 1991. V. 18. P. 1797–1800.
11. *Belov A.V. et al.* Cosmic ray anisotropy before and during the passage of major solar wind disturbances / A.V. Belov, J.W. Bieber, E.A. Eroshenko, P. Evenson, R. Pyle, V.G. Yanke // Adv. Space Res. 2003. V. 31. № 4. P. 919–924.

12. *Shimelevich M.I., Obornev E.A.* Application of the neural network method for approximating inverse operators in electromagnetic sounding problems // Izv. Universities Geology and exploration. 1999. № 2. P. 102–106.

13. *Baldin N.P.* Investigation of forecasting convergence by neural networks with feed-back // Machine learning and data analysis. 2011. V. 1. № 1. P. 61–76.

14. *Golovko V.A.* Neural networks: training, organization and application. Moscow: IPRZhR, 2001.

15. *Mandrikova O.V., Polozov Yu.A., Solovev I.S., Fetisova(Glushkova) N.V., Zalyaev T.L., Kupriyanov M.S., Dmitriev A.V.* Methods of Analysis of Geophysical Data during Increased Solar Activity // Pattern recognition and image analysis (advances in mathematical theory and applications). 2016. V. 26. № 2, P. 406–418. https://doi.org/10.1134/S1054661816020103

16. *Mandrikova O.V. et al.* Analysis of the Cosmic Rays dynamics on the basis of Neural Networks / O.V. Mandrikova, T.L. Zalyaev, B.S. Mandrikova, M.S. Kupriyanov // Proceedings of 2018 21th IEEE International Conference on Soft Computing and Measurements (SCM 2018). 2018. V. 361. P. 683–686.

17. *Mandrikova O.V., Geppener V.V., Mandrikova B.S.* Method of analysis of cosmic ray data based 363 on neural networks of LVQ // J. Physics: Conference Series (JPCS). 2019. V. 1374. https://doi.org/10.1088/1742-6596/1368/5/052026

18. *Chui C.K.* An introduction in wavelets. New York: Academic Press, 1992. P. 264.

19. *Mallat S.* A wavelet tour of signal processing. London: Academic Press, 1999. P. 620.

20. *Mandrikova O.V., Zhizhikina E.A.* Automatic method for estimation of geomagnetic field state // Computer Optics. Number Special. 2016. V. 39. № 3. P. 420–428.

21. *Mandrikova O.V., Solovev I.S., Zalyaev T.L.* Methods of analysis of geomagnetic field variations and cosmic ray data // Earth Planet Space. 2014. V. 66. https://doi.org/10.3711186/s40623-014-0148-0

22. *Mandrikova O., Polozov Yu., Fetisova N., Zalyaev T.* Analysis of the dynamics of ionospheric parameters during periods of increased solar activity and magnetic storms // Journal of Atmospheric and Solar-Terrestrial Physics. 2018. V. 181. P. 116–126. https://doi.org/10.1016/j.jastp.2018.10.019

23. *Mandrikova O., Polozov Yu., Geppener V.* Method of ionospheric data analysis based on a combination of wavelet transform and neural networks // Procedia Engineering. 2017. V. 201. P. 756–766. https://doi.org/10.1016/j.proeng.2017.09.622

24. *Burikov S.A., Efitorov A.O., Dolenko T.A., Shirokiy V.R., Dolenko S.A.* Solving inverse problems of Raman spectroscopy of aqueous salt solutions using wavelet-neural networks // Siberian Journal of Physics. 2018. V. 13. № 3. P. 101–109.

25. *Mandrikova O.V.* Approximation and Analysis of Ionospheric Parameters Based on a Combination of Wavelet Transformation and Neural Networks Groups / O.V Mandrikova, Yu.A. Polozov // Informatsionnye tekhnologii. 2014. № 7. P. 61–65.

26. *Daubechies I.* Ten Lectures on wavelets. SIAM, Philadelphia. 1992.

27. *Hammer B., Villmann T.* Generalized relevance learning vector quantization // Neural Networks. 2002. V. 5. P. 1059–1068.

28. *Mertens J.-F., Neyman A.* Minimax Theorems for Undiscounted Stochastic Games // Game Theory and Mathematical Economics. 1981. P. 83–87.

29. *Kohonen T.* "Self-organizing maps". 3 Ed. Tokyo: Springer, 2001. P. 501.

30. *Bertin E., Bischof H., Bertolino P.* Voronoi pyramids controlled by Hopfield neural networks // Comput. Vision-Image Understand. 1996. V. 63. № 3. P. 462–475.

31. Indices of geomagnetic activity [Электронный ресурс] – Режим доступа: http://geobrk.adm.yar.ru/database/indices/index?lang=ru (11.11.2019).

32. Forecast of space weather according to the data of Federov Institute of Applied [Электронный ресурс] – Режим доступа:http://ipg.geospace.ru (01.12.2018).

33. NASA Interface to produce plots listings or output files from OMNI [Электронный ресурс] – Режим доступа: https://omniweb.gsfc.nasa.gov/form/dx1.html (11.11.2019).

34. *Mandrikova O.V., Polozov Yu.A., Mandrikova B.S.* Analysis of cosmic ray dynamics and ionospheric parameters during increased solar activity and magnetic storms, E3S Web of Conferences, 2019. V. 127. https://doi.org/10.1051/e3sconf/201912702002

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1149–1161

ИНФОРМАТИКА

УДК 519.72

АНАЛИЗ ВЫБОРА АПРИОРНОГО РАСПРЕДЕЛЕНИЯ ДЛЯ СМЕСИ ЭКСПЕРТОВ[1)]

© 2021 г. А. В. Грабовой[1,*], В. В. Стрижов[1,2,**]

[1] 141701 Долгопрудный, М.о., Институтский пер., 9, Московский физико-технический институт, Россия

[2] 119333 Москва, ул. Вавилова, 40, ВЦ РАН им. А.А. Дородницына ФИЦ ИУ РАН, Россия

*e-mail: grabovoy.av@phystech.edu

**e-mail: strijov@phystech.edu

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Исследуются свойства смеси экспертов. Смесь экспертов – это ансамбль локальных аппроксимирующих моделей, которые являются экспертами и шлюзовой функцией, которая взвешивает данные экспертов. В качестве экспертов рассматриваются линейные модели, а в качестве шлюзовой функции – нейронная сеть с функцией softmax на последнем слое. Анализируются разные априорные распределения для каждого эксперта. Предложен метод, который учитывает связь между априорными распределениями разных экспертов. Для поиска оптимальных параметров локальных моделей и шлюзовой функции используется EM-алгоритм. Рассматривается задача распознавания окружностей на изображении. Каждый эксперт аппроксимирует одну окружность на изображении: находит координаты центра окружности и радиус окружности. Для анализа предложенного метода проводится вычислительный эксперимент на синтетических и реальных данных. В качестве реальных данных используются изображения радужки глаза, которые применяются в задачах распознавания радужки глаза. Библ. 23. Фиг. 13. Табл. 1.

Ключевые слова: смесь экспертов, байесовский выбор модели, априорное распределение.

DOI: 10.31857/S0044466921070073

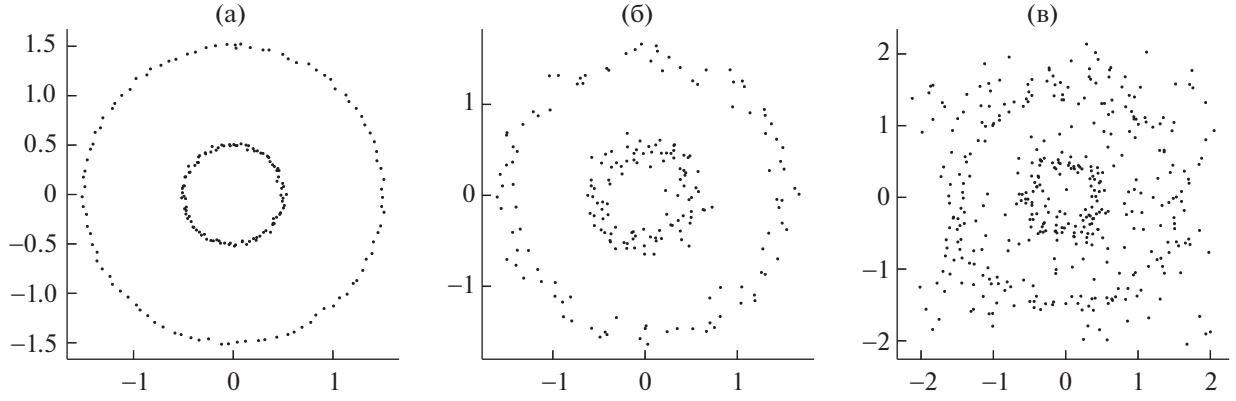
1. ВВЕДЕНИЕ

В работе исследуется проблема построения смеси экспертов. Смесь экспертов – это мультимодель, которая состоит из множества локальных моделей, называемых экспертами и шлюзовой функцией. Смесь экспертов использует шлюзовую функцию для взвешивания прогнозов каждого эксперта. Весовые коэффициенты шлюзовой функции зависят от объекта, для которого проводится прогноз. Примерами мультимоделей являются бэггинг, градиентный бустинг (см. [1]) и случайный лес (см. [2]). В [3] предполагается, что вклад каждого эксперта в ответ зависит от объекта из набора данных.

Основной проблемой построения мультимоделей является то, что ансамбль зависит от начальной инициализации параметров. Для улучшения устойчивости мультимодели предлагается использовать вероятностную постановку задачи для поиска оптимальных параметров шлюзовой функции и параметров локальной модели. В данной работе задается априорное распределение на параметры локальных моделей, а также предлагается учесть зависимость априорных распределений для разных моделей.

В настоящей работе решается задача поиска окружностей на бинаризованном изображении. Предполагается, что радиусы окружностей различаются значимо, а также, что центры почти сов-

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



Фиг. 1. Пример окружностей с разным уровнем шума: (а) – окружности без шума, (б) – окружности с зашумленным радиусом, (в) – окружности с зашумленным радиусом, а также с равномерным шумом по всему изображению.

падают. Пример изображений показан на фиг. 1. В качестве экспертов рассматриваются линейные модели – каждая модель аппроксимирует одну окружность. В качестве шлюзовой функции рассматривается двухслойная нейронная сеть.

Большое количество работ в области построения смеси экспертов посвящены выбору шлюзовой функции: используется softmax, процесс Дирихле (см. [4]), нейронная сеть (см. [5]) с функцией softmax на последнем слое. Ряд работ посвящен выбору моделей в качестве отдельных экспертов. В качестве модели эксперта в [6], [7] рассматривается линейная модель, в [8], [9] – модель SVM. В [3] представлен обзор методов и моделей в задачах смеси экспертов.

Смесь экспертов имеет множество приложений в прикладных задачах. Работы [10]–[12] посвящены применению смеси экспертов в задачах прогнозирования временных рядов. В [13] предложен метод распознавания рукописных цифр. Метод распознавания текстов с помощью смеси экспертов иследуется в [14], распознавание речи – в [15]–[17]. В [18] исследуется смесь экспертов для задачи распознавания трехмерных движений человека. В [19] описаны работы по исследованию обнаружения радужки глаза на изображении. В [20], [21], в частности, описаны методы выделения границ радужки и зрачка.

2. ПОСТАНОВКА ЗАДАЧИ АППРОКСИМАЦИИ ПАРАМЕТРОВ ОКРУЖНОСТИ

Задано бинарное изображение

$$\mathbf{M} \in \{0, 1\}^{m_1 \times m_2},$$

где 1 – это черный пиксель, который принадлежит рассматриваемой фигуре на изображении, а 0 – белый пиксель, который является фоном изображения. Пример изображения показан на фиг. 1.

Изображение $\mathbf{M}$ отображается в множество координат $\mathbf{C} = \{x_i, y_i\}_{i=1}^{N}$. Координата (x_i, y_i) является координатой i-го черного пикселя на изображении $\mathbf{M}$:

$$\mathbf{C} \in \mathbb{R}^{N \times 2},$$

где N – число черных пикселей.

Обозначим через (x_0, y_0) центр окружности, а r – радиус окружности. Координаты $(x_i, y_i) \in \mathbf{C}$ – это геометрическое место точек, которое удовлетворяет системе уравнений

$$(x_i - x_0)^2 + (y_i - y_0)^2 = r^2 + \varepsilon_i, \quad i \in \{1, 2, \ldots, N\},$$

где $\varepsilon_i \in \mathcal{N}(0, \beta^{-1})$ – невязка i-го уравнения, которая является следствием шума на изображении. Раскрыв скобки, получим

$$(2x_0) \cdot x_i + (2y_0) \cdot y_i + (r^2 - x_0^2 - y_0^2) \cdot 1 = x_i^2 + y_i^2 - \varepsilon_i. \tag{2.1}$$

Выражение (2.1) переписывается в задачу линейной регрессии следующим образом:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathbf{R}^n} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2, \quad \mathbf{X} = [\mathbf{C}, 1], \quad \mathbf{y} = [x_1^2 + y_1^2, \ldots, x_N^2 + y_N^2]^{\mathrm{T}}. \tag{2.2}$$

Используя вектор параметров $\mathbf{w} = [w_1, w_2, w_3]^{\mathrm{T}}$, получаем параметры окружности x_0, y_0, r:

$$x_0 = \frac{w_1}{2}, \quad y_0 = \frac{w_2}{2}, \quad r = \sqrt{w_3 + x_0^2 + y_0^2}.$$

Решая уравнения (2.2), находим параметры единственной окружности на изображении. В случае, когда на изображении несколько окружностей, предлагается использовать смесь экспертов, которая состоит из линейных моделей – экспертов. Каждый эксперт описывает одну окружность на изображении.

3. ПОСТАНОВКА ЗАДАЧИ ПОСТРОЕНИЯ СМЕСИ ЭКСПЕРТОВ

Обобщим подход аппроксимации одной окружности на изображении на случай, когда на изображении несколько окружностей. Пусть изображение состоит из K окружностей, тогда множество черных пикселей $\mathbf{C}$ представляется в виде

$$\mathbf{C} = \coprod_{k=1}^{K} \mathbf{C}'_k,$$

где $\mathbf{C}'_k$ – множество точек, принадлежащих k-й окружности. Множеству точек $\mathbf{C}'_k \subset \mathbf{C}$ соответсвует задача линейной регрессии для выборки $\mathbf{X}'_k \subset \mathbf{X}$, $\mathbf{y}'_k \subset \mathbf{y}$. Модель $\mathbf{g}_k$, аппроксимирующая k-ю подвыборку $\mathbf{X}'_k$, $\mathbf{y}'_k$, является локальной моделью для выборки $\mathbf{X}$, $\mathbf{y}$.

Определение 1. Модель $\mathbf{g}$ называется *локальной моделью* для выборки $\mathbf{X}$, $\mathbf{y}$, если $\mathbf{g}$ аппроксимирует некоторое непустое подмножество $\mathbf{X}'$, $\mathbf{y}'$ этой выборки.

Определение 2. Мультимодель $\mathbf{f}$ называется *смесью экспертов*, если

$$\mathbf{f} = \sum_{k=1}^{K} \pi_k \mathbf{g}_k(\mathbf{w}_k), \quad \pi_k(\mathbf{x}, \mathbf{V}) : \mathbb{R}^{n\times|\mathbf{V}|} \to [0, 1], \quad \sum_{k=1}^{K} \pi_k(\mathbf{x}, \mathbf{V}) = 1, \tag{3.1}$$

где $\mathbf{g}_k$ является k-й локальной моделью, π_k – шлюзовая функция, вектор $\mathbf{w}_k$ – параметр k-й локальной модели, а $\mathbf{V}$ – параметры шлюзовой функции.

В данной работе в качестве локальных моделей рассматриваются линейные модели. В качестве шлюзовой функции рассматривается двухслойный перцептрон:

$$\mathbf{g}_k(\mathbf{x}) = \mathbf{w}_k^{\mathrm{T}}\mathbf{x}, \quad \pi(\mathbf{x}, \mathbf{V}) = \mathbf{softmax}(\mathbf{V}_1^{\mathrm{T}}\sigma(\mathbf{V}_2^{\mathrm{T}}\mathbf{x})), \tag{3.2}$$

где $\mathbf{V} = \{\mathbf{V}_1, \mathbf{V}_2\}$ – множество параметров шлюзовой функции.

Предлагается использовать вероятностный подход для описания смеси экспертов. Вводится предположение, что $\mathbf{y}$ является случайным вектором, который задается плотностью распределения $p(\mathbf{y}|\mathbf{X})$. Предполагается, что плотность распределения $p(\mathbf{y}|\mathbf{X}, \mathbf{f})$ аппроксимирует истинную плотность распределения $p(\mathbf{y}|\mathbf{X})$:

$$p(\mathbf{y}|\mathbf{X}, \mathbf{f}) = \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k p_k(y_i|\mathbf{g}_k(\mathbf{x}_i)) \right), \tag{3.3}$$

где $\mathbf{f}$ – смесь экспертов, а $\mathbf{g}_k, \pi$ определяются выражением (3.2).

Пусть $\mathbf{w}_k$ является случайным вектором, который задается плотностью распределения $p^k(\mathbf{w}_k)$. Получим совместное распределение параметров локальных моделей и вектора ответов:

$$p(\mathbf{y}, \mathbf{W}|\mathbf{X}, \mathbf{V}) = \prod_{k=1}^{K} p^k(\mathbf{w}_k) \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k p_k(y_i|\mathbf{w}_k, \mathbf{x}_i) \right), \tag{3.4}$$

где $\mathbf{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$. Оптимальные параметры находятся с помощью максимизации правдоподобия:

$$\hat{\mathbf{V}}, \hat{\mathbf{W}} = \arg\max_{\mathbf{V},\mathbf{W}} p(\mathbf{y}, \mathbf{W} | \mathbf{X}, \mathbf{V}).$$

4. ВЕРОЯТНОСТНАЯ ПОСТАНОВКА СМЕСИ ЭКСПЕРТОВ

Для построения смеси экспертов (3.1), (3.4) введем следующие вероятностные предположения о данных (2.2):

(i) правдоподобие $p_k(y_i | \mathbf{w}_k, \mathbf{x}_i) = \mathcal{N}(y_i | \mathbf{w}_k^{\mathrm{T}} \mathbf{x}_i, \beta^{-1})$, где параметр β является уровнем шума,

(ii) априорное распределение параметров $p^k(\mathbf{w}_k) = \mathcal{N}(\mathbf{w}_k | \mathbf{w}_k^0, \mathbf{A}_k)$, где $\mathbf{w}_k^0$ – вектор размерности $n \times 1$, а $\mathbf{A}_k$ – ковариационная матрица размерности $n \times n$,

(iii) регуляризация априорного распределения $p(\varepsilon_{k,k'} | \Xi) = \mathcal{N}(\varepsilon_{k,k'} | 0, \Xi)$, где Ξ – ковариационная матрица, а $\varepsilon_{k,k'} = \mathbf{w}_k^0 - \mathbf{w}_{k'}^0$.

Предположение (i) задает априорное предположение о распределениии вектора параметров локальной модели $\mathbf{w}_k$. Априорное распределение задает ограничения на локальную модель. Например, если $\mathbf{w}_k^0 = [0, 0, 1]$, то k-я локальная модель аппроксимирует окружность с параметрами $x_0 = 0$, $y_0 = 0$, $r = 1$ с большей вероятностью.

Предположение (iii) задает регуляризацию априорных распределений. Она учитывает связь между априорными ограничениями разных локальных моделей. Например, если $\mathrm{diag}(\Xi) = [0.001, 0.001, 1]$, то центры разных окружностей совпадают.

Используя предположения (i)–(iii) и выражение (3.4), получаем полное правдоподобие:

$$p(\mathbf{y}, \mathbf{W} | \mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \Xi, \beta) = \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k N(y_i | \mathbf{w}_k^{\mathrm{T}} \mathbf{x}_i, \beta^{-1}) \right) \prod_{k=1}^{K} \mathcal{N}(\mathbf{w}_k | \mathbf{w}_k^0, \mathbf{A}_k) \prod_{k,k'=1}^{K} \mathcal{N}(\varepsilon_{k,k'} | 0, \Xi), \tag{4.1}$$

где $\mathbf{A} = \{\mathbf{A}_1, \cdots, \mathbf{A}_K\}$.

Введем бинарную матрицу $\mathbf{Z}$. Элемент матрицы $z_{ik} = 1$ тогда и только тогда, когда i-й объект аппроксимируется k-й локальной моделью. Подставляя бинарную матрицу $\mathbf{Z}$ в выражении (4.1), а также взяв логарифм, получаем

$$\begin{aligned} \log p(\mathbf{y}, \mathbf{Z}, \mathbf{W} | \mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \Xi, \beta) = {} & \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left[\log \pi_k(\mathbf{x}_i, \mathbf{V}) - \frac{\beta}{2} (y_i - \mathbf{w}_k^{\mathrm{T}} \mathbf{x}_i)^2 + \frac{1}{2} \log \frac{\beta}{2\pi} \right] + \\ & + \sum_{k=1}^{K} \left[-\frac{1}{2} (\mathbf{w}_k - \mathbf{w}_k^0)^{\mathrm{T}} \mathbf{A}_k^{-1} (\mathbf{w}_k - \mathbf{w}_k^0) + \frac{1}{2} \log \det \mathbf{A}_k^{-1} - \frac{n}{2} \log 2\pi \right] + \\ & + \sum_{k=1}^{K} \sum_{k'=1}^{K} \left[-\frac{1}{2} (\mathbf{w}_k^0 - \mathbf{w}_{k'}^0)^{\mathrm{T}} \Xi^{-1} (\mathbf{w}_k^0 - \mathbf{w}_{k'}^0) + \frac{1}{2} \log \det \Xi - \frac{n}{2} \log 2\pi \right]. \end{aligned} \tag{4.2}$$

Получаем новую задачу оптимизации обоснованности. Функция обоснованности получается при интегрировании выражения (4.2) по параметрам $\mathbf{W}$, $\mathbf{Z}$:

$$\mathbf{V}, \mathbf{W}^0, \mathbf{A}, \beta = \arg\max_{\mathbf{V},\mathbf{W}^0,\mathbf{A},\beta} \int_{\mathbf{W},\mathbf{Z}} \log p(\mathbf{y}, \mathbf{Z}, \mathbf{W} | \mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \Xi, \beta) d\mathbf{W} d\mathbf{Z}. \tag{4.3}$$

5. ЕМ-АЛГОРИТМ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ОПИМИЗАЦИИ

Рассмотрим вариационную плотность $q(\mathbf{W}, \mathbf{Z})$ для параметров $\mathbf{W}$, $\mathbf{Z}$. Тогда функция обоснованности принимает следующий вид:

$$\begin{aligned} \log p(\mathbf{y} | \mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \Xi, \beta) &= \int_{\mathbf{W},\mathbf{Z}} q(\mathbf{W}, \mathbf{Z}) \log p(\mathbf{y} | \mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \Xi, \beta) d\mathbf{W} d\mathbf{Z} = \\ &= \int_{\mathbf{W},\mathbf{Z}} q(\mathbf{W}, \mathbf{Z}) \log \frac{p(\mathbf{y}, \mathbf{W}, \mathbf{Z} | \mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \Xi, \beta)}{p(\mathbf{W}, \mathbf{Z} | \mathbf{y}, \mathbf{X}, \mathrm{V}, \mathbf{A}, \mathbf{W}^0, \Xi, \beta)} d\mathbf{W} d\mathbf{Z} = \end{aligned}$$

$$
\begin{aligned}
&= \int\limits_{\mathbf{W},\mathbf{Z}} q(\mathbf{W},\mathbf{Z}) \log \frac{p(\mathbf{y},\mathbf{W},\mathbf{Z}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\Xi,\beta) q(\mathbf{W},\mathbf{Z})}{p(\mathbf{W},Z|\mathbf{y},\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\Xi,\beta) q(\mathbf{W},\mathbf{Z})} d\mathbf{W} d\mathbf{Z} = \\
&= \int\limits_{\mathbf{W},\mathbf{Z}} q(\mathbf{W},\mathbf{Z}) \frac{p(\mathbf{y},\mathbf{W},\mathbf{Z}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\Xi,\beta)}{q(\mathbf{W},\mathbf{Z})} d\mathbf{W} d\mathbf{Z} + \\
&\quad + \int\limits_{\mathbf{W},\mathbf{Z}} q(\mathbf{W},\mathbf{Z}) \frac{q(\mathbf{W},\mathbf{Z})}{p(\mathbf{W},\mathbf{Z}|\mathbf{y},\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\Xi,\beta)} d\mathbf{W} d\mathbf{Z} = \\
&= \mathcal{L}(q,\mathbf{V},\mathbf{W}^0,\mathbf{A},\beta) + D_{KL}(q(\mathbf{W},\mathbf{Z}) \| p(\mathbf{W},\mathbf{Z}|\mathbf{y},\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\Xi,\beta)).
\end{aligned} \tag{5.1}
$$

Используя (5.1), получаем нижнюю оценку обоснованности:

$$\log p(\mathbf{y}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\Xi,\beta) \geq \mathcal{L}(q,\mathbf{V},\mathbf{W}^0,\mathbf{A},\beta),$$

где $\mathcal{L}(q,\mathbf{V},\mathbf{W}^0,\mathbf{A},\beta)$ называется нижней оценкой обоснованности.

Используем EM-алгоритм (см. [22], [23]) для решения оптимизационной задачи (4.3). Заметим, что EM-алгоритм вместо оптимизации $\log p(\mathbf{y}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\Xi,\beta)$ оптимизирует нижнюю оценку $\mathcal{L}(q,\mathbf{V},\mathbf{W}^0,\mathbf{A},\beta)$.

E-шаг. E-шаг решает следующую оптимизационую задачу:

$$\mathcal{L}(q,\mathbf{V},\mathbf{W}^0,\mathbf{A},\beta) \to \max_{q(\mathbf{W},\mathbf{Z})},$$

где параметры $\mathbf{V}$, $\mathbf{W}^0$, $\mathbf{A}$, β являются зафиксированными.

Пусть совместное распределение $q(\mathbf{Z},\mathbf{W})$ удовлетворяет условию независимости $q(\mathbf{Z},\mathbf{W}) = q(\mathbf{Z})q(\mathbf{W})$ (см. [23]). Далее символом $\propto$ обозначим то, что обе стороны выражения равны с точностью до аддитивной константы. Сначала найдем распределение $q(\mathbf{Z})$:

$$
\begin{aligned}
&\log q(\mathbf{Z}) = \mathrm{E}_{q/\mathbf{Z}} \log p(\mathbf{y},\mathbf{Z},\mathbf{W}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\Xi,\beta) \propto \\
&\propto \sum_{i+1}^{N}\sum_{k=1}^{K} z_{ik}\left[\log \pi_k(\mathbf{x}_i,\mathbf{V}) - \frac{\beta}{2}(y_i^2 - \mathbf{x}_i^{\mathrm{T}}\mathrm{E}\mathbf{w}_k + \mathbf{x}_i^{\mathrm{T}}\mathrm{E}\mathbf{w}_k\mathbf{w}_k^{\mathrm{T}}\mathbf{x}_i) + \frac{1}{2}\log\frac{\beta}{2\pi}\right], \\
&p(z_{ik}=1) = \frac{\exp\left[\log \pi_k(\mathbf{x}_i,\mathbf{V}) - \frac{\beta}{2}(\mathbf{x}_i^{\mathrm{T}}\mathrm{E}\mathbf{w}_k\mathbf{w}_k^{\mathrm{T}}\mathbf{x}_i - \mathbf{x}_i^{\mathrm{T}}\mathrm{E}\mathbf{w}_k)\right]}{\sum\limits_{k'=1}^{K}\exp\left[\log \pi_{k'}(\mathbf{x}_i,\mathbf{V}) - \frac{\beta}{2}(\mathbf{x}_i^{\mathrm{T}}\mathrm{E}\mathbf{w}_{k'}\mathbf{w}_{k'}^{\mathrm{T}}\mathbf{x}_i - \mathbf{x}_i^{\mathrm{T}}E\mathbf{w}_{k'})\right]}.
\end{aligned} \tag{5.2}
$$

Используя выражения (5.2), получаем, что распределение $q(z_{ik})$ является бернулевским распределением с параметром z_{ik}, которое задается выражением (5.2). Далее найдем распределение $q(\mathbf{W})$:

$$
\begin{aligned}
&\log q(\mathbf{W}) = \mathrm{E}_{q/\mathbf{W}} \log p(\mathbf{y},\mathbf{Z},\mathbf{W}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\Xi,\beta) \propto \\
&\propto \sum_{i=1}^{N}\sum_{k=1}^{K} \mathrm{E}z_{ik}\left[\log \pi_k(\mathbf{x}_{i,\mathbf{V}}) - \frac{\beta}{2}(y_i - \mathbf{w}_k^{\mathrm{T}}\mathbf{x}_i)^2 + \frac{1}{2}\log\frac{\beta}{2\pi}\right] + \\
&+ \sum_{k=1}^{K}\left[-\frac{1}{2}(\mathbf{w}_k - \mathbf{w}_k^0)^{\mathrm{T}}\mathbf{A}_k^{-1}(\mathbf{w}_k - \mathbf{w}_k^0) + \frac{1}{2}\log\det\mathbf{A}_k^{-1} - \frac{n}{2}\log 2\pi\right] \propto \\
&\propto \sum_{k=1}^{K}\left[\mathbf{w}_k^{\mathrm{T}}\left(\mathbf{A}_k^{-1}\mathbf{w}_k^0 + \beta\sum_{i=1}^{N}\mathbf{x}_i y_i \mathrm{E}z_{ik}\right) - \frac{1}{2}\mathbf{w}_k^{\mathrm{T}}\left(\mathbf{A}_k^{-1} + \beta\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right)\mathbf{w}_k\right].
\end{aligned} \tag{5.3}
$$

Используя выражение (5.3), получаем, что распределение $q(\mathbf{w}_k)$ является нормальным распределением со средним $\mathbf{m}_k$ и ковариационной матрицей $\mathbf{B}_k$:

$$\mathbf{m}_k = \mathbf{B}_k\left(\mathbf{A}_k^{-1}\mathbf{w}_k^0 + \beta\sum_{i=1}^{N}\mathbf{x}_i y_i \mathrm{E}z_{ik}\right), \quad \mathbf{B}_k = \left(\mathbf{A}_k^{-1} + \beta\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\mathrm{E}z_{ik}\right)^{-1}.$$

М-шаг. М-шаг решает следующую оптимизационную задачу:

$$\mathscr{L}(q, \mathbf{V}, \mathbf{W}^0, \mathbf{A}, \beta) \to \max_{\mathbf{V}, \mathbf{W}^0, \mathbf{A}, \beta},$$

где $q(\mathbf{W}, \mathbf{Z})$ является известной плотностью распределения. Распределение $q(\mathbf{Z}, \mathbf{W})$ является фиксированным, в то время как вариационная нижняя оценка $\mathscr{L}(\mathbf{V}, \mathbf{W}^0, \mathbf{A}, \beta)$ максимизируется по параметрам $\mathbf{V}$, $\mathbf{W}^0$, $\mathbf{A}$, β:

$$\begin{aligned}
&\mathscr{L}(\mathbf{V}, \mathbf{W}^0, \mathbf{A}, \beta) = \mathrm{E}_q \log p(\mathbf{y}, \mathbf{Z}, \mathbf{W} \,|\, \mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \Xi, \beta) = \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} \mathrm{E} z_{ik} \left[\log \pi_k(\mathbf{x}_i, \mathbf{V}) - \frac{\beta}{2} \mathrm{E}(y_i - \mathbf{w}_k^{\mathrm{T}} \mathbf{x}_i)^2 + \frac{1}{2} \log \frac{\beta}{2\pi} \right] + \\
&+ \sum_{k=1}^{K} \left[-\frac{1}{2} \mathrm{E}(\mathbf{w}_k - \mathbf{w}_k^0)^{\mathrm{T}} \mathbf{A}_k^{-1} (\mathbf{w}_k - \mathbf{w}_k^0) + \frac{1}{2} \log \det \mathbf{A}_k^{-1} - \frac{n}{2} \log 2\pi \right] + \\
&+ \sum_{k=1}^{K} \sum_{k'=1}^{K} \left[-\frac{1}{2} (\mathbf{w}_k^0 - \mathbf{w}_{k'}^0)^{\mathrm{T}} \Xi^{-1} (\mathbf{w}_k^0 - \mathbf{w}_{k'}^0) + \frac{1}{2} \log \det \Xi - \frac{n}{2} \log 2\pi \right].
\end{aligned} \tag{5.4}$$

Для нахождения оптимального параметра $\mathbf{V}$ используется градиентный метод оптимизации, который сходится к некоторому локальному экстремуму. Используя выражения (5.4), получаем оптимальное значение параметра $\mathbf{A}_k$:

$$\begin{aligned}
&\frac{\partial \mathscr{L}(\mathbf{V}, \mathbf{W}^0, \mathbf{A}, \beta)}{\partial \mathbf{A}_k^{-1}} = \frac{1}{2} \mathbf{A}_k - \frac{1}{2} \mathrm{E}(\mathbf{w}_k - \mathbf{w}_k^0)(\mathbf{w}_k - \mathbf{w}_k^0)^{\mathrm{T}} = 0, \\
&\mathbf{A}_k = \mathrm{E}\mathbf{w}_k \mathbf{w}_k^{\mathrm{T}} - \mathbf{w}_k^0 \mathrm{E}\mathbf{w}_k^{\mathrm{T}} - \mathrm{E}\mathbf{w}_k \mathbf{w}_k^{0\mathrm{T}} + \mathbf{w}_k^0 \mathbf{w}_k^{0\mathrm{T}}.
\end{aligned}$$

Аналогично получаем оптимальные значения для параметра β и для параметров $\mathbf{w}_k^0$:

$$\begin{aligned}
&\frac{\partial \mathscr{L}(\mathbf{V}, \mathbf{W}^0, \mathbf{A}, \beta)}{\partial \beta} = \sum_{k=1}^{K} \sum_{i=1}^{N} \left(\frac{1}{\beta} \mathrm{E} z_{ik} - \frac{1}{2} \mathrm{E} z_{ik} [y_i^2 - 2 y_i \mathbf{x}_i^{\mathrm{T}} \mathrm{E}\mathbf{w}_k + \mathbf{x}_i^{\mathrm{T}} \mathbf{w}_k \mathbf{w}_k^{\mathrm{T}} \mathbf{x}_i] \right) = 0, \\
&\frac{1}{\beta} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} [y_i^2 - 2 y_i \mathbf{x}_i^{\mathrm{T}} \mathrm{E}\mathbf{w}_k + \mathbf{x}_i^{\mathrm{T}} \mathrm{E}\mathbf{w}_k \mathbf{w}_k^{\mathrm{T}} \mathbf{x}_i] \mathrm{E} z_{ik}, \\
&\frac{\partial \mathscr{L}(\mathbf{V}, \mathbf{W}^0, \mathbf{A}, \beta)}{\partial \mathbf{w}_k^0} = \mathbf{A}_k^{-1} (\mathrm{E}\mathbf{w}_k - \mathbf{w}_k^0) + \Xi \sum_{k'=1}^{K} [\mathbf{w}_{k'}^0 - \mathbf{w}_k^0] = 0, \\
&\mathbf{w}_k^0 = [\mathbf{A}_k^{-1} + (K-1)\Xi]^{-1} \left(\mathbf{A}_k^{-1} \mathrm{E}\mathbf{w}_k + \Xi \sum_{k'=1, k' \neq k}^{K} \mathbf{w}_{k'}^0 \right).
\end{aligned} \tag{5.5}$$

Выражения (5.2)–(5.5) задают итеративную процедуру, которая сходится к некоторому локальному максимуму оптимизационной задачи (4.3).

6. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Для анализа качества различных мультимоделей для аппроксимации окружности проводится вычислительный эксперимент. В эксперименте рассматриваются следующие мультимодели: мультимодель $\mathbf{f}_1$ без использования априорных распределений, мультимодель $\mathbf{f}_2$, которая использует априорные распределения (6.2) для параметров, и мультимодель $\mathbf{f}_3$, которая использует регуляризацию априорных распределений. Точность аппроксимации мультимодели $\mathbf{f}_i$ задается следующим образом:

$$\mathscr{S}_{\mathbf{f}_i} = \sum_{k=1}^{K} (x_0^k - x_{\mathrm{pr}}^k)^2 + (y_0^k - y_{\mathrm{pr}}^k)^2 + (r^k - r_{\mathrm{pr}}^k)^2, \tag{6.1}$$

где x_0^k, y_0^k, r^k – истинный центр и радиус для k-й окружности соответственно, x_{pr}^k, y_{pr}^k, r_{pr}^k – предсказанные центр и радиус для k-й окружности соответственно.

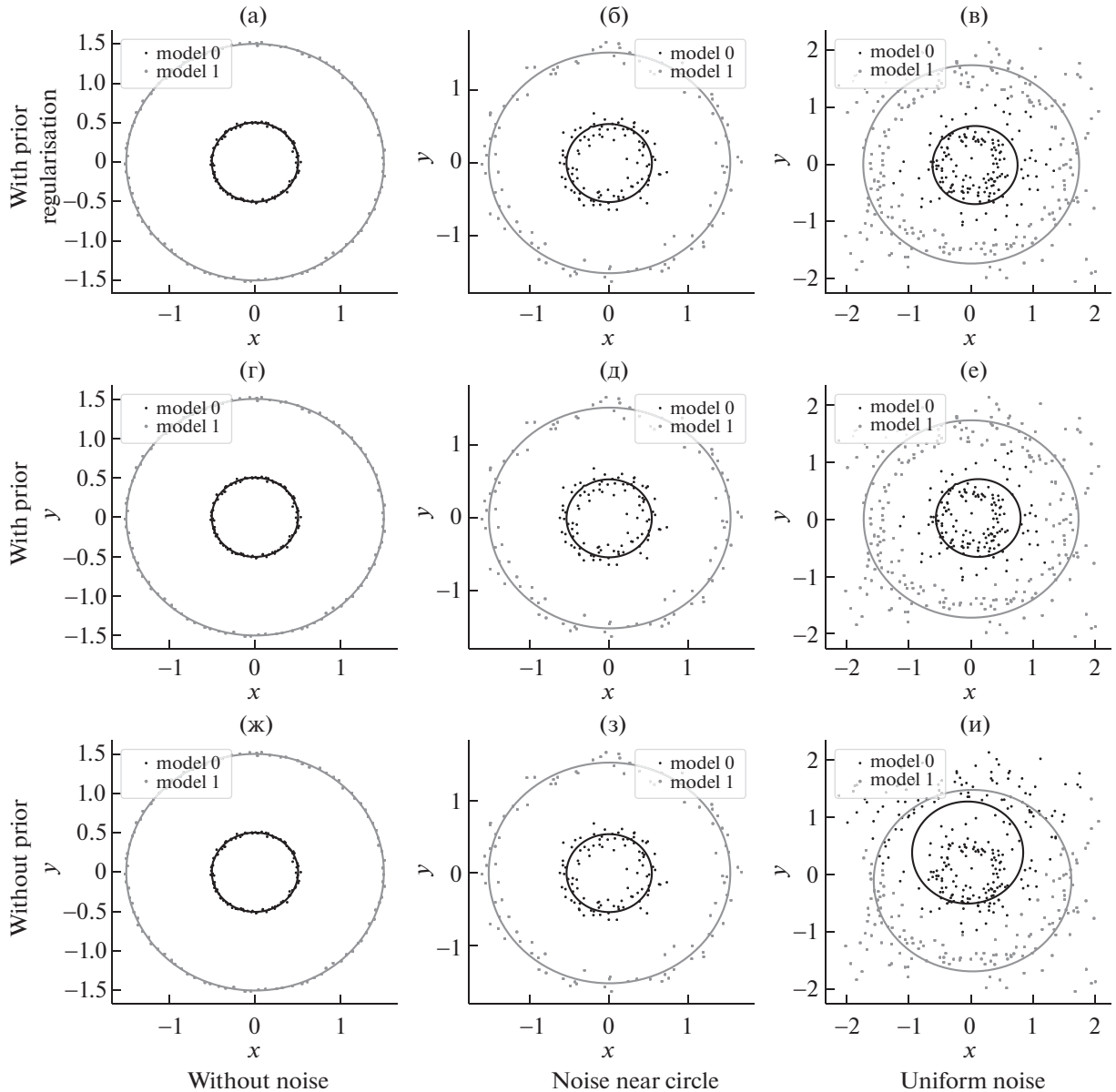
Таблица 1. Качество аппроксимации (6.1) для всех мультимоделей

Выборка	$\mathscr{S}_{\mathbf{f}_1}$	$\mathscr{S}_{\mathbf{f}_2}$	$\mathscr{S}_{\mathbf{f}_3}$
Synthetic 1	10^{-5}	10^{-5}	10^{-5}
Synthetic 2	0.6	10^{-3}	10^{-3}
Synthetic 3	0.6	10^{-3}	10^{-3}

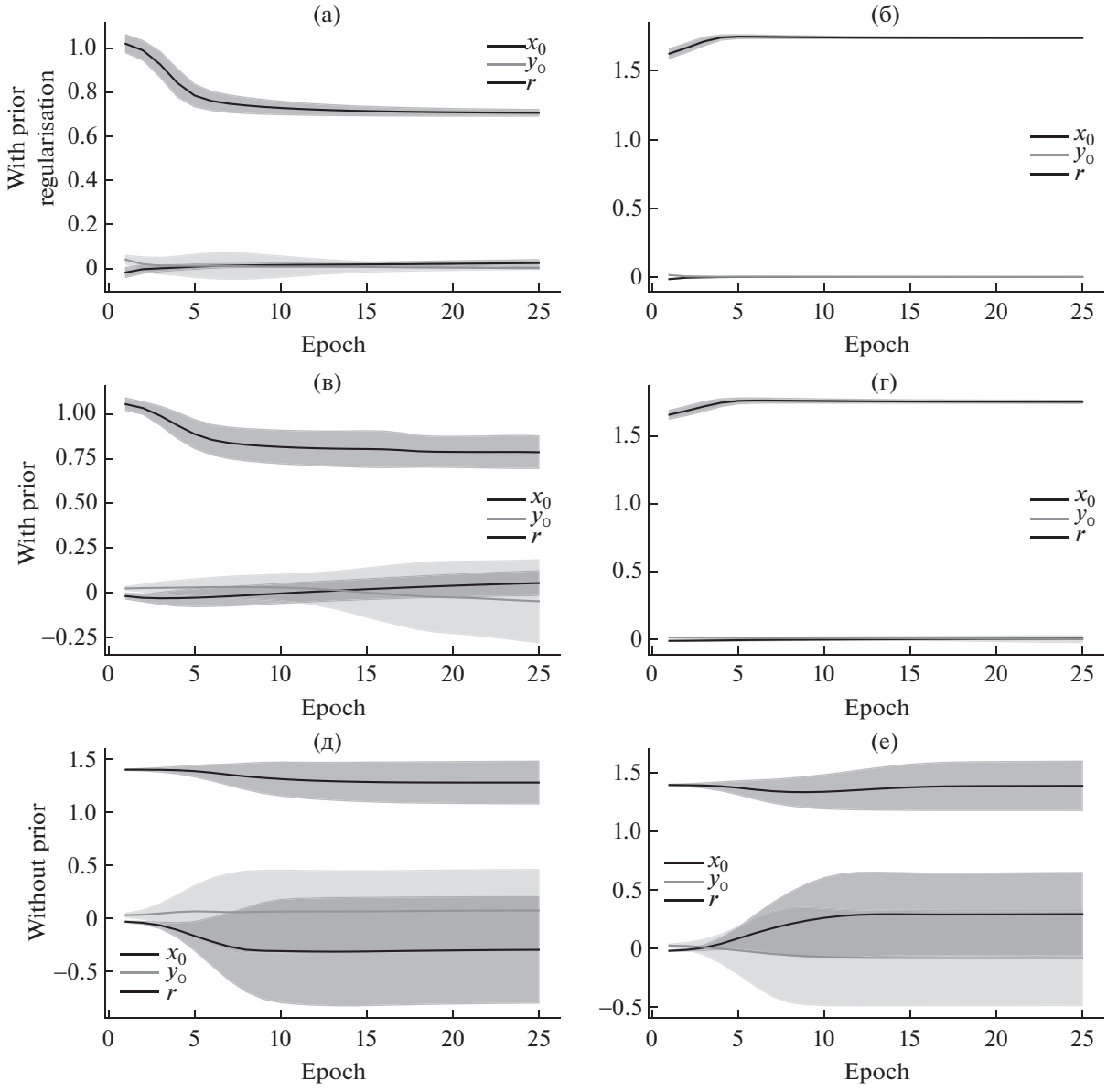
Для сравнения моделей с разными вероятностными предположениями используется правдоподобие (3.3). В вычислительном эксперименте используется следующее априорное распределение:

$$p^1(\mathbf{w}_1) \sim \mathcal{N}(\mathbf{w}_1^0, \mathbf{I}), \quad p^2(\mathbf{w}_2) \sim \mathcal{N}(\mathbf{w}_2^0, \mathbf{I}), \tag{6.2}$$

где $\mathbf{w}_1^0 = [0, 0, 0.1]$, $\mathbf{w}_2^0 = [0, 0, 2]$.



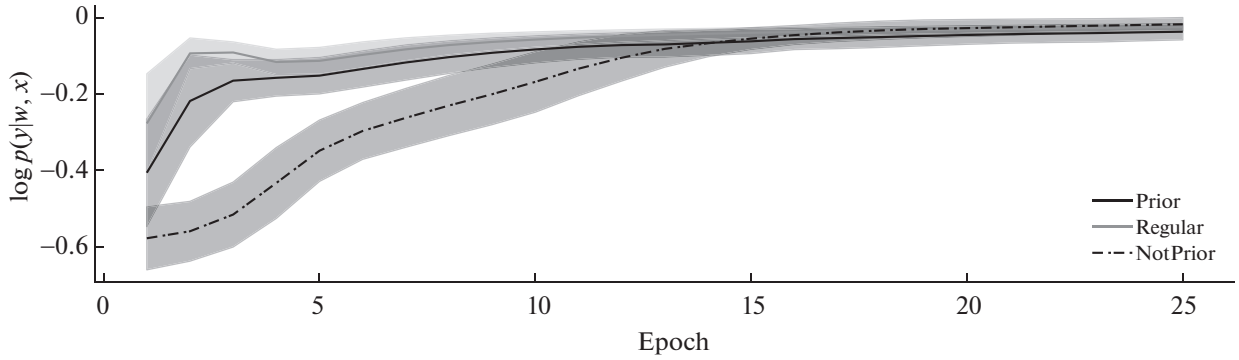
Фиг. 2. Мультимодель в зависимости от разных априорных предположений и в зависимости от разного уровня шума: (а)–(в) – модель с регуляризацией априорных распределений, (г)–(е) – модель с заданными априорными распределениями на параметрах локальных моделей, (ж)–(и) – модель без заданных априорных предположений.



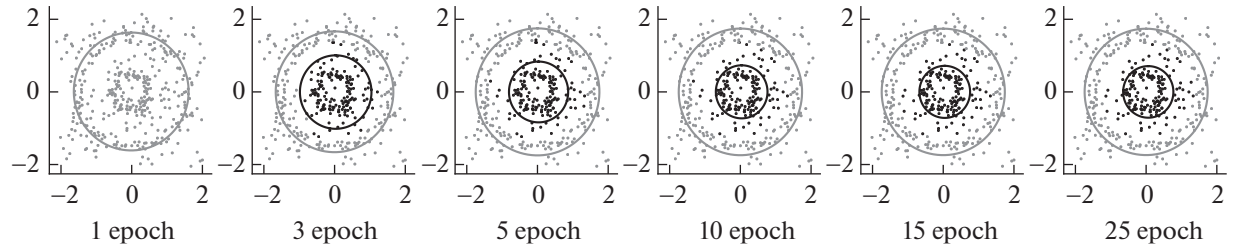
Фиг. 3. Зависимости центра и радиуса окружностей от номера итерации: (а), (б) – модель с регуляризацией априорных распределений; (в), (г) – модель с заданными априорными распределениями на параметры моделей; (д), (е) – модель без задания априорных распределений.

Синтетические данные с разным типом шума в изображении. В вычислительном эксперименте сравнивается качество следующих мультимоделей $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$ на синтетических данных. Синтетические данные являются двумя концентрическими окружностями с разным уровнем шума. Выборка Synthetic 1 является изображением без шума, выборка Synthetic 2 – изображением с зашумленным радиусом окружности, а выборка Synthetic 3 – изображением с равномерным шумом. На фиг. 2 показаны результаты для мельтимоделей $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$. Все модели оптимизировались с помощью 50 итераций EM-алгоритма. Мультимодели $\mathbf{f}_2$, $\mathbf{f}_3$ аппроксимируют окружности лучше, чем мультимодель $\mathbf{f}_1$. В табл. 1 показано качество аппрроксимации (6.1) для всех мультимоделей.

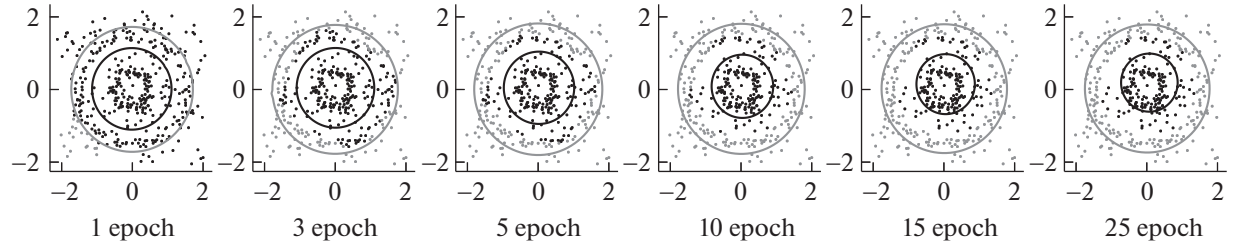
Анализ сходимости на синтетической выборке. Данная часть эксперимента анализирует качество сходимости EM-алгоритма для разных мультимоделей $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$. Анализ всех мультимоделей проводится на выборке Synthetic 3.



Фиг. 4. Зависимости логарифма правдоподобия (3.3) от номера итерации.



Фиг. 5. Визуализации процесса сходимости мультимодели с использованием априорной регуляризации.



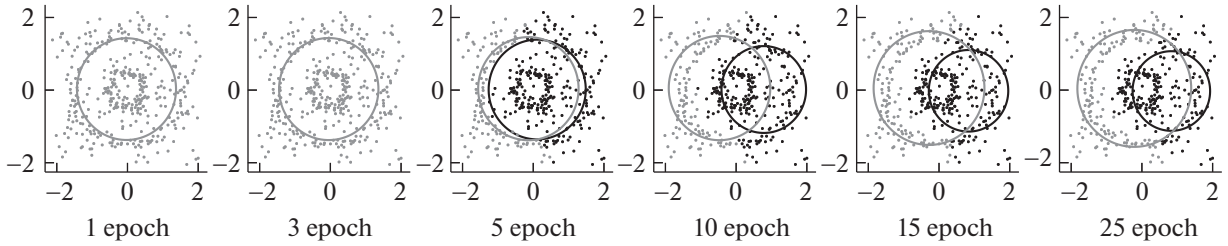
Фиг. 6. Визуализации процесса сходимости мультимодели с использованием априорного распределения.

На фиг. 3 показана зависимость предсказанных центра и радуса в зависимости от номера итерации EM-алгоритма. Мультимодель $\mathbf{f}_2$, которая использует априорное распределение, аппроксимирует окружность лучше мультимодели $\mathbf{f}_1$, которая не использует никакого априорного распределения. Мультимодель $\mathbf{f}_3$, которая использует регуляризатор априорных распределений, является более стабильной, чем мультимодель $\mathbf{f}_2$.

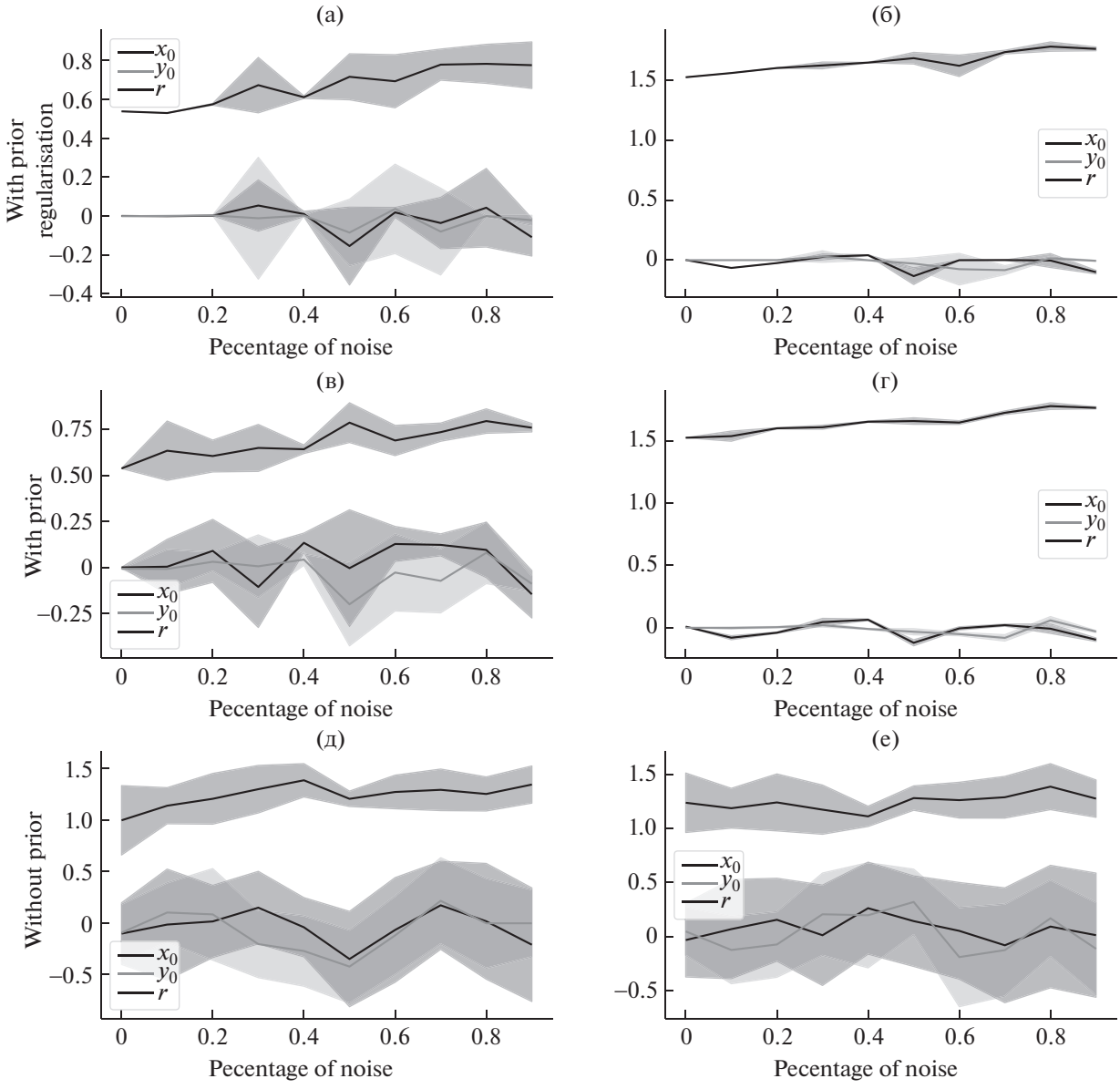
На фиг. 4 показана зависимость логарифма правдоподобия (3.3) от номера итерации EM-алгоритма. Логарифм правдоподобия мультимоделей $\mathbf{f}_2$, $\mathbf{f}_3\mathbf{f}_2, \mathbf{f}_3$ растет быстрее, чем логарифм правдоподобия мультимодели $\mathbf{f}_1$. После 20-й итерации все мультимодели имеют одинаковое правдоподобие.

На фиг. 5–7 показан процесс сходимости для разных мультимоделей $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$. На фиг. 7 показана мультимодель $\mathbf{f}_1$, которая аппроксимирует окружности не верно. На фиг. 5, 6 показаны мультимодели $\mathbf{f}_2$, $\mathbf{f}_3$, которые аппроксимируют окружности верно.

Вычислительный эксперимент показывает, что мультимодели $\mathbf{f}_2$, $\mathbf{f}_3$ которые используют априорные распределения на параметры экспертов, аппроксимируют окружности лучше, чем мультимодель $\mathbf{f}_1$, которая работает без априорных распределений.

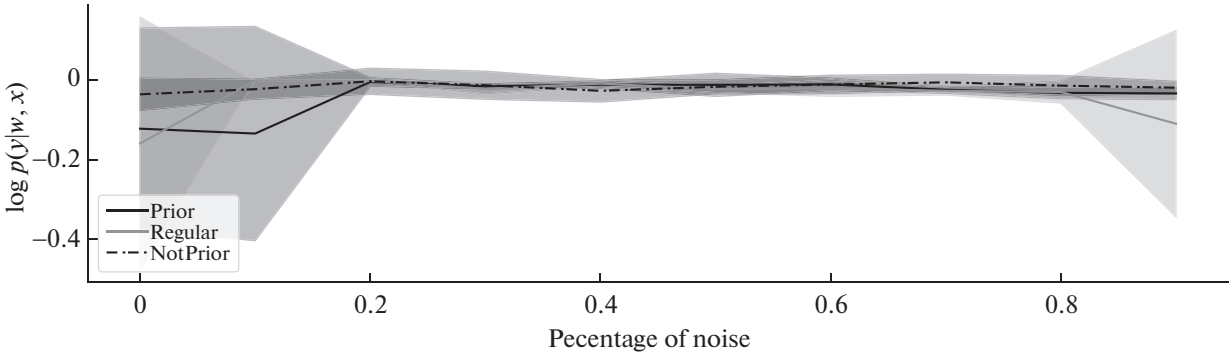


Фиг. 7. Визуализации процесса сходимости мультимодели без использования априорного распределения.

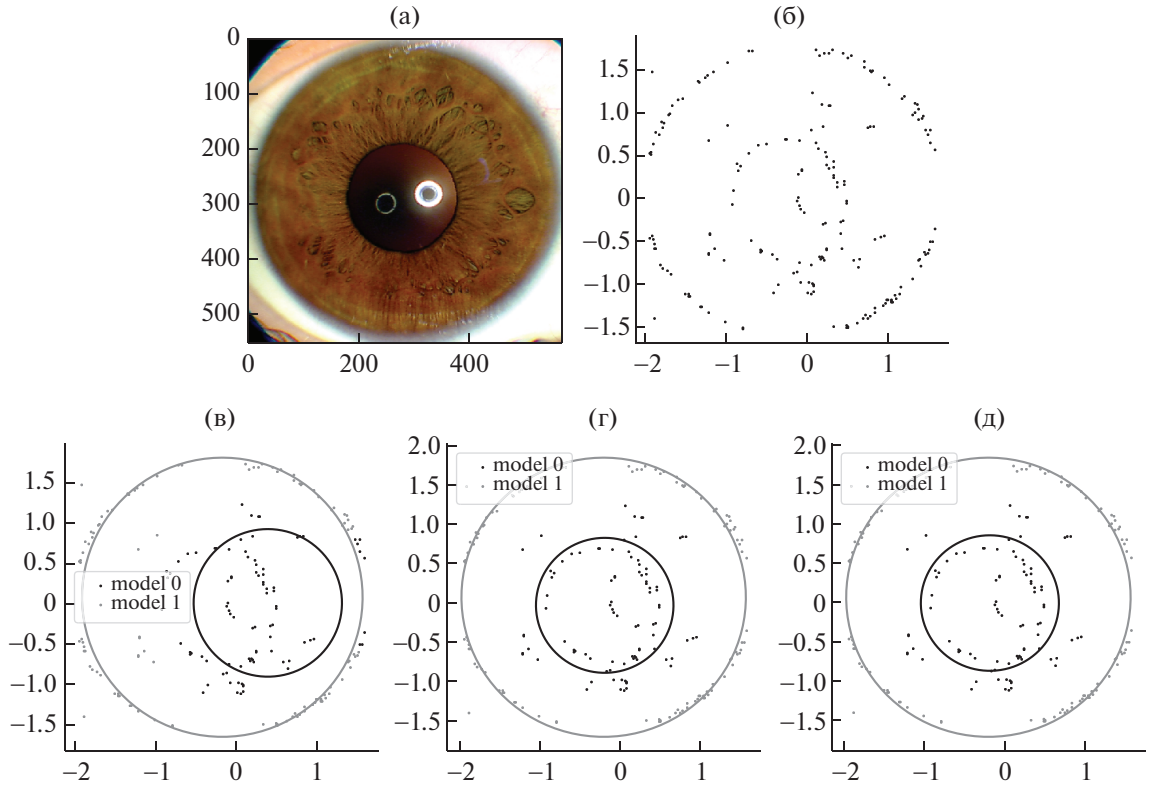


Фиг. 8. Зависимости центра и радиуса окружностей от номера итерации: (а), (б) – модель с регуляризацией априорных распределений; (в), (г) – модель с заданными априорными распределениями на параметры моделей; (д), (е) – модель без задания априорных распределений.

Анализ мультимоделей в зависимости от уровня шума. Данная часть эксперимента анализирует зависимость разных мультимоделей $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$ от уровня шума. Анализ всех мультимоделей проводится на выборке Synthetic 1 с добавлением разного уровня шума. Минимальный уровень шума

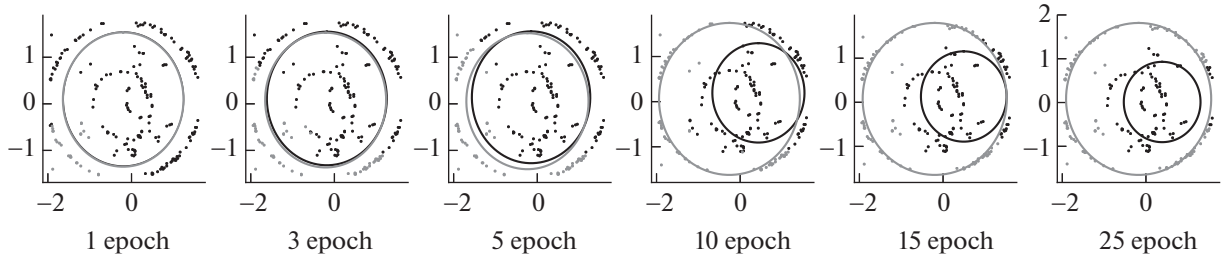


Фиг. 9. Зависимости логарифма правдоподобия (3.3) от уровня шума.

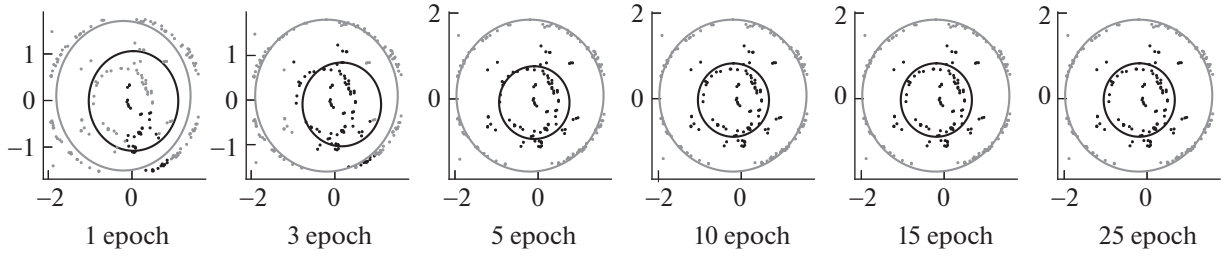


Фиг. 10. Мультимодель в зависимости от разных априорных предположений на реальном изображении: (а) – исходное изображение, (б) – бинаризованное изображение, (в) – мультимодель без априорных предположений, (г) – мультимодель с априорными распределениями на параметрах локальных моделей, (д) – мультимодель с регуляризацией на априорных распределениях параметров локальных моделей.

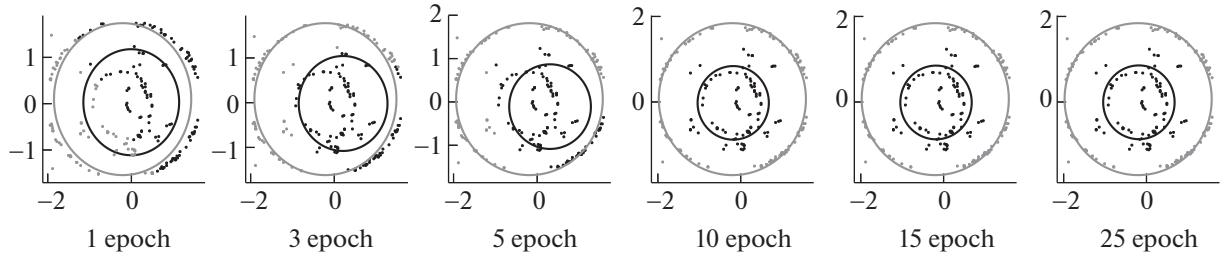
равен 0, когда число шумовых точек равно 0. Максимальный уровень шума равен 1, когда число шумовых точек равно числу точек на изображении. На фиг. 8 показаны график зависимости центра окружности и ее радиус в зависимости от уровня шума. Из графика следует, что радиус окружности увеличивается при увеличении уровня шума. Мультимодели $\mathbf{f}_2$, $\mathbf{f}_3$ аппроксимируют центр окружности верно, но мультимодель $\mathbf{f}_3$ более устойчива к шуму. На фиг. 9 показана зависимость логарифма правдоподобия (3.3) от уровня шума. Из графика следует, что логарифм правдоподобия (3.3) эквивалентный для всех мультимоделей, но на фиг. 8 видно, что качество аппроксимации (6.1) зависит от мультимодели. Данная часть вычислительного эксперимента показывает, что мультимодель $\mathbf{f}_3$ с регуляризацией априорного распределения является более устойчивой к шуму, чем остальные.



Фиг. 11. Визуализации процесса сходимости мультимодели без использования априорного распределения.



Фиг. 12. Визуализации процесса сходимости мультимодели с использованием априорного распределения.



Фиг. 13. Визуализации процесса сходимости мультимодели с использованием априорной регуляризации.

Реальные данные. Настоящая часть эксперимента анализирует разные мультимодели $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$ на реальной выборке. На фиг. 10 показан результат работы разных мультимоделей. Мультимодель $\mathbf{f}_1$ не верно аппроксимирует меньшую окружность. Мультимодели $\mathbf{f}_2$, $\mathbf{f}_3$ аппроксимируют обе окружности верно.

На фиг. 11–13 показан процесс аппроксимации для разных мультимоделей $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$.

Данная часть эксперимента показывает, что мультимодели $\mathbf{f}_2$, $\mathbf{f}_3$ аппроксимируют окружности на реальных изображениях лучше, чем мультимодель $\mathbf{f}_1$.

7. ЗАКЛЮЧЕНИЕ

В настоящей работе сравниваются мультимодели, которые используют различные априорные предположения. Для анализа проводился вычислительный эксперимент на концентрических окружностях с разным уровнем шума. Для аппроксимации окружности на изображении использовалась линейная модель. Для взвешивания ответов разных линейных моделей использовалась шлюзовая функция, которая является двухслойным перцептроном с функцией softmax на последнем слое. В вычислительном эксперименте сравниваются мультимодели, которые используют априорное распределение и которые его не используют. Мультимодели, которые используют априорные распределения, имеют большую точность аппроксимации, чем мультимодель, которая не использует априорные распределения.

Также был проведен эксперимент по исследованию различных способов регуляризации априорных распределений параметров локальных моделей. В эксперименте показано, что в случае, когда регуляризация задана, мультимодель находит окружности более устойчиво. В эксперименте было показано, что все мультимодели являются чувствительными к выбросам. Для решения данной задачи предлагается использовать еще одну локальную модель, которая будет аппроксимировать шум.

В дальнейшем планируется улучшить мультимодель с помощью задания априорного распределения на шлюзовую функцию. Планируется рассмотреть в качестве моделей не только модели, которые описывают данные, но и модель, которая аппроксимирует шум в данных. Предполагается, что число шумовых точек мало, поэтому требуется задать априорное распределение, которое учитывает данную информацию.

СПИСОК ЛИТЕРАТУРЫ

1. *Tianqi C., Carlos G.* XGBoost: A Scalable Tree Boosting System // Proceed. 22nd ACM SIGKDD Internat. Conf. Knowledge Discovery and Data Mining. 2016.
2. *Xi C., Hemant I.* Random Forests for Genomic Data Analysis // Genomics. 2012. Iss. 99. № 6. P. 323–329.
3. *Esen Y.S., Wilson J., Gader P.D.* Twenty Years of Mixture of Experts // IEEE Transact. Neural Networks and Learn. Syst. 2012. Iss. 23. № 8. P. 1177–1193.
4. *Rasmussen C.E., Ghahramani Z.* Infinite Mixtures of Gaussian Process Experts // Adv. Neural Informat. Proc. Syst. 14. 2002. P. 881–888.
5. *Shazeer N., Mirhoseini A., Maziarz K.* Outrageously large neural networks: the sparsely-gated mixture-of-experts layer // Internat. Conf. Learn. Representat. 2017.
6. *Jordan M.I.* Hierarchical mixtures of experts and the EM algorithm // Neural Comput. 1994. V. 6. № 2. P. 181–214.
7. *Jordan M.I., Jacobs R.A.* Hierarchies of adaptive experts // Adv. Neural Informat. Proc. Syst. 1991. P. 985–992.
8. *Lima C., Coelho A., Zuben F.J.* Hybridizing mixtures of experts with support vector machines: Investigation into nonlinear dynamic systems identification // Inf. Sci. 2007. V. 177. № 10. P. 2049–2074.
9. *Cao L.* Support vector machines experts for time series forecasting // Neurocomputing. 2003. V. 51. P. 321–339.
10. *Yumlu M.S., Gurgen F.S., Okay N.* Financial time series prediction using mixture of experts // Proc. 18th Int. Symp. Comput. Inf. Sci. 2003. P. 553–560.
11. *Cheung Y.M., Leung W.M., Xu L.* Application of mixture of experts model to financial time series forecasting // Proc. Int. Conf. Neural Netw. Signal Process. 1995. P. 1–4.
12. *Weigend A.S., Shi S.* Predicting daily probability distributions of S&P500 returns // J. Forecast. 2000. V. 19. № 4. P. 375–392.
13. *Ebrahimpour R., Moradian M.R., Esmkhani A., Jafarlou F.M.* Recognition of Persian handwritten digits using characterization loci and mixture of experts // J. Digital Content Technol. Appl. 2009. V. 3. № 3. P. 42–46.
14. *Estabrooks A., Japkowicz N.* A mixture-of-experts framework for text classification // Proc. Workshop Comput. Natural Lang. Learn., Assoc. Comput. Linguist. 2001. P. 1–8.
15. *Mossavat S., Amft O., Petkov Vries B., Kleijn W.* A Bayesian hierarchical mixture of experts approach to estimate speech quality // Proc. 2nd Int. Workshop Qual. Multimedia Exper. 2010. P. 200–205.
16. *Peng F., Jacobs R.A., Tanner M.A.* Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition // J. Amer. Stat. Assoc. 1996. V. 91. № 435. P. 953–960.
17. *Tuerk A.* The state based mixture of experts HMM with applications to the recognition of spontaneous speech. Ph.D. thesis. Cambridge: Univ. Cambridge, 2001.
18. *Sminchisescu C., Kanaujia A., Metaxas D.* Discriminative density propagation for visual tracking // IEEE Trans. Pattern Anal. Mach. Intell. 2007. V. 29. № 11. P. 2030–2044.
19. *Bowyer K., Hollingsworth K., Flynn P.* A Survey of Iris Biometrics Research: 2008–2010.
20. *Matveev I.* Detection of iris in image by interrelated maxima of brightness gradient projections // Appl. Comput. Math. 2010. V. 9. № 2. P. 252–257.
21. *Matveev I., Simonenko I.* Detecting precise iris boundaries by circular shortest path method // Pattern Recognit. and Image Anal. 2014. V. 24. P. 304–309.
22. *Dempster A.P., Laird N.M., Rubin D.B.* Maximum Likelihood from Incomplete Data via the EM Algorithm // J. the Royal Statist. Soc. Ser. B (Methodological). 1977. V. 39. № 1 P. 1–38.
23. *Bishop C.* Pattern Recognition and Machine Learning. Berlin: Springer, 2006. P. 758.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1162–1171

ИНФОРМАТИКА

УДК 519.72

РАСПОЗНАВАНИЕ КВАЗИПЕРИОДИЧЕСКОЙ ПОСЛЕДОВАТЕЛЬНОСТИ, ВКЛЮЧАЮЩЕЙ НЕИЗВЕСТНОЕ ЧИСЛО НЕЛИНЕЙНО-РАСТЯНУТЫХ ЭТАЛОННЫХ ПОДПОСЛЕДОВАТЕЛЬНОСТЕЙ[1)]

© 2021 г. А. В. Кельманов[1,2], Л. В. Михайлова[1,*], П. С. Рузанкин[1,2,**], С. А. Хамидуллин[1]

[1] 630090 Новосибирск, пр-т акад. Коптюга, 4, Ин-т матем. им. С.Л. Соболева, Россия

[2] 630090 Новосибирск, ул. Пирогова, 2, Новосибирский гос. ун-т, Россия

*e-mail: mikh@math.nsc.ru

**e-mail: ruzankin@math.nsc.ru

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Рассматривается неизученная экстремальная задача, которая индуцируется одной из задач помехоустойчивого распознавания квазипериодической последовательности, а именно, задачей распознавания последовательности Y длины N как последовательности, порожденной некоторой последовательностью U, принадлежащей заданному конечному множеству W (алфавиту) последовательностей. Каждая последовательность U из W порождает экспоненциальное по мощности множество $\mathscr{X}(U)$ последовательностей, объединяющее все последовательности длины N, которые в качестве подпоследовательностей включают переменное число допустимых квазипериодических (флуктуационных) повторов последовательности U. Каждый квазипериодический повтор порождается допустимыми преобразованиями последовательности U, а именно, сдвигами и растяжениями. Задача распознавания состоит в выборе последовательности U из W и аппроксимации последовательности Y элементом X из множества $\mathscr{X}(U)$ последовательностей. Критерием аппроксимации является минимум суммы квадратов расстояний между элементами последовательностей. Мы показываем, что рассматриваемая задача эквивалентна задаче суммирования элементов двух числовых последовательностей, в которой требуется минимизировать сумму неизвестного числа M слагаемых, каждое из которых является разностью невзвешенной автосвертки растянутой на переменную длину последовательности U (путем кратных повторов ее элементов) и взвешенной свертки этой растянутой последовательности с подпоследовательностью из Y. Мы доказываем, что рассматриваемая экстремальная задача и вместе с ней задача распознавания разрешимы за полиномиальное время. Примерами численного моделирования проиллюстрирована применимость алгоритма к решению модельных прикладных задач помехоустойчивой обработки ECG-подобных и PPG-подобных квазипериодических сигналов (electrocardiogram-like and photoplethysmogram-like signals). Библ. 9. Фиг. 5.

Ключевые слова: числовые последовательности, распознавание, квазипериодическая последовательность, полиномиальная разрешимость, разность взвешенных сверток.

DOI: 10.31857/S0044466921070097

ВВЕДЕНИЕ

В работе рассматривается неизученная экстремальная задача, которая индуцируется одной из задач помехоустойчивого распознавания квазипериодической последовательности. Цель работы – доказательство полиномиальной разрешимости данной задачи и построение алгоритма, гарантирующего получение оптимального решения. Поводом для проведения исследований послужило отсутствие каких-либо эффективных (полиномиальных) вычислительных алгоритмов с априорно гарантированными оценками точности для ее решения.

[1)]Работа выполнена при финансовой поддержке РФФИ, проекты 19-07-00397 и 19-01-00308, программы ФНИ РАН, проект 0314-2019-0015, а также программы Top-5-100 Минобрнауки РФ.

Рассматриваемая задача актуальна для помехоустойчивого мониторинга природных объектов, типичное состояние которых во времени квазипериодически повторяется с флуктуациями. То есть расстояние между двумя последовательными повторами лежит в заданном интервале, а типичное состояние допускает некоторую вариативность от повтора к повтору. А именно, для прикладных задач, когда требуется помимо обнаружения этих типовых повторов идентифицировать (распознать) либо сам объект, либо состояние объекта среди множества допустимых.

Подобный характер повторяемости состояний типичен, в первую очередь, для биомедицинских задач. В частности, задач анализа и распознавания ECG и PPG сигналов. Поэтому для иллюстрации работы алгоритма далее приведены примеры модельных прикладных задач распознавания ECG и PPG-подобных квазипериодических сигналов.

1. ФОРМУЛИРОВКА ЗАДАЧИ, ЕЕ ИСТОКИ И ТРАКТОВКИ

Рассматриваемая экстремальная задача имеет следующую формулировку.

Задача 1. *Дано:* числовая последовательность $Y = (y_1,\ldots,y_N)$, совокупность $W = \{U^{(1)},\ldots,U^{(K)} | U^{(k)} = (u_1^{(k)},\ldots,u_{q_k}^{(k)}), k = 1,\ldots,K\}$, натуральные числа $T_{\max}$ и ℓ. *Найти:* числовую последовательность $U = (u_1,\ldots,u_{q(U)}) \in W$, набор $\mathcal{M} = \{n_1,\ldots n_m,\ldots\}$ номеров последовательности Y, набор $\mathcal{P} = \{p_1,\ldots,p_m,\ldots\}$ натуральных чисел, набор $\mathcal{J} = \{J^{(1)},\ldots,J^{(m)},\ldots\}$ сжимающих отображений, в котором $J^{(m)} : \{1,\ldots,p_m\} \to \{1,\ldots,q(U)\}$, а также размерность M этих наборов, доставляющих минимум целевой функции

$$F(\mathcal{U},\mathcal{M},\mathcal{P},\mathcal{J}) = \sum_{m=1}^{M}\sum_{i=1}^{p_m}\{u^2_{J^{(m)}(i)} - 2y_{n_m+i-1}u_{J^{(m)}(i)}\}, \tag{1.1}$$

при ограничениях

$$q(U) \le p_m \le \ell \le T_{\max} \le N, \quad m = 1,\ldots,M, \tag{1.2}$$

$$p_{m-1} \le n_m - n_{m-1} \le T_{\max}, \quad m = 2,\ldots,M, \tag{1.3}$$

$$p_M \le N - n_M + 1, \tag{1.4}$$

на элементы искомых наборов $\mathcal{M}$, $\mathcal{P}$, и при ограничениях

$$\begin{gathered} J^{(m)}(1) = 1, \quad J^{(m)}(p_m) = q(U), \\ 0 \le J^{(m)}(i) - J^{(m)}(i-1) \le 1, \quad i = 2,\ldots,p_m, \quad m = 1,\ldots,M, \end{gathered} \tag{1.5}$$

на элементы искомых сжимающих отображений.

Приведем несколько трактовок задачи 1.

Из формулировки задачи 1 и вида целевой функции (1.1) следует, что задача 1 – задача об оптимальном (в смысле минимума (1.1)) суммировании элементов двух последовательностей, одна из которых Y задана, а другая U принадлежит заданному множеству последовательностей. Эквивалентная перезапись целевой функции (1.1) в виде

$$F(\mathcal{U},\mathcal{M},\mathcal{P},\mathcal{J}) = \sum_{m=1}^{M}\left\{\sum_{i=1}^{p_m}u^2_{J^{(m)}(i)} - 2\sum_{i=1}^{p_m}y_{n_m+i-1}u_{J^{(m)}(i)}\right\}$$

позволяет трактовать задачу 1 как задачу минимизации суммы разностей взвешенных сверток. Действительно, при каждом $m = 1,\ldots,M$ выражение $\sum_{i=1}^{p_m}u^2_{J^{(m)}(i)}$ в фигурных скобках – невзвешенная автосвертка последовательности $u_{J^{(m)}(i)}$, $i = 1,\ldots,p_m$, полученной из элементов U растяжением путем дублирования элементов, а выражение $\sum_{i=1}^{p_m}y_{n_m+i-1}u_{J^{(m)}(i)}$ – свертка этой растянутой последовательности с подпоследовательностью из Y, имеющей ту же длину p_m (коэффициент 2 – вес этой свертки).

Другая возможная трактовка – задача совместного выбора последовательности $U \in W$ и аппроксимации последовательности Y последовательностью $X \in \mathcal{X}(U)$, по критерию минимума суммы квадратов расстояний между элементами Y и X, т.е. задача

$$\|Y - X\|^2 \to \min_{U, \mathcal{X}(U)}. \tag{1.6}$$

Здесь $\mathcal{X}(U)$, $U \in W$, – множество допустимых аппроксимирующих последовательностей, порожденных последовательностью U. Каждый элемент $X = (x_1, \ldots, x_N) \in \mathcal{X}(U)$ однозначно определяется наборами $\mathcal{M}$, $\mathcal{P}$, $\mathcal{J}$, удовлетворяющими ограничениям (1.2)–(1.5), по правилу

$$x_n = \sum_{m=1}^{M} h^{(m)}_{n-n_m+1}, \quad n = 1, \ldots, N, \tag{1.7}$$

где

$$h^{(m)}_i = \begin{cases} u_{J^{(m)}(i)}, & \text{если} \quad i = 1, \ldots, p_m, \\ 0, & \text{если} \quad i < 1, \quad i > p_m, \end{cases} \quad m = 1, \ldots, M. \tag{1.8}$$

Равенство (1.8) устанавливает связь между элементами последовательностей $h^{(m)}_i$ и U. Из этого равенства видно, что каждая из последовательностей $h^{(m)}_i$ – растянутая до длины p_m последовательность U, а кратность дублирования ее элементов определяется формулой

$$k^{(m)}_t = |\{i \,|\, J^{(m)}(i) = t, i \in \{1, \ldots, p_m\}\}|, \quad t = 1, \ldots, q(U), \tag{1.9}$$

причем

$$p_m = k^{(m)}_1 + \ldots + k^{(m)}_{q(U)}, \quad m = 1, \ldots, M. \tag{1.10}$$

Формула (1.7) – сумма M растянутых последовательностей вида (1.8). Таким образом, последовательность X включает в себя M повторов растянутых последовательностей U. Значение индекса $n = n_m$, $n_m \in \mathcal{M}$, определяет начальный номер m-го повтора, а значение $p = p_m$, $p_m \in \mathcal{P}$, и отображение $J = J^{(m)}$, $J^{(m)} \in \mathcal{J}$, – длину этого повтора и кратности дублирования элементов из U в этом повторе.

Неформально можно сказать, что каждая последовательность $X \in \mathcal{X}(U)$ включает в себя неизвестное число допустимых квазипериодических повторов последовательности U. При этом каждый повтор определяется 1) сдвигом U на переменную величину, которая между соседними повторами не превышает $T_{\max} \le N$; 2) допустимым растяжением последовательности U, путем дублирования ее компонент.

Из (1.7) и (1.8) следует, что для каждого $U \in W$ между элементами множества $\mathcal{X}(U)$ и наборами $\mathcal{M}$, $\mathcal{P}$, $\mathcal{J}$, удовлетворяющими ограничениям (1.2)–(1.5), существует взаимооднозначное соответствие, т.е. $X = X(U, \mathcal{M}, \mathcal{P}, \mathcal{J})$, поэтому задачу (1.6) можно эквивалентно записать в виде

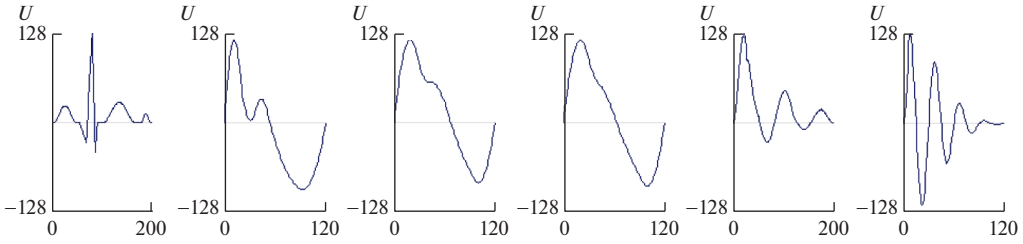
$$\|Y - X\|^2 = \|Y - X(U, \mathcal{M}, \mathcal{P}, \mathcal{J})\|^2 \to \min_{U, \mathcal{M}, \mathcal{P}, \mathcal{J}}. \tag{1.11}$$

Наконец, преобразовав $\|Y - X\|^2$ с учетом (1.7) и (1.8), имеем

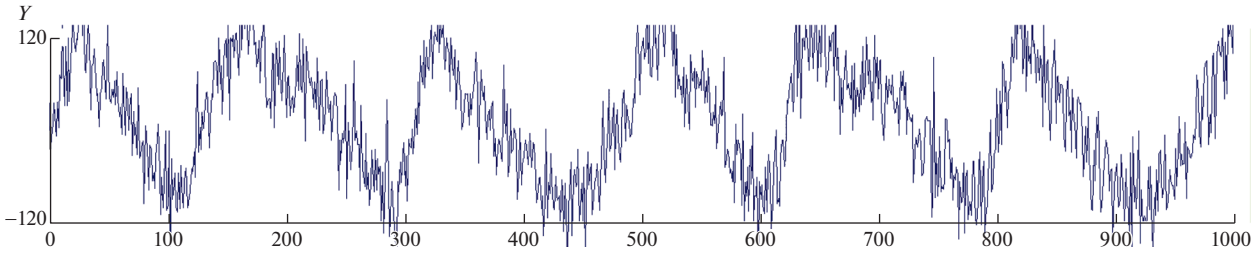
$$\sum_{n=1}^{N} (x_n - y_n)^2 = \sum_{n=1}^{N} y_n^2 + \sum_{m=1}^{M} \sum_{i=1}^{p_m} \{u^2_{J^{(m)}(i)} - 2y_{n_m+i-1} u_{J^{(m)}(i)}\}.$$

Легко видеть, что первое слагаемое в правой части этого равенства – константа, не зависящая от переменных задачи 1, а второе совпадает с целевой функцией задачи 1. Поэтому задача (1.11), а вместе с ней и задача (1.6), эквивалентна оптимизационной задаче 1.

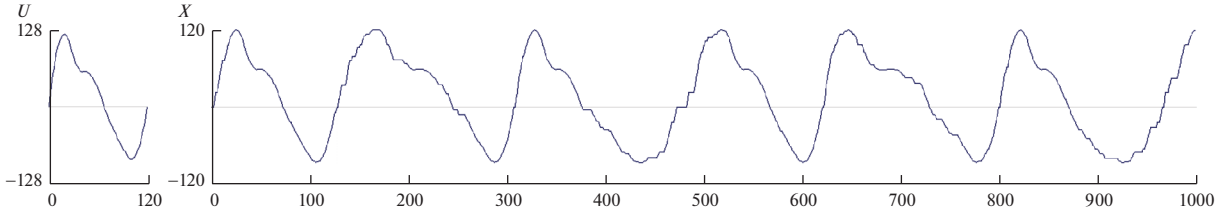
Чтобы проиллюстрировать данную интерпретацию, на фиг. 1 и 2 в виде графиков изображены входные данные задачи 1. Здесь и далее считаем, что номерам элементов последовательностей из алфавита и последовательности Y соответствуют дискретно-временные отсчеты непрерывных сигналов. На фиг. 1 приведен пример алфавита W, содержащего 6 последовательностей, на фиг. 2 – пример последовательности Y, порожденной одной из последовательностей алфавита.



Фиг. 1. Пример входных значений задачи 1. Алфавит W.



Фиг. 2. Пример входных значений задачи 1. Последовательность Y, подлежащая обработке.



Фиг. 3. Пример последовательности U из алфавита W и одной из последовательностей $X \in \mathcal{X}(U)$.

На фиг. 3а изображена последовательность U из алфавита W, на фиг. 3б – пример последовательности $X \in \mathcal{X}(U)$, построенной по правилам (1.7), (1.8), (1.9), (1.10) для некоторых допустимых значений $\{p_m\}$ и $\{k_t^{(m)}\}$. Ступенчатость сигнала на этой фигуре обусловлена кратными повторениями элементов из U.

2. БЛИЗКИЕ ПО ПОСТАНОВКЕ ЗАДАЧИ

Задача 1 является обобщением ранее исследованных задач распознавания. Частный случай задачи 1, в котором $q_1 = \ldots = q_K = q$, $p_m = q$, $J^{(m)}(i) = i$, $m = 1,\ldots,M$, исследовался в [1]. В [2] была рассмотрена модификация этого частного случая, когда M является частью входа задачи. В этом частном случае и его модификации, как и в задаче 1, требуется распознать квазипериодическую последовательность, но все повторы порождающей последовательности в ней идентичны. В [1] и [2] были построены алгоритмы, позволяющие получить оптимальное решение за время $\mathbb{O}(KT_{\max}N)$ и $\mathbb{O}(KMT_{\max}N)$. Другой частный случай, когда $K = 1$, был исследован в [3]. В этом частном случае алфавит состоит из единственной последовательности, поэтому задачу можно интерпретировать как задачу аппроксимации. В этой работе приведен точный алгоритм, позволяющий получить точное решение за время $\mathbb{O}(T_{\max}^3 N)$.

Близкими в постановочном плане являются задачи распознавания квазипериодических последовательностей по усеченным данным [4] и [5]. В этих задачах, как и в задаче 1, предполагалось, что распознаваемая квазипериодическая последовательность включает в себя искаженные допустимым образом повторы некоторой последовательности, принадлежащей заданному мно-

жеству. В отличие от задачи 1, здесь предполагалось, что множество допустимых преобразований последовательности состоит из всевозможных ее подпоследовательностей, полученных удалением некоторого количества начальных и конечных отсчетов. Для этих задач в цитируемых работах получены точные полиномиальные алгоритмы для их решения.

Алгоритм решения задачи 1 является подходящим инструментом для решения прикладных задач распознавания и анализа сигналов, имеющих квазипериодическую структуру в виде флуктуирующих повторов участков сигнала. Распознавание и анализ таких сигналов актуальны для различных приложений, имеющих дело с обработкой импульсных сигналов, полученных от природных источников: биомедицинские, геофизические и т.п. Далее, в разделе численное моделирование, будут приведены примеры обработки медицинских сигналов, имеющих вид квазипериодических последовательностей (ECG, PPG).

3. РАЗМЕР МНОЖЕСТВА ДОПУСТИМЫХ РЕШЕНИЙ

Приведенная выше интерпретация позволяет заметить, что число допустимых решений задачи 1 совпадает с мощностью множества $\mathscr{X} = \bigcup_{U\in W} \mathscr{X}(U)$. Из комбинаторных соображений легко видеть, что при каждом $U \in W$ справедлива оценка

$$\left|\mathscr{X}(U)\right| \le (N - q(U) + 1) \sum_{M=1}^{M_{\max}(U)} q(U)^{(T_{\max}-q(U))M} (T_{\max} - q(U) + 1)^{2M-1} \le$$
$$\le q(U)^{(T_{\max}-q(U))M_{\max}(U)} (N - q(U) + 1)(T_{\max} - q(U) + 1)^{2M_{\max}(U)-1} M_{\max}(U),$$

где $M_{\max}(U) = \lfloor N/q(U) \rfloor$ – максимально возможное число повторов последовательности U в Y. Данная верхняя оценка позволяет оценить влияние, которое параметры задачи 1 оказывают на мощность множества допустимых решений.

С другой стороны, за исключением тривиального случая $T_{\max} = q(U)$ справедлива нижняя оценка

$$\left|\mathscr{X}(U)\right| \ge 2^{\left\lfloor \frac{N-q(U)+1}{q(U)+1} \right\rfloor}, \quad U \in W.$$

Отсюда следует, что

$$\left|\mathscr{X}\right| = \sum_{U\in W} \left|\mathscr{X}(U)\right| \ge K 2^{\left\lfloor \frac{N-q_{\max}+1}{q_{\max}+1} \right\rfloor},$$

где $q_{\max} = \max\limits_{U\in W} q(U)$, а K – мощность алфавита последовательностей. Это означает, что если $q_{\max}$ ограничено некоторой константой (что типично для приложений), мощность множества $\mathscr{X}$ допустимых решений задачи 1 экспоненциально растет с ростом N. Очевидно, что перебрать напрямую элементы этого множества за приемлемое время вряд ли удастся, так как N – часть входа задачи. Несмотря на этот экспоненциальный рост, ниже будет приведен алгоритм, позволяющий получить оптимальное решение за полиномиальное время.

4. ОСНОВЫ АЛГОРИТМА

Для построения алгоритма решения задачи 1 нам потребуется следующая вспомогательная

Задача 2. *Дано:* числовые последовательности $Y = (y_1, \ldots, y_N)$, $U = (u_1, \ldots, u_{q(U)})$, и натуральные числа $T_{\max}$, ℓ. *Найти:* набор $\mathscr{M} = \{n_1, \ldots n_m, \ldots\}$ номеров последовательности Y, набор $\mathscr{P} = \{p_1, \ldots, p_m, \ldots\}$ натуральных чисел, набор $\mathscr{J} = \{J^{(1)}, \ldots, J^{(m)}, \ldots\}$ сжимающих отображений, в котором $J^{(m)} : \{1, \ldots, p_m\} \to \{1, \ldots, q(U)\}$, а также размерность M этих наборов, которые минимизируют целевую функцию

$$G(\mathscr{M}, \mathscr{P}, \mathscr{J}) = F(\bullet|U) \tag{4.1}$$

при ограничениях (1.2), (1.3), (1.4) на элементы искомых наборов $\mathscr{M}$, $\mathscr{P}$, и при ограничениях (1.5) на элементы искомых сжимающих отображений.

Для полноты изложения приведем алгоритм ее решения, базирующийся на результатах, полученных в [3]. Проанализировав ограничения (1.2), (1.3), (1.4), входящие в условия задачи 2, легко видеть, что справедливо

Утверждение 1. *Пусть выполнены условия задачи* 2*, тогда для компонент наборов* $\mathcal{M}$ *и* $\mathcal{P}$ *справедливо*:

1) $n_m \in \omega$, $m = 1,\ldots,M$, *где*

$$\omega = \{1,\ldots,N - q(U) + 1\};$$

2) $n_m \in \omega^+$, $m = 2,\ldots,M$, *где*

$$\omega^+ = \{q(U) + 1,\ldots,N - q(U) + 1\};$$

3) *если* $n_m = n$, $m = 1,\ldots M$, $n \in \omega$, *то* $p_m \in \delta(n)$, *где*

$$\delta(n) = \{q(U),\ldots,\min\{\ell, N - n + 1\}\};$$

4) *если* $n_m = n$, $m = 2,\ldots,M$, $n \in \omega^+$, *то* $n_{m-1} \in \gamma(n)$, *где*

$$\gamma(n) = \{k \,|\, \max\{n - T_{\max}, 1\} \le k \le n - q(U)\};$$

5) *если* $n_m = n$ *и* $n_{m-1} = j$, $m = 2,\ldots,M$, $n \in \omega^+$, $j \in \gamma(n)$ *то* $p_{m-1} \in \theta(n, j)$, *где*

$$\theta(n, j) = \{q(U),\ldots,\min\{\ell, n - j\}\}.$$

Опираясь на это утверждение, выпишем алгоритм решения задачи 2.

Алгоритм $\mathcal{A}_1$

Вход: Y, U, $T_{\max}$ и ℓ.

Прямой ход алгоритма.

Шаг 1. Для каждого $n = 1,\ldots,N - q(U) + 1$ выполним:

Для каждого $p = q(U),\ldots,\min\{\ell, N - n + 1\}$ выполним:

(1) Вычислим

$$w_{s,t} = u_t^2 - 2y_{n+s-1}u_t, \quad s = 1,\ldots,p, \quad t = 1,\ldots,q(U).$$

(2) Найдем значение W^* по формулам

$$W^* = W_{p,q(U)},$$

$$W_{s,t} = \min\{W_{s-1,t}, W_{s-1,t-1}\} + w_{s,t}, \quad s = 1,\ldots,p, \quad t = 1,\ldots,q(U),$$

$$W_{s,t} = \begin{cases} 0, & s = 0, \quad t = 0, \\ +\infty, & s = 0, \quad t = 1,\ldots,q(U), \\ +\infty, & s = 1,\ldots,p, \quad t = 0, \end{cases}$$

и последовательность $J^*(1),\ldots,J^*(p)$ по формулам

$$J^*(p) = q(U);$$

$$J^*(i-1) = \begin{cases} J^*(i), & \text{если} \quad W_{i-1,J^*(i)} \le W_{i-1,J^*(i)-1}, \\ J^*(i) - 1, & \text{если} \quad W_{i-1,J^*(i)} > W_{i-1,J^*(i)-1}, \end{cases}$$

$$i = p, p-1,\ldots,2.$$

(3) Положим $W^*(n, p) = W^*$; $J^*(n, p) = \{J^*(i), i = 1,\ldots,p\}$.

Шаг 2. Вычислим совокупность значений $G(n, p)$, $p = q(U),\ldots,\ell$, $n = 1,\ldots,N - p + 1$, по формулам

$$G(n, p) = \begin{cases} W^*(n, p), & n \in \omega, \quad p \in \delta(n), \\ \min\{0, \min\limits_{j \in \gamma(n)} \min\limits_{i \in \theta(n,j)} G(j, i)\} + W^*(n, p)), & n \in \omega^+, \quad p \in \delta(n), \end{cases}$$

а так же значение G^* по формуле

$$G^* = \min_{n\in\omega}\min_{p\in\delta(n)} G(n, p).$$

Положим $F_A = G^*$.

Обратный ход алгоритма.

Шаг 3. Вычислим $\pi(n)$ и $I(n)$, $n = 1,\ldots, N - q(U) + 1$, по формулам

$$I(n) = \begin{cases} 0, & \text{если} \quad n \in \omega\backslash\omega^+, \\ 0, & \text{если} \quad \min\limits_{j\in\gamma(n)}\min\limits_{i\in\theta(n,j)} G(j,i) \geq 0, \quad n \in \omega^+, \\ \arg\min\limits_{j\in\gamma(n)}\left\{\min\limits_{i\in\theta(n,j)} G(j,i)\right\}, & \text{если} \quad \min\limits_{j\in\gamma(n)}\min\limits_{i\in\theta(n,j)} G(j,i) < 0, \quad n \in \omega^+, \end{cases}$$

и

$$\pi(n) = \begin{cases} 0, & \text{если} \quad I(n) = 0, \\ \arg\min\limits_{i\in\theta(n,I(n))} G(I(n), i), & \text{если} \quad I(n) > 0, \quad n \in \omega. \end{cases}$$

Шаг 4. Найдем компоненты вспомогательных наборов $(\pi_1,\ldots,\pi_M)$ и $(\nu_1,\ldots,\nu_M)$ и их размерность M по формулам

$$\pi_1 = \arg\min_{p\in\{q(U),\ldots,\ell\}}\left\{\min_{n\in\{1,\ldots,N-p+1\}} G(n, p)\right\},$$

$$\nu_1 = \arg\min_{n\in\{1,\ldots,N-\pi_1+1\}} G(n, \pi_1),$$

$$\pi_m = \pi(\nu_{m-1}), \quad \nu_m = I(\nu_{m-1}), \quad m = 2,\ldots, M,$$

где M – наименьшее значение m, при котором $\pi(\nu_m) = 0$.

Шаг 5. Положим: $M_{A_1} = M$, $\mathcal{M}_{A_1} = \{\nu_M,\ldots,\nu_1\}$, $\mathcal{P}_{A_1} = \{\pi_M,\ldots,\pi_1\}$, $\mathcal{J}_{A_1} = \{J^*(\nu_M,\pi_M),\ldots,J^*(\nu_1,\pi_1)\}$.

Выход: M_{A_1}, $\mathcal{M}_{A_1}$, $\mathcal{P}_{A_1}$, $\mathcal{J}_{A_1}$ и G_{A_1}.

Утверждение 2 (см. [3]). *Алгоритм $\mathcal{A}_1$ находит точное решение задачи* 2 *за время* $\mathbb{O}(T_{\max}^3 N)$.

5. АЛГОРИТМ

Опираясь на вспомогательную задачу 2 и утверждение 2, сформулируем алгоритм решения задачи 1.

Алгоритм $\mathcal{A}$

Вход: Y, W, $T_{\max}$ и ℓ.

Шаг 1. Для каждого $U \in W$ выполним алгоритм $\mathcal{A}_1$ с входными данными Y, U, $T_{\max}$ и ℓ. Положим $M(U) = M_{A_1}$, $\mathcal{M}(U) = \mathcal{M}_{A_1}$, $\mathcal{P}(U) = \mathcal{P}_{A_1}$, $\mathcal{J}(U) = \mathcal{J}_{A_1}$, $G_U = G_{A_1}$.

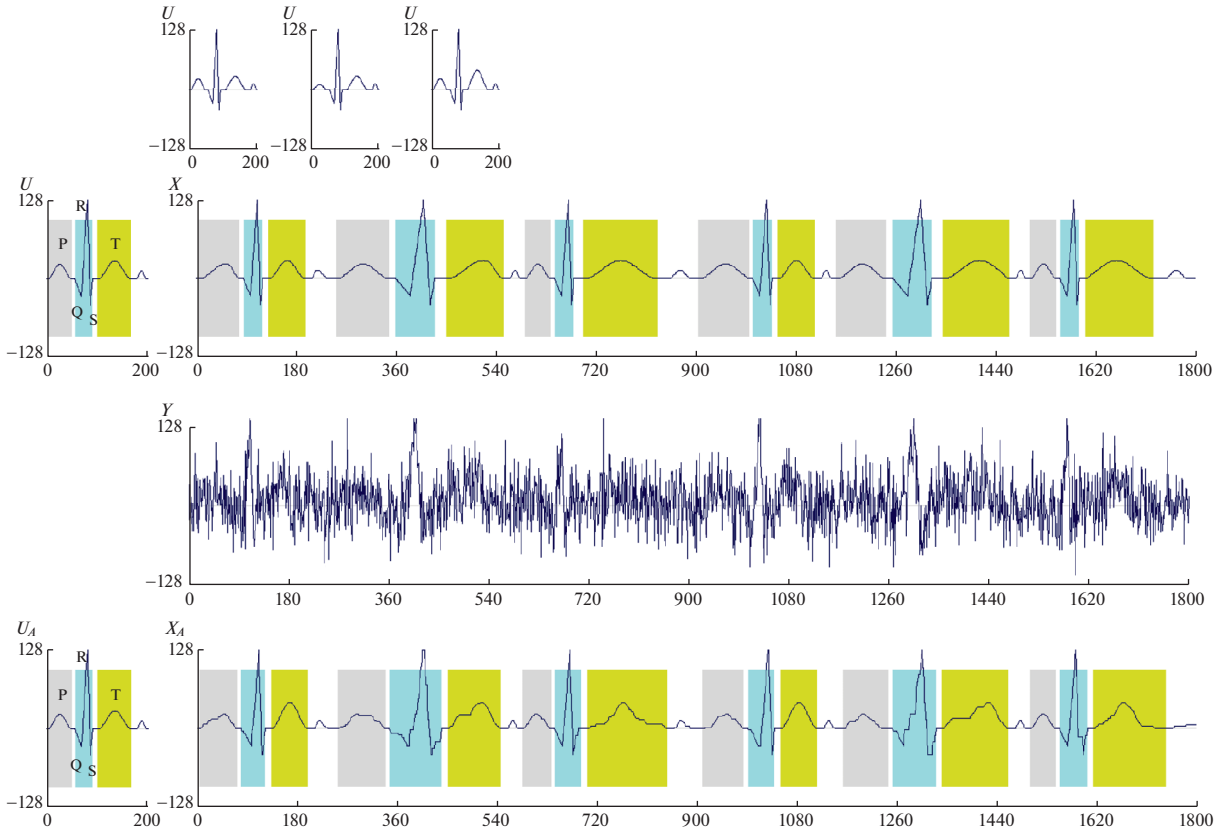
Шаг 2. Вычислим

$$\begin{aligned} F_A &= \min_{U\in W} G_U, \\ U_A &= \arg\min_{U\in W} G_U. \end{aligned} \tag{5.1}$$

Шаг 3. Положим $M_A = M(U_A)$, $\mathcal{M}_A = \mathcal{M}(U_A)$, $\mathcal{P}_A = \mathcal{P}(U_A)$, $\mathcal{J}_A = \mathcal{J}(U_A)$.

Выход: U_A, M_A, $\mathcal{M}_A$, $\mathcal{P}_A$, $\mathcal{J}_A$ и F_A.

Теорема 1. *Алгоритм $\mathcal{A}$ находит точное решение задачи* 1 *за время* $\mathbb{O}(KT_{\max}^3 N)$.



Фиг. 4. Пример 1. Распознавание ECG-подобной последовательности.

Доказательство. Оптимальность решения, найденного алгоритмом $\mathscr{A}$, следует из Утверждения 2 и следующей цепочки равенств:

$$F^* \overset{1}{=} \min_{U,\mathcal{M},\mathcal{P},\mathcal{J}} F(U,\mathcal{M},\mathcal{P},\mathcal{J}) \overset{2}{=} \min_{U} \min_{\mathcal{M},\mathcal{P},\mathcal{J}} F(U,\mathcal{M},\mathcal{P},\mathcal{J}) \overset{3}{=}$$

$$\overset{3}{=} \min_{U} \left\{ \min_{\mathcal{M},\mathcal{P},\mathcal{J}} F(U,\mathcal{M},\mathcal{P},\mathcal{J}|U) \right\} \overset{4}{=} \min_{U} \min_{\mathcal{M},\mathcal{P},\mathcal{J}} G(\mathcal{M},\mathcal{P},\mathcal{J}) \overset{5}{=} \min_{U} G_U \overset{6}{=} F_A.$$

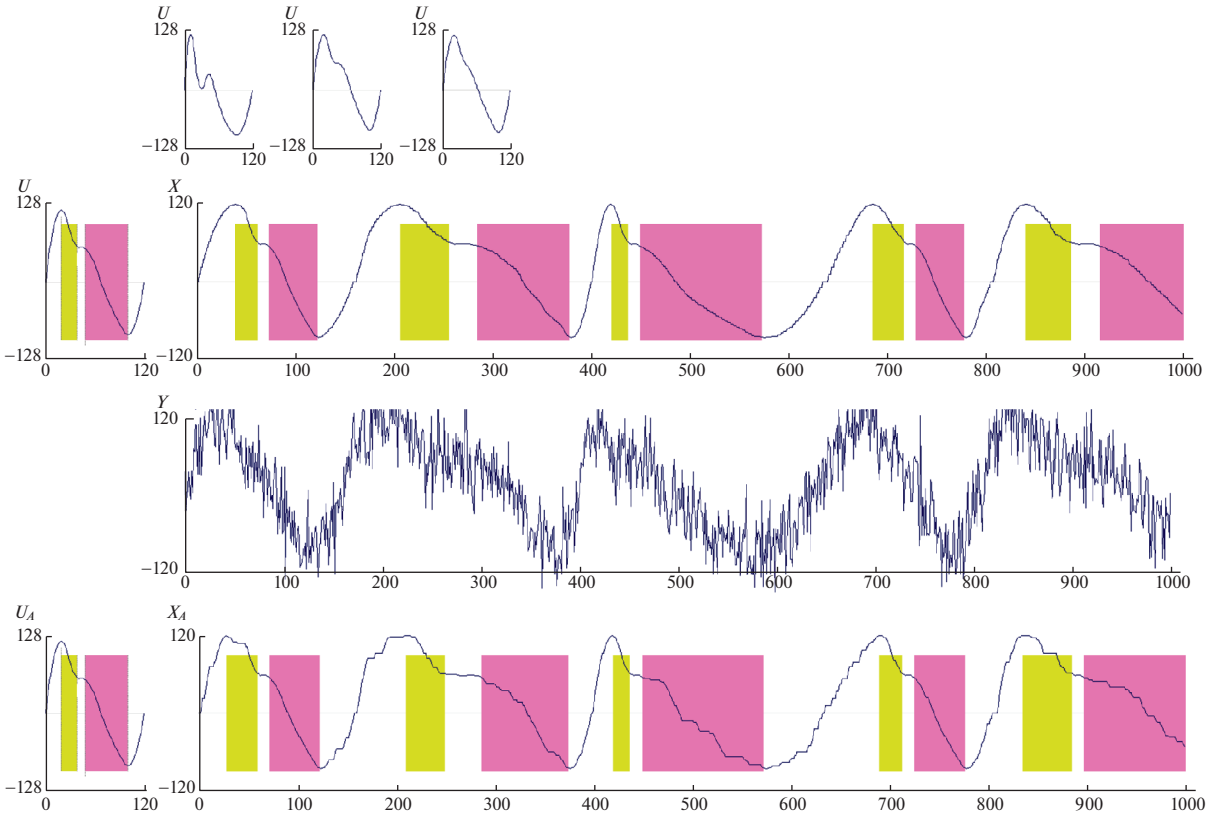
В этой цепочке равенство 1 – определение оптимального решения радачи 1. Равенство 2 следует из вида целевой функции (1.1). Равенство 3 очевидно. Равенство 4 следует из формулы (4.1), равенство 5 – определение оптимального решения задачи 2 при каждом фиксированном значении U. Наконец, равенство 6 следует из формулы (5.1) в пошаговой записи алгоритма $\mathscr{A}$.

В соответствии с уверждением 2, время работы алгоритма $\mathscr{A}_1$ оценивается величиной $\mathbb{O}(T_{\max}^3 N)$. На шаге 1 алгоритма $\mathscr{A}$ требуется K раз выполнить алгоритм $\mathscr{A}_1$, поэтому трудоемкость шага 1 алгоритма $\mathscr{A}$ есть величина $\mathbb{O}(KT_{\max}^3 N)$. Шаг 2, очевидно, требует $\mathbb{O}(K)$ операций, а шаг 3 – константного числа операций. Суммируя временные затраты на всех шагах алгоритма, получаем итоговую трудоемкость алгоритма $\mathscr{A}$. Теорема доказана.

Замечание 1. Если $T_{\max}$ является частью входа задачи, а мощность K алфавита – фиксированный параметр, то время работы алгоритма равно $\mathbb{O}(N^4)$. Таким образом, алгоритм $\mathscr{A}$ является полиномиальным.

6. ПРИМЕРЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

Для иллюстрации работоспособности алгоритма приведем два примера обработки последовательностей (временных рядов), которые можно интерпретировать как квазипериодические последовательности флуктуирующих ECG-подобных (фиг. 4) и PPG-подобных (фиг. 5) импульсов



Фиг. 5. Пример 2. Распознавание PPG-подобной последовательности.

при наличии аддитивной помехи. В действительности, с математической точки зрения не имеет значения, какие именно последовательности включены в состав алфавита W. Выбор именно ECG и PPG-подобных сигналов обусловлен желанием проиллюстрировать применимость алгоритма для биомедицинских приложений. Форма этих импульсов, а так же характерные участки и значимые точки описаны экспертами, см., например, [6]–[9]. На фиг. 4 (ECG-импульс) этими участками являются P-Q-R-S-T-U, на фиг. 5 (PPG-импульс) – систолическая точка, дикротическая точка, диастолическая точка. Раскраска фигур маркирует характерные участки или интервалы между значимыми точками.

Рассмотрим подробнее фиг. 4, иллюстрирующую помехоустойчивую обработку ECG-подобного сигнала. В верхнем ряду изображен алфавит. Ниже слева изображена последовательность U – одна из последовательностей из алфавита. Справа от нее изображена программно-сгенерированная последовательность – квазипериодическая последовательность, порожденная флуктуирующими повторами импульса U. Под модельной последовательностью изображена последовательность Y – результат сложения модельной последовательности и последовательности независимых одинаково распределенных гауссовских случайных величин с нулевым математическим ожиданием. Важно отметить, что только Y и W являются входными данными для алгоритма. Последовательность U и модельная последовательность X, изображенные во втором ряду фигуры, приведены для иллюстрации; эти данные недоступны.

В нижней части фигуры представлены результаты работы алгоритма $\mathcal{A}$: слева последовательности U_A, которая является оценкой для модельной, ненаблюдаемой последовательности. Справа изображены компоненты последовательности X_A, восстановленной по формулам (1.7) и (1.8) с использованием четырех наборов U_A, $\mathcal{M}_A$, $\mathcal{P}_A$ и $\mathcal{J}_A$, полученных в результате работы алгоритма $\mathcal{A}$. Данные на фиг. 4 получены при $K = 3$, $q(U) = 203$, $U \in W$, $T_{\max} = 370$, $N = 1800$, максимальная амплитуда сигнала -128, уровень шума $\sigma = 35$.

Фигура 5 показывает результаты обработки PPG-подобного сигнала. Эта фигура имеет такую же структуру, как и фиг. 4. Данные на фиг. 5 получены при $K = 3$, $q(U) = 120$, $U \in W$, $T_{\max} = 240$, $N = 1000$, максимальная амплитуда сигнала – 128, уровень шума $\sigma = 35$.

Приведенные примеры показывают, что построенный алгоритм позволяет с вполне приемлемым качеством обрабатывать данные в виде квазипериодических последовательностей флуктуирующих импульсов. Во-первых, неизвестная последовательность U и определенная алгоритмом последовательность U_A совпадают в обоих примерах (распознавание произведено корректно). Во-вторых, визуальное сравнение графиков ненаблюдаемой последовательности X и восстановленной последовательности X_A демонстрирует лишь незначительные отклонения одного графика от другого и практически точное совпадение всех характерных точек.

ЗАКЛЮЧЕНИЕ

В работе показано, что одна из неисследованных задач дискретной оптимизации полиномиально разрешима. Полиномиальная разрешимость доказана конструктивно, т.е. построен алгоритм, гарантирующий получение оптимального решения задачи и получена полиномиальная оценка его трудоемкости.

Результаты численного моделирования продемонстрировали, что предложенный алгоритм может служить подходящим инструментом для решения задач помехоустойчивого распознавания и анализа квазипериодических импульсных последовательностей. Приведены примеры обработки ECG и PPG-подобных сигналов.

Остается неизученной модификация задачи 1, в которой число суммируемых сверток является частью входа задачи. Значительный математический интерес так же представляет дискретная экстремальная задача, когда алфавит последовательностей не задан, т.е. требуется по входной последовательности Y распознать последовательность U как элемент континуального множества числовых последовательностей, имеющих фиксированную конечную длину. Исследование этих задач представляется делом ближайшей перспективы.

СПИСОК ЛИТЕРАТУРЫ

1. *Kel'manov A.V., Khamidullin S.A., Okol'nishnikova L.V.* Recognition of a quasiperiodic sequence containing identical subsequences-fragments // Pattern Recognition and Image Analysis. 2004. V. 14. № 1. P. 72–83.
2. *Kel'manov A.V., Khamidullin S.A.* Recognizing a quasiperiodic sequence composed of a given number of identical subsequences // Pattern Recognition and Image Analysis. 2000. V. 10. № 1. P. 127–142.
3. *Kel'manov A.V., Khamidullin S.A., Mikhailova L.V., Ruzankin P.S.* Polynomial-time solvability of one optimization problem induced by processing and analyzing quasiperiodic ECG and PPG signals // Communications in Computer and Information Science. 2020. V. CCIS 1145. P. 88–101.
4. *Kel'manov A.V., Khamidullin S.A.* Algorithm of recognition of a quasiperiodic sequence composed of a given number of truncated subsequences // Pattern Recognition and Image Analysis. 2001. V. 11. № 1. P. 43–46.
5. *Кельманов А.В., Хамидуллин С.А.* Распознавание числовой последовательности по фрагментам квазипериодически повторяющейся эталонной последовательности // Сиб. ж. индустр. матем. 2004. Т. 7. № 2. С. 68–87.
6. *Rajni R., Kaur I.* Electrocardiogram signal analysis – an overview // Int. J. Comput. Appl. 2013. V. 84. № 7. P. 22–25.
7. *Al-Ani M.S.* ECG Waveform Classification Based on P-QRS-T Wave Recognition // UHD Journal of Science and Technology. 2018. V. 2. № 2.
8. *Shelley K., Shelley S.* Pulse Oximeter Waveform: Photoelectric Plethysmography. In: Carol Lake, Hines R., Blitt C. (eds). Clinical Monitoring. W.B. Saunders Company. 2001. P. 420–428.
9. *Elgendi M.* On the analysis of fingertip photoplethysmogram signals // Current Cardiology Reviews. 2012. V. 8. № 1. P. 14–25.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1172–1178

ИНФОРМАТИКА

УДК 519.87

НЕЙРОННАЯ СЕТЬ С ГЛАДКИМИ ФУНКЦИЯМИ АКТИВАЦИИ И БЕЗ УЗКИХ ГОРЛОВИН ПОЧТИ НАВЕРНОЕ ЯВЛЯЕТСЯ ФУНКЦИЕЙ МОРСА[1)]

© 2021 г. С. В. Курочкин

109028 Москва, Покровский бул., 11, Нац. исследовательский ун-т "Высшая школа экономики", Россия

e-mail: skurochkin@hse.ru

Поступила в редакцию 15.12.2019 г.
Переработанный вариант 15.12.2019 г.
Принята к публикации 15.09.2020 г.

Доказано, что нейронная сеть с функциями активации типа сигмоидной является функцией Морса для почти всех, в смысле меры Лебега, наборов своих параметров (весов) в случае, когда архитектура сети не предусматривает сужений – слоев, в которых количество нейронов меньше, чем в соседних. На примерах показано, что требование отсутствия горловин является существенным. Библ. 16. Фиг. 1.

Ключевые слова: нейронная сеть, функции Морса.

DOI: 10.31857/S0044466921070103

1. ВВЕДЕНИЕ

Искусственные нейронные сети стали весьма распространенным и во многих случаях эффективным инструментом для решения различных задач анализа данных. Возможность с их помощью распознавать/аппроксимировать сложные нелинейные зависимости в данных подтверждена практикой. Теоретическим подкреплением такой достаточно универсальной применимости нейронных сетей выступает так называемая теорема Цыбенко (см. [1], [2]) – ряд результатов, полученных независимо различными авторами в конце 1980-х годов, по смыслу близких к классической теореме Соуна–Вейерштрасса в применении к конкретному множеству аппроксимируемых и запасу аппроксимирующих функций.

Предметом настоящей работы является теоретическое обоснование другого реально наблюдаемого и используемого свойства функций, получаемых в результате аппроксимации точечных или дискретных данных посредством нейросетей: возможности получать информацию об исследуемом объекте, анализируя структуру линий уровня и/или индексы критических точек аппроксимирующей функции. Целесообразность такого подхода проявляется, например, в задачах анализа изображений: согласно современным представлениям, как зрение человека (см. [3]), так и наиболее продвинутые системы машинного зрения (см. [4, гл. 4, 5]) существенно используют анализ контуров. Пример результата такого типа описан в [5], где предложен метод распознавания гомотопического типа объекта через степень аппроксимирующего отображения.

Среди всех вообще дифференцируемых функций нескольких переменных (и функций на многообразиях) функции Морса выделяются именно регулярным устройством своих линий уровня, их перестройкой при изменении уровня, а в количестве и индексах их критических точек содержится важная информация (см., например, [6]–[8]). Свойство быть функцией Морса является свойством общего положения: такие функции образуют открытое всюду плотное множество в пространстве дифференцируемых функций (точные формулировки см. в указанных и многих других текстах по теории Морса). Для применения в прикладных задачах, где пространство всевозможных функций описывается конечным (возможно, как в случае нейронных сетей, очень большим) числом параметров, такой результат представляется недостаточным. Желательно иметь уверенность в том, что в данном пространстве дополнение к функциям Морса имеет нулевую меру. Практически это будет означать, что при решении реальной задачи функции, получаемые на всех шагах так называемого обучения нейронной сети (и, разумеется, сама обученная сеть), будут функциями Морса, и это даст дополнительные возможности для анализа.

[1)]Результаты работы получены в рамках НИР, реализуемой в ЦХАБД МГУ им. М.В. Ломоносова.

В данной работе получено условие на архитектуру нейронной сети, при котором для почти всех наборов параметров (так называемых весов) реализуемое сетью отображение будет функцией Морса. Смысл условия в том, что в сети не должно быть узких горловин (bottleneck) – когда в каком-то слое количество нейронов строго меньше, чем в слоях по обе стороны от данного. Сети с горловиной (обычно, одной) используются в специальных целях, в частности, как автокодировщики, когда требуется понизить размерность задачи путем выбора меньшего количества признаков (features), чем размерность входного вектора. Сети без горловин являются обычной практикой, они же фигурируют в теоремах об универсальной аппроксимации (см. [1]). Также на примере показано, что уже простейшая сеть с горловиной может не быть функцией Морса для множества параметров положительной меры.

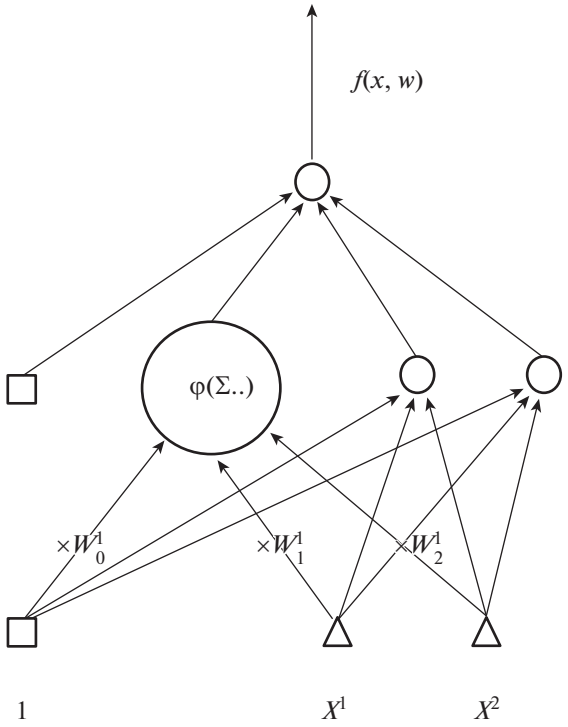
Возможно, наиболее близким по смыслу известным автору результатом является утверждение, что почти все (в смысле меры Лебега в пространстве коэффициентов) многочлены нескольких переменных являются функциями Морса (см. [9], [10]).

Структура работы следующая. В разд. 2 даны терминология по нейронным сетям и необходимые сведения из дифференциальной топологии. В разд. 3 сформулирован основной результат. Далее в основном тексте дано его доказательство на примере сети классической архитектуры – с одним скрытым слоем. Основная идея работает уже на этом случае, и ход рассуждения можно проследить наглядно. Доказательство в общем случае технически несколько сложнее и вынесено в Приложение. В Заключении сформулированы выводы и возможные открытые вопросы. В Приложении также приведены доказательства основной теоремы и двух лемм.

2. ТЕРМИНОЛОГИЯ И ПОСТАНОВКА ЗАДАЧИ

2.1. Нейронные сети

В математических терминах нейронная сеть – это вещественнозначная функция нескольких вещественных переменных, являющаяся композицией (последовательным применением) нескольких отображений вида: аффинное преобразование, затем покоординатное применение фиксированной нелинейной функции (так называемой функции активации). Коэффициенты аффинных преобразований являются настраиваемыми параметрами сети и называются весами. Иногда те из них, которые являются свободными членами соответствующих выражений, называются порогами или смещениями. Каждое отдельное взятие аффинной формы с последующим преобразованием посредством функции активации называется нейроном. В качестве функций активации могут выбираться различные варианты. На первом этапе развития теории использовались ступенчатые функции (Хевисайда), что давало возможность строить универсальные классификаторы, однако итоговое отображение получается разрывным, и задача нахождения наилучших весов имеет экспоненциальную сложность. Затем было предложено использовать сглаженные аналоги, в частности: сигмоидную, или логистическую $\varphi(x) = 1/[1 + \exp(-x)]$, арктангенс, гиперболический тангенс и некоторые другие. Использование дифференцируемых функций активации, вместе с простым способом вычисления градиента по весам от функции ошибок сети (так называемый метод обратного распространения ошибки), дало существенное продвижение и обусловило по сей день широкое применение именно таких сетей. При этом, как было доказано теоретически и как показала практика, конкретный выбор функции активации, хотя и доступен в применяемых программных реализациях, не влияет на результат сколько-нибудь существенным образом, важны лишь общие свойства таких функций: дифференцируемость, строгая монотонность, ограниченность, унимодальность первой производной. Также в последнее время в связи с резко возросшим объемом задач анализа дискретных данных используются недифференцируемые функции, например, $\mathrm{Re}\, LU(x) = \max(0, x)$. На фиг. 1 представлен пример нейронной сети, на вход которой подается двумерный вектор, его координаты преобразуются независимо тремя нейронами промежуточного, или скрытого, слоя, выходы которых, в свою очередь, суммируются и преобразуются выходным нейроном (архитектура 2-3-1). Как это обычно делается для единообразия и наглядности, на фиг. 1 добавлены условные входы, тождественно равные единице, которые умножатся на пороги нейронов. В результате получается функция $f(x, w)$ двух переменных $x = (x^1, x^2)$, зависящая от $3 \times 3 + 4 = 13$ параметров-весов (объединенных в вектор w). Большое количество настраиваемых параметров характерно для нейронных сетей вообще и особенно для таких, где аффинное+нелинейное преобразование (называется слоем сети) последовательно делается много раз (такие сети называются глубокими или глубинными). Задача нахождения наилучшего (или удовлетворительного) набора весов ставится как задача минимизации ошибки аппроксимации на заданном (обучающем) наборе данных, который содержит входные векторы x_k, $k = 1, 2, \ldots, N$, и соответствующие им целевые значения y_k, N – количество наблюдений. Эта задача глобальной безусловной невыпуклой оптимизации ре-



Фиг. 1. Пример искусственной нейронной сети. Для одного из нейронов скрытого слоя (выделен) показан принцип преобразования входного сигнала.

шается методами типа градиентного спуска, иногда второго порядка (с использованием гессиана), с регуляризацией, препятствующей чрезмерной подгонке к данным (переобучение, overfitting), и в сочетании с методами стохастической оптимизации. Подробно тема изложена во многих источниках (см., например, [11]). Итерации процесса оптимизации называются шагами обучения, которое для больших задач может занимать длительное время даже на мощных процессорах. Естественно, при таком подходе внимания требуют методы эффективного вычисления градиента и гессиана (см., например, [12]). Однако и то и другое берется по отношению к весам w, а не по аргументу x, и соответствующие результаты не удается применить к рассматриваемой здесь задаче.

2.2. Сведения из дифференциальной топологии

Здесь кратко сформулированы начальные понятия и результаты дифференциальной топологии, необходимые для изложения. Подробно материал изложен во многих высококачественных текстах (см., например, [6]–[8]). Пусть $U \subset \mathbb{R}^n$ – область, $f : U \to \mathbb{R}$ – дифференцируемая функция. Точка $x \in U$ называется регулярной, если градиент f в этой точке не равен нулю, и критической – в противном случае. Число $y \in \mathbb{R}$ является критическим значением для f, если $y = f(x)$ для некоторой критической точки x. Если в $f^{-1}(y)$ нет критических точек (в частности, если прообраз пуст), то такое значение y называется регулярным для f. Все остальные значения являются критическими. Важный результат – теорема Сарда: множество критических значений имеет меру ноль. Критическая точка называется невырожденной, если в этой точке гессиан f является невырожденной матрицей. Невырожденные критические точки изолированы. В окрестности такой точки после подходящей замены координат в векторе x функция представляется в виде $f(x) = -(x^1)^2 - \ldots - (x^q)^2 + (x^{q+1})^2 + \ldots + (x^n)^2$ (лемма Морса), число q называется индексом этой критической точки. Если функция имеет только невырожденные критические точки, то она называется функцией Морса. Функции Морса существуют и всюду плотны в пространстве дифференцируемых функций. В силу своих хороших дифференциальных свойств, отмеченных во Введении, они представляют вполне прикладной интерес. Но с наибольшей силой это понятие работает, когда функция определена не на подмножестве $\mathbb{R}^n$, а на многообразии: количество и индексы критических точек произвольной функции Морса связаны с топологией многообразия.

Вопрос, рассматриваемый в данной работе, формулируется так: дана архитектура нейронной сети; можно ли, и при каких условиях, утверждать, что для почти всех наборов весов (в смысле меры Лебега в пространстве весов) соответствующая сеть является функцией Морса. В следующем разделе будет получен такой критерий, а также рассмотрены контрпримеры.

Очевидно, что “почти всех” нельзя заменить на “всех” – любая нейронная со всеми весами, равными нулю, дает на выходе константу.

3. КРИТЕРИЙ ТОГО, ЧТО НЕЙРОННАЯ СЕТЬ ЯВЛЯЕТСЯ ФУНКЦИЕЙ МОРСА

Теорема. *Пусть $f(x,w)$ – нейронная сеть с произвольным количеством слоев и нейронов в слоях, функциями активации φ типа сигмоидной и условием, что в ней нет такого промежуточного слоя, количество нейронов в котором строго меньше, чем в некоторых слоях по обе стороны от него (условие отсутствия горловины; при этом входной вектор в данном случае также считается слоем с количеством нейронов, равным размерности пространства признаков). Считаем, что $f : U \times W \to \mathbb{R}$, $x \in U \subset \mathbb{R}^n$, $w \in W \subset \mathbb{R}^p$, U, W – области соответственно в пространстве признаков и пространстве весов. Тогда для почти всех w для любого x частная производная $\partial^2 f(x,w)/\partial w \partial x$, рассматриваемая как линейное отображение $\partial(\partial f(x,w)/\partial x)/\partial w : \mathbb{R}^p \to \mathbb{R}^n$ (пространство $\mathbb{R}^n$ отождествляется со своим сопряженным) является сюръекцией в точке (x,w).*

Доказательство. Здесь будет рассмотрен случай сети архитектуры 2-3-1 (фиг. 1). Доказательство для общего случая вынесено в Приложение.

На вход i-го, $i = 1,2,3$, нейрона скрытого слоя подается вектор (x^1, x^2). От него берутся аффинные формы $s^i = w_0^i + w_1^i x_1 + w_2^i x_2$, $i = 1,2,3$, затем результат покоординатно преобразуется функцией активации φ. От полученных выходов y^1, y^2, y^3 берется аффинная форма $\tilde{s} = \tilde{w}_0 + \tilde{w}_1 y_1 + \tilde{w}_2 y_2 + \tilde{w}_3 y_3$ и окончательно, $f(x,w) = \varphi(\tilde{s})$. Имеем

$$\frac{\partial f(x,w)}{\partial x} = \varphi'(\tilde{s})(\tilde{w}_1, \tilde{w}_2, \tilde{w}_3)\,\mathrm{diag}(\varphi'(s^1), \varphi'(s^2), \varphi'(s^3))\begin{pmatrix} w_1^1 & w_2^1 \\ w_1^2 & w_2^2 \\ w_1^3 & w_2^3 \end{pmatrix} \tag{1}$$

(произведение скаляра, 1×3-строки, 3×3-матрицы и 3×2-матрицы, результат – 1×2-строка). Тогда, например, для производной по порогу первого нейрона скрытого слоя имеем

$$\frac{\partial^2 f(x,w)}{\partial w_0^1 \partial x} = \frac{\partial}{\partial s^1}[\varphi'(\tilde{s})(\tilde{w}_1, \tilde{w}_2, \tilde{w}_3)\,\mathrm{diag}(\varphi'(s^1), \varphi'(s^2), \varphi'(s^3))]\begin{pmatrix} w_1^1 & w_2^1 \\ w_1^2 & w_2^2 \\ w_1^3 & w_2^3 \end{pmatrix}. \tag{2}$$

Для краткости записи выражение в квадратных скобках (это 1×3-строка) обозначим через $u(x,w)$. Далее

$$\frac{\partial^2 f(x,w)}{\partial w_1^1 \partial x} = x^1 \frac{\partial}{\partial s^1} u(x,w)\begin{pmatrix} w_1^1 & w_2^1 \\ w_1^2 & w_2^2 \\ w_1^3 & w_2^3 \end{pmatrix} + u(x,w)\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{3}$$

и аналогично для производной по w_2^1 и для производных по весам двух других нейронов 1-го (скрытого) слоя. Вычитая выражение (2), домноженное на x^i, из выражения (3) для соответствующего веса, получаем, что уже только вариация весов w_i^j скрытого слоя позволяет получать в образе касательного пространства w при отображении $\partial^2 f(x,w)/\partial w \partial x$ всевозможные строки вида uA, где A – произвольная 3×2-матрица. При этом всегда $\varphi' \neq 0$ и для почти всех (в смысле меры Лебега) наборов весов $\tilde{w}_i$ строка $(\tilde{w}_1, \tilde{w}_2, \tilde{w}_3)$, а тем самым и строка u, ненулевая. Следовательно, для почти всех наборов весов в виде uA можно представить любую 1×2-строку.

Далее потребуется следующий факт из дифференциальной топологии.

Лемма 1. *Пусть U, W – области соответственно в $\mathbb{R}^n$ и $\mathbb{R}^p$, $p \geq n$, $f(x,w)$ – дифференцируемая функция, $f : U \times W \to \mathbb{R}$. Обозначим $f_w(x) = f(x,w)$, $f_w : U \to \mathbb{R}$, $df_w : U \to \mathbb{R}^n$ – ее производная по x,*

$$df_w(x) = \left(\frac{\partial f_w(x)}{\partial x_1}, \ldots, \frac{\partial f_w(x)}{\partial x_n} \right),$$

и $F(x,w) = df_w(x)$, $F : U \times W \to \mathbb{R}^n$. Пусть известно, что в любой точке (x,w) производная от F по w является сюръекцией. Тогда для почти всех w (в смысле обычной меры Лебега в $\mathbb{R}^p$) f_w является функцией Морса.

Это утверждение может быть получено из [4, теорема 1.2.4]. Для замкнутости изложения в Приложении приведено краткое доказательство.

Непосредственное применение леммы 1 дает

Следствие. Нейронная сеть, удовлетворяющая условиям теоремы, для почти всех наборов весов w является функцией Морса.

Замечание 1. Приведенное выше доказательство годится не только для конкретной сети на фиг. 1, но и для любой сети с одним промежуточным слоем: при произвольной размерности входного вектора и любом количестве нейронов в слое.

Следующий пример демонстрирует связь между наличием горловины и возможностью существования вырожденных критических точек.

Контрпример. Рассмотрим сеть архитектуры 2-1-2-1:

$$f(x,w) = \varphi(\hat{w}_0 + \hat{w}_1\varphi(\tilde{w}_0^1 + \tilde{w}_1^1\varphi(w_0 + w_1x^1 + w_2x^2)) + \hat{w}_2\varphi(\tilde{w}_0^2 + \tilde{w}_1^2\varphi(w_0 + w_1x^1 + w_2x^2))).$$

Поскольку при преобразовании в первом слое размерность входного вектора понижается, все критические точки f, если они есть, обязаны быть вырожденными. Пусть в качестве функции активации φ будет, например, сигмоидная, и рассмотрим следующий набор весов: $w_0 = 0$, $w_1 = 1$, $w_2 = 0$, $\tilde{w}_0^1 = -10$, $\tilde{w}_1^1 = 1$, $\tilde{w}_0^2 = 10$, $\tilde{w}_1^2 = -1$, $\hat{w}_0 = 0$, $\hat{w}_1 = 1$, $\hat{w}_2 = 1$. Тогда точка $x = (0,0)$ является локальным минимумом, который устойчив, т.е. существует и мало меняется, при произвольных малых возмущениях всех весов. Таким образом, в пространстве весов существует множество положительной меры такое, что все соответствующие сети имеют вырожденную критическую точку.

Замечание 2. Из этого примера можно сконструировать другие, демонстрирующие различные дифференциальные свойства нейронных сетей как отображений.

1. Если убрать один слой со стороны входа, т.е. взять 1-2-1-сеть, то она почти для всех весов будет функцией Морса и при этом для множества весов положительной меры будет иметь критические точки.

2. Если убрать еще один слой, то полученная 2-1-сеть для всех ненулевых наборов весов будет иметь только регулярные значения.

4. ЗАКЛЮЧЕНИЕ

Использование дифференциальных свойств функций для исследования топологии объекта, на котором они заданы или который они аппроксимируют, успешно применяется в математике, прежде всего, дифференциальной топологии и геометрии, и в прикладных областях, таких так топологический анализ данных. Искусственные нейронные сети являются в настоящее время общеупотребительным универсальным аппроксиматором. Полученный в данной работе критерий того, что сеть является функцией Морса, дает теоретическое основание для примерения дифференциально-топологических методов в широком классе прикладных задач. Дальнейшие вопросы теоретического свойства, по-видимому, могут быть связаны с получением различных явных соотношений между дифференциальными характеристиками сети как функции, топологической (конкретно, клеточной) структурой поверхностей уровня и топологической структурой исследуемого объекта.

ПРИЛОЖЕНИЕ

Доказательство леммы 1

Точка $\hat{x}$ является невырожденной критической точкой для f_w если и только если: 1) $df_w(\hat{x}) = 0$, и 2) $ddf_w(\hat{x})$ – сюръекция (и тогда биекция); здесь $ddf_w(\hat{x}) : \mathbb{R}^n \to \mathbb{R}^n$ – вторая поизводная от df_w по x, взятая в точке $\hat{x}$.

Из предположения относительно функции F следует, что $0 \in \mathbb{R}^n$ является ее регулярным значением. Пусть $Z = F^{-1}(0,0)$, тогда Z – подмногообразие в $U \times W$ размерности p (см., например, [8, теорема 15.3]).

Пусть $\pi : Z \to W$, $\pi(x,w) = w$ – ограничение на Z естественной проекции. Предположим, что некоторое w является регулярным значением π. Возможно одно из двух:

– либо w не принадлежит образу π, это означает, что f_w не имеет критических точек и потому является функцией Морса;

– либо принадлежит. Тогда пусть $\hat{x}$ – одна из таких точек, что $df_w(\hat{x}) = 0$. Касательное пространство к Z в точке $(\hat{x}, w)$ совпадает с ядром оператора производной от F по совокупности переменных x, w в этой точке. Сужение на это подпространство производной отображения π является сюръекцией (w – регулярное значение π). Одновременно производная F по совокупности x, w в этой (и в любой) точке также является сюръекцией. Несложный аргумент из линейной алгебры показывает, что в таком случае оператор производной от F по x (т.е. второй производной от f) в этой точке также должен быть сюръекцией. Следовательно, $0 \in \mathbb{R}^n$ является регулярным значением df_w, т.е. f_w для такого w является функцией Морса. Выше в качестве w было взято произвольное регулярное значение отображения π. По теореме Сарда, дополнение к множеству регулярных значений имеет меру ноль.

Доказательство теоремы. Общий случай

По правилу дифференцирования сложной функции, производная $\partial f(x,w)/\partial x$ представляется (ср. с (1)) в виде произведения нескольких (по числу слоев) матричных сомножителей вида $\Psi(s_{(t)})W_{(t)}$, где t – номер слоя, $W_{(t)}$ – матрица весов этого слоя (без порогов), $\Psi(s_{(t)}) = \operatorname{diag}(\varphi'(s^j_{(t)}))$ – диагональная матрица с положительными элементами, зависящими, через свои непосредственные аргументы $s^j_{(t)}$, от аргумента x и весов текущего и предшествующих, но не последующих, слоев сети:

$$\frac{\partial f(x,w)}{\partial x} = [\Psi(s_{(L)})W_{(L)}]\ldots[\Psi(s_{(1)})W_{(1)}]. \tag{4}$$

При этом крайний левый сомножитель, соответствующий выходному нейрону сети, имеет формат строки.

Множество $\mathbb{W}$ таких наборов весов w, что все матрицы $W_{(k)}$, $k = 1,2,\ldots,L$, имеют полный ранг, является в пространстве весов дополнением к объединению конечного числа многообразий меньшей размерности (см. [8, теорема 17.3]), т.е. множеству меры ноль. При этом, как и сами матрицы $W_{(k)}$, $\mathbb{W}$ не зависит от x. Далее считаем, что рассмотрение проводится поочередно для каждой из компонент $\mathbb{W}$.

Обозначим $A_{(k)}(x,w) = \Psi(s_{(k)})W_{(k)}$, выделим в представлении (4) один из сомножителей $A_{(k)}$ и, имея в виду цель – эпиморфность производной по весам, рассмотрим последствия малых вариаций весов данного (k-го) слоя:

$$\frac{\partial^2 f(x,w)}{\partial w^j_{k0}\partial x} = \left[\frac{\partial}{\partial s^j_k}u(x,w)\right]W_{(k)}\ldots,$$

где

$$u(x,w) = A_{(L)}(x,w)\ldots A_{(k+1)}(x,w)\Psi(s_{(k)})$$

и многоточием обозначены предшествующие (по ходу сигнала в сети) члены, которые не зависят от весов текущего слоя. Аналогично, для производных по весам из матрицы $W_{(k)}$:

$$\frac{\partial^2 f(x,w)}{\partial w^j_{ki}\partial x} = \left\{\left[\frac{\partial}{\partial s^j_k}u(x,w)\right]W_{(k)} + u(x,w)E^j_i\right\}\ldots,$$

где E^j_i – матрица, в которой элемент (i,j) равен единице, а остальные нулю. Отсюда следует, что, используя различные вариации весов k-го слоя, можно получать возмущения градиента сети вида

$$A_{(L)}(x,w)\ldots A_{(k+1)}(x,w)ZA_{(k-1)}(x,w)\ldots A_{(1)}(x,w),$$

где Z – произвольная матрица соответствующего размера (здесь учтено, что $\Psi(s_{(k)})$ всегда обратима).

Замечание 3. Доказательство частного случая (см. разд. 3), где сомножителей всего два, на этом месте заканчивается применением элементарных соображений с рангами матриц.

Далее потребуется факт из линейной алгебры.

Лемма 2. *Матричное уравнение $ADX + YDB = C$, где D – положительная диагональная матрица и все размеры матриц считаются согласованными, разрешимо относительно X, Y для тех и только тех C, которые, рассматриваемые как линейные отображения, отображают ядро B в образ A.*

Доказательство. Уравнение $AX - YB = C$ разрешимо, если и только если $(E - AA^+)C(E - B^+B) = 0$, где A^+ – обобщенная обратная для матрицы A по Муру–Пенроузу (см. [15]). При этом (см., например, [16, теорема 8.6.1.1.]) AA^+ – это ортогональный проектор на образ A, а B^+B – на ортогональное дополнение к ядру B. Остается заметить, что присутствие матрицы D не меняет ни ядра B, ни образа A. Лемма доказана.

Полагая последовательно $k = L, L-1, \ldots, 1$, рассмотрим на предмет эпиморфности производной по w произведения $A_{(L)}(x,w)\ldots A_{(k)}(x,w)$. При $k = L$ эпиморфности, очевидно, имеет место. При домножении на каждую очередную матрицу $A_{(k-1)}$ могут представиться следующие случаи.

1. $\prod_{j=k,\ldots L} A_{(j)} \neq 0$ и горизонтальный размер матрицы $A_{(k-1)}$ не меньше вертикального. Тогда из элементарных соотношений для рангов $\prod_{j=k-1,\ldots L} A_{(j)} \neq 0$ и из леммы 2, возмущениями весов слоев с $k-1$ по L можно получить произвольное возмущение результата, т.е. эпиморфность сохраняется.

2. $\prod_{j=k,\ldots L} A_{(j)} \neq 0$ и горизонтальный размер матрицы $A_{(k-1)}$ меньше вертикального. Тогда из леммы 2 эпиморфность сохраняется, но может оказаться, что $\prod_{j=k-1,\ldots L} A_{(j)} = 0$.

3. $\prod_{j=k,\ldots L} A_{(j)} = 0$ и горизонтальный размер матрицы $A_{(k-1)}$ не больше вертикального. Тогда эпиморфность сохраняется и $\prod_{j=k-1,\ldots L} A_{(j)} = 0$.

4. $\prod_{j=k,\ldots L} A_{(j)} = 0$ и горизонтальный размер матрицы $A_{(k-1)}$ больше вертикального. Тогда $\prod_{j=k-1,\ldots L} A_{(j)} = 0$ и образ производной совпадает с линейной оболочкой строк матрицы $A_{(k-1)}$, т.е. эпиморфность нарушается.

В итоге, для того чтобы эпиморфность производной нарушилась, необходимо и достаточно, чтобы сначала встретился случай типа 2 и затем типа 4.

СПИСОК ЛИТЕРАТУРЫ

1. *Cybenko G.V.* Approximation by Superpositions of a Sigmoidal function // Math. Control Signals Systems. 1989. V. 2. № 4. P. 303–314.
2. *Pinkus A.* Approximation theory of the MLP model in neural networks // Acta Numerica. 1999. V. 8. P. 143–195.
3. *Dagnelie G.* (ed.) Visual Prosthetics. Springer, 2011.
4. *Szeliski R.* Computer Vision. Springer, 2011.
5. *Курочкин С.В.* Распознавание гомотопического типа объекта с помощью дифференциально-топологических инвариантов аппроксимирующего отображения // Компьютерная оптика. 2019. Т 43. 4. (в печати)
6. *Хирш М.* Дифференциальная топология М.: Мир, 1979.
7. *Постников М.М.* Введение в теорию Морса. М.: Наука, 1971.
8. *Прасолов В.В.* Элементы комбинаторной и дифференциальной топологии. М.: МЦНМО, 2014.
9. *Le C.* A note on optimization with Morse polynomials // Commun. Korean Math. Soc. 2018. V. 33. № 2. P. 671–676.
10. *Banyaga A., Hurtubise D.* Lectures on Morse Homology. Kluwer Texts Math. Sci. V. 29, Kluwer Acad. Publ. Dordrecht, 2004.
11. *Гудфеллоу Я., Бенджио И., Курвилль А.* Глубокое обучение. М.: ДМК Пресс, 2017.
12. *Bishop C.* Exact Calculation of the Hessian Matrix for the Multi-layer Perceptron // Neur. Computat. 1992. V. 4. № 4. P. 494–501.
13. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning. N.Y.: Springer, 2009. ISBN 978-0-387-84857-0.
14. *Nicolaescu L.* An Invitation to Morse Theory. Springer, 2011. ISBN 978-1-4614-1105-5
15. *Baksalary J.K., Kala R.* The matrix equation $AX - YB = C$ // Linear Algebra and its Applicat. 1979. V. 30. P. 41–43.
16. *Прасолов В.В.* Задачи и теоремы линейной алгебры. М.: МЦНМО, 2015.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1179–1191

ИНФОРМАТИКА

УДК 519.72

МЕТРИЧЕСКИЙ ПОДХОД НАХОЖДЕНИЯ ПРИБЛИЖЕННЫХ РЕШЕНИЙ ЗАДАЧ ТЕОРИИ РАСПИСАНИЙ[1)]

© 2021 г. А. А. Лазарев[1,*], Д. В. Лемтюжникова[1], Н. А. Правдивец[1]

[1] *117997 Москва, ул. Профсоюзная, 65, ИПУ РАН, Россия*

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

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Вводятся функции метрики для разных классов задач теории расписаний для одного прибора. Показано, как с помощью введенных функций находятся приближенные решения NP-трудных задач. Величина метрики находится в результате решения задачи линейного программирования, ограничениями которой являются системы линейных неравенств полиномиальных или псевдополиномиальных разрешимых случаев исследуемых задач. Фактически находится проекция во введенной метрике решаемого примера на разрешимые подслучаи задачи. Библ. 23. Фиг. 1. Табл. 3.

Ключевые слова: теория расписаний, метрика, аппроксимация, методы оптимизации.

DOI: 10.31857/S0044466921070127

1. ВВЕДЕНИЕ

Подавляющее большинство задач теории расписаний являются NP-трудными, поэтому поиск приближенных полиномиальных алгоритмов является актуальной задачей.

Существуют два типа методов решения таких задач: точные и приближенные (см. [1]). К первой группе относятся целочисленное линейное программирование (см. [2]), динамическое программирование (см. [3]), метод ветвей и границ (см. [4]), локальный элиминационный алгоритм (см. [5]) и т.д. В этом случае оптимальное значение целевой функции вычисляется без каких-либо ошибок, но вычисления требуют больших затрат времени и памяти. Приближенные методы, такие как генетические алгоритмы (см. [6]), алгоритмы муравьиной колонии (см. [7]), алгоритмы пчелиного роя (см. [8]), табу-поиск (см. [9]) и многие другие, гораздо быстрее получают приближенное решение, но, как правило, не имеют оценок отклонения значения целевой функции от оптимального (см. [10]).

В настоящей работе мы описываем общий приближенный подход, который называется *метрический* (см. [11]). Данный подход позволяет находить приближенное решение с гарантированной погрешностью целевой функции, не превышающей значения функции метрики.

Идея метрического подхода заключается в следующем. Пример задачи A характеризуется точкой, координаты которой соответствуют входным данным задачи. Для этой задачи рассмотрим все известные полиномиальные и псевдополиномиальные алгоритмы. Затем вводим некоторую метрику, значение которой фактически ограничивает сверху оценку абсолютной погрешности целевой функции. С помощью этой метрики находим полиномиально или псевдополиномиально разрешимый пример задачи B с наименьшим расстоянием от данного примера A во введенной метрике. Для этого мы формулируем и решаем задачу линейного программирования. Другими словами, строим проекцию начальной точки A на подпространство задач.

На данный момент для этого подхода получены следующие результаты. В [12] проводится сравнение допустимых областей и дается понятие полиномиальной меры неразрешимости для задачи теории расписаний с одним прибором $1|r_j|L_{\max}$. Возможно, его можно уменьшить еще больше, найдя новые подпространства полиномиально разрешимых примеров. В [13], [14] получены соответствующие метрики для задач теории расписаний с функцией суммарного запаздывания.

[1)]Работа выполнена при частичной финансовой поддержке РФФИ (коды проектов № 18-31-00458, № 20-58-S52006).

2. ПОСТАНОВКА ЗАДАЧИ СУММАРНОГО ЗАПАЗДЫВАНИЯ ДЛЯ ОДНОГО ПРИБОРА

Необходимо обслужить n требований на одном приборе. Прерывания обслуживания требований, искусственные простои прибора при обслуживании и обслуживание более одного требования в любой момент времени запрещены. Требования пронумерованы числами $1, 2, \ldots, n$. Множество $N = \{1, 2, \ldots, n\}$ назовем *множеством требований*.

Для требования $j \in N$ заданы следующие параметры: продолжительность обслуживания $p_j > 0$ и директивный срок окончания обслуживания d_j. Задан момент начала обслуживания t_0, с которого прибор готов начать обслуживание требований. Все требования поступают на обслуживание одновременно в момент времени t_0.

Расписание обслуживания требований множества N задается кусочно-постоянной непрерывной слева функцией $s : \mathbb{R} \to \{0, 1, \ldots, n\}$. Если $s(t) = 0$, то в момент времени t прибор простаивает; если $s(t) = j$, $j \in N$, то в момент времени t прибор обслуживает требование j. Поскольку в рассматриваемой задаче требования поступают на обслуживание одновременно, обслуживаются на приборе без прерываний и искусственных простоев прибора, то расписание однозначно задается перестановкой π элементов множества N. Далее в работе понятие расписания и перестановки множества требований будем отождествлять и называть *расписанием* перестановку π.

Индивидуальный пример исследуемой задачи (далее *пример*) с заданными множеством требований N, продолжительностями обслуживания p_j, директивными сроками d_j и моментом начала обслуживания t_0 будем обозначать через $I = \langle \{p_j, d_j\}_{j \in N}, t_0 \rangle$. В случае, если параметры требований фиксированы (однозначно определены), для обозначения примера I будем использовать запись $\{N, t_0\}$. Множество всех $n!$ расписаний для примера I будем обозначать через $\Pi(I)$.

Через $C_j(\pi)$ будем обозначать момент окончания обслуживания требования j при расписании π. Моменты окончания обслуживания требований при расписании $\pi = (j_1, \ldots, j_n)$ вычисляются следующим образом:

$$C_{j_1}(\pi) = t_0 + p_{j_1},$$

$$C_{j_k}(\pi) = C_{j_{k-1}}(\pi) + p_{j_k}, \quad k = 2, 3, \ldots, n.$$

Если обслуживание требования i предшествует обслуживанию требования j при расписании π (т.е. выполняется $C_i(\pi) < C_j(\pi)$), то будем использовать запись $(i \to j)_\pi$. Используя это обозначение, момент окончания обслуживания требования $j \in N$ при расписании π можно записать как $C_j(\pi) = t_0 + \sum_{i \in N : (i \to j)_\pi} p_i + p_j$. Обслуживание всех требований примера I завершается в момент времени $t_0 + \sum_{j \in N} p_j$ при любом расписании $\pi \in \Pi(I)$.

Под *запаздыванием* требования $j \in N$ при расписании π будем понимать величину

$$T_j(\pi) = \max\{0, C_j(\pi) - d_j\}.$$

Суммарное запаздывание требований при расписании π определяется как

$$F(\pi) = \sum_{j=1}^{n} T_j(\pi).$$

Задача минимизации суммарного запаздывания заключается в построении оптимального расписания $\pi^* \in \Pi(I)$, при котором выполняется неравенство $F(\pi^*) \le F(\pi)$ для всех расписаний $\pi \in \Pi(I)$. Отметим, что данная задача является NP-трудной (см. [15]). Известен псевдополиномиальный алгоритм ее решения (см. [16]) трудоемкости $O\left(n^4 \sum_{j \in N} p_j\right)$ операций, в случае когда $p_j \in \mathbb{Z}^+ \ \forall j \in N$, основанный на методе динамического программирования.

Через $\Pi^*(I)$ будем обозначать множество всех оптимальных расписаний для примера I. Для обозначения величины оптимального суммарного запаздывания примера I будем использовать запись $F^*(I)$.

Покажем, что без ограничения общности рассуждений можно предполагать, что рассматриваются только те примеры I, для которых выполняется

$$d_j \in \left[t_0; t_0 + \sum_{i\in N} p_i\right] \quad \forall j \in N. \tag{2.1}$$

В случае, если для примера I условия (2.1) не выполняются, построим пример $I' = \left\langle \{p_j, d'_j\}_{j\in N}, t_0 \right\rangle$ такой, что

$$d'_j = \min\left\{\max\{t_0, d_j\},\ t_0 + \sum_{i\in N} p_i\right\} \quad \forall j \in N.$$

Если для некоторого требования $j \in N$ выполняется $d_j > t_0 + \sum_{i\in N} p_i$, то имеем $d'_j = t_0 + \sum_{i\in N} p_i$. Это означает, что требование j не запаздывает при любом расписании π для обоих примеров I и I'. Таким образом, можно исключить такое требование j из рассмотрения и после построения оптимального расписания для редуцированного примера добавить требование j на последнюю позицию в построенное расписание.

Если для некоторого требования $j \in N$ выполняется $d_j < t_0$, то имеем $d'_j = t_0$, и требование j запаздывает при любом расписании π. Лоулером в [16] была доказана следующая

Теорема 1 (см. [16]). *Пусть π^* – оптимальное расписание примера $I_1 = \left\langle \{p_j, d_j\}_{j\in N}, t_0 \right\rangle$. Выберем величины d'_j так, что*

$$\min\{d_j, C_j(\pi^*)\} \le d'_j \le \max\{d_j, C_j(\pi^*)\} \quad \forall j \in N. \tag{2.2}$$

Тогда любое оптимальное расписание примера $I_2 = \left\langle \{p_j, d'_j\}_{j\in N}, t_0 \right\rangle$ будет оптимальным и для примера I_1.

В нашем случае условия (2.2) для примеров I и I' выполняются, поскольку при любом расписании π справедливо неравенство $t_0 < C_j(\pi) \le t_0 + \sum_{i\in N} p_i$, $j \in N$. Следовательно, любое оптимальное расписание для примера I' будет оптимальным и для примера I. Таким образом, параметры требований примера I' удовлетворяют условиям (2.1) и выполняется $\Pi^*(I') \subseteq \Pi^*(I)$.

Необходимо заметить, что не любое оптимальное расписание примера I будет оптимальным и для примера I'.

Как следствие теоремы 1, может быть сформулирована

Лемма 1. *Для любого оптимального расписания π и любого незапаздывающего требования j верно: все требования, которые обслуживаются в интервале $[C_j(\pi), d_j]$ не запаздывают.*

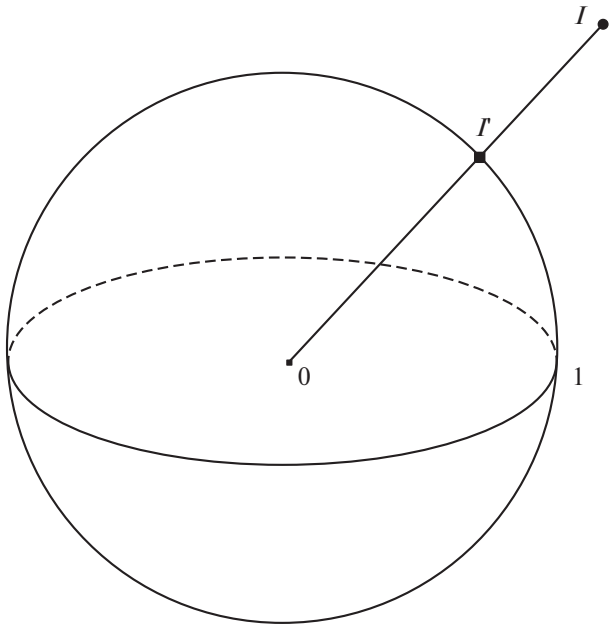
Данная лемма легко доказывается от противного.

Два примера $I = \left\langle \{p_j, d_j\}_{j\in N}, t_0 \right\rangle$ и $I' = \left\langle \{p'_j, d'_j\}_{j\in N}, t'_0 \right\rangle$ будем называть *равными*, если множества оптимальных расписаний для обоих примеров совпадают, т.е. $\Pi^*(I) = \Pi^*(I')$. Через $T'_j(\pi)$ будем обозначать значение запаздывания требования j при некотором расписании π, вычисленное для примера I'. В рамках данного исследования будут полезны следующие случаи равенства примеров.

(1) Примеры I и I' равны, если $p'_j = p_j$, $d'_j = d_j + C$, $j \in N$, и $t'_0 = t_0 + C$, где C – произвольная константа. Действительно, в этом случае для любого расписания π и любого требования $j \in N$, обслуживаемого при расписании π, выполняется

$$T'_j(\pi) = \max\left\{0, t_0 + C + \sum_{i:\,(i\to j)_\pi} p_i + p_j - (d_j + C)\right\} =$$

$$= \max\left\{0, t_0 + \sum_{i:\,(i\to j)_\pi} p_i + p_j - d_j\right\} = \max\{0, C_j(\pi) - d_j\} = T_j(\pi).$$



Фиг. 1. Схематичное изображение проецирования примера I на единичную сферу в $(2n+1)$-мерном евклидовом пространстве. Примеры I и I' равны.

Следовательно, $F^*(I) = F^*(I')$ и $\Pi^*(I) = \Pi^*(I')$. В этом смысле без потери общности момент начала обслуживания требований можно считать равным нулю: $t_0 = 0$. Если требуется выполнение некоторых условий относительно директивных сроков, например $d_j \in \mathbb{Z}^+$, то это обычно достигается путем изменения момента начала обслуживания t_0. Поэтому в работе мы будем предполагать, что t_0 – произвольная величина.

(2) Примеры I и I' равны, если $p'_j = \alpha p_j$, $d'_j = \alpha d_j$, $j \in N$, и $t'_0 = \alpha t_0$, где $\alpha > 0$ – произвольная положительная константа. Действительно, в этом случае для любого расписания π и любого требования $j \in N$, обслуживаемого при расписании π, выполняется

$$T'_j(\pi) = \max\left\{0, \alpha t_0 + \sum_{i:\,(i\to j)_\pi} \alpha p_i + \alpha p_j - \alpha d_j\right\} = \alpha \max\left\{0, t_0 + \sum_{i:\,(i\to j)_\pi} p_i + p_j - d_j\right\} = \alpha T_j(\pi).$$

Следовательно, $F^*(I) = \alpha F^*(I')$ и $\Pi^*(I) = \Pi^*(I')$. Рассмотрим любой пример I как точку в $(2n+1)$-мерном евклидовом пространстве с координатами $(t_0, p_1, \ldots, p_n, d_1, \ldots, d_n)$. Тогда все примеры (точки), расположенные на луче, исходящем из нуля пространства, равны между собой (фиг. 1). Поэтому можно предполагать, что рассматриваются только те примеры (точки), которые лежат на поверхности единичной сферы в рассматриваемом пространстве. Однако в общем случае данный подход неприемлем при построении оптимального расписания некоторым алгоритмом, для работы которого необходимо, чтобы продолжительности обслуживания требований были целочисленными величинами. Поэтому ограничимся выбором константы $\alpha = 1/\text{НОД}(p_1, \ldots, p_n)$, где НОД – наибольший общий делитель.

Необходимо также отметить, что для любых двух равных примеров I и I' выполняется свойство: если требование при оптимальном расписании π примера I запаздывает (не запаздывает), то и в примере I' при расписании π требование запаздывает (не запаздывает). Доказательство этого можно провести от противного.

Введем некоторые дополнительные обозначения и понятия.

Расписание π будем называть SPT-расписанием (Shortest Processing Time) и обозначать через π_{spt}, если требования упорядочены при данном расписании в порядке неубывания продолжительностей обслуживания, т.е. для любых двух требований i и j таких, что $p_i < p_j$, выполняется $(i \to j)_{\pi_{spt}}$.

Расписание π будем называть EDD-расписанием (Earliest Due Date) и обозначать через π_{edd}, если требования упорядочены при данном расписании в порядке неубывания директивных сроков, т.е. для любых двух требований i и j таких, что $d_i < d_j$, выполняется $(i \to j)_{\pi_{edd}}$.

Через $\{\pi\}$ будем обозначать множество требований, упорядоченных при расписании π. В случае $\{\pi\} \neq N$ ($\{\pi\} \subset N$) расписание π будем называть *частичным* расписанием.

Запись $(i \to j)_\pi$ при необходимости будет расширена до $(i \to j \to k)_\pi$ для обозначения порядка обслуживания более, чем двух требований при некотором расписании π. Также будем использовать записи $(j \to N')_\pi$ или $(N' \to N'')_\pi$ для обозначения порядка обслуживания между некоторыми подмножествами требований.

Для обозначения структуры расписания π будем использовать запись $\pi = (\pi_1, \ldots, \pi_m)$, где расписание π_i, $i = 1, 2, \ldots, m$, является частичным расписанием (подрасписанием расписания π). При этом для частичных расписаний π_i выполняется $\{\pi\} = \bigcup_{i=1}^{m} \{\pi_i\}$, $\{\pi_i\} \cap \{\pi_j\} = \emptyset$ для $i \neq j$, $(\{\pi_1\} \to \ldots \to \{\pi_m\})_\pi$, и требования каждого частичного расписания π_i, $i = 1, 2, \ldots, m$, упорядочены в том же порядке, что и при расписании π. Для обозначения структуры расписания π относительно некоторого требования $j \in \{\pi\}$ будем использовать запись $\pi = (\pi_1, j, \pi_2)$.

Пользуясь нотацией, предложенной в [17], для обозначения исследуемой задачи будем использовать запись $1\|\sum T_j$. Рассматриваемая задача $1\|\sum T_j$ является NP-трудной в обычном смысле (см. [15]).

В заключение данного раздела приведем пример связи рассматриваемой задачи для одного прибора $1\|\sum T_j$ с некоторой задачей оптимального упорядочения набора требований, возникающей на практике. Рассмотрим некоторый обслуживающий центр (ОЦ), процесс работы которого разбит на циклы, состоящие из двух этапов. На первом этапе центр набирает заявки и на втором этапе осуществляет их исполнение. В ОЦ имеется m однородных исполнителей, каждая заявка может быть исполнена любым из них. Исполнители обладают в общем случае различной производительностью и выполняют в любой момент времени не более одной заявки. Выполнение заявки не может быть прервано, т.е., если исполнитель начинает выполнение некоторой заявки, то он выполняет ее до конца. При приеме заявки ОЦ и клиент договариваются о сумме, которую ОЦ получит за выполнение заявки, а также о моменте времени, к которому заявка должна быть исполнена. При выполнении некоторой заявки позднее обговоренного момента времени ОЦ выплачивает клиенту фиксированный для любой заявки штраф за каждую единицу времени, прошедшую после этого момента. Целью составления расписания работы ОЦ является минимизация суммы штрафных выплат по всем принятым к исполнению заявкам.

Данная задача может быть сформулирована в терминах теории расписаний следующим образом. Задано множество требований $N = \{1, 2, \ldots, n\}$ и множество параллельных приборов $M = \{1, 2, \ldots, m\}$. Каждое требование $j \in N$ имеет директивный срок d_j, к которому желательно завершить обслуживание данного требования, и продолжительность p_{ji} обслуживания требования j на приборе $i \in M$. Требования обслуживаются на приборах без прерываний, в любой момент времени любой прибор обслуживает не более одного требования. Расписание обслуживания требований π строится с момента времени $t_0 = 0$ и для данной задачи может быть задано в виде набора из m перестановок $\pi_1, \ldots, \pi_m$, где π_i, $i \in M$, задает порядок обслуживания требований множества $\{\pi_i\}$ на приборе i. При этом $\{\pi_1\}, \ldots, \{\pi_m\}$ задает разбиение множества требований N (т.е. $\bigcup_{i \in M} \{\pi_i\} = N$ и $\{\pi_i\} \cap \{\pi_j\} = \emptyset$ для $i, j \in M$, $i \neq j$). Момент окончания $C_j(\pi)$ требования $j \in N$ при расписании π вычисляется описанным выше образом. Необходимо построить такое расписание обслуживания требований множества N на приборах из множества M, при котором минимизируется суммарное запаздывание требований, т.е. $\sum_{j \in N} \max\{0, C_j(\pi) - d_j\} \to \min$.

Имея некоторый алгоритм, основанный, например, на идее метода ветвей и границ, для задачи суммарного запаздывания для параллельных приборов было бы желательно иметь некоторый способ вычисления нижних оценок оптимального значения целевой функции. Это может быть сделано с помощью использования методов решения и построения оценок для примеров задачи $1\|\sum T_j$. Однако алгоритм решения задачи суммарного запаздывания для параллельных приборов может быть организован так, что сначала строится разбиение множества требований N на m подмножеств, которые будут обслуживаться на соответствующих приборах, а затем решается m независимых примеров задачи $1\|\sum T_j$. В этом случае исходная задача для многих приборов допускает разбиение на несколько независимых подзадач для одного прибора.

3. МЕТРИКА ДЛЯ ЗАДАЧИ $1|r_j|\sum T_j$

3.1. Постановка задачи

Имеется множество N, состоящее из n требований, которые необходимо обслужить на одном приборе. Прибор готов начать обслуживание в момент времени $t_0 = 0$ и не может обслуживать более одного требования одновременно. Прерывания при обслуживании запрещены. Для каждого требования $j \in N$ заданы: момент поступления r_j, продолжительность обслуживания p_j и директивный срок d_j. Расписание $\pi = \{j_1,\ldots,j_n\}$ определяет порядок, в котором обслуживаются требования. Естественно рассматривать ранние расписания, при которых

$$C_{j_1}(\pi) = \max\{t_0, r_{j_1}\} + p_{j_1},$$
$$C_{j_k}(\pi) = \max\{r_{j_k}, C_{j_{k-1}}(\pi)\} + p_{j_k}, \quad k = 2,3,\ldots,n,$$

где $C_j(\pi)$ – момент окончания обслуживания требования j при расписании π. Необходимо построить оптимальное расписание π^*, минимизирующее целевую функцию – суммарное запаздывание $\sum_{j\in N} T_j(\pi)$, где $T_j(\pi) = \max\{0, C_j(\pi) - d_j\}$ – запаздывание требования j при расписании π. В дальнейшем, если из контекста ясно, о каком расписании идет речь, то зависимость от π может опускаться. Данная задача является NP–трудной (см. [15]) и обозначается $1|r_j|\sum T_j$ (см. [17]).

Задача $1|r_j|\sum T_j$ полностью характеризуется $3n$ параметрами – моментами поступления, продолжительностями обслуживания и директивными сроками каждого из n требований. Будем говорить, что задан пример A задачи, если задано $3n$ параметров $\{r_j^A, p_j^A, d_j^A, j = 1,2,\ldots,n\}$, характеризующих задачу.

Для частного случая $r_j = 0$, $j \in N$, задачи минимизации суммарного запаздывания ранее был предложен полиномиальный алгоритм нахождения приближенного решения с относительной погрешностью $\varepsilon > 0$ сложности $O\left(\frac{n^7}{\varepsilon}\right)$ операций (см. [18]). Позднее М.Я. Ковалев усилил оценку до $O(n^6 \log n + n^6/\varepsilon)$ операций (см. [19]). Также для этого случая известен псевдополиномиальный алгоритм сложности $O\left(n^4\sum p_j\right)$ (см. [16]). В случае, если

$$p_1 \geq \ldots \geq p_n,$$
$$d_1 \leq \ldots \leq d_n,$$

сложность псевдополиномиального алгоритма может быть уменьшена до $O\left(n^2\sum p_j\right)$ операций (см. [20]). Для случая $1|r_j, p_j = p|\sum T_j$ известен полиномиальный алгоритм сложности $O(n^7)$ операций, предложенный Ф. Баптисте (см. [21]), на основе метода динамического программирования.

Предлагается подход приближенного решения задачи $1|r_j|\sum T_j$ с гарантированной погрешностью, основанный на введении метрики для пространства параметров задачи, и рассматриваются возможности применения данного подхода к другим задачам теории расписаний. Затем описываются численные эксперименты, проведенные для проверки предложенного метода.

3.2. Метрика для пространства параметров

Задача $1|r_j|\sum T_j$ полностью характеризуется $3n$ параметрами, что позволяет нам рассматривать примеры задачи, как точки в $3n$-мерном пространстве параметров $\Omega = \{r_1,\ldots,r_n,p_1,\ldots,p_n,d_1,\ldots,d_n\}$.

Лемма 2. *Пусть примеры A и B имеют одинаковые продолжительности обслуживания и директивные сроки*:

$$p_j^A = p_j^B, \quad d_j^A = d_j^B, \quad j \in N.$$

Тогда для любого расписания π

$$\left|\sum_{j\in N} T_j^A(\pi) - \sum_{j\in N} T_j^B(\pi)\right| \le n\max_{j\in N}|r_j^A - r_j^B|.$$

Доказательство. Используя определение запаздывания и известное неравенство

$$|\max\{a,b\} - \max\{c,d\}| \le \max\{|a-c|,|b-d|\} \quad \forall a,b,c,d \in \mathbb{R},$$

получаем

$$\left|\sum_{j\in N} T_j^A - \sum_{j\in N} T_j^B\right| \le \sum_{j\in N}|C_j^A - C_j^B + d_j^B - d_j^A| \le \sum_{j\in N}|C_j^A - C_j^B| + \sum_{j\in N}|d_j^A - d_j^B|. \tag{3.3}$$

Или, учитывая равенство директивных сроков,

$$\left|\sum_{j\in N} T_j^A(\pi) - \sum_{j\in N} T_j^B(\pi)\right| \le \sum_{j\in N}|C_j^A - C_j^B|. \tag{3.4}$$

Учитывая свойства раннего расписания, заметим, что

$$|C_{j_1}^A - C_{j_1}^B| = |r_{j_1}^A - r_{j_1}^B| \le \max_{j\in N}|r_j^A - r_j^B|,$$

$$|C_{j_k}^A - C_{j_k}^B| \le \max\{|r_{j_k}^A - r_{j_k}^B|, |C_{j_{k-1}}^A - C_{j_{k-1}}^B|\} \le \max_{j\in N}|r_j^A - r_j^B|,$$

$$k = 2,3,\ldots,n.$$

Отсюда и из неравенства (3.4) получаем утверждение леммы.

Лемма 3. *Пусть примеры A и B имеют одинаковые моменты поступления и директивные сроки*:

$$r_j^A = r_j^B, \quad d_j^A = d_j^B, \quad j \in N.$$

Тогда для любого расписания π

$$\left|\sum_{j\in N} T_j^A(\pi) - \sum_{j\in N} T_j^B(\pi)\right| \le n\sum_{j\in N}|p_j^A - p_j^B|.$$

Доказательство. При условиях леммы также выполняется неравенство (3.4):

$$\left|\sum_{j\in N} T_j^A - \sum_{j\in N} T_j^B\right| \le \sum_{j\in N}|C_j^A - C_j^B|.$$

Учитывая свойства раннего расписания и равенство моментов поступления, получаем

$$|C_{j_1}^A - C_{j_1}^B| = |p_{j_1}^A - p_{j_1}^B| \le \sum_{j\in N}|p_j^A - p_j^B|,$$

$$|C_{j_k}^A - C_{j_k}^B| \le |p_{j_k}^A - p_{j_k}^B| + |C_{j_{k-1}}^A - C_{j_{k-1}}^B| \le \sum_{j\in N}|p_j^A - p_j^B|, \quad k = 2,3,\ldots,n.$$

Отсюда и из неравенства (3.4) получаем утверждение леммы.

Лемма 4. *Пусть примеры A и B имеют одинаковые моменты поступления и продолжительности обслуживания*:

$$r_j^A = r_j^B, \quad p_j^A = p_j^B, \quad j \in N.$$

Тогда для любого расписания π

$$\left|\sum_{j\in N} T_j^A(\pi) - \sum_{j\in N} T_j^B(\pi)\right| \le \sum_{j\in N} |d_j^A - d_j^B|.$$

Доказательство. При условиях леммы $C_{j_k}^A = C_{j_k}^B$, $k \in N$, и неравенство (3.3) имеет вид

$$\left|\sum_{j\in N} T_j^A - \sum_{j\in N} T_j^B\right| \le \sum_{j\in N} |d_j^A - d_j^B|,$$

т.е. утверждение леммы выполняется.

Теорема 2. *Функция, определенная на пространстве примеров* $\Omega\times\Omega$,

$$\rho(A,B) = n\max_{j\in N}|r_j^A - R_j^B| + n\sum_{j\in N}|p_j^A - p_j^B| + \sum_{j\in N}|d_j^A - d_j^B|$$

удовлетворяет аксиомам метрики.

Доказательство. Очевидно, что функция $\rho(A,B)$ симметрична и неотрицательна, причем $\rho(A,B) = 0$ тогда и только тогда, когда $A = B$. Неравенство треугольника выполняется в силу свойств модуля суммы.

Лемма 5. *Для любых примеров* A *и* B *и любого расписания* π *справедливо неравенство*

$$\left|\sum_{j\in N} T_j^A(\pi) - \sum_{j\in N} T_j^B(\pi)\right| \le \rho(A,B). \tag{3.5}$$

Доказательство. Пусть пример C имеет такие же моменты поступления и продолжительности обслуживания, как и пример A, а директивные сроки, как у примера B. Пусть далее пример D имеет моменты поступления, как у примера A, а директивные сроки и продолжительности обслуживания, как у примера B, тогда, используя леммы 2–4, получаем

$$\left|\sum_{j\in N} T_j^A - \sum_{j\in N} T_j^B\right| \le \left|\sum_{j\in N} T_j^B - \sum_{j\in N} T_j^D\right| + \left|\sum_{j\in N} T_j^D - \sum_{j\in N} T_j^C\right| + \left|\sum_{j\in N} T_j^A - \sum_{j\in N} T_j^C\right| \le$$
$$\le n\max_{j\in N}\left|r_j^A - r_j^B\right| + n\sum_{j\in N}\left|p_j^A - p_j^B\right| + \sum_{j\in N}\left|d_j^A - d_j^B\right| = \rho(A,B).$$

3.3. Метод изменения параметров

Теорема 3. *Пусть* π^A *и* π^B *– оптимальные расписания для примеров* A *и* B*, тогда*

$$\sum_{j\in N} T_j^A(\pi^B) - \sum_{j\in N} T_j^A(\pi^A) \le 2\rho(A,B). \tag{3.6}$$

Доказательство. Используя лемму 4, получаем

$$\sum_{j\in N} T_j^A(\pi^B) - \sum_{j\in N} T_j^A(\pi^A) = \left[\sum_{j\in N} T_j^A(\pi^B) - \sum_{j\in N} T_j^B(\pi^B)\right] + \left[\sum_{j\in N} T_j^B(\pi^B) - \sum_{j\in N} T_j^B(\pi^A)\right] +$$
$$+ \left[\sum_{j\in N} T_j^B(\pi^A) - \sum_{j\in N} T_j^A(\pi^A)\right] \le 2\rho(A,B).$$

Доказанная теорема позволяет использовать новый метод решения задачи $1|r_j|\sum T_j$, названный методом изменения параметров. Метод состоит в том, чтобы использовать оптимальное расписание некоторого полиномиально или псевдополиномиально решаемого примера B в качестве расписания для исходного примера A. Теорема 3 позволяет оценить сверху абсолютную погрешность значения целевой функции такого решения с помощью функции $\rho(A,B)$. Естественно конструировать пример B так, чтобы минимизировать функцию $\rho(A,B)$. Таким образом, задача $1|r_j|\sum T_j$ заменяется задачей на минимизацию функции метрики.

Рассмотрим случай, когда искомый пример B должен принадлежать некоторому полиномиально или псевдополиномиально разрешимому классу примеров, задаваемому системой неравенств

$$\mathcal{A} \cdot R^B + \mathcal{B} \cdot P^B + \mathcal{C} \cdot D^B \le H, \tag{3.7}$$

где $R^B = (r_1^B, \ldots, r_n^B)^{\mathrm{T}}$, $P^B = (p_1^B, \ldots, p_n^B)^{\mathrm{T}}$, $D^B = (d_1^B, \ldots, d_n^B)^{\mathrm{T}}$, $p_j^B \ge 0$, $r_j^B \ge 0$, $j \in N$, $^{\mathrm{T}}$ — символ транспонирования, $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ — матрицы $m \times n$, H — столбец из m элементов.

В этом случае задача минимизации функции метрики может быть поставлена как задача линейного программирования:

$$\min n(y^r - x^r) + n\sum_{j\in N}(y_j^p - x_j^p) + \sum_{j\in N}(y_j^d - x_j^d),$$

при условиях

$$\begin{aligned}
&x^r \le r_j^A - r_j^B \le y^r,\\
&x_j^p \le p_j^A - p_j^B \le y_j^p,\\
&x_j^d \le d_j^A - d_j^B \le y^d,\\
&r_j^B \ge 0, \quad p_j^B \ge 0, \quad j \in N,\\
&\mathcal{A} \cdot R^B + \mathcal{B} \cdot P^B + \mathcal{C} D^B \le H.
\end{aligned}$$

Неизвестными в данной задаче являются $7n + 2$ переменных:

$$r_j^B, p_j^B, d_j^B, x_j^p, y_j^p, x_j^d, y_j^d, x^r, y^r, j \in N.$$

Тем не менее сепарабельность функции $\rho(A, B)$ во многих случаях позволяет находить ее минимум гораздо проще, без использования методов линейного программирования. Необходимо заметить, что для всех известных полиномиальных и псевдополиномиальных разрешимых случаев задачи выполняется система линейных неравенств типа (3.7).

3.4. Применение метода изменения параметров для других задач теории расписаний

Описанный метод не является жестко привязанным к виду целевой функции, что позволяет использовать его для решения других задач теории расписаний. Теорему 3 можно обобщить на случай общего вида целевой функции $F(\pi)$.

Теорема 4. *Пусть $F(\pi)$ — некоторая регулярная целевая функция, а $\rho(A, B)$ — функция метрики для любых A, B, π, удовлетворяющая неравенству*

$$|F^A(\pi) - F^B(\pi)| \le \rho(A, B). \tag{3.8}$$

Пусть далее π^A и π^B — оптимальные расписания для примеров A и B, тогда

$$F^A(\pi^B) - F^A(\pi^A) \le 2\rho(A, B).$$

Доказательство. Доказательство теоремы 4 повторяет доказательство теоремы (3.3) с заменой $\sum_{j\in N} T_j$ на F.

Таким образом, для применения метода изменения параметров достаточно построить функцию $\rho(A, B)$, удовлетворяющую неравенству (3.8). Такие функции были построены ранее для задач $1\|\sum T_j$ [14] и $1|r_j|L_{\max}$ (см. [22]). Здесь мы приведем вариант построения таких функций для общих случаев аддитивной и “минимаксной” целевой функции.

Лемма 6 (см. [13]). *В случае аддитивной целевой функции вида*

$$F(\pi) = \sum_{j\in N} \phi_j(\pi, r_1, \ldots, r_n, p_1, \ldots, p_n, d_j)$$

Таблица 1. Классы примеров, использованные в численных экспериментах

Класс примеров	Метрика между примером B класса и произвольным примером A
$\{\mathcal{PR} : p_j = p, r_j = r, j \in N\}$	$\rho(A,B) = n\sum_{j=1}^{n}\left\|p_j^A - p\right\| + n\max_{j\in N}\left\|r_j^A - r\right\|$
$\{\mathcal{PD} : p_j = p, d_j = d, j \in N\}$	$\rho(A,B) = n\sum_{j\in N}\left\|p_j^A - p\right\| + \sum_{j\in N}\left\|d_j^A - d\right\|$
$\{\mathcal{RD} : r_j = r, d_j = d, j \in N\}$	$\rho(A,B) = n\max_{j\in N}\left\|r_j^A - r\right\| + \sum_{j\in N}\left\|d_j^A - d\right\|$
$\{\mathcal{P} : p_j = p, j \in N\}$	$\rho(A,B) = n\sum_{j\in N}\left\|p_j^A - p\right\|$
$\{\mathcal{R}0 : r_j = 0, j \in N\}$	$\rho(A,B) = n\max_{j\in N}\left\|r_j^A - r\right\|$

функция

$$\rho(A,B) = \sum_{j\in N}\sum_{i\in N}(R_{ji}|r_j^A - r_j^B| + P_{ji}|p_j^A - p_j^B|) + \sum_{j\in N} D_j|d_j^A - d_j^B| \tag{3.9}$$

удовлетворяет неравенству (3.8). *Здесь R_{ji}, P_{ji} — константы Липшица для функции ϕ_i по переменным r_j и p_j, D_j — константа Липшица для функции ϕ_j по переменной $d_j, i, j \in N$.*

Лемма 7 (см. [13]). *В случае "минимаксной" целевой функции вида*

$$F(\pi) = \max_{j\in N}\phi_j(\pi, r_1, \ldots, r_n, p_1, \ldots, p_n, d_j)$$

функция

$$\rho(A,B) = \sum_{j\in N}(R_j|r_j^A - r_j^B| + P_j|p_j^A - p_j^B|) + D\max_{j\in N}|d_j^A - d_j^B| \tag{3.10}$$

удовлетворяет неравенству (3.8). *Здесь R_j, P_j — наибольшие константы Липшица из констант для функций ϕ_i по переменным r_j и p_j, D — наибольшая из констант Липшица для функций ϕ_j по переменным $d_j, i, j \in N$.*

Заметим, что функции (3.9) и (3.10) сепарабельны, что значительно облегчает нахождение их минимумов.

3.5. Численные эксперименты

Для определения эффективности предложенной схемы была проведена серия численных экспериментов. Классы, в которых проводился поиск полиномиально разрешимых примеров, представлены в табл. 1.

В первых трех классах решением является расписание, упорядоченное по неубыванию свободного параметра. Алгоритмы решения задач последних двух классов представлены в [21] и [23] и имеют сложности $O(n^7)$ и $O\left(n^4\sum p_j\right)$ операций соответственно.

Для нахождения в указанных классах полиномиально разрешимого примера B, ближайшего к заданному примеру, необходимо найти минимум функций:

$$f(r) = n\max_{j\in N}|r_j^A - r|, \tag{3.11}$$

$$g(p) = n\sum_{j=1}^{n}|p_j^A - p|, \tag{3.12}$$

$$h(d) = \sum_{j\in N}|d_j^A - d|. \tag{3.13}$$

Лемма 8. 1. *Минимум функции* (3.11) *достигается в точке* $r = (r^A_{\max} + r^A_{\min})/2$, *где* $r^A_{\max} = \max\limits_{j\in N} r^A_j$, $r^A_{\min} = \min\limits_{j\in N} r^A_j$.

2. *Минимум функции* (3.12) *достигается в некоторой точке* $p \in \{p^A_1, \ldots, p^A_n\}$.

3. *Минимум функции* (3.13) *достигается в некоторой точке* $d \in \{d^A_1, \ldots, d^A_n\}$.

Доказательство. Функция $f(r)$ представима в виде

$$n \max_{j\in N} |r^A_j - r| = n \max\{r - r^A_{\min}, r^A_{\max} - r\} = n\left(\frac{r^A_{\max} - r^A_{\min}}{2} + \left|r - \frac{r^A_{\max} + r^A_{\min}}{2}\right|\right)$$

и, очевидно, имеет минимум в точке $\dfrac{r^A_{\max} + r^A_{\min}}{2}$.

Пусть функция $g(p)$ имеет минимум в точке p_0, тогда либо $f'(p_0) = 0$, либо $p_0 \in \{p^A_1, \ldots, p^A_n\}$. Поскольку $g(p)$ – кусочно-линейная функция, обращение ее производной в нуль означает, что функция является константой на некотором интервале $[p^A_k, p^A_{k+1}]$, $k = 1, 2, \ldots, n-1$, а значит, граничные точки p^A_k и p^A_{k+1} также являются точками минимума.

Последнее утверждение леммы о минимуме функции $h(d)$ доказывается аналогично.

Было проведено несколько серий экспериментов. Во всех сериях использовались примеры с параметрами, распределенными равномерно на интервалах [1,100] для p^A_j, $\left[p_j, \sum_{j\in N} p_j\right]$ для d^A_j и $[0, d_j - p_j]$ для r^A_j, $j \in N$. В первой серии экспериментов оценивалась величина различия между правой и левой частями неравенства (3.5). Данная величина позволяет оценить погрешность метода. Для каждого $n = 10, 20, \ldots, 100$ генерировалось 10 000 пар примеров. Использованные в экспериментах расписания генерировались случайным образом. Для каждой пары вычислялась величина $\dfrac{\left|\sum_{j\in N} T^A_j - \sum_{j\in N} T^B_j\right|}{\rho(A, B)}$. Также для определения параметров, имеющих наибольшее влияние на функцию метрики, вычислялись процентные величины вкладов частей метрики, зависящих от продолжительностей обслуживания, директивных сроков и моментов поступления.

Результаты представлены в табл. 2. Среднее значение $\dfrac{\left|\sum_{j\in N} T^A_j - \sum_{j\in N} T^B_j\right|}{\rho(A, B)}$ меняется от 5 до 10% при росте n, а части функции метрики, зависящие от продолжительностей обслуживания, директивных сроков и моментов поступления дают вклады приблизительно 35, 20 и 25% в общую величину функции соответственно.

Вторая серия экспериментов проводилась для проверки метода изменения параметров по следующей схеме. Рассматривались значения $n = 4, 5, \ldots, 10$, для каждого n генерировались по 10000 примеров. К каждому примеру применялась вышеописанная схема для нахождения приближенного решения со значением целевой функции F_e, затем с помощью алгоритма ветвей и границ искалось точное решение с оптимальным значением целевой функции F^*. Далее вычислялось Δ – отношение реальной погрешности схемы $\delta = F_e - F^*$ к ее верхней оценке, определяемой неравенством (3.6):

$$\Delta = \frac{F_e - F^*}{2\rho(A, B)}.$$

Было обнаружено, что если поиск полиномиально разрешимого примера ведется в классе RD, то средняя погрешность решения растет от 20 до 30% от верхней оценки (3.6) при увеличении n. Это показывает, что расписание по возрастанию продолжительностей обслуживание плохо применимо для примеров с выбранным распределением параметров. Для остальных классов сред-

Таблица 2. Средняя разница между целевыми функциями и доли составных частей метрики

n	$\frac{\left\lvert\sum T_j^A - T_j^B\right\rvert}{\rho}$	$\frac{\rho_r}{\rho}$	$\frac{\rho_p}{\rho}$	$\frac{\rho_d}{\rho}$
10	11.7%	35.6%	42.3%	20.6%
20	10.4%	39.7%	39.4%	19.4%
40	8.9%	42.4%	37.4%	18.6%
60	7.8%	43.6%	36.6%	18.3%
80	7.3%	44.4%	34.4%	18%
100	6.7%	44.9%	35.7%	17.9%

Таблица 3. Средняя экспериментальная погрешность в процентах от теоретической

n	$\mathcal{PR}$	$\mathcal{PD}$	$\mathcal{RD}$	$\mathcal{P}$	$\mathcal{R}0$
4	2.5%	4.6%	20.8%	1.8%	2.9%
5	2.6%	4.8%	23.1%	1.9%	2.8%
6	2.6%	4.6%	24.6%	1.9%	2.7%
7	2.6%	4.7%	26%	1.9%	2.5%
8	2.5%	4.6%	27%	2%	2.3%
9	2.4%	4.7%	27.9%	2%	2.2%
10	2.4%	4.6%	28.6%	1.9%	2.1%

няя погрешность решения не зависит от n и составляет несколько процентов от максимальной теоретической. Столь малая погрешность обусловлена тем, что примерно в 20% случаев метод изменения параметров давал точное решение задачи. Зависимость средней ошибки Δ от n представлена в табл. 3 (см. [13]).

СПИСОК ЛИТЕРАТУРЫ

1. *Brucker P., Knust S.* Complex scheduling. Berlin Heidelberg: Springer-Verlag, 2011.
2. *Wagner H.M.* An integer linear–programming model for machine scheduling // Naval Res. Logist. Quart. 1959. V. 6. № 2. P. 131–140.
3. *Rothkopf M.H.* Scheduling independent tasks on parallel processors // Management Sci. 1966. V. 12. № 5. P. 437–447.
4. *Ignall E., Schrage L.* Application of the branch and bound technique to some flow-shop scheduling problems // Operat. Res. 1965. V. 13. № 3. P. 400–412.
5. *Lemtyuzhnikova D., Leonov V.* Large-scale problems with quasi-block matrices // J. of Comp. and Syst. Sci. Int. 2019. V. 58. № 4. P. 571–578.
6. *Werner F.* A Survey of Genetic Algorithms for Shop Scheduling Problems // P. Siarry: Heuristics: Theory and Applications, Nova Sci. Publ., 2013. P. 161–222.
7. *Zuo L., Shu L., Dong S., Zhu C., Hara T.* A multi-objective optimization scheduling method based on the ant colony algorithm in cloud computing // IEEE Access. 2015. V. 3. P. 2687–2699.
8. *Pan Q.* An effective co-evolutionary artificial bee colony algorithm for steelmaking-continuous casting scheduling // Europ. J. Operat. Res. 2016. V. 250. № 3. 702–714.
9. *Sels V., Coelho J., Dias A., Vanhoucke M.* Hybrid tabu search and a truncated branch-and-bound for the unrelated parallel machine scheduling problem // Comput. Operat. Res. 2015. V. 53. P. 107–117.
10. *Lazarev A.* Estimation of absolute error in scheduling problems of minimizing the maximum lateness // Dokl. Math. 2007. V. 76. P. 572–574.
11. *Лазарев А.А.* Теория расписаний: методы и алгоритмы. М.: ИПУ РАН, 2019. 408 с.
12. *Лазарев А.А., Архипов Д.И.* Оценка абсолютной погрешности и полиномиальной разрешимости для классической NP-трудной задачи теории расписаний // Докл. АН. 2018. Т. 480. № 5. С. 523–527.
13. *Лазарев А.А., Коренев П.С., Сологуб А.А.* Метрика для задачи минимизации суммарного запаздывания // Управление большими системами. 2015. Вып. 57. С. 123–137.

14. *Лазарев А.А., Кварацхелия А.Г.* Метрики в задачах теории расписаний // Докл. АН. 2010. Т. 432. № 6. С. 746–749.
15. *Du J., Leung J.Y.-T.* Minimizing total tardiness on one processor is *NP*-hard // Math. Operat. Res. 1990. V. 15. P. 483–495.
16. *Lawler E.L.* A pseudopolynomial algorithm for sequencing jobs to minimize total tardiness // Ann. Discrete Math. 1977. V. 1. P. 331–342.
17. *Graham R.L., Lawler E.L., Lenstra J.K., Rinnooy Kan A.H.G.* Optimization and approximation in deterministic sequencing and scheduling: a survey // Ann. Discrete Math. 1979. V. 5. P. 287–326.
18. *Lawler E.L.* A fully polynomial approximation scheme for the total tardiness problem // Operat. Res. Lett. 1982. V. 1. № 6. P. 207–208.
19. *Kovalyov M.Y.* Improving the complexities of approximation algorithms for optimization problems // Operat. Res. Lett. 1995. V. 17. P. 85–87.
20. *Lazarev A.A., Werner F.* Algorithms for special single machine total tardiness problem and an application to the even-odd partition problem // Math. and Comp. Model. 2009. № 49. P. 2078–2089.
21. *Baptiste Ph.* Scheduling equal-length jobs on identical parallel machines // Discret. Appl. Math. 2000. № 103. P. 21–32.
22. *Лазарев А.А., Садыков Р.Р., Севастьянов С.В.* Схема приближенного решения проблемы $1|r_j|L_{\max}$ // Дискретный анализ и исслед. операций. 2006. Сер. 2. Т. 13. № 1. С. 57–76.
23. *Лазарев А.А., Гафаров Е.Р.* Теория расписаний. Минимизация суммарного запаздывания для одного прибора // М.: Научное издание. ВЦ им. А.А. Дородницына РАН, 2006. 134 с.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1192–1205

ИНФОРМАТИКА

УДК 519.72

О СООТНОШЕНИИ ВЗАИМНОЙ ИНФОРМАЦИИ И ВЕРОЯТНОСТИ ОШИБКИ В ЗАДАЧЕ КЛАССИФИКАЦИИ ДАННЫХ[1)]

© 2021 г. А. М. Ланге[1,], М. М. Ланге[1,*], С. В. Парамонов[1,***]**

[1] *119333 Москва, ул. Вавилова, 40, ФИЦ ИУ РАН, Россия*

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

***e-mail: lange_mm@ccas.ru*

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

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Исследуется модель классификации данных на основе зависимости средней взаимной информации между предъявляемыми объектами и принимаемыми решениями от вероятности ошибки. Оптимизация модели заключается в нахождении обменного соотношения “взаимная информация–вероятность ошибки” между наименьшей средней взаимной информацией и вероятностью ошибки, которое аналогично известной функции “скорость–погрешность” (rate distortion function) для модели кодирования сообщений с допустимой погрешностью, переданных по каналу с искажениями. Строится нижняя граница введенного соотношения, которая дает нижнюю оценку вероятности ошибки классификации на заданном множестве объектов при любом фиксированном значении средней взаимной информации. Приводится обобщение соотношения “взаимная информация–вероятность ошибки” и его нижней границы для ансамбля источников. Полученные границы полезны для оценивания избыточности вероятности ошибки решающих алгоритмов с заданными наборами разделяющих функций. Библ. 11. Фиг. 4.

Ключевые слова: классификация, ансамбль источников, вероятность ошибки, взаимная информация, соотношение “взаимная информация–вероятность ошибки”, нижняя граница, разделяющая функция, решающий алгоритм, избыточность вероятности ошибки.

DOI: 10.31857/S0044466921070115

1. ВВЕДЕНИЕ

В теории кодирования источников с допустимой погрешностью Шенноном введена функция “скорость–погрешность” (rate distortion function) [1], которая при заданной погрешности дает нижнюю границу скорости кодирования, либо при заданной скорости определяет нижнюю границу погрешности для всевозможных способов (алгоритмов) кодирования. Функция “скорость–погрешность” определяется параметрами источника и метрикой погрешности и не зависит от выбранного алгоритма кодирования. Поэтому качество любого алгоритма кодирования может быть оценено избыточностью скорости кода относительно нижней границы при заданной погрешности, либо избыточностью погрешности кода относительно нижней границы при заданной скорости.

Следуя результатам теории кодирования источников, для модели классификации данных целесообразно найти аналогичную функцию “взаимная информация–вероятность ошибки” в виде зависимости наименьшей средней взаимной информации между множеством классифицируемых объектов и множеством решений о классах этих объектов от заданной допустимой вероятности ошибки. Такая функция дает нижнюю границу вероятности ошибки на заданном множестве объектов при фиксированной средней взаимной информации и, следовательно, позволяет оценить избыточность вероятности ошибки решающего алгоритма, реализуемого на заданном наборе разделяющих функций.

[1)]Работа выполнена при частичной финансовой поддержке РФФИ (код проекта 18-07-01231).

Известны работы, в которых приводятся теоретические оценки точности для классификаторов с различными решающими правилами [2], [3]. Однако эти работы не содержат общего подхода к построению нижней границы вероятности ошибки, не зависящей от конкретных решающих алгоритмов. Поскольку многие современные модели анализа данных и машинного обучения широко используют методы теории информации [4], естественно предложить новый подход для построения нижней границы вероятности ошибки классификации данных с использованием средней взаимной информации между множеством классифицируемых данных и множеством возможных решений о классах предъявляемых данных. Цель такого подхода заключается в нахождении наименьшей средней взаимной информации как убывающей функции заданной допустимой вероятности ошибки. Очевидно, что при фиксированной средней взаимной информации такая функция дает наименьшую вероятность ошибки.

Для получения нижней границы вероятности ошибки классификации на заданном множестве объектов предлагается использовать теоретико-информационную модель на основе известной схемы кодирования источника с допустимой погрешностью при наличии канала наблюдения с шумом [5]. В предлагаемой модели метки классов и классифицируемые объекты рассматриваются как входные и выходные данные канала наблюдения, вероятностные характеристики которого определяются условными по классам вероятностями или их аппроксимациями, построенными с использованием метрики на множестве объектов. Средняя погрешность между исходными метками классов и решениями измеряется в метрике Хемминга и эквивалентна вероятности ошибки. Для такой модели вводится функция “взаимная информация–вероятность ошибки” в форме зависимости наименьшей средней взаимной информации, содержащейся в множестве объектов относительно множества принимаемых решений, от вероятности ошибки и строится нижняя граница этой функции. Предлагаемая граница анонсирована в работах [6], [7] в форме обобщения нижней границы Шеннона для функции “скорость–погрешность” в схеме кодирования независимых символов конечного алфавита с допустимой средней погрешностью в метрике Хемминга. В настоящей работе дается строгое доказательство предложенной нижней границы для обменного соотношения “взаимная информация–вероятность ошибки”.

В настоящей работе получены нижние границы соотношения “взаимная информация–вероятность ошибки” для множества объектов с заданной метрикой и для ансамбля множеств различной модальности. Показана возможность уменьшения вероятности ошибки за счет увеличения средней взаимной информации между ансамблем данных и множеством классов с увеличением числа источников в ансамбле. Построенные границы достигаются на байесовском решающем алгоритме с разделяющими функциями, заданными апостериорными вероятностями классов [8]. Приведены численные реализации нижних границ функции “взаимная информация–вероятность ошибки” для множества подписей, множества лиц и для множества составных объектов, образованных парами “лицо, подпись”.

2. ТЕОРЕТИКО-ИНФОРМАЦИОННАЯ МОДЕЛЬ КЛАССИФИКАЦИИ И ЗАДАЧА ИССЛЕДОВАНИЯ

Пусть $\Omega = \{\omega_1,\ldots,\omega_c\}$, $c \geq 2$, – множество классов с априорными вероятностями $P(\omega_i), i = 1,\ldots,c$, и $\mathbf{X}$ – множество объектов с условными по классам вероятностями $P(\mathbf{x}|\omega_i)$ $\forall \mathbf{x} \in \mathbf{X}$, $i = 1,\ldots,c$. Будем считать, что множества Ω и $\mathbf{X}$ являются входными и выходными данными преобразования $\Omega \to \mathbf{X}$. Множество объектов $\mathbf{X}$ и множество решений $\hat{\Omega} = \{\omega_j = 1,\ldots,c\}$ о классах этих объектов являются входом и выходом преобразования $\mathbf{X} \to \hat{\Omega}$. Пусть $\mathrm{X}^N = (\mathbf{x}_1,\ldots,\mathbf{x}_N)$ – блок объектов $\mathbf{x}_n \in \mathbf{X}$, $n = 1,\ldots,N$, и $\mathbf{X}^N$ – множество всевозможных блоков длины N. Условные по классам вероятности блоков $\mathrm{X}^N \in \mathbf{X}^N$ являются характеристиками канала наблюдения $\Omega \to \mathbf{X}^N$ и образуют множество распределений

$$\mathrm{P} = \left\{ P(\mathrm{X}^N|\omega_i) = \prod_{n=1}^{N} P(\mathbf{x}_n|\omega_i) : \sum_{\mathrm{X}^N \in \mathbf{X}^N} P(\mathrm{X}^N|\omega_i) = 1, \;\; i = 1,\ldots,c \right\},$$

а условные вероятности решений $\omega_j \in \hat{\Omega}$ для каждого предъявляемого блока являются характеристиками тест-канала $\mathbf{X}^N \to \hat{\Omega}$ и образуют множество распределений

$$\mathrm{Q} = \left\{ Q(\omega_j|\mathrm{X}^N) : \sum_{j=1}^{c} Q(\omega_j|\mathrm{X}^N) = 1, \ \ \forall \mathrm{X}^N \in \mathbf{X}^N \right\}.$$

В принятых обозначениях множества Ω и $\hat{\Omega}$ состоят их одних и тех же элементов, однако вероятности элементов множества $\hat{\Omega}$ могут отличаться от априорных вероятностей соответствующих элементов множества Ω. Множества Ω, $\mathbf{X}^N$, $\hat{\Omega}$ совместно с распределениями P и Q порождают схему классификации

$$\Omega \xrightarrow{\mathrm{P}} \mathbf{X}^N \xrightarrow{\mathrm{Q}} \hat{\Omega}, \tag{1}$$

в которой распределения P считаются известными, а распределения Q подлежат оптимизации.

Для оптимизации множества распределений Q в (1) введем функционалы средней взаимной информации $I_{\mathrm{Q}}(\mathbf{X}^N;\hat{\Omega}) \geq 0$ и вероятности ошибки $E_{\mathrm{Q}}(\mathbf{X}^N,\hat{\Omega}) \geq 0$, зависящие от Q. Согласно [1], средняя взаимная информация имеет вид

$$\begin{aligned} I_{\mathrm{Q}}(\mathbf{X}^N;\hat{\Omega}) &= \sum_{\mathrm{X}^N \in \mathbf{X}^N} P(\mathrm{X}^N) \sum_{j=1}^{c} Q(\omega_j|\mathrm{X}^N) \ln \frac{Q(\omega_j|\mathrm{X}^N)}{Q(\omega_j)} = -\sum_{j=1}^{c} Q(\omega_j) \ln Q(\omega_j) + \\ &+ \sum_{\mathrm{X}^N \in \mathbf{X}^N} P(\mathrm{X}^N) \sum_{j=1}^{c} Q(\omega_j|\mathrm{X}^N) \ln Q(\omega_j|\mathrm{X}^N) = H(\hat{\Omega}) - H(\hat{\Omega}|\mathbf{X}^N), \end{aligned} \tag{2}$$

где

$$P(\mathrm{X}^N) = \sum_{i=1}^{c} P(\omega_i) P(\mathrm{X}^N|\omega_i),$$

$$Q(\omega_j) = \sum_{\mathrm{X}^N \in \mathbf{X}^N} P(\mathrm{X}^N) Q(\omega_j|\mathrm{X}^N)$$

соответственно безусловные вероятности блоков $\mathrm{X}^N \in \mathbf{X}^N$ и решений $\omega_j \in \hat{\Omega}$, а $H(\hat{\Omega})$ и $H(\hat{\Omega}|\mathbf{X}^N)$ – соответственно безусловная и условная энтропии на множестве решений $\hat{\Omega}$, причем $H(\hat{\Omega}) \geq H(\hat{\Omega}|\mathbf{X}^N)$. Средняя вероятность ошибки определяется средним значением индикатора $[\omega_j \neq \omega_i]$ и имеет вид

$$\begin{aligned} E_{\mathrm{Q}}(\mathbf{X}^N,\hat{\Omega}) &= \sum_{i=1}^{c} P(\omega_i) \sum_{\mathrm{X}^N \in \mathbf{X}^N} P(\mathrm{X}^N|\omega_i) \sum_{j=1}^{c} Q(\omega_j|\mathrm{X}^N)[\omega_j \neq \omega_i] = \\ &= \sum_{\mathrm{X}^N \in \mathbf{X}^N} P(\mathrm{X}^N) \sum_{j=1}^{c} Q(\omega_j|\mathrm{X}^N) \sum_{i=1}^{c} P(\omega_i|\mathrm{X}^N)[\omega_j \neq \omega_i], \end{aligned} \tag{3}$$

где

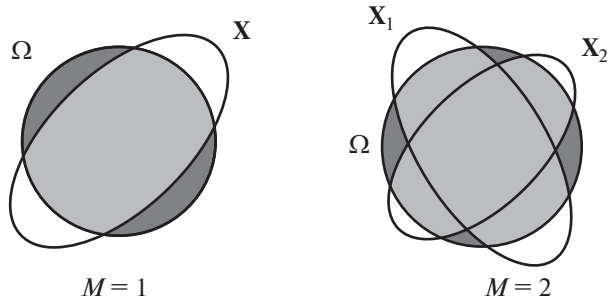
$$\sum_{i=1}^{c} P(\omega_i|\mathrm{X}^N)[\omega_j \neq \omega_i] = 1 - P(\omega_j|\mathrm{X}^N)$$

есть вероятность ошибки на решении ω_j по блоку объектов X^N.

Используя функционалы (2) и (3), введем функцию “взаимная информация–вероятность ошибки”

$$R(\varepsilon) = \min_{N} \min_{\mathrm{Q}:\, E_{\mathrm{Q}}(\mathbf{X}^N,\hat{\Omega}) \leq \varepsilon} I_{\mathrm{Q}}(\mathbf{X}^N;\hat{\Omega}), \tag{4}$$

где внутренний минимум берется по всевозможным множествам распределений Q при $\varepsilon > 0$ и $N \geq 1$. Задача состоит в нахождении нижней границы функции $R(\varepsilon)$ для источника с заданным



Фиг. 1. Интерпретация уменьшения минимальной вероятности ошибки с увеличением числа источников.

множеством объектов $\mathbf{X}$ и для ансамбля множеств $\mathbf{X}^M = \mathbf{X}_1 \ldots \mathbf{X}_M$, порождаемых $M \geq 2$ источниками различной модальности.

Следует отметить, что функция $R(\varepsilon)$ убывает с ростом ε и, следовательно, при любом фиксированном значении дает наименьшую вероятность ошибки, которая может быть достигнута при заданной средней взаимной информации между входом и выходом классификатора. В общем случае минимальная вероятность ошибки соответствует наибольшему значению функции вида (1), которая может быть введена для $M \geq 1$ источников. Наибольшее значение такой функции должно определяться средней взаимной информацией $I(\mathbf{X}^M;\Omega)$ между ансамблем $\mathbf{X}^M$ и множеством классов Ω. Иллюстрация зависимости минимальной вероятности ошибки от числа источников M дана на фиг. 1. Поскольку средняя взаимная информация $I(\mathbf{X}^M;\Omega) = H(\Omega) - H(\Omega \mid \mathbf{X}^M)$ является теоретико-информационной мерой пересечения $\Omega \cap \mathbf{X}^M$ (“серая область”), а условная энтропия $H(\Omega|\mathbf{X}^M)$ – теоретико-информационной мерой разности $\Omega \backslash \mathbf{X}^M$ (“черная область”), то наименьшая вероятность ошибки должна уменьшаться с уменьшением $H(\Omega|\mathbf{X}^M)$ и, соответственно, с ростом $I(\mathbf{X}^M;\Omega)$. С увеличением числа M декоррелированных источников “серая область” увеличивается, а “черная область” уменьшается, что должно привести к снижению наименьшей вероятности ошибки. Формальное обоснование зависимости минимальной вероятности ошибки от энтропии $H(\Omega|\mathbf{X}^M)$ или взаимной информации $I(\mathbf{X}^M;\Omega)$ будет дано в разд. 3 и 4.

3. НИЖНЯЯ ГРАНИЦА ФУНКЦИИ $R(\varepsilon)$ ДЛЯ ЗАДАННОГО ИСТОЧНИКА

Построение нижней границы $\underline{R}(\varepsilon) \leq R(\varepsilon)$ базируется на технике, предложенной в работе [5]. Пусть $\Omega^* = \{\omega_k, k = 1,\ldots,c\}$ – множество решений о классах, получаемых на преобразовании $\mathbf{X}^{N^*} \to \Omega^*$ с наименьшей вероятностью ошибки $\varepsilon_{\min}$, которая реализуется на байесовском решающем правиле для блоков $\mathrm{X}^{N^*} \in \mathbf{X}^{N^*}$ длины N^* (см. [8]). Введенное преобразование порождает разбиение множества $\mathbf{X}^{N^*}$ на непересекающиеся области решений $\mathbf{X}_k^{N^*} \to \omega_k$, $k = 1,\ldots,c$, которые соответствуют классам множества Ω^*. В полученном разбиении k-я область имеет вероятность

$$P^*(\omega_k) = \sum_{\mathrm{X}^{N^*} \in \mathbf{X}_k^{N^*}} P(\mathrm{X}^{N^*})$$

и для любого блока $\mathrm{X}^{N^*} \in \mathbf{X}_k^{N^*}$ условная вероятность решения $\omega_j \in \hat{\Omega}$ одинакова и равна $Q(\omega_j|\mathrm{X}^{N^*}) = Q^*(\omega_j|\omega_k)$. С учетом условных распределений

$$Q^* = \left\{ Q^*(\omega_j|\omega_k) \colon \sum_{j=1}^{c} Q^*(\omega_j|\omega_k) = 1,\ k = 1,\ldots,c \right\}$$

и $N = N^*$ средняя взаимная информация (2) преобразуется к виду

$$I_{\mathrm{Q}}(\mathbf{X}^{N^*};\hat{\Omega}) = I_{\mathrm{Q}^*}(\Omega^*;\hat{\Omega}) = \sum_{k=1}^{c} P^*(\omega_k)\sum_{j=1}^{c} Q^*(\omega_j|\omega_k)\ln\frac{Q^*(\omega_j|\omega_k)}{Q(\omega_j)}\ , \tag{5}$$

а вероятность ошибки (3) удовлетворяет неравенству

$$E_{\mathrm{Q}}(\mathbf{X}^{N^*},\hat{\Omega}) \le E(\mathbf{X}^{N^*},\Omega^*) + E_{\mathrm{Q}^*}(\Omega^*,\hat{\Omega}) = \varepsilon_{\min} + \sum_{k=1}^{c} P^*(\omega_k)\sum_{j=1}^{c} Q^*(\omega_j|\omega_k)[\omega_j \ne \omega_k]. \tag{6}$$

Соотношения (5) и (6) сводят нахождение границы $\underline{R}(\varepsilon)$ к вычислению минимума

$$\min_{\mathrm{Q}^*:\ E_{\mathrm{Q}^*}(\Omega^*,\hat{\Omega})\le\varepsilon-\varepsilon_{\min}} I_{\mathrm{Q}^*}(\Omega^*;\hat{\Omega}). \tag{7}$$

Минимум (7) следует из известной нижней границы Шеннона [1] в схеме кодирования $\Omega^* \to \hat{\Omega}$ с допустимой погрешностью в метрике Хемминга $[\omega_k \ne \omega_j]$ при условии, что средняя погрешность не превосходит величины $\varepsilon - \varepsilon_{\min} \ge 0$. Полученная граница сформулирована в следующей теореме.

Теорема. *Нижняя граница функции $R(\varepsilon)$ имеет вид*

$$\underline{R}(\varepsilon) = I(\mathbf{X};\Omega) - h(\varepsilon-\varepsilon_{\min}) - (\varepsilon-\varepsilon_{\min})\ln(c-1), \quad \varepsilon_{\min} \le \varepsilon \le \varepsilon_{\max},$$

где $h(z) = -z\ln z - (1-z)\ln(1-z)$, $\underline{R}(\varepsilon_{\min}) = I(\mathbf{X};\Omega)$, $\underline{R}(\varepsilon_{\max}) = 0$ и $I(\mathbf{X};\Omega)$ — средняя взаимная информация между множествами $\mathbf{X}$ и Ω.

Доказательство. Применение нижней границы Шеннона для минимума в (7) дает

$$R_L(\varepsilon) = H(\Omega^*) - H_s(\hat{\Omega}|\Omega^*), \tag{8}$$

где

$$H(\Omega^*) = -\sum_{k=1} P^*(\omega_k)\ln P^*(\omega_k),$$

$$H_s(\hat{\Omega}|\Omega^*) = -\sum_{k=1}^{c} P^*(\omega_k)\sum_{j=1}^{c} Q_s^*(\omega_j|\omega_k)\ln Q_s^*(\omega_j|\omega_k)\ ,$$

а условные распределения имеют параметрическую форму

$$\mathrm{Q}_s^* = \left\{ Q_s^*(\omega_j|\omega_k) = \frac{e^{-s[\omega_j\ne\omega_k]}}{\sum_{i=1}^{c} e^{-s[\omega_i\ne\omega_k]}},\ j = 1,\ldots,c\ ;\ k = 1,\ldots,c \right\}$$

с параметром $s > 0$. Значение параметра следует из уравнения

$$\sum_{k=1}^{c} P^*(\omega_k)\sum_{j=1}^{c} Q_s^*(\omega_j|\omega_k)[\omega_k \ne \omega_j] = \varepsilon - \varepsilon_{\min},$$

которое дает решение $e^{-s} = (\varepsilon-\varepsilon_{\min})/(c-1)\big(1-(\varepsilon-\varepsilon_{\min})\big)$ и условные вероятности

$$Q_s^*(\omega_j|\omega_k) = \begin{cases} (\varepsilon-\varepsilon_{\min})/(c-1), & j \ne k, \\ 1-(\varepsilon-\varepsilon_{\min}), & j = k, \end{cases}$$

на которых достигается условная энтропия

$$H_s(\hat{\Omega}|\Omega^*) = h(\varepsilon-\varepsilon_{\min}) + (\varepsilon-\varepsilon_{\min})\ln(c-1). \tag{9}$$

В точке $\varepsilon = \varepsilon_{\min}$ соотношение (9) дает $H_s(\hat{\Omega}|\Omega^*) = 0$ и правая часть в (8) равна $H(\Omega^*)$.

Согласно теореме кодирования для источника [1], энтропия выхода преобразования $\mathbf{X}^{N^*} \to \Omega^*$ удовлетворяет неравенству $H(\Omega^*) \ge R(\varepsilon_{\min})$. Из (4) имеем $R(\varepsilon_{\min}) = I_{\mathrm{Q}}(\mathbf{X}^{N^*};\Omega^*)$, где Q — множество распределений апостериорных вероятностей (см. [8])

$$Q(\omega_k|\mathrm{X}^{N^*}) = P(\omega_k)P(\mathrm{X}^{N^*}|\omega_k)/P(\mathrm{X}^{N^*}), \quad k = 1,\ldots,c, \quad \forall \mathrm{X}^{N^*} \in \mathbf{X}^{N^*}.$$

В этом случае $I_{\mathrm{Q}}(\mathbf{X}^{N^*};\Omega^*) = I(\mathbf{X}^{N^*};\Omega)$, а $I(\mathbf{X}^{N^*};\Omega)$ – средняя взаимная информации между выходом и входом канала наблюдения в схеме (1). Поскольку для любого ансамбля $\mathbf{X}^N = \mathbf{X}_1 \ldots \mathbf{X}_N$ справедливо неравенство $\max_{n=1}^{N} I(\mathbf{X}_n;\Omega) \le I(\mathbf{X}^N;\Omega)$, которое в случае тождественных множеств $\mathbf{X}_n = \mathbf{X}$, $n = 1,\ldots,N$, выполняется со знаком равенства, имеем $I(\mathbf{X}^{N^*};\Omega) = I(\mathbf{X};\Omega)$. С учетом сделанных замечаний получаем для энтропии $H(\Omega^*)$ следующую нижнюю оценку:

$$H(\Omega^*) \ge R(\varepsilon_{\min}) = I(\mathbf{X};\Omega). \tag{10}$$

Замены условной и безусловной энтропий в (8) правыми частями соотношений (9) и (10) завершают доказательство теоремы.

Граница $\underline{R}(\varepsilon)$ убывает с увеличением ε и достигает наибольшего значения $I(\mathbf{X};\Omega) = H(\Omega) - H(\Omega|\mathbf{X})$ в точке $\varepsilon = \varepsilon_{\min}$. Здесь

$$H(\Omega) = -\sum_{i=1}^{c} P(\omega_i)\ln P(\omega_i)$$

есть энтропия множества Ω и

$$H(\Omega|\mathbf{X}) = -\sum_{\mathbf{x}\in\mathbf{X}} P(\mathbf{x})\sum_{i=1}^{c} P(\omega_i|\mathbf{X})\ln P(\omega_i|\mathbf{X})$$

есть условная энтропия множества Ω при заданном множестве $\mathbf{X}$, причем $H(\Omega|\mathbf{X}) \le H(\Omega)$.

Следует отметить, что сформулированная в теореме граница $\underline{R}(\varepsilon)$ является обобщением нижней границы Шеннона в схеме кодирования независимых дискретных сообщений с ограниченной средней погрешностью, измеряемой в метрике Хемминга [1]. В схеме Шеннона условные вероятности $P(\mathbf{x}|\omega_i) = [\mathbf{x} = \mathbf{x}_i]$, $i = 1,\ldots,c$, принимают значения 1 и 0. В этом случае согласно формуле Байеса $Q(\omega_j|\mathbf{x}) = [\omega_j = \omega_i]$ и, следовательно, $H(\Omega|\mathbf{X}) = 0$ и $I(\mathbf{X};\Omega) = H(\Omega)$, что дает $\varepsilon_{\min} = 0$ и $\underline{R}(\varepsilon_{\min}) = H(\Omega)$. В общем случае $\varepsilon_{\min} \ge 0$, а наибольшая вероятность ошибки равна

$$\varepsilon_{\max} = (c-1)\min_{i=1}^{c} P(\omega_i)$$

и преобразуется к виду $\varepsilon_{\max} = (c-1)/c$ при равномерном априорном распределении.

Минимальная вероятность ошибки $\varepsilon_{\min}$ зависит от величины условной энтропии $H(\Omega|\mathbf{X}) \ge 0$. Поскольку при значениях $\varepsilon_{\min} \ge 0$ граница $\underline{R}(\varepsilon)$ достигает нулевого значения в точке $\varepsilon = \varepsilon_{\max}$, справедливо равенство

$$h(\varepsilon_{\max}) - h(\varepsilon_{\max} - \varepsilon_{\min}) + \varepsilon_{\min}\ln(c-1) = H(\Omega|\mathbf{X}),$$

которое с учетом разложения

$$h(\varepsilon_{\max}) - h(\varepsilon_{\max} - \varepsilon_{\min}) = h'(\varepsilon_{\max})\varepsilon_{\min} + \frac{1}{2}\left|h''(\varepsilon_{\max})\right|\varepsilon_{\min}^2 + O(\varepsilon_{\min}^3)$$

при малых значениях $\varepsilon_{\min}$ преобразуется к виду

$$\frac{1}{2}\left|h''(\varepsilon_{\max})\right|\varepsilon_{\min}^2 + \left(h'(\varepsilon_{\max}) + \ln(c-1)\right)\varepsilon_{\min} = H(\Omega|\mathbf{X}).$$

Решение полученного уравнения относительно $\varepsilon_{\min}$ дает при значениях производных $h'(\varepsilon_{\max}) = \ln\left((1-\varepsilon_{\max})/\varepsilon_{\max}\right)$ и $\left|h''(\varepsilon_{\max})\right| = \left(\varepsilon_{\max}(1-\varepsilon_{\max})\right)^{-1}$ асимптотическую оценку

$$\begin{aligned}\varepsilon_{\min} = \varepsilon_{\max}(1-\varepsilon_{\max})&\left(\ln^2\left((c-1)\frac{(1-\varepsilon_{\max})}{\varepsilon_{\max}}\right) + 2\frac{H(\Omega|\mathbf{X})}{\varepsilon_{\max}(1-\varepsilon_{\max})}\right)^{1/2} - \\ &- \varepsilon_{\max}(1-\varepsilon_{\max})\ln\left((c-1)\frac{(1-\varepsilon_{\max})}{\varepsilon_{\max}}\right),\end{aligned} \tag{11}$$

которая в случае равновероятных классов имеет вид

$$\varepsilon_{\min} = \frac{(c-1)}{c}\left(2\frac{H(\Omega|\mathbf{X})}{(c-1)}\right)^{1/2}. \tag{12}$$

Оценки (11) и (12) демонстрируют уменьшение вероятности ошибки $\varepsilon_{\min}$ с уменьшением условной энтропии $H(\Omega|\mathbf{X})$ и, следовательно, с увеличением средней взаимной информации $I(\mathbf{X};\Omega)$ при фиксированной энтропии $H(\Omega)$. При этом $H(\Omega|\mathbf{X}) = 0$ обеспечивает $\varepsilon_{\min} = 0$.

4. ОБОБЩЕНИЕ ДЛЯ АНСАМБЛЯ ИСТОЧНИКОВ

Будем считать, что ансамбль $\mathbf{X}^M = \mathbf{X}_1 \ldots \mathbf{X}_M$ порождает составные объекты $\mathbf{x}^M = (\mathbf{x}_1, \ldots, \mathbf{x}_M)^t$, содержащие M объектов одного класса, по одному от каждого источника $\mathbf{x}_m \in \mathbf{X}_m$, $m = 1, \ldots, M$, и каждый составной объект $\mathbf{x}^M$ образует столбец размера M. Набор из N составных объектов образует блок столбцов $\mathrm{X}^{MN} = (\mathbf{x}_1^M, \ldots, \mathbf{x}_N^M)$, $\mathbf{x}_n^M \in \mathbf{X}^M$, $n = 1, \ldots, N$; всевозможные блоки X^{MN} размера $M \times N$ образуют множество $\mathbf{X}^{MN}$. В этом случае канал наблюдения $\Omega \to \mathbf{X}^{MN}$ задается условными распределениями вероятностей блоков X^{MN}:

$$\mathrm{P}^M = \left\{ P(\mathrm{X}^{MN}|\omega_i) = \prod_{n=1}^{N} P(\mathbf{x}_n^M|\omega_i) : \sum_{\mathrm{X}^{MN} \in \mathbf{X}^{MN}} P(\mathrm{X}^{MN}|\omega_i) = 1,\ i = 1, \ldots, c \right\},$$

а тест-канал $\mathbf{X}^{MN} \to \hat{\Omega}$ определяется условными распределениями вероятностей решений:

$$\mathrm{Q}^M = \left\{ Q(\omega_j|\mathrm{X}^{MN}) : \sum_{j=1}^{c} Q(\omega_j|\mathrm{X}^{MN}) = 1,\ \ \forall \mathrm{X}^{MN} \in \mathbf{X}^{MN} \right\}$$

по блокам составных объектов.

Множества Ω, $\mathbf{X}^{MN}$, $\hat{\Omega}$ совместно с распределениями P^M и Q^M порождают схему классификации

$$\Omega \xrightarrow{\mathrm{P}^M} \mathbf{X}^{MN} \xrightarrow{\mathrm{Q}^M} \hat{\Omega},$$

которая аналогична схеме (1). Распределения P^M и Q^M дают среднюю взаимную информацию $I_{\mathrm{Q}^M}(\mathbf{X}^{MN};\hat{\Omega})$ и вероятность ошибки $E_{\mathrm{Q}^M}(\mathbf{X}^N,\hat{\Omega})$ в форме функционалов (2) и (3). Для заданного значения $\varepsilon > 0$ эти функционалы позволяют ввести невозрастающую функцию

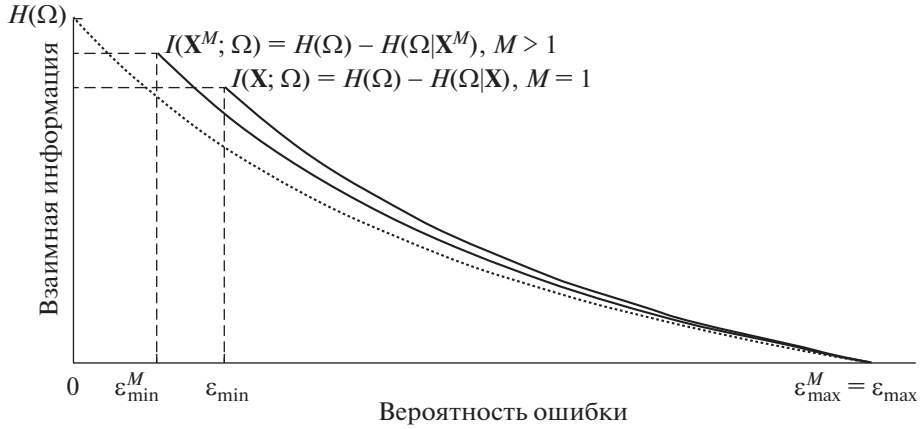
$$R_M(\varepsilon) = \min_N \min_{\mathrm{Q}^M:\, E_{\mathrm{Q}^M}(\mathbf{X}^{MN},\hat{\Omega}) \le \varepsilon} I_{\mathrm{Q}^M}(\mathbf{X}^{MN};\hat{\Omega}) \tag{13}$$

в форме, аналогичной (4), причем $R_M(\varepsilon) = R(\varepsilon)$, когда $M = 1$. Применение техники, использованной в разд. 3 для построения границы $\underline{R}(\varepsilon) \le R(\varepsilon)$, дает для функции (13) нижнюю границу

$$\underline{R}_M(\varepsilon) = I(\mathbf{X}^M;\Omega) - h(\varepsilon - \varepsilon_{\min}^M) - (\varepsilon - \varepsilon_{\min}^M)\ln(c-1), \quad \varepsilon_{\min}^M \le \varepsilon \le \varepsilon_{\max}^M, \tag{14}$$

которая по форме аналогична границе $\underline{R}(\varepsilon)$, приведенной в теореме. В границе (14) $I(\mathbf{X}^M;\Omega)$ – средняя взаимная информация между ансамблем $\mathbf{X}^M$ и множеством Ω, $\underline{R}_M(\varepsilon_{\min}^M) = I(\mathbf{X}^M;\Omega)$ и $\underline{R}_M(\varepsilon_{\max}^M) = 0$. Для минимальной вероятности ошибки $\varepsilon_{\min}^M$ сохраняются асимптотические оценки вида (11) и (12) с заменой энтропии $H(\Omega|\mathbf{X})$ величиной $H(\Omega|\mathbf{X}^M)$ и заменой вероятности ошибки $\varepsilon_{\max}$ величиной $\varepsilon_{\max}^M$. При любом $M \ge 1$ значение максимальной вероятности ошибки равно $\varepsilon_{\max}^M = (c-1)\min_{i=1}^c P(\omega_i)$.

Уменьшение минимальной вероятности ошибки $\varepsilon_{\min}^M$ с ростом числа декоррелированных множеств $\mathbf{X}_m$, $m = 1, \ldots, M$, имеет формальное объяснение. Пусть множество $\mathbf{X}_1$ обеспечивает наибольшую среднюю взаимную информацию $I(\mathbf{X}_1;\Omega) = \max_{m=1}^M I(\mathbf{X}_m;\Omega)$.



Фиг. 2. Характер границ $\underline{R}_M(\varepsilon)$ для одного и нескольких источников.

С учетом симметрии средней взаимной информации имеем

$$I(\mathbf{X}^M;\Omega) = I(\Omega;\mathbf{X}^M) = I(\Omega;\mathbf{X}_1) + \sum_{m=2}^{M} I(\Omega;\mathbf{X}_m|\mathbf{X}_{m-1}\ldots\mathbf{X}_1),$$

где

$$I(\Omega;\mathbf{X}_m|\mathbf{X}_{m-1}\ldots\mathbf{X}_1) = H(\mathbf{X}_m|\mathbf{X}_{m-1}\ldots\mathbf{X}_1) - H(\mathbf{X}_m|\mathbf{X}_{m-1}\ldots\mathbf{X}_1\Omega) \geq 0,$$

$$H(\mathbf{X}_m|\mathbf{X}_{m-1}\ldots\mathbf{X}_1) = -\sum_{\mathbf{x}_1\in\mathbf{X}_1}\ldots\sum_{\mathbf{x}_m\in\mathbf{X}_m} P(\mathbf{x}_1,\ldots,\mathbf{x}_m)\ln P(\mathbf{x}_m|\mathbf{x}_{m-1},\ldots,\mathbf{x}_1),$$

$$H(\mathbf{X}_m|\mathbf{X}_{m-1}\ldots\mathbf{X}_1\Omega) = -\sum_{i=1}^{c} P(\omega_i)\sum_{\mathbf{x}_1\in\mathbf{X}_1}\ldots\sum_{\mathbf{x}_m\in\mathbf{X}_m} P(\mathbf{x}_1,\ldots,\mathbf{x}_m|\omega_i)\ln P(\mathbf{x}_m|\mathbf{x}_{m-1},\ldots,\mathbf{x}_1,\omega_i).$$

Декорреляция источников предполагает строгую положительность условной средней взаимной информации $I(\Omega;\mathbf{X}_m|\mathbf{X}_{m-1}\ldots\mathbf{X}_1) > 0$, $m = 2,\ldots,M$, и, следовательно, увеличение средней взаимной информации $I(\mathbf{X}^M;\Omega)$ и уменьшение условной энтропии $H(\Omega|\mathbf{X}^M) = H(\Omega) - I(\mathbf{X}^M;\Omega)$ с ростом M. Поэтому из (11) и (12) следует, что увеличение числа декоррелированных источников приводит к уменьшению $\varepsilon^M_{\min}$.

Характер границ $\underline{R}_M(\varepsilon)$ при значениях $M \geq 1$ показан на фиг. 2 сплошными кривыми. Для сравнения дана точечная кривая нижней границы Шеннона.

5. УСЛОВНЫЕ ПО КЛАССАМ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ОБЪЕКТОВ

Вычисление средней взаимной информации $I(\mathbf{X};\Omega)$, которая используется в нижней границе $\underline{R}_M(\varepsilon)$, требует знания совместного распределения вероятностей $\{P(\omega_i,\mathbf{x}) = P(\omega_i)P(\mathbf{x}|\omega_i),\ \forall\mathbf{x}\in\mathbf{X},\ i = 1,\ldots,c\}$ на произведении $\Omega\times\mathbf{X}$. Априорные вероятности считаются известными, а условные по классам вероятности могут быть определены с помощью метрики $d(\mathbf{x},\hat{\mathbf{x}}) \geq 0$ в L_p, $p \geq 1$, для любой пары объектов $\mathbf{x}\in\mathbf{X}$, $\hat{\mathbf{x}}\in\mathbf{X}$.

Полагая, что $\mathbf{x}_i\in\mathbf{X}_i, i = 1,\ldots,c$, являются “центрами” непересекающихся кластеров $\mathbf{X}_i : \bigcup_{i=1}^{c}\mathbf{X}_i = \mathbf{X}$ и объекты каждого класса обладают свойством компактности, определим условные по классам вероятности

$$P(\mathbf{x}|\omega_i) = \frac{e^{-\nu_i d^p(\mathbf{x},\mathbf{x}_i)}}{\sum\limits_{\mathbf{x}\in\mathbf{X}} e^{-\nu_i d^p(\mathbf{x},\mathbf{x}_i)}}, \quad i = 1,\ldots,c, \tag{15}$$

с параметрами $v_i > 0, i = 1 = 1,\ldots,c$. Параметры распределений (15) могут быть оценены с использованием расстояний между объектами из $\mathbf{X}$ и "центрами"

$$\mathbf{x}_i = \arg\min_{\hat{\mathbf{x}}\in\mathbf{X}_i} \sum_{\mathbf{x}\in\mathbf{X}_i} d^p(\mathbf{x}, \hat{\mathbf{x}}), \quad i = 1,\ldots,c, \tag{16}$$

которые обеспечивают наименьшие рассеяния объектов внутри кластеров.

Для нахождения параметров $v_i, i = 1,\ldots,c$, воспользуемся плотностями распределения

$$p(\theta_i) = \frac{v_i^{1/p} e^{-v_i\theta_i^p}}{\Gamma(1+1/p)}, \quad i = 1,\ldots,c,$$

случайных величин $\theta_i = d(\mathbf{x},\mathbf{x}_i)$. Используя плотность $p(\theta_i)$ и требуя максимума вероятности

$$\Pr\{|\theta_i - \mu_i| \le \alpha_i\mu_i\} = \int_{\mu_i(1-\alpha_i)}^{\mu_i(1+\alpha_i)} p(\theta_i)d\theta_i = \frac{1}{\Gamma(1/p)}\left(\Gamma\left(1/p, v_i\mu_i^p(1-\alpha_i)^p\right) - \Gamma\left(1/p, v_i\mu_i^p(1+\alpha_i)^p\right)\right) \to \max_{v_i}$$

значений θ_i в α_i-окрестности среднего значения $\mu_i > 0$ (указанная вероятность выражена в терминах неполных гамма-функций), получаем

$$v_i = \frac{1}{\mu_i^p\left((1+\alpha_i)^p - (1-\alpha_i)^p\right)} \ln\frac{1+\alpha_i}{1-\alpha_i}, \quad i = 1,\ldots,c, \tag{17}$$

где $0 < \alpha_i < 1$.

Параметры μ_i и α_i в (17) определяются статистиками внутриклассовых расстояний $d(\mathbf{x},\mathbf{x}_i)$ в виде выборочных средних

$$\mu_i = \frac{1}{\|\mathbf{X}_i\|}\sum_{\mathbf{x}\in\mathbf{X}_i} d(\mathbf{x},\mathbf{x}_i), \quad i = 1,\ldots,c,$$

и выборочных дисперсий

$$\sigma_i^2 = \frac{1}{\|\mathbf{X}_i\|}\sum_{\mathbf{x}\in\mathbf{X}_i} |d(\mathbf{x},\mathbf{x}_i) - \mu_i|^2, \quad i = 1,\ldots,c.$$

Используя величины

$$\alpha_i^* = 2\mu_i \Big/ \left(\mu_i + \left(\mu_i^2 + \sigma_i^2\right)^{1/2}\right) < 1, \quad i = 1,\ldots,c, \quad \delta = 1 - c^{-1}\sum_{i=1}^{c}\alpha_i^*,$$

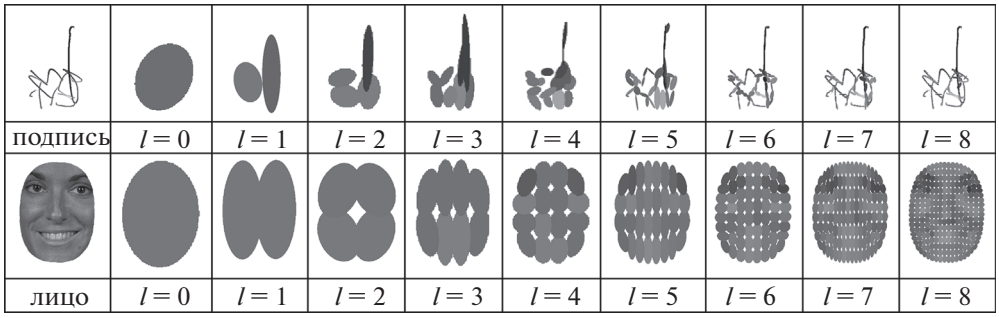
вычисляются значения $\alpha_i = \alpha_i^* + 2(1-\alpha_i^*)\tau_s - (1-\alpha_i^*)\tau_s^2 < 1$ в точках $\tau_s = \sum_{k=1}^{s} 2^{-k}$, $s = 1,2,\ldots$, и соответствующие растущие значения параметра v_i вида (17) и $\underline{R}(\varepsilon_{\max}) < 0$ до выполнения условия $|\underline{R}(\varepsilon_{\max})| \le \delta$ при достаточно малом пороге δ, когда $\mu_i \ge \sigma_i$. Для вычисления значений $\underline{R}(\varepsilon_{\max})$ используются асимптотические оценки $\varepsilon_{\min}$ вида (11). Истинное значение $\varepsilon_{\min}$ вычисляется коррекцией оценки $\varepsilon_{\min}$ при наибольшем значении v_i путем сдвига вправо до достижения равенства $\underline{R}(\varepsilon_{\max}) = 0$.

Полученные на множествах объектов от различных источников вероятности (15) с учетом оценок (16) и (17) порождают условные по классам распределения

$$\left\{P(\mathbf{x}_m|\omega_i) = \frac{e^{-v_{im}d^p(\mathbf{x}_m,\mathbf{x}_{im})}}{\sum\limits_{\mathbf{x}_m\in\mathbf{X}_m} e^{-v_{im}d^p(\mathbf{x}_m,\mathbf{x}_{im})}} : \sum_{\mathbf{x}_m\in\mathbf{X}_m} P(\mathbf{x}_m|\omega_i) = 1,\ i = 1,\ldots,c\right\}, \quad m = 1,\ldots,M, \tag{18}$$

на множествах $\mathbf{X}_m, m = 1,\ldots,M$, и условные по классам распределения

$$\left\{P(\mathbf{x}^M|\omega_i) = \prod_{m=1}^{M} P(\mathbf{x}_m|\omega_i) : \sum_{\mathbf{x}^M\in\mathbf{X}^M} P(\mathbf{x}^M|\omega_i) = 1,\ i = 1,\ldots,c\right\} \tag{19}$$



Фиг. 3. Примеры древовидных представлений подписи и лица.

на ансамбле $\mathbf{X}^M = \mathbf{X}_1 \ldots \mathbf{X}_M$. Условные распределения (18) и (19) совместно с априорным распределением классов позволяют вычислить среднюю взаимную информацию $I(\mathbf{X}_m; \Omega)$ для каждого источника $\mathbf{X}_m, m = 1, \ldots, M$, и среднюю взаимную информацию $I(\mathbf{X}^M; \Omega)$ для ансамбля источников $\mathbf{X}^M = \mathbf{X}_1 \ldots \mathbf{X}_M$ и, следовательно, получить численные реализации нижней границы вида (14) при значениях $M \geq 1$.

6. ЭКСПЕРИМЕНТАЛЬНЫЕ СООТНОШЕНИЯ “ВЗАИМНАЯ ИНФОРМАЦИЯ–ВЕРОЯТНОСТЬ ОШИБКИ” ДЛЯ МНОЖЕСТВА ПОДПИСЕЙ, МНОЖЕСТВА ЛИЦ И ДЛЯ АНСАМБЛЯ ЭТИХ ИСТОЧНИКОВ

В численном эксперименте использованы множества подписей и лиц, заданные полутоновыми изображениями размера 256 × 256 с 8-битовым кодированием элементов. Каждое изображение содержит один информативный объект (лицо или подпись). Множества лиц и подписей содержат по 1000 объектов от 25 персон ($c = 25$), по 40 реализаций от каждой персоны. Априорное распределение классов принято равномерным. Информативные объекты заданы древовидными представлениями в виде структурированных наборов эллиптических примитивов [9]. Различие любой пары объектов на множестве их древовидных представлений определяется расстоянием в метрике $L_p, p \geq 1$, которое является модификацией расстояния в метрике L_1, введенного в [9]. Цель вычислительного эксперимента состояла в получении реализаций границы $\underline{R}(\varepsilon)$ на множестве лиц и множестве подписей с использованием матриц расстояний, заданных ресурсами [10] и [11], а также в вычислении границы $\underline{R}_M(\varepsilon)$ на ансамбле этих источников.

Примеры древовидных представлений подписи и лица даны на фиг. 3. Информативные уровни $l = 1, \ldots, 8$ представлений содержат по 2^l эллиптических примитивов. Параметры примитива нулевого уровня используются для нормировки параметров примитивов последующих уровней. Нормированные примитивы всех уровней задаются в собственных координатных осях примитива нулевого уровня и имеют номера соответствующих им вершин бинарного дерева. Построение примитивов в собственных осях нулевого уровня и нормировка параметров примитивов обеспечивают инвариантность представлений к сдвигу, повороту, масштабу и уровню яркости объектов. Приведенные представления могут быть построены для любого объекта, заданного на изображении односвязным или многосвязным набором пикселей, который имеет идентифицируемые собственные оси.

В соответствии с выбранным способом представления, объект $\mathbf{x} \in \mathbf{X}$ преобразуется в набор эллиптических примитивов

$$\mathbf{x}^L = \{\mathrm{E}_n = (\mathbf{r}_n, \mathbf{u}_n, \mathbf{v}_n, z_n)\}, \tag{20}$$

образующих бинарное дерево глубины L. Глубина дерева определяется наибольшим уровнем разрешения; эллиптический примитив E_n соответствует n-й вершине дерева и определяется вектором центра $\mathbf{r}_n$, векторами большой и малой полуосей $\mathbf{u}_n$, $\mathbf{v}_n$ и средней яркостью z_n фрагмента, аппроксимируемого примитивом. Номера вершин формируются согласно правилу: каждая промежуточная вершина с номером n порождает пару вершин следующего уровня с номерами $2n + 1$,

$2n+2$; вершина нулевого уровня имеет номер $n=0$. При указанной нумерации примитив E_n с номером n находится в дереве на уровне $l=\lfloor \log_2(n+1)\rfloor$.

Используя представления вида (20), введем на множестве $\mathbf{X}$ расстояние в метрике L_p, $p\geq 1$, между любой парой объектов $\mathbf{x}\in\mathbf{X}$ и $\hat{\mathbf{x}}\in\hat{\mathbf{X}}$. Примитивы $\mathrm{E}_n\in\mathbf{x}^L$ и $\hat{\mathrm{E}}_n\in\hat{\mathbf{x}}^L$ с одинаковыми номерами будем считать соответственными, так что множество пар $(\mathrm{E}_n,\hat{\mathrm{E}}_n)$ образует пересечение $\mathbf{x}^L\cap\hat{\mathbf{x}}^L$. На пересечении любой пары представлений $\mathbf{x}^L$ и $\hat{\mathbf{x}}^L$ определим их p-степенные различия по векторам центров $(\mathbf{r}_n,\hat{\mathbf{r}}_n)$, векторам полуосей $(\mathbf{u}_n,\hat{\mathbf{u}}_n)$, $(\mathbf{v}_n,\hat{\mathbf{v}}_n)$ и уровням яркости$(z_n,\hat{z}_n)$:

$$\rho_r^p(\mathbf{x}^L,\hat{\mathbf{x}}^L)=\sum_{n\,:\,(\mathrm{E}_n,\hat{\mathrm{E}}_n)\in\mathbf{x}^L\cap\hat{\mathbf{x}}^L}\lambda_n\|\mathbf{r}_n-\hat{\mathbf{r}}_n\|^p,$$

$$\rho_{uv}^p(\mathbf{x}^L,\hat{\mathbf{x}}^L)=\sum_{n\,:\,(\mathrm{E}_n,\hat{\mathrm{E}}_n)\in\mathbf{x}^L\cap\hat{\mathbf{x}}^L}\lambda_n\left(\|\mathbf{u}_n-\hat{\mathbf{u}}_n\|^p+\|\mathbf{v}_n-\hat{\mathbf{v}}_n\|^p\right),$$

$$\rho_z^p(\mathbf{x}^L,\hat{\mathbf{x}}^L)=\sum_{n\,:\,(\mathrm{E}_n,\hat{\mathrm{E}}_n)\in\mathbf{x}^L\cap\hat{\mathbf{x}}^L}\lambda_n\|z_n-\hat{z}_n\|^p,$$

где

$$\lambda_n=\frac{\lfloor\log_2(n+1)\rfloor 2^{-\lfloor\log_2(n+1)\rfloor}}{\sum\limits_{n\,:\,(\mathrm{E}_n,\hat{\mathrm{E}}_n)\in\mathbf{x}^L\cap\hat{\mathbf{x}}^L}\lfloor\log_2(n+1)\rfloor 2^{-\lfloor\log_2(n+1)\rfloor}}$$

есть вес n-й пары соответственных примитивов. Введенные различия дают средние значения

$$\sigma_r^p(\hat{\mathbf{x}}^L)=\frac{1}{\|\mathbf{X}\|}\sum_{\mathbf{x}^L:\,\mathbf{x}\in\mathbf{X}}\rho_r^p(\mathbf{x}^L,\hat{\mathbf{x}}^L),$$

$$\sigma_{uv}^p(\hat{\mathbf{x}}^L)=\frac{1}{\|\mathbf{X}\|}\sum_{\mathbf{x}^L:\,\mathbf{x}\in\mathbf{X}}\rho_{uv}^p(\mathbf{x}^L,\hat{\mathbf{x}}^L),$$

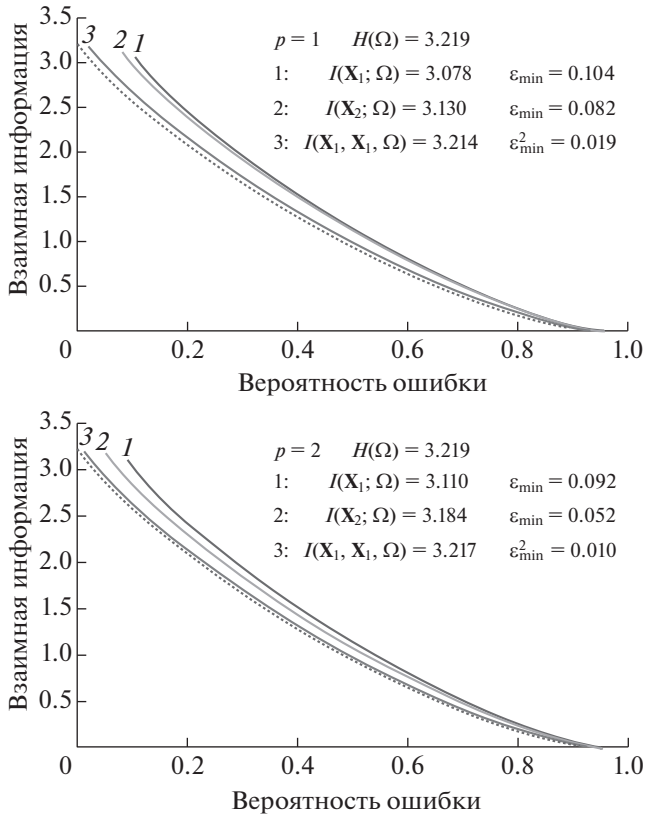
$$\sigma_z^p(\hat{\mathbf{x}}^L)=\frac{1}{\|\mathbf{X}\|}\sum_{\mathbf{x}^L:\,\mathbf{x}\in\mathbf{X}}\rho_z^p(\mathbf{x}^L,\hat{\mathbf{x}}^L),$$

на множестве объектов $\mathbf{X}$ относительно объекта $\hat{\mathbf{x}}\in\mathbf{X}$. Указанные характеристики позволяют ввести расстояние в метрике L_p, $p\geq 1$, между объектами $\mathbf{x}$ и $\hat{\mathbf{x}}$ в форме

$$d(\mathbf{x},\hat{\mathbf{x}})=\left(\frac{\rho_r^p(\mathbf{x}^L,\hat{\mathbf{x}}^L)}{\sigma_r^p(\hat{\mathbf{x}}^L)}+\frac{\rho_{uv}^p(\mathbf{x}^L,\hat{\mathbf{x}}^L)}{\sigma_{uv}^p(\hat{\mathbf{x}}^L)}+\frac{\rho_z^p(\mathbf{x}^L,\hat{\mathbf{x}}^L)}{\sigma_z^p(\hat{\mathbf{x}}^L)}\right)^{1/p}. \tag{21}$$

Расстояния вида (21) на множествах $\mathbf{X}_m$, $m=1,\ldots,M$, совместно с представителями классов (16) и оценками параметров (17) полностью определяют для указанных источников условные по классам распределения вида (18), а также условные по классам распределения вида (19) для ансамбля этих источников. В экспериментах использованы расстояния в метрике L_p с параметром $p=1$ и $p=2$.

Графики границ $\underline{R}_M(\varepsilon)$, вычисленных на множестве лиц $\mathbf{X}_1$ и множестве подписей $\mathbf{X}_2$, а также на ансамбле $\mathbf{X}_1\mathbf{X}_2$, даны на фиг. 4. Для сравнения приведена точечная кривая нижней границы Шеннона. Графики демонстрируют существенное уменьшение минимальной вероятности ошибки на ансамбле по сравнению с аналогичными вероятностями ошибки на множествах лиц или подписей. Необходимо отметить, что для рассматриваемых источников расстояние в метрике L_2 обеспечивает возможность достижения большей точности по сравнению с расстоянием в метрике L_1.



Фиг. 4. Реализации нижних границ функции “взаимная информация–вероятность ошибки” для множества лиц $\mathbf{X}_1$ (кривая *1*), множества подписей $\mathbf{X}_2$ (кривая *2*) и ансамбля $\mathbf{X}_1\mathbf{X}_2$ (кривая *3*).

7. ИЗБЫТОЧНОСТЬ ВЕРОЯТНОСТИ ОШИБКИ РЕШАЮЩЕГО АЛГОРИТМА

Полученная нижняя граница $\underline{R}_M(\varepsilon)$ позволяет оценить избыточность вероятности ошибки алгоритма, принимающего решения по блокам X^{MN} из $N \geq 1$ объектов ($M \geq 1$) с использованием набора разделяющих функций

$$\mathrm{G} = \left\{ g_j(\mathrm{X}^{MN}),\ \mathrm{X}^{MN} \in \mathbf{X}^{MN};\ j = 1,\ldots,c \right\}.$$

Разделяющие функции дают значения правдоподобия решений $\omega_j \in \hat{\Omega}$, $j = 1,\ldots,c$, по предъявляемому блоку X^{MN} и позволяют минимизировать вероятность ошибки при использовании решающего правила $j^* = \arg\max_{j=1}^{c} g_j(\mathrm{X}^{MN})$. В общем случае разделяющие функции отличаются от апостериорных вероятностей, поэтому реализуемая на них вероятность ошибки превосходит минимальную вероятность ошибки байесовского алгоритма.

Решающее правило на наборе разделяющих функций G порождает разбиение множества $\mathbf{X}^{MN}$ на непересекающиеся области решений $\mathbf{X}_j^{MN} \to \omega_j$, $j = 1,\ldots,c$, которые имеют вероятности

$$Q_{\mathrm{G}}(\omega_j) = \sum_{\mathrm{X}^{MN} \in \mathbf{X}_j^{MN}} P(\mathrm{X}^{MN}), \quad j = 1,\ldots,c.$$

Тогда набор G дает энтропию множества решений

$$H_{\mathrm{G}} = -\sum_{j=1}^{c} Q_{\mathrm{G}}(\omega_j) \ln Q_{\mathrm{G}}(\omega_j)$$

и вероятность ошибки

$$\varepsilon_{\mathrm{G}} = 1 - \sum_{\mathrm{X}^{MN} \in \mathbf{X}^{MN}} P(\mathrm{X}^{MN}) Q(\omega_{j^*} | \mathrm{X}^{MN}) [j^* = \arg\max_{j=1}^{c} g_j(\mathrm{X}^{MN})],$$

где

$$Q\left(\omega_j | \mathrm{X}^{MN}\right) = \frac{g_j(\mathrm{X}^{MN})}{\sum_{i=1}^{c} g_i(\mathrm{X}^{MN})}$$

есть оценка апостериорной вероятности решения ω_j по блоку составных объектов X^{MN}.

Согласно теореме кодирования [1] справедливо неравенство $H_{\mathrm{G}} \geq R_M(\varepsilon_{\mathrm{G}})$. Тогда пара H_{G}, ε_{G} позволяет ввести избыточность $r_{\mathrm{G}} = \varepsilon_{\mathrm{G}} - \varepsilon$ вероятности ошибки ε_{G} относительно значения

$$\varepsilon = \underline{R}_M^{-1}(H_{\mathrm{G}})[H_{\mathrm{G}} < I(\mathbf{X}^M; \Omega)] + \varepsilon_{\min}^M [H_{\mathrm{G}} \geq I(\mathbf{X}^M; \Omega)],$$

где $\underline{R}_M^{-1}(H_{\mathrm{G}})$ – значение обратной функции от нижней границы $\underline{R}_M(\varepsilon) = H_{\mathrm{G}}$.

В случае принятия решений по блокам объектов длины $N \geq 1$, наименьшая вероятность ошибки ε_{G} реализуется на байесовском алгоритме, разделяющие функции которого дают апостериорные вероятности $P(\omega_j | \mathrm{X}^{MN})$, $j = 1, \ldots, c$ (см. [8]). В этом случае $H_{\mathrm{G}} \geq I(\mathbf{X}^M; \Omega)$ и избыточность $r_{\mathrm{G}} = \varepsilon_{\mathrm{G}} - \varepsilon_{\min}^M \geq 0$. При оптимальной длине блоков $N = N^*$ байесовский алгоритм достигает минимальной вероятности ошибки $\varepsilon_{\min}^M$ и обеспечивает нулевую избыточность.

8. ЗАКЛЮЧЕНИЕ

Исследована теоретико-информационная модель классификации данных, для которой введено обменное соотношение между наименьшей средней взаимной информацией множества классифицируемых объектов относительно множества решений по классам и заданной допустимой вероятностью ошибки. Соотношение “взаимная информация–вероятность ошибки” определено для заданного источника данных и для ансамбля источников, и является аналогом известной в теории информации функции “скорость–погрешность” (rate distortion function) для модели кодирования с заданной точностью при наличии канала наблюдения с шумом. Построена нижняя граница функции “взаимная информация–вероятность ошибки”, которая является обобщением нижней границы Шеннона. Полученная граница зависит от априорного распределения классов и условных по классам распределений на заданном множестве объектов и не зависит от решающего алгоритма. При любом фиксированном значении средней взаимной информации нижняя граница функции “взаимная информация–вероятность ошибки” дает потенциально достижимую вероятность ошибки классификации по объектам источника или по составным объектам от ансамбля источников. Показана возможность оценивания избыточности вероятности ошибки относительно нижней границы для решающего алгоритма с заданным набором разделяющих функций. Предложенная методика вычисления избыточности позволяет оценить качество любых разделяющих функций, построенных независимо от априорных вероятностей классов и условных по классам распределений предъявляемых объектов.

СПИСОК ЛИТЕРАТУРЫ

1. *Gallager R.G.* Information Theory and Reliable Communication. New York: Wiley & Sons, Inc. 1968.
2. *Kuncheva L.I., Whitaker C.J., Shipp C.A., Duin R.P.W.* Limits on the majority vote accuracy in classifier fusion // Pattern Analysis and Applicat.2003. Vol. 6. P. 22–31. https://doi.org/10.1007/s10044-002-0173-7
3. *Lam L., Suen C.Y.* Application of majority voting to pattern recognition: An analysis of its behavior and performance // IEEE Transactions on Systems, Man, and Cybernetics. 1997. Vol. 27(5). P. 553–568. https://doi.org/10.1109/3468.618255

4. *MacKay D.J.C.* Information Theory, Inference, and Learning Algorithms. C.U.P. 2003.
5. *Dobrushin R.L., Tsybakov B.S.* Information transmission with additional noise // IRE Trans. Information Theory. 1962. Vol. 8(5). P. 293–304.
https://doi.org/10.1109/TIT.1962.1057738
6. *Lange M., Ganebnykh S., Lange A.* An Information Approach to Accuracy Comparison for Classification Schemes in an Ensemble of Data Sources // Communications in Comput. and Informat. Sci. CCIS: Springer. 2019. Vol. 794. P. 28–43.
https://doi.org/10.1007/978-3-030-35400-8_3
7. *Lange M.M., Ganebnykh S.N., Lange A.M.* On an information-theoretical lower bound to a classification error probability // Math. Methods for Pattern Recognition: Book of abstracts of the 19th Russian Conference. Moscow: MMPR-2019. P. 59–61.
8. *Duda R.O., Hart P.E., Stork D.G.* Pattern Classification, 2nd ed. New York: Wiley & Sons, Inc. 2001.
9. *Lange M.M., Ganebnykh S.N.* On fusion schemes for multiclass object classification with reject in a given ensemble of sources // J. of Physics: Conference Series. 2018. Vol. 1096. No. 012048.
https://doi.org/10.1088/1742-6596/1096/1/012048
10. Матрицы расстояний изображений подписей [Электронный ресурс]. Режим доступа: http://sourceforge.net/projects/distance-matrices-signature (Июнь, 2020).
11. Матрицы расстояний изображений лиц [Электронный ресурс]. Режим доступа: http://sourceforge.net/projects/distance-matrices-face (Июнь, 2020).

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1206–1219

ИНФОРМАТИКА

УДК 519.72

АППРОКСИМИРУЕМОСТЬ ЗАДАЧИ МАРШРУТИЗАЦИИ ТРАНСПОРТА С ОГРАНИЧЕННЫМ ЧИСЛОМ МАРШРУТОВ В МЕТРИЧЕСКИХ ПРОСТРАНСТВАХ ФИКСИРОВАННОЙ РАЗМЕРНОСТИ УДВОЕНИЯ[1)]

© 2021 г. Ю. Ю. Огородников[1,2,*], М. Ю. Хачай[1,2,3,]**

[1] 620990 Екатеринбург, ул. Софьи Ковалевской, 16, ФГБУ ИММ им. Н.Н. Красовского УрО РАН, Россия

[2] 620075 Екатернбург, пр-т Ленина, 51, Уральский федеральный ун-т, Россия

[3] 644050 Омск, пр-т Мира, 11, Омский гос. техн. ун-т, Россия

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

***e-mail: mkhachay@imm.uran.ru*

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Задача маршрутизации транспорта ограниченной грузоподъемности (Capacitated Vehicle Routing Problem, CVRP) – одна из классических проблем комбинаторной оптимизации, обладающая широким спектром важных практических приложений в исследовании операций. Как и большинство известных комбинаторных задач, CVRP NP-трудна в сильном смысле и сохраняет труднорешаемость даже на евклидовой плоскости. В метрической постановке задача CVRP APX-полна, что исключает ее аппроксимацию с произвольной заданной точностью в классе алгоритмов полиномиальной трудоемкости (в рамках гипотезы $P \neq NP$). В то же время для случая конечномерных евклидовых пространств подход, опирающийся на работы С. Ароры, А. Дас и К. Матье, позволил обосновать аппроксимируемость задачи в классе квазиполиномиальных и даже полиномиальных приближенных схем. В данной работе впервые удалось распространить этот подход на существенно более широкий класс метрических пространств с фиксированной размерностью удвоения. Показано, что задача CVRP, сформулированная в таком пространстве, обладает квазиполиномиальной приближенной схемой каждый раз, когда число маршрутов в ее оптимальном решении ограничено сверху полиномом от логарифма длины записи условия задачи. Библ. 37.

Ключевые слова: задача маршрутизации транспорта с ограничением на грузоподъемность (CVRP), квазиполиномиальная приближенная схема (QPTAS), метрическое пространство, размерность удвоения.

DOI: 10.31857/S0044466921070140

ВВЕДЕНИЕ

Задача маршрутизации транспортных средств ограниченной грузоподъемности (Capacitated Vehicle Routing Problem, CVRP) является одной из наиболее известных и активно изучаемых комбинаторных задач, обладающей широким спектром значимых приложений в области исследования операций (см. [1]–[3]). По-видимому, впервые постановка задачи маршрутизации была сформулирована Г. Данцигом и Дж. Рамсером в классической работе [4], посвященной математическому моделированию процесса снабжением топливом сети заправочных станций.

Как и для большинства современных задач комбинаторной оптимизации, исследования в области алгоритмического анализа CVRP традиционно развиваются в рамках следующих основных направлений. Первое направление основано на редукции исходной задачи к подходящей постановке задачи целочисленного (смешанного) программирования с последующим поиском оптимального решения последней с помощью той или иной модификации метода ветвей и границ (см. обзор в [3]). К сожалению, несмотря на стремительные темпы развития вычислитель-

[1)]Работа выполнена в рамках исследований, проводимых в Уральском математическом центре при финансовой поддержке Минобрнауки России и при финансовой поддержке РФФИ (код проекта 19-07-01243).

ной техники и очевидные успехи последних лет в области совершенствования алгоритмов (см., например, [5]–[8]), практическая применимость данного подхода по-прежнему ограничена постановками достаточно скромного размера ввиду известной NP-трудности задачи CVRP.

Широкий спектр современных эвристических алгоритмов и метаэвристик составляет основу второго направления. Наибольшего успеха в области эффективной аппроксимируемости задачи удалось достичь в классах методов локального поиска (см. [9], [10]), поиска с запретами (см. [11]), переменных окрестностей (VNS) (см. [12], [13]), методов машинного обучения (см. [14]), эволюционных (см. [15]) и биоинспирированных алгоритмов (см. [16], [17]), а также их комбинаций (см. [18], [19]). Нередко эвристические алгоритмы демонстрируют потрясающую производительность, эффективно находя близкие к оптимальным или даже оптимальные решения для отдельных постановок CVRP чрезвычайно большого размера. Тем не менее в отсутствие теоретически обоснованных гарантий применение этих алгоритмов сопряжено с дополнительными трудозатратами, связанными с численным оцениванием их точности и возможной дополнительной настройкой внутренних параметров при переходе к каждому новому классу постановок.

Перечисленные аргументы подтверждают актуальность третьего направления, связанного с аппроксимируемостью задачи в классе эффективных алгоритмов с теоретическими оценками точности и трудоемкости. Как известно, задача CVRP NP-трудна в сильном смысле, являясь обобщением классической задачи коммивояжера (TSP), и сохраняет труднорешаемость (при условии, что грузоподъемность q является частью входа) даже на евклидовой плоскости (см. [20]). Задача не аппроксимируема в общем случае (при $P \neq NP$), APX-полна при произвольной метрике (см. [21], [22]) и остается APX-трудной при произвольной фиксированной грузоподъемности $q \geq 3$.

Наибольших успехов в области аппроксимируемости задачи CVRP удалось достичь в конечномерных числовых пространствах. Известные результаты в этой области восходят к классическим работам М. Хаймовича, А. Ринной Кана (см. [22]) и С. Ароры (см. [23]). На данный момент наиболее общим результатом для задачи CVRP на евклидовой плоскости является квазиполиномиальная приближенная схема (QPTAS) А. Дас и К. Матье (см. [24]). Вводя ограничение на рост грузоподъемности q, удается обосновать и полиномиальные приближенные схемы (PTAS), среди которых рекордным на данный момент является алгоритм (см. [25]), позволяющий за полиномиальное время найти $(1+\varepsilon)$-приближенное решение задачи при условии $q \leq 2^{\log^{\delta(\varepsilon)} n}$. Подход, предложенный в [22], удалось распространить на модификации задачи в числовых пространствах произвольной фиксированной размерности (см. [26]–[28]), учитывающие дополнительные ограничения на временные промежутки обслуживания (см. [29], [30]) и неоднородность спроса (см. [31]).

Так или иначе до последнего времени полиномиальные и квазиполиномиальные приближенные схемы удавалось обосновать лишь для геометрических постановок задачи CVRP, за исключением, быть может, немногочисленных специальных случаев, описанных в [32], [33]. Долгое время до появления пионерских работ К. Талвара (см. [34]) и Я. Бартала и соавт. (см. [35]), аналогичным образом складывалась ситуация и с аппроксимируемостью близкой к CVRP задачи коммивояжера. Предложенный в этих работах подход позволил распространить классический результат С. Ароры (см. [23]) на существенно более широкий класс постановок задачи коммивояжера, задаваемых в метрических пространствах произвольной фиксированной размерности удвоения (о существовании полиномиальных приближенных схем (PTAS) для TSP в $\mathbb{R}^d$).

В данной статье в развитие близкого подхода к подходу А. Дас и К. Матье впервые обосновывается возможность построения квазиполиномиальной схемы для задачи CVRP, сформулированной в метрическом пространстве произвольной фиксированной размерности удвоения. Рассуждения проведены для частого случая задачи, стесненной дополнительным ограничением на число маршрутов в ее оптимальном решении.

1. ПОСТАНОВКИ ЗАДАЧ

Содержательная постановка задачи CVRP может быть задана следующим образом. Имеется множество потребителей X, каждый из которых обладает единичным спросом на однородную продукцию, хранящуюся на складе y. На складе базируются идентичные транспортные средства фиксированной грузоподъемности q, используемые для удовлетворения потребительского спроса. Задача состоит в построении набора циклических маршрутов, удовлетворяющего сово-

купный потребительский спрос и минимизирующего суммарные транспортные издержки так, чтобы каждый из маршрутов начинался и завершался на складе y и удовлетворял ограничению нагрузоподъемность q.

Для последующих построений наряду с классической постановкой задачи CVRP нам потребуется ее обобщение, известное в литературе как задача маршрутизации транспорта с ограниченной грузоподъемностью и неоднородным разделяемым спросом (CVRP-SD).

Задача CVRP-SD. Пусть заданы полный взвешенный граф $G = (X \cup \{y\}, E, D, w)$ и натуральное число q. Здесь $X = \{x_1 \ldots, x_n\}$ – множество потребителей, y – склад, неотрицательная весовая функция $D\colon X \to \mathbb{Z}_+$ задает объем спроса каждого из потребителей, симметричная весовая функция $w\colon E \to \mathbb{R}_+$ сопоставляет произвольной паре вершин $\{u, v\} \subset X \cup \{y\}$ транспортные издержки $w(u, v)$, связанные с непосредственной перевозкой по ребру $\{u, v\} \in E$, а q – верхняя оценка грузоподъемности используемых транспортных средств.

Маршрутом называется упорядоченная пара $\mathcal{R} = (\pi, S_{\mathcal{R}})$, в которой $\pi = y, x_{i_1}, \ldots, x_{i_t}, y$ – (не обязательно простой) цикл в графе G, а функция $S_{\mathcal{R}}\colon X \to \mathbb{Z}_+$ задает распределение спроса потребителей, удовлетворяемого данным маршрутом.

Стоимость (вес) маршрута $\mathcal{R}$ определяется выражением

$$w(\mathcal{R}) = w(y, x_{i_1}) + w(x_{i_1}, x_{i_2}) + \cdots + w(x_{i_{t-1}}, x_{i_t}) + w(x_{i_t}, y).$$

Маршрут $\mathcal{R} = (\pi, S_{\mathcal{R}})$ называется *допустимым*, если

$$S_{\mathcal{R}}(x)\begin{cases} \le D(x), & x \in \{x_{i_1}, \ldots, x_{i_t}\}, \\ = 0, & \text{в противном случае,} \end{cases} \quad \text{и} \quad \sum_{x \in X} S_{\mathcal{R}}(x) \le q.$$

Задача CVRP-SD заключается в построении семейства $\mathfrak{S}$ допустимых маршрутов минимальной суммарной стоимости, удовлетворяющего совокупный потребительский спрос:

$$\begin{aligned} w(\mathfrak{S}) &\equiv \sum_{\mathcal{R} \in \mathfrak{S}} w(\mathcal{R}) \to \min, \\ \sum_{\mathcal{R} \in \mathfrak{S}} S_{\mathcal{R}}(x) &= D(x), \quad x \in X. \end{aligned} \tag{1}$$

Постановка классической задачи CVRP легко может быть получена из постановки (1) задачи CVRP-SD введением дополнительного ограничения $D(x) \equiv 1$.

Если функция w удовлетворяет неравенству треугольника, т.е. для произвольного подмножества $\{v_1, v_2, v_3\} \subset X \cup \{y\}$ справедливо соотношение $w(v_1, v_2) \le w(v_1, v_3) + w(v_3, v_2)$, то вершины графа G принято называть *точками*, величину $w(u, v)$ – *расстоянием* между точками u и v, стоимость $w(\mathcal{R})$ произвольного маршрута $\mathcal{R}$ – его *длиной*, а соответствующую постановку задачи – *метрической*.

В данной работе мы ограничимся рассмотрением исключительно метрических постановок задачи CVRP. Более того, задавшись произвольным натуральным числом $d > 1$, мы потребуем, чтобы каждая рассматриваемая постановка обладала следующими свойствами:

1) пара (Z, ρ), в которой $Z = X \cup \{y\}$, а $\rho|_E \equiv w$, является конечным метрическим пространством размерности удвоения d;

2) число маршрутов хотя бы одного из оптимальных решений задачи ограничено сверху величиной $\mathrm{polylog}(n)$, являющейся полиномом от логарифма размера входных данных.

Всюду ниже мы договоримся не различать весовую функцию w и порождаемую ей метрику ρ и использовать обозначения $\mathrm{CVRP}(Z, w, q)$ и $\mathrm{CVRP}^*(Z, w, q)$ для постановки задачи CVRP, задаваемой графом $G = (X \cup \{y\}, E, w)$ и грузоподъемностью q, и ее оптимального значения соответственно. Для задачи CVRP-SD вводим обозначения $\mathrm{CVRP\text{-}SD}(Z, D, w, q)$ и $\mathrm{CVRP\text{-}SD}^*(Z, D, w, q)$ по аналогии.

Определение 1. *Квазиполиномиальной приближенной схемой* (QPTAS) для комбинаторной задачи минимизации называется параметризованное семейство алгоритмов, содержащее для каждого фиксированного значения $\varepsilon > 0$ алгоритм, находящий для произвольной постановки задачи $(1 + \varepsilon)$-приближенное решение за время, ограниченное сверху квазиполиномом $O(n^{\mathrm{poly}\log n})$ от размера записи ее условия (здесь и ниже мы следуем известной RAM-модели вычислений, пред-

полагающей константную трудоемкость произвольной элементарной операции над вещественными числами).

Основным результатом данной работы является квазиполиномиальная приближенная схема для рассматриваемой постановки задачи.

2. МЕТРИЧЕСКИЕ ПРОСТРАНСТВА ФИКСИРОВАННОЙ РАЗМЕРНОСТИ УДВОЕНИЯ

Нам потребуется несколько известных понятий и вспомогательных результатов.

Определение 2. *Метрическим шаром* с центром $z_0 \in Z$ радиуса R называется множество $B(z_0, R) = \{z \in Z : \rho(z_0, z) \le R\}$.

Определение 3 (см., например, [36]). Говорят, что метрическое пространство (Z, ρ) обладает *размерностью удвоения* d, если для произвольных $z_0 \in Z$ и $R > 0$ найдется число $M \le 2^d$ и точки $z_1, \ldots, z_M \in Z$ такие, что $B(z_0, R) \subseteq \bigcup_{j=1}^{M} B(z_j, R/2)$.

Нетрудно убедиться, что произвольное пространство l_p^d обладает размерностью удвоения $O(d)$. Однако известно, что класс метрических пространств конечной размерности существенно шире семейства конечномерных числовых пространств (см., например, [37]).

Пусть $Z' \subset Z$ произвольное подпространство пространства Z размерности удвоения d. Через $\Delta = \Delta_\rho(Z') = \sup\{\rho(u, v) : u, v \in Z'\}$ и $\alpha = \alpha_\rho(Z') = \inf\{\rho(u, v) : \{u, v\} \subset Z'\}$ обозначим верхнюю и нижнюю грань попарных расстояний между элементами Z'.

Лемма 1 (см. [34]). *Пусть* $0 < \alpha \le \Delta < \infty$. *Подпространство* Z' *конечно*,

$$|Z'| \le \left(\frac{2\Delta}{\alpha}\right)^d.$$

В данной статье мы ограничиваемся исключительно рассмотрением конечных метрических пространств (Z, ρ), порождаемых полными взвешенными графами $G = (Z, E, w)$. Пусть $U \subseteq Z$ – произвольное непустое подмножество вершин графа G, $\mathrm{MST}(U)$ – остовное дерево минимального веса подграфа $G\langle U \rangle$, индуцированного подмножеством U, и $R = R(U)$ – радиус минимального объемлющего U метрического шара $B(z, R)$ с центром в подходящей точке $z \in Z$.

Нам потребуется следующая техническая лемма (см., например, [34]), которую для полноты изложения мы приводим с доказательством.

Лемма 1. *Пусть* (Z, ρ) – *конечное метрическое пространство размерности удвоения* $d > 1$. *Для произвольного непустого подмножества* $U \subseteq Z$ *радиуса* R

$$w(\mathrm{MST}(U)) \le 12R|U|^{1-1/d}. \tag{2}$$

Доказательство. Для произвольного $\varnothing \ne V \subseteq U$ обоснуем соотношение

$$w(\mathrm{MST}(V)) \le 12R(V)(|V|^{1-1/d} - 1). \tag{3}$$

Доказательство ведем индукцией по $|V|$.

База: $|V| \le 2$. При $|V| = 1$ неравенство (3), очевидно, справедливо, так как

$$R(V) = w(\mathrm{MST}(V)) = 0.$$

Рассмотрим случай $V = \{u, v\}$. Поскольку $2^{1-1/d} - 1 \ge \sqrt{2} - 1 > 0.4$ при произвольном $d \ge 2$,

$$w(\mathrm{MST}(V)) = \rho(u, v) \le 2R(V) < 4.8R(V) < 12R(V)(|V|^{1-1/d} - 1).$$

Шаг индукции: Пусть $|V| \ge 3$. По предположению индукции соотношение (3) верно для произвольного непустого подмножества $V' \subset U$, $|V'| < |V|$. Обоснуем справедливость неравенства (3) для подмножества V.

Пусть B – минимальный метрический шар (радиуса $R(V)$), содержащий подмножество V. По условию для некоторого $l \le 2^d$ найдутся шары $B_1,\ldots,B_l \subset Z$ радиуса $R(V)/2$ такие, что $\bigcup_{j=1}^{l} B_j \supseteq B$. Введя обозначение $V_j = B_j \cap V$, без ограничения общности полагаем $V_i \neq \varnothing$ и $V_j \cap V_k = \varnothing$ для произвольных i и $j \neq k$.

Кроме того, мы всегда можем полагать $l \ge 3$. В самом деле, $l > 1$ в силу минимальности шара B. Поскольку $|V| \ge 3$, при $l = 2$ хотя бы одно из подмножеств V_1 или V_2, например, V_1 не является синглетоном и может быть дополнительно разбито на два непустых подмножества.

Таким образом, построено разбиение множества V не менее чем на три непустых дизъюнктных подмножества $V_1,\ldots,V_l$. По построению для каждого элемента разбиения V_j справедливо неравенство $R(V_j) \le R(V)/2$. Следовательно,

$$w(\mathrm{MST}(V_j)) \le 12R(V_j)(|V_j|^{1-1/d} - 1) \le 6R(V)(|V_j|^{1-1/d} - 1)$$

по предположению индукции.

Зафиксируем произвольную систему представителей $H = \{v_j \in V_j : j \in \{1,2,\ldots,l\}\}$ и рассмотрим дерево

$$T = \mathrm{MST}(H) \cup \bigcup_{j=1}^{l} \mathrm{MST}(V_j).$$

По построению $w(\mathrm{MST}(H)) \le 2R(V)(l-1)$, поскольку $\Delta(V) \le 2R(V)$. Объединяя оценки, имеем

$$w(T) = \sum_{j=1}^{l} w(\mathrm{MST}(V_j)) + w(\mathrm{MST}(H)) < 6R(V)\sum_{j=1}^{l}(|V_j|^{1-1/d} - 1) + 2lR(V) \le 6R(V)\sum_{j=1}^{l}|V_j|^{1-1/d} -$$

$$- 4lR(V) \le 6R(V)\sum_{j=1}^{l}\left(\frac{|V|}{l}\right)^{1-1/d} - 4lR(V) = 6R(V)l^{1/d}|V|^{1-1/d} - 4lR(V) \le 12R(V)|V|^{1-1/d} - 12R(V).$$

Последнее неравенство следует из условий $l \le 2^d$ и $l \ge 3$. Индуктивный переход обоснован. Для завершения доказательства леммы достаточно рассмотреть случай $V = U$.

3. АППРОКСИМАЦИОННАЯ СХЕМА ДЛЯ CVRP

Приведенные в данном разделе рассуждения развивают известный подход С. Ароры (см. [23]) к построению PTAS для евклидовой задачи коммивояжера и его обобщение (см. [35], [34]) на случай метрических пространств фиксированной размерности удвоения. Структура предлагаемой аппроксимационной схемы состоит из нескольких стадий и представима в виде приведенной ниже последовательности.

1. **Стадия предварительной обработки и округления**, в рамках которой для заданного $\varepsilon > 0$ исходной задаче сопоставляется вспомогательная постановка более простой структуры так, что произвольное $(1+\varepsilon)$-приближенное решение полученной *округленной* задачи соответствует подходящему $(1 + O(\varepsilon))$-приближенному решению исходной.

2. **Стадия рандомизированной иерархической кластеризации**, в рамках которой для заданных случайных значений параметров строится совокупность вложенных друг в друга разбиений множества $X \cup \{y\}$. В кластере произвольного уровня выбирается подмножество специальных точек-*порталов*. Далее, следуя подходу [35], мы показываем, что для поиска искомого приближенного решения округленной задачи достаточно ограничиться семействами так называемых *сетевых* маршрутов, пересекающих границу произвольного кластера ограниченное число раз, причем исключительно в порталах.

3. **Стадия поиска семейства сетевых маршрутов минимального веса**. Методом динамического программирования построенной случайной реализации иерархической кластеризации сопоставляется допустимое решение, состоящее из сетевых маршрутов минимальной суммарной стоимости.

4. **Стадия дерандомизации**. Показывается, что по аналогии со схемой К. Талвара (см. [34]), предложенный рандомизированный алгоритм может быть эффективно дерандомизирован.

3.1. Предварительная обработка и округление

Первая стадия алгоритма развивает подход к преобразованию исходной задачи, впервые реализованный С. Аророй при построении PTAS для евклидовой задачи коммивояжера (см. [23]), и состоит из двух этапов.

Пусть как и ранее $\Delta = \Delta_w(Z) = \max\{w(u,v)\colon u,v \in Z = X \cup \{y\}\}$ – диаметр множества Z. Без ограничения общности полагаем выполненным соотношение $\Delta = n/\varepsilon$, поскольку в противном случае исходной задаче CVRP(Z,E,w) всегда может быть сопоставлена эквивалентная (с точки зрения совпадения оптимальных множеств) постановка CVRP(Z,E,w') с весовой функцией

$$w'(u,v) = w(u,v)\frac{n}{\varepsilon\cdot\Delta},$$

обладающая этим свойством.

Постановку искомой *округленной* задачи удобно описывать в терминах метрических сетей.

Определение 4. Пусть Z' – произвольное непустое подпространство метрического пространства (Z,ρ). Для заданного $\delta > 0$ подмножество $N \subseteq Z'$ называется δ-*сетью подпространства* Z', если

1) для произвольного $u \in Z'$ найдется такой элемент $v = v(u) \in N$, что $\rho(u,v) \le \delta$;

2) для произвольного подмножества $\{v_1,v_2\} \subset N$ расстояние $\rho(v_1,v_2) > \delta$.

Пусть $N_1' = \{\xi_1,\ldots,\xi_J\}$ – 1-сеть множества X. Положив $N_1 = N_1' \cup \{y\}$, сопоставим исходной задаче CVRP(Z,w,q) округленную постановку CVRP-SD(N_1,D,w_1,q) по следующему правилу:

1) произвольным образом разрешая возможную неоднозначность, построим отображение $\xi\colon X \to N_1'$ так, чтобы неравенство $w(x,\xi(x)) \le 1$ выполнялось для произвольного $x \in X$;

2) спрос произвольного узла сети $\xi_j \in N_1'$ определим соотношением $D(\xi_j) = |\xi^{-1}(\xi_j)|$;

3) в качестве весовой функции w_1 выберем сужение функции w на сеть N_1.

Справедливы следующие соотношения, связывающие оптимальные значения рассматриваемых задач.

Лемма 2.

$$\text{CVRP}^*(Z,w,q) - 2n \le \text{CVRP-SD}^*(N_1,D,w_1,q) \le \text{CVRP}^*(Z,w,q) + 2n.$$

Доказательство. 1. Для обоснования верхней оценки рассмотрим произвольное оптимальное решение $\mathfrak{S} = \{\mathcal{R}\}$ задачи CVRP(Z,w,q). Каждому маршруту $\mathcal{R} = (\pi, S_{\mathcal{R}}) \in \mathfrak{S}$, $\pi = y, x_{i_1},\ldots,x_{i_t},y$, сопоставим маршрут $\overline{\mathcal{R}} = (\overline{\pi}, S_{\overline{\mathcal{R}}})$, в котором $\overline{\pi} = y,\xi(x_{i_1}),\ldots,\xi(x_{i_t}),y$, а распределение $S_{\overline{\mathcal{R}}}\colon N_1' \to \mathbb{Z}_+$ определяется соотношением

$$S_{\overline{\mathcal{R}}}(\xi_j) = \sum_{\xi(x)=\xi_j} S_{\mathcal{R}}(x).$$

Полученное в результате семейство маршрутов $\overline{\mathfrak{S}} = \{\overline{\mathcal{R}}\}$, очевидно, является допустимым решением округленной задачи CVRP-SD(N_1,D,w_1,q). Оценим его вес $w_1(\overline{\mathfrak{S}}) = \sum_{\overline{\mathcal{R}}\in\overline{\mathfrak{S}}} w_1(\overline{\mathcal{R}})$. По выбору отображений w_1, ξ и неравенству треугольника

$$w_1(\overline{\mathcal{R}}) = w_1(y,\xi(x_{i_1})) + \sum_{j=1}^{t} w_1(\xi(x_{i_j}),\xi(x_{i_{j+1}})) + w_1(\xi(x_{i_t}),y) \le$$

$$\le w(y,x_{i_1}) + \sum_{j=1}^{t} w(x_{i_j},x_{i_{j+1}}) + w(x_{i_t},y) + 2\sum_{j=1}^{t} w(x_{i_j},\xi(x_{i_j})) \le w(\mathcal{R}) + 2t.$$

Следовательно,

$$\text{CVRP-SD}^*(N_1,D,w_1,q) \le w_1(\overline{\mathfrak{S}}) = \sum w_1(\overline{\mathcal{R}}) \le \sum w(\mathcal{R}) + 2n =$$
$$= w(\mathfrak{S}) + 2n = \text{CVRP}^*(Z,w,q) + 2n,$$

поскольку произвольный потребитель $x \in X$ посещается в точности одним из маршрутов $\mathcal{R}$ оптимального решения $\mathfrak{S}$.

2. Обоснование нижней оценки проведем по аналогии. Зафиксируем произвольное оптимальное решение $\overline{\mathfrak{S}} = \{\overline{\mathcal{R}}_1,\ldots,\overline{\mathcal{R}}_K\}$ задачи CVRP-SD(N_1, D, w_1, q). По определению объем спроса, удовлетворяемый маршрутом $\overline{\mathcal{R}} = (\overline{\pi}, S_{\overline{\mathcal{R}}}) \in \overline{\mathfrak{S}}$ в произвольном узле ξ сети N_1', определяется соотношением $S_{\overline{\mathcal{R}}}(\xi)$ так, что

$$\sum_{\xi\in N_1'} S_{\overline{\mathcal{R}}}(\xi) \le q.$$

Произвольному $j \in \{1,2,\ldots,J\}$ сопоставим отображение $\eta_j : X_j \to \overline{\mathfrak{S}}$ так, что

$$\left|\eta_j^{-1}(\overline{\mathcal{R}})\right| = S_{\overline{\mathcal{R}}}(\xi_j), \quad 1 \le j \le J, \quad \overline{\mathcal{R}} \in \overline{\mathfrak{S}}.$$

Далее, произвольному маршруту $\overline{\mathcal{R}} = (\overline{\pi}, S_{\overline{\mathcal{R}}}) \in \overline{\mathfrak{S}}$, $\overline{\pi} = y,\xi_{j_1},\ldots,\xi_{j_t},y$, сопоставим маршрут $\mathcal{R} = \mathcal{R}(\overline{\mathcal{R}})$, начинающийся (и заканчивающийся) на складе y и последовательно обходящий вершины подмножеств $\eta_{j_1}^{-1}(\overline{\mathcal{R}}),\ldots,\eta_{j_t}^{-1}(\overline{\mathcal{R}})$. Полученное в результате семейство маршрутов $\mathfrak{S} = \{\mathcal{R}_1,\ldots,\mathcal{R}_K\}$, очевидно, является допустимым решением задачи CVRP(Z, w, b). Оценим его стоимость. По построению сети N_1' и в силу неравенства треугольника

$$w(\mathcal{R}_k) \le w_1(\overline{\mathcal{R}}_k) + 2\sum_{i=1}^{t} S_{\overline{\mathcal{R}}_k}(\xi_{j_i}),$$

следовательно,

$$w(\mathfrak{S}) = \sum_{k=1}^{K} w(\mathcal{R}) \le \sum_{k=1}^{K} w_1(\overline{\mathcal{R}}_k) + \sum_{k=1}^{K}\sum_{j=1}^{J} S_{\overline{\mathcal{R}}_k}(\xi_j) = w_1(\overline{\mathfrak{S}}) + 2n, \tag{4}$$

откуда

$$\text{CVRP}^*(Z, w, q) \le w(\mathfrak{S}) \le \text{CVRP-SD}^*(N_1, D, w_1, q) + 2n.$$

Лемма доказана.

Заметим, что процедуры построения сети N_1', сопоставления исходной задаче CVRP(Z, w, q) округленной постановки CVRP-SD(N_1, D, w_1, q), а также восстановления решения $\mathfrak{S}$, соответствующего решению $\overline{\mathfrak{S}}$, очевидно, могут быть проведены за время, ограниченное сверху полиномом от n.

В качестве простого следствия убедимся в том, что произвольное приближенное решение задачи CVRP-SD(N_1, D, w_1, q) соответствует подходящему приближенному решению исходной задачи CVRP(Z, w, q). В самом деле, пусть $\overline{\mathfrak{S}}$ – произвольное $(1+\varepsilon)$-приближенное решение задачи CVRP-SD(N_1, D, w_1, q). Применяя подход, описанный в доказательстве п. 2 леммы 3, сопоставим решению $\overline{\mathfrak{S}}$ семейство маршрутов $\mathfrak{S}$.

Следствие 1. Семейство $\mathfrak{S}$ является $(1 + O(\varepsilon))$-приближенным решением CVRP(Z, w, q).

Доказательство. По построению $\mathfrak{S}$ – допустимое решение задачи CVRP(Z, w, q), стоимость $w(\mathfrak{S})$ которого определяется соотношением (4). Учитывая

$$w_1(\overline{\mathfrak{S}}) \le (1+\varepsilon)\text{CVRP-SD}^*(N_1, D, w_1, q) \quad \text{и} \quad \Delta_w(Z) = n/\varepsilon,$$

имеем

$$w(\mathfrak{S}) \le (1+\varepsilon)\text{CVRP-SD}^*(N_1, D, w_1, q) + 2n \le (1+\varepsilon)(\text{CVRP}^*(Z, w, q) + 2n) + 2n \le$$
$$\le (1+\varepsilon)\text{CVRP}^*(Z, w, q) + 2\Delta_w(Z)\varepsilon(2+\varepsilon) = (1 + O(\varepsilon))\text{CVRP}^*(Z, w, q),$$

так как неравенство треугольника, очевидно, влечет $2\Delta_w(Z) \le \text{CVRP}^*(Z, w, q)$.

Результаты данного раздела свидетельствуют о том, что при поиске $(1 + O(\varepsilon))$-приближенного решения метрической задачи CVRP без ограничения общности можно полагать, что диаметр $\Delta_w(Z) = n/\varepsilon$, расстояние между произвольными попарно различными потребителями u и v удовлетворяет соотношению $w(u, v) = \rho(u, v) > 1$, а объем спроса произвольного потребителя измеряется подходящим натуральным числом.

3.2. Рандомизированная иерархическая кластеризация

Задавшись произвольным значением параметра $s \geq 6$, положим $L = \lceil \log_s \Delta_w(Z) \rceil = O(\log n - \log \varepsilon)$. На каждом уровне $l = 0,1,\ldots,L+1$ зафиксируем произвольную s^{L-l}-сеть $N(l)$ множества Z. По выбору L сеть $N(0)$ – синглетон. Без ограничения общности полагаем $N(l) \subset N(l+1)$ при произвольном l и $N(L+1) = Z$.

Следуя подходу, предложенному в [34], последовательно строим рандомизированную иерархическую кластеризацию множества Z индукцией по $l = 0,1,\ldots,L+1$. На уровне $l = 0$ имеем единственный кластер C_1^0.

Пусть кластеризация реализована на всех уровнях, включая $l \leq L$, и, в частности, на уровне l разбиение Z имеет вид $Z = C_1^l \cup \ldots \cup C_K^l$. Кластеризацию на уровне $l+1$ строим путем разбиения каждого кластера C_j^l в отдельности. Задавшись на множестве $\{h_1,\ldots,h_{t_{l+1}}\}$ элементов s^{L-l-1}-сети $N(l+1)$ случайным порядком σ, произвольному $h_{\sigma(i)} \in N(l+1)$ сопоставляем подмножество

$$C_{ji}^{l+1} = B(h_{\sigma(i)}, \mu \cdot s^{L-l-1}) \cap C_j^l \setminus \bigcup_{k=1}^{i-1} C_{jk}^{l+1},$$

где μ – случайная величина, равномерно распределенная на $[1,2)$, после чего формируем искомую кластеризацию Z на уровне $l+1$ из непустых подмножеств C_{ji}^{l+1}.

В силу нашего допущения кластеры уровня $L+1$ – одноэлементные, а общее число кластеров на всех уровнях не превосходит $(n+1)(L+1) = O(n \log n)$ (при произвольном фиксированном $\varepsilon > 0$).

Как будет показано ниже, при поиске $(1+\varepsilon)$-приближенного решения исследуемой задачи можно ограничиться решениями, состоящими из так называемых *сетевых* маршрутов, пересекающих границу произвольного кластера не слишком часто, причем исключительно в специально заданных точках, именуемых *порталами*. Нам понадобится несколько определений и вспомогательных утверждений, доказательства которых легко могут быть получены из результатов [35].

Определение 5. Маршрут $\mathcal{R} = (\pi, S_{\mathcal{R}})$ называется *сетевым* по отношению к иерархии сетей $N(l)$, $l = 0,1,\ldots,L+1$, при заданном $\varepsilon > 0$, если для произвольного ребра $\{u,v\}$ длины $\rho(u,v)$ цикла π неравенство

$$s^{L-l} \leq \varepsilon \cdot \rho(u,v) < s^{L-l+1}$$

влечет принадлежность обеих вершин u и v сети $N(l)$.

Определение 6. Пусть M – степень s, удовлетворяющая соотношению

$$\frac{M}{s} \leq \frac{dL}{\varepsilon} < M. \tag{5}$$

Порталом кластера C_j^l называется произвольная принадлежащая ему точка – представитель s^{L-l}/M-сети.

Пусть C_j^l – произвольный кластер уровня $l > 0$. Будем говорить, что маршрут $\mathcal{R} = (\pi, S_{\mathcal{R}})$ пересекает границу кластера C_j^l, если цикл π содержит ребро $\{u,v\}$ такое, что $|\{u,v\} \cap C_j^l| = 1$.

Определение 7. Маршрут $\mathcal{R}$ называется *r-легким*, если для произвольного кластера C_j^l уровня $l > 0$ число пересечений его границы маршрутом $\mathcal{R}$ не превосходит r.

Лемма 3. *Пусть* $\varepsilon \in (0,1/8)$. *Произвольному маршруту* $\mathcal{R} = (\pi, S_{\mathcal{R}})$ *может быть сопоставлен подходящий сетевой маршрут* $\tilde{\mathcal{R}} = (\tilde{\pi}, S_{\tilde{\mathcal{R}}})$, $S_{\tilde{\mathcal{R}}} = S_{\mathcal{R}}$, *стоимость которого удовлетворяет соотношению*

$$w(\tilde{\mathcal{R}}) \leq (1+16\varepsilon) w(\mathcal{R}). \tag{6}$$

Лемма 4 позволяет произвольному допустимому решению исходной задачи сопоставить близкое по стоимости решение, целиком состоящее из сетевых маршрутов. Из утверждения следующей леммы следует, что для произвольного $r \geq 2$ без ограничения общности можно полагать, что все маршруты полученного в результате решения являются r-легкими.

Лемма 4. *Пусть $r \geq 2$ и $C' \subset C_j^l$, $|C'| > \overline{r} > r$, – множество точек пересечения границы кластера C_j^l некоторым маршрутом $\mathcal{R} = (\pi, S_{\mathcal{R}})$. Найдется сетевой маршрут $\tilde{\mathcal{R}} = (\tilde{\pi}, S_{\tilde{\mathcal{R}}})$, $S_{\tilde{\mathcal{R}}} = S_{\mathcal{R}}$, пересекающий границу кластера C_j^l дважды, стоимость которого*

$$w(\tilde{\mathcal{R}}) \leq w(\mathcal{R}) + 4w(\mathrm{MST}(C')). \tag{7}$$

Заметим, что утверждения лемм 4 и 5 справедливы, вообще говоря, для произвольной метрики. В пространствах размерности удвоения $d > 1$ оценка (7) может быть приведена в более конкретной форме. В самом деле, по построению кластер C_j^l лежит в некотором метрическом шаре, радиус которого не превосходит $2s^{L-l}$. Оценивая сверху $w(\mathrm{MST}(C'))$ по лемме 2, получаем

$$w(\tilde{\mathcal{R}}) = w(\mathcal{R}) + O\left(s^{L-l}\overline{r}^{(1-1/d)}\right). \tag{8}$$

Рассуждая по аналогии, оценим число порталов m в произвольном кластере C_j^l. Действительно, по определению, порталами в кластере C_j^l являются представители s^{L-l}/M-сети, следовательно, по лемме 1

$$m < \left(2\frac{4s^{L-l}}{s^{L-l}/M}\right)^d = (8M)^d = O\left(\left(\frac{d(\log n - \log\varepsilon)}{\varepsilon}\right)^d\right)$$

в пространстве произвольной размерности удвоения $d > 1$ при произвольном фиксированном $\varepsilon > 0$.

Пусть далее $\mathfrak{S}$ – произвольное допустимое решение задачи CVRP-SD(N_1, D, w_1, q), а $\overline{\mathfrak{S}}$ – соответствующее ему решение, состоящее из сетевых r-легких маршрутов, полученных из маршрутов $\mathfrak{S}$ последовательным применением лемм 4 и 5. Приведенная ниже структурная теорема позволяет оценить среднее (относительно случайной иерархической кластеризации) значение стоимости полученного решения.

Теорема 1 (структурная). *Пусть* $0 < \varepsilon < 1/8$, $d > 1$ *и* $r = m$.

$$E(w(\overline{\mathfrak{S}})) = (1 + O(\varepsilon))w(\mathfrak{S}).$$

Доказательство. Без ограничения общности полагаем, что исходное решение $\mathfrak{S}$ состоит из сетевых маршрутов, так как в противном случае последовательное применение леммы 4 к каждому из несетевых маршрутов решения $\mathfrak{S}$ позволяет сопоставить ему допустимое решение $\tilde{\mathfrak{S}}$ стоимости $w(\tilde{\mathfrak{S}}) = (1 + O(\varepsilon))w(\mathfrak{S})$, обладающее этим свойством. Предположим, что не все маршруты решения $\mathfrak{S}$ являются r-легкими. Пусть $\overline{r} > r$ – число пересечений одного из таких маршрутов $\mathcal{R} \in \mathfrak{S}$ границы произвольного кластера C_j^l на уровне $l > 0$. По лемме 5 (в силу соотношения (8)) маршруту $\mathcal{R}$ можно сопоставить сетевой маршрут $\overline{\mathcal{R}}$, стоимость которого превышает стоимость исходного маршрута на $O\left(s^{L-l}\overline{r}^{(1-1/d)}\right)$. Таким образом, прирост стоимости, приходящийся на одно из $\overline{r} > r = m$ ребер, пересекающих границу C_j^l, составит

$$O\left(\frac{s^{L-l}\overline{r}^{(1-1/d)}}{\overline{r}}\right) = O\left(\frac{s^{L-l}}{r^{1/d}}\right) = O\left(\frac{s^{L-l}}{M}\right) = O\left(\frac{s^{L-l}\varepsilon}{dL}\right),$$

где последнее равенство следует непосредственно из соотношения (5).

В свою очередь, из результатов (см. [34]) следует, что вероятность пересечения произвольным ребром $\{u, v\}$ границы кластера на уровне l не превосходит $O(d/s^{L-l})\rho(u, v)$. Следовательно, математическое ожидание прироста стоимости сетевого маршрута на уровне l, порождаемого его произвольным ребром $\{u, v\}$, составит $O(\varepsilon)\rho(u, v)/L$. Суммируя по $l = 1, 2, \ldots, L + 1$ и пользуясь линейностью математического ожидания, завершаем доказательство теоремы.

Как и в [23], [35], [24] доказанная выше структурная теорема играет ключевую роль в обосновании аппроксимационных свойств предлагаемого алгоритма. В самом деле, применяя утвер-

ждение теоремы 1 к произвольному оптимальному решению $\mathfrak{S}$ исследуемой задачи, получаем соотношение

$$\mathrm{CVRP}^*(Z, w, q) \le E(w(\widetilde{\mathfrak{S}})) \le (1 + O(\varepsilon))\mathrm{CVRP}^*(Z, w, q).$$

В следующем разделе описывается алгоритм построения сетевого r-легкого решения минимального веса для заданной случайной реализации иерархической кластеризации.

3.3. Динамическое программирование

Зададимся случайной иерархической кластеризацией. Методом динамического программирования построим допустимое решение задачи, состоящее из сетевых r-легких маршрутов, обладающих минимальной суммарной стоимостью. Для описания структуры таблицы динамического программирования нам потребуется несколько вспомогательных понятий и определений.

Пусть маршрут $\mathcal{R}$ пересекает границу кластера C уровня $l > 0$ в (необязательно различных) порталах p^{in} и p^{out} так, что его связный фрагмент

$$p^{\mathrm{in}}, x_{i_1}, \ldots, x_{i_r}, p^{\mathrm{out}} \tag{9}$$

полностью содержится в данном кластере. Произвольный максимальный по включению фрагмент (9) договоримся называть *сегментом* маршрута $\mathcal{R}$, *пересекающим* кластер C (или просто пересекающим сегментом). С точки зрения кластеризации более высокого уровня $l - 1$, произвольный сегмент, пересекающий кластер C, однозначно определяется кортежем

$$(p^{\mathrm{in}}, p^{\mathrm{out}}, Sup, Dep),$$

где p^{in} и p^{out} – упомянутые выше порталы “входа” и “выхода”, $0 \le Sup \le q$ – объем спроса, удовлетворяемый данным сегментом внутри кластера C, а $Dep \in \{0,1\}$ – индикатор, указывающий посещение им склада.

Конфигурацией, ассоциируемой с заданным кластером C произвольного уровня $l \ge 0$, назовем конечную последовательность

$$\mathfrak{C} = ((p_i^{\mathrm{in}}, p_i^{\mathrm{out}}, Sup_i, Dep_i),\ i = 1, 2, \ldots, T_{\mathfrak{C}}), \tag{10}$$

каждый элемент которой определяет сегмент какого-либо сетевого r-легкого маршрута, пересекающий кластер C. Длина $T_{\mathfrak{C}}$ конфигурации $\mathfrak{C}$ – неотрицательное целое число, не превосходящее $|\mathfrak{S}^*|r$, где $|\mathfrak{S}^*|$ – размер (число маршрутов) произвольного оптимального решения $\mathfrak{S}^*$ задачи. Верхняя граница следует из определения r-легкого маршрута, который не может пересекать границу произвольного кластера более r раз. В случае $T_{\mathfrak{C}} = 0$ конфигурация $\mathfrak{C}$, называется *пустой* и описывает ситуацию, при которой ни один из маршрутов не пересекает границ ассоциированного кластера (соответствующую, например, кластеру наивысшего уровня $l = 0$).

Конфигурация $\mathfrak{C}$ называется *допустимой* для кластера C, если найдется *подходящее* семейство $\tilde{\mathfrak{S}} = \tilde{\mathfrak{S}}(C, \mathfrak{C})$ допустимых сетевых r-легких маршрутов (возможно, включающее маршруты, целиком содержащиеся в кластере C), удовлетворяющих совокупный спрос потребителей из C, такое, что сегменты маршрутов, пересекающие кластер C, взаимно однозначно соответствуют элементам конфигурации $\mathfrak{C}$.

Внутренней стоимостью семейства маршрутов $\tilde{\mathfrak{S}}$ (относительно кластера C) назовем суммарный вес сегментов входящих в него маршрутов, пересекающих кластер C, дополненный весом маршрутов, внутренних относительно данного кластера.

Ячейка таблицы динамического программирования индексируется упорядоченной парой $(C, \mathfrak{C})$, в которой C – кластер произвольного уровня l, а $\mathfrak{C}$ – произвольная допустимая конфигурация. Значение ячейки $(C, \mathfrak{C})$ – минимальная внутренняя стоимость подходящего семейства маршрутов.

Заполнение таблицы динамического программирования производится рекурсивно в порядке убывания l. Ячейки базового $(L + 1)$-уровня, соответствующие одноэлементным кластерам, заполняются очевидным образом. Предполагая заполненными все ячейки, соответствующие кла-

стерам, расположенным на уровнях $l+1,\ldots,L+1$, опишем вычисление произвольной ячейки $(C_j^l,\mathfrak{C})$, порождаемой кластером C_j^l уровня l.

В самом деле, пусть $C_j^l = C_{j1}^{l+1}\cup\ldots\cup C_{jk}^{l+1}$ разбиение кластера C_j^l, построенное на этапе иерархической кластеризации. Через

$$Ch = (C_{j1}^{l+1},\mathfrak{C}_1),\ \ldots,\ (C_{jk}^{l+1},\mathfrak{C}_k) \tag{11}$$

обозначим произвольный набор порождаемых ими ячеек таблицы.

Для описания совместимости набора ячеек Ch с вычисляемой ячейкой $(C_j^l,\mathfrak{C})$ и последующей *склейки* сегментов нам потребуется несколько дополнительных понятий.

Профилем конкатенации, описывающим порядок склейки (сегмента) маршрута, назовем конечную последовательность $\varphi = ((p_t^1, p_t^2, sup_t, dep_t)\colon t = 1,2,\ldots,\theta_\varphi)$, в которой $p_t^1, p_t^2 \in C_j^l \cap N(l+1)$, а остальные компоненты кортежей имеют тот же смысл, что и в определении конфигурации.

Получающийся в результате склейки сегмент r-легкого маршрута может пересекать границы кластеров $C_{j1}^{l+1},\ldots,C_{jk}^{l+1}$, поэтому $\theta_\varphi \le kr$ для произвольного профиля φ.

Интерфейсом назовем произвольную конечную совокупность (не обязательно попарно различных) профилей конкатенации $\Im = (\varphi_1,\ldots,\varphi_\nu)$.

Будем говорить, что интерфейс $\Im$ *согласован* с вычисляемой ячейкой $(C_j^l,\mathfrak{C})$ и набором Ch, если выполнены следующие условия:

1) каждый кортеж произвольной конфигурации $\mathfrak{C}_i$ встречается в профилях интерфейса $\Im$ в точности один раз;

2) для произвольного профиля конкатенации φ, входящего в интерфейс $\Im$,

$$\sum_{t=1}^{\theta_\varphi} sup_t \le q;$$

3) суммарное значение удовлетворяемого спроса sup_j всеми профилями конкатенации, входящими в интерфейс $\Im$, совпадает с совокупным объемом спроса потребителей кластера C_j^l:

$$\sum_{\tau=1}^{\nu}\sum_{t=1}^{\theta_{\varphi_\tau}} sup_t = \sum_{x\in C_j^l} D(x);$$

4) произвольному кортежу $(p_i^{\text{in}}, p_i^{\text{out}}, Sup_i, Dep_i)$ конфигурации $\mathfrak{C}$ однозначно соответствует содержащийся в интерфейсе $\Im$ профиль конкатенации $\varphi = \varphi(i)$ такой, что

$$p_1^1 = p_i^{\text{in}},\quad p_{\theta_\varphi}^2 = p_i^{\text{out}},\quad \sum_{t=1}^{\theta_\varphi} sup_t = Sup_i,\quad \bigvee_{t=1}^{\theta_\varphi} dep_t = Dep_i;$$

5) если профиль конкатенации φ не соответствует никакому кортежу конфигурации $\mathfrak{C}$, то $\bigvee_{t=1}^{\theta_\varphi} dep_t = 1$.

По определению для совместимости набора ячеек Ch с ячейкой $(C_j^l,\mathfrak{C})$ необходимо и достаточно существование хотя бы одного согласованного интерфейса $\Im$.

Пусть далее $w(\Im)$ – суммарная стоимость (сегментов) маршрутов, получаемых в процессе склейки в соответствии с профилями конкатенации, входящими в согласованный интерфейс $\Im$. Сопоставив произвольному набору ячеек Ch стоимость

$$w(Ch) = \min\left\{w(\Im)\colon \Im \text{ согласован с набором } Ch \text{ и ячейкой } (C_j^l,\mathfrak{C})\right\},$$

значение, сохраняемое в ячейке $(C_j^l,\mathfrak{C})$, определим по формуле

$$w(C_j^l,\mathfrak{C}) = \min\{w(Ch)\colon \text{набор } Ch \text{ совместим с ячейкой } (C_j^l,\mathfrak{C})\}.$$

В случае, если ни один из наборов (11) не совместим с ячейкой $(C_j^l, \mathfrak{C})$, она исключается из рассмотрения, а конфигурация $\mathfrak{C}$ называется недопустимой для кластера C_j^l.

Минимизация по всем допустимым конфигурациям кластера C_1^0, расположенного на верхнем уровне рекурсии,

$$\min\{w(C_1^0, \mathfrak{C}): \mathfrak{C} - \text{допустима}\}$$

и последующее восстановление маршрутов методом обратной прогонки завершают поиск искомого сетевого r-легкого решения исходной задачи, соответствующего заданной случайной реализации иерархической кластеризации.

3.4. Дерандомизация

Следуя схеме доказательства, проведенного в [34], нетрудно убедиться, что для построения рандомизированной иерархической кластеризации (п. 3.2) достаточно $O(\log(n+d)\log\log n) + 2^{O(d)}$ случайных бит. Таким образом, наивная дерандомизация описанного в п. 3.3 алгоритма, основанная на полном переборе всех элементарных исходов, обладает полиномиальной трудоемкостью, как и в случае с классической схемой С. Ароры (см. [23]).

4. ВЕРХНЯЯ ОЦЕНКА ТРУДОЕМКОСТИ

Очевидно, что операция сопоставления исходной задаче округленной постановки (описанная в п. 3.1), может быть проведена за время $O(n)$. Случайная реализация иерархической кластеризации исходной постановки также обладает не более чем полиномиальной трудоемкостью. Оценим трудоемкость алгоритма динамического программирования.

В процессе построения иерархической кластеризации может быть построено до $O(n\log n)$ кластеров, с каждом из которых может быть ассоциировано до $(2m^2q)^{|\mathfrak{S}^*|r}$ конфигураций, где $\mathfrak{S}^*$ – произвольное оптимальное решение исследуемой задачи. В самом деле, по определению каждая конфигурация $\mathfrak{C}$ определяется соотношением (10), в котором произвольный кортеж $(p_i^{in}, p_i^{out}, Sup_i, Dep_i)$ может быть выбран не более чем $2m^2q$ способами.

Учитывая естественное ограничение $q \le n$, получаем верхнюю оценку размера таблицы динамического программирования:

$$O(n\log n)(2m^2n)^{|\mathfrak{S}^*|r}.$$

По построению вычисление каждой ячейки $(C_j^l, \mathfrak{C})$ таблицы сопряжено с полным перебором произвольных наборов дочерних ячеек (11), число которых не превосходит

$$(m^2n)^{|\mathfrak{S}^*|r\cdot 2^{O(d)}}.$$

Обработка произвольного набора (11) связана с перебором всевозможных интерфейсов $\mathfrak{I}$, число которых, в свою очередь, ограничено сверху:

$$(m^2n)^{|\mathfrak{S}^*|r(r+1)2^{O(d)}}.$$

Проверка согласованности произвольного интерфейса $\mathfrak{I}$, очевидно, может быть произведена за время, органиченное сверху его размером, т.е.

$$O(|\mathfrak{S}^*|\cdot(r+1)\cdot 2^{O(d)}\cdot r) = \mathrm{poly}(n),$$

где $\mathrm{poly}(n)$ обозначает некоторый полином от размера входных данных.

Соответственно, общая трудоемкость схемы с учетом выполнения условий $r = m$ и $|\mathfrak{S}^*| = \mathrm{polylog}n$ составит

$$\mathrm{poly}(n)\cdot(m^2n)^{\mathrm{polylog}\,n\cdot m^2\cdot 2^{O(d)}}, \quad \text{где} \quad m = O\left(\left(\frac{d(\log n - \log\varepsilon)}{\varepsilon}\right)^d\right).$$

Таким образом, нами обоснован основной результат данной статьи.

Теорема 2. $(1 + O(\varepsilon))$*-приближенное решение задачи CVRP, сформулированной в метрическом пространстве произвольной размерности удвоения* $d > 1$ *и обладающей оптимальным решением с числом маршрутов, не превосходящим* $\operatorname{polylog} n$*, может быть найдено рандомизированным алгоритмом за время* $\operatorname{poly}(n) \cdot (m^2 n)^{\operatorname{polylog} n \cdot m^2 \cdot 2^{O(d)}}$*, где* $m = O\left(\left(\frac{d(\log n - \log \varepsilon)}{\varepsilon}\right)^d\right)$*. Алгоритм допускает полиномиальную дерандомизацию.*

5. ЗАКЛЮЧЕНИЕ

В данной работе обосновывается аппроксимационная схема для метрической постановки классической задачи CVRP. Показано, что описанный в статье алгоритм реализует квазиполиномиальную приближенную схему (QPTAS) при условии, что задача сформулирована в метрическом пространстве произвольной фиксированной размерности удвоения $d > 1$ и обладает оптимальным решением, число маршрутов которого не превосходит $\operatorname{polylog} n$. Второе условие выполнено всякий раз, когда грузоподъемность $q = \Omega(n/\operatorname{polylog} n)$. Отметим, что конечной размерностью удвоения обладает широкий класс метрик, порождаемых большими графами, возникающими, в том числе, в процессе описания современных социальных сетей. Поэтому применимость полученных в статье результатов не ограничивается исключительно транспортными задачами.

Тем не менее вопрос об аппроксимируемости общих постановок задачи CVRP, не стесненных дополнительными ограничениями на рост грузоподъемности, сформулированных в метрических пространствах такого типа, в классе квазиполиномиальных приближенных схем до сих пор остается открытым. Авторы планируют вернуться к нему в одной из ближайших работ.

СПИСОК ЛИТЕРАТУРЫ

1. *Demir E., Huckle K., Syntetos A., Lahy A., Wilson M.* Vehicle routing problem: past and puture, Cham: Springer Inter. Publ., 2019. P. 97–117.
2. *Laporte G.* Fifty years of vehicle routing // Transport. Sci. 2009. V. 43. P. 408–416.
3. *Toth P., Vigo D.* Vehicle routing: problems, methods and applications, 2nd Ed. MOS-Siam Ser. on Optimizat. SIAM, 2014. P. 53–85.
4. *Dantzig G., Ramser J.* The truck dispatching problem // Management Sci. 1959. V. 6. Iss. 1. P. 80–91.
5. *Drexl M.* Branch-and-cut algorithms for the vehicle routing problem with trailers and transshipments // Networks. 2014. V. 63. № 1. P. 119–133.
6. *Hokama P., Miyazawa F.K., Xavier E.C.* A branch-and-cut approach for the vehicle routing problem with loading constraints // Expert Systems with Appl. 2016. V. 47. P. 1–13.
7. *Pecin D., Pessoa A., Poggi M., Uchoa E.* Improved branch-cut-and-price for capacitated vehicle routing // Math. Program. Comp. 2017. V. 9. № 1. P. 61–100.
8. *Pessoa A.A., Sadykov R., Uchoa E.* Enhanced branch-cut-and-price algorithm for heterogeneous fleet vehicle routing problems // Europ. J. of Operat. Res. 2018. V. 270. № 2. P. 530–543.
9. *Arnold F., Sorensen K.* Knowledge-guided local search for the vehicle routing problem // Comput. Operat. Res. 2019. V. 105. P. 32–46.
10. *Avdoshin S., Beresneva E.* Local search metaheuristics for Capacitated Vehicle Routing Problem: a comparative study // Proceed. Inst. for System Program. of RAS. 2019. V. 331. P. 121–138.
11. *Qiu M., Fu Zh., Eglese R., Tang Q.* A Tabu Search algorithm for the vehicle routing problem with discrete split deliveries and pickups // Comp. Operat. Res. 2018. V. 100. P. 102–116.
12. *Frifita S., Masmoudi M.* VNS methods for home care routing and scheduling problem with temporal dependencies, and multiple structures and specialties // Inter. Transact. in Operat. Res. 2020. V. 27. № 1. P. 291–313.
13. *Polat O.* A parallel variable neighborhood search for the vehicle routing problem with divisible deliveries and pickups // Comp. Operat. Res. 2017. V. 85. P. 71–86.
14. *Nazari M., Oroojlooy A., Takac M., Snyder L.V.* Reinforcement learning for solving the vehicle routing problem. Proceed. of the 32nd Inter. Conf. on Neural Inform. Process. Syst., Curran Associates Inc., Red Hook, NY, USA, 2018. P. 9861–9871.
15. *Vidal Th., Crainic T.G., Gendreau M., Prins Ch.* A hybrid genetic algorithm with adaptive diversity management for a large class of vehicle routing problems with time windows // Comput. Oper. Res. 2013. V. 40. № 1. P. 475–489.

16. *Necula R., Breaban M., Raschip M.* Tackling dynamic vehicle routing problem with time windows by means of ant colony system // 2017 IEEE Congress on Evolutionary Comput. (CEC), 2017. P. 2480–2487.

17. *Su-Ping Y., Wei-Wei M.* An improved ant colony optimization for VRP with time windows // Inter. J. of Signal Proc., Image Proc. and Pattern Recognit. 2016. V. 9. P. 327–334.

18. *Chen, Gui, Ding, Zhou*. Optimization of transportation routing problem for fresh food by improved ant colony algorithm based on tabu search // Sustainability. 2019. V. 11.

19. *Nalepa J., Blocho M.* Adaptive memetic algorithm for minimizing distance in the vehicle routing problem with time windows // Soft Comput. 2016. V. 20. № 6. P. 2309–2327.

20. *Papadimitriou Ch*. Euclidean TSP is NP-complete // Theor. Comput. Sci. 1977. V. 4. Iss. 3. P. 237–244.

21. *Asano T., Katoh N., Tamaki H., Tokuyama T.* Covering points in the plane by K-tours: towards a polynomial time approximation scheme for general K // Proceed. of the Twenty-ninth Ann. ACM Symp. on Theory of Comput., STOC '97, ACM, El Paso, Texas, USA, 2017. P. 275–283.

22. *Haimovich M., Rinnooy Kan A.H.G.* Bounds and heuristics for capacitated routing problems // Math. of Operat. Res. 1985. V. 10. № 4. P. 527–542.

23. *Arora S.* Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems // J. of the ACM. 1998. V. 45. Iss. 5. P. 753–782.

24. *Das A., Mathieu C.* A Quasipolynomial Time Approximation Scheme for Euclidean Capacitated Vehicle Routing // Algorithmica. 2015. V. 73. P. 115–142.

25. *Adamaszek A., Czumaj A., Lingas A.* PTAS for k-tour cover problem on the plane rof moderately large values of k // Inter. J. of Foundat. of Comp. Sci. 2010. V. 21. № 6. P. 893–904.

26. *Khachai M.Yu., Dubinin R.D.* Approximability of the vehicle routing problem in finite-dimensional euclidean spaces // Proceed. of the Steklov Instit. of Math. 2017. V. 297. № 1. P. 117–128.

27. *Khachay M., Dubinin R.* PTAS for the Euclidean capacitated vehicle routing problem in R^d // LNCS, V. 9869, Discrete Opt. and Operat. Res.: 9th Inter. Conf., DOOR 2016, Vladivostok, Russia, September 19–23, 2016. P. 193–205.

28. *Khachay M., Zaytseva H.* Polynomial time approximation scheme for single-depot Euclidean capacitated vehicle routing problem // LNCS, V. 9486. Combinatorial Opt. and Appl.: 9th Inter. Conf., COCOA 2015, Houston, TX, USA, December 18–20, 2015. P. 178–190.

29. *Khachay M., Zaytseva H.* Polynomial time approximation scheme for single-depot Euclidean capacitated vehicle routing problem // LNCS, V. 9486. Combinatorial Opt. and Appl.: 9th Inter. Conf., COCOA 2015, Houston, TX, USA, December 18–20, 2015. P. 178–190.

30. *Khachay M.Yu., Ogorodnikov Yu.Yu.* Efficient PTAS for the Euclidean CVRP with time windows // LNCS, V. 11179. Anal. of Images, Social Networks and Texts – 7th Inter. Conf., AIST 2018. Springer Inter. Publ., 2018. P. 318–328.

31. *Khachay M.Yu., Ogorodnikov Yu.Yu*. Approximation scheme for the capacitated vehicle routing problem with time windows and non-uniform demand // LNCS, V. 11548. Math. Optimizat. Theory and Operations Res. 18th Inter. Conf. (MOTOR 2019). Springer Inter. Publ., 2019. P. 309–327.

32. *Becker A., Klein P.N., Schild A.* A PTAS for bounded-capacity vehicle routing in planar graphs // Algorithms and Data Structures. Springer Inter. Publ., 2019. P. 99–111.

33. *Khachai M.Y., Ogorodnikov Y.Y.* Haimovich–Rinnooy Kan polynomial-time approximation scheme for the CVRP in metric spaces of a fixed doubling dimension // Trudy Instituta Matematiki i Mekhaniki UrO RAN. 2019. V. 25. Iss. 4. P. 235–248.

34. *Talwar K.* Bypassing the embedding: algorithms for low dimensional metrics // Proceed. of the Thirty-Sixth Ann. ACM Symp. on Theory of Comp. Associat. for Comp. Machinery, New York, NY, USA, 2004. P. 281–290.

35. *Bartal Y., Gottlieb L.A., Krauthgamer R.* The traveling salesman problem: low-dimensionality implies a polynomial time approximation scheme // SIAM J. on Comp. 2016. V. 45. № 4. P. 1563–1581.

36. *Abraham I., Bartal Y., Neiman O.* Advances in metric embedding theory // Adv. in Math. 2011. V. 228. № 6. P. 3026–3126.

37. *Gupta A., Krauthgamer R., Lee J.R.* Bounded geometries, fractals, and low-distortion embeddings // 44th Ann. IEEE Symp. on Foundat. of Comp. Sci., 2003. P. 534–543.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 7, с. 1220–1232

ИНФОРМАТИКА

УДК 519.72

ОБ ОДНОМ ПОДХОДЕ К СТАТИСТИЧЕСКОМУ МОДЕЛИРОВАНИЮ ТРАНСПОРТНЫХ ПОТОКОВ[1)]

© 2021 г. В. М. Старожилец[1],*, Ю. В. Чехович[1],**

[1] *119333 Москва, Вавилова, 42, ФИЦ ИУ РАН, Россия*

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

***e-mail: chehovich@forecsys.ru*

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Предлагается статистическая модель транспортных потоков, предназначенная для моделирования движения транспортных средств на автомагистралях значительной протяженности. Модель симулирует движение групп транспортных средств по магистрали с использованием фундаментальной диаграммы поток-плотность на выбранном участке автодороги для расчета скорости группы, размер группы считается линейно зависящим от ее скорости. Предложенный авторами подход к моделированию позволяет совместить достоинства макроскопического и микроскопического моделирования. А именно моделировать движение на транспортных системах мегаполисов с высокой точностью и низкими требованиями к вычислительным мощностям. В статье описаны принципы моделирования, приведены алгоритмы пересчета состояний модели, приведены результаты вычислительных экспериментов для подтверждения работоспособности и адекватности результатов модели для различных конфигураций дорожно-транспортной сети. Фундаментальная диаграмма в приведенных экспериментах строится по данным дорожных датчиков Центра организации дорожного движения. Библ. 20. Фиг. 5.

Ключевые слова: моделирование транспортных потоков, фундаментальная диаграмма потоков, группы автомобильно-транспортных средств (АТС).

DOI: 10.31857/S0044466921070152

1. ВВЕДЕНИЕ

Данная работа посвящена описанию модели, предназначенной для моделирования потоков в автомобильно-транспортной сети с использованием анонимных данных с GPS-треков и дорожных датчиков, а также полученных видеосъемкой. Процедура комплексирования данных из GPS-треков и дорожных датчиков подробно рассмотрена в [1]. Также проводятся эксперименты на синтетических данных с целью показать адекватность поведения модели.

Моделирование транспортных потоков основано на их сходстве с жидкой или газовой средой. В частности, базовая модель Лайтхилла–Уизема–Ричардса (Lighthill–Whitham–Richards, LWR) [2]–[4] основана на предположении о существовании взаимно однозначной зависимости между скоростью и плотностью потока автомобильно-транспортных средств(АТС) и сохранении числа АТС в транспортной сети. В современном макроскопическом подходе транспортный поток описывается нелинейной системой гиперболических дифференциальных уравнений в частных производных второго порядка в различных постановках [5]–[12].

В современных исследованиях также пытаются учесть разнородность транспортных стредств в потоке АТС. В работе [13] детально рассматривается движение транспортного потока, состоящего из автомобилей, автобусов, двухколесных и трехколесных мотоциклов на двухполосной дороге. В [14] рассматривается смешанный поток из велосипедов и автомобилей. В [15], [16] для той же задачи моделирования смешанного потока используются клеточные автоматы.

Основное отличие данной модели от уже представленных в том, что рассматривается движение неразделимых групп автомобилей по магистрали (которые, однако, могут соединяться меж-

[1)]Работа выполнена при финансовой поддержке РФФИ (код прокта 17-07-01574).

ду собой) вместо движения самих автомобилей, считая скорость всех транспортных средств в группе одинаковой. Зависимость скорости групп АТС от плотности автомобилей на рассматриваемом участке автомобильно-транспортной сети рассчитывается на основе построенной по историческим данным фундаментальной диаграммы потока для данного участка [17].

Хотя модель использует довольно грубые приближения из-за использования группы АТС как базовой единицы моделирования, будет показано, что групп АТС достаточно для получения результатов, хорошо совпадающих с реальными измерениями, при любых режимах автомагистрали [18].

2. СТРУКТУРА МОДЕЛИ

2.1. Внутренние свойства модели

Транспортная сеть представляет собой связный ориентированный граф $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, где $\mathbf{V}$ – множество вершин, $\mathbf{E} = \{(i, j)\}$ – множество ветвей графа. На граф также накладываются ограничения на максимальную и минимальную степень вершин $d(i)$: $\min(d(i)) = 1$ и $\max(d(i)) = 3$. Также $\forall i : d(i) > 1 \to \exists j, l \in \mathbf{V} : (j,i), (i,l) \in \mathbf{E}$, т.е. не существует вершины, в которой только заканчиваются несколько ветвей, и не существует вершины, в которой только начинаются несколько ветвей.

Определим теперь все типы вершин графа в зависимости от их степеней.

1. $d(i) = 1$. В данном случае существует два варианта:

(а) $i : \exists(i, j) \in \mathbf{E}$. Такие вершины будем называть *вершинами-въездами*. Вершины-въезды образуют множество $\mathbf{V}_{\text{in}}$ и являются источниками автомобильно-транспортных средств(АТС) в рассматриваемой модели.

(б) $i : \exists(j, i) \in \mathbf{E}$. Такие вершины будем называть *вершинами-съездами*. Вершины-съезды образуют множество $\mathbf{V}_{\text{out}}$ и являются стоками автомобилей в рассматриваемой модели.

2. $d(i) = 2$. Это внутренние вершины модели, образующие множество $\mathbf{V}_{\text{int}}$.

3. $d(i) = 3$ – вершины-центры перекрестков дорожно-транспортной сети. Данные вершины также входят в множество $\mathbf{V}_{\text{int}}$, но образуют еще два подмножества.

(а) Если $\exists(i, j) \in \mathbf{E}$, $(i, k) \in \mathbf{E}$: $j \neq k$, то такие вершины образуют множество $\mathbf{V}_{\text{sep}}$ – вершины, в которых происходит разделение потоков в дорожно-транспортной сети.

(б) Если $\exists(j, i) \in \mathbf{E}$, $(k, i) \in \mathbf{E}$: $j \neq k$, то такие вершины образуют множество $\mathbf{V}_{\text{mer}}$ – вершины, в которых происходит слияние потоков в дорожно-транспортной сети.

Таким образом, $\mathbf{V} = \mathbf{V}_{\text{int}} \cup \mathbf{V}_{\text{out}} \cup \mathbf{V}_{\text{in}}$ – все вершины распределены по трем непересекающимся группам. Вершины же перекрестки с $d(i) = 3$ дополнительно разделены по типу перекрестка, причем $V_{\text{sep}} \cap V_{\text{mer}} = \emptyset$. Ввиду того, что в данной работе рассматриваются только автомагистрали, то вершины с $d(i) = 4$ и более не встречаются.

Разделим схожим образом ребра, инцидентные этим вершинам. Рассмотрим для этого некоторое ребро (i, j).

1. Если $i \in \mathbf{V}_{\text{in}}$, то ребро (i, j) – это ребро-въезд. Такие ребра образуют множество въездов $\mathbf{E}_{\text{in}}$.

2. Если $j \in \mathbf{V}_{\text{out}}$, то ребро (i, j) – это ребро-съезд. Такие ребра образуют множество съездов $\mathbf{E}_{\text{out}}$.

3. Если $i \in \mathbf{V}_{\text{int}}$ и $j \in \mathbf{V}_{\text{int}}$, то ребро (i, j) – это внутреннее ребро модели. Такие ребра образуют множество внутренних ребер модели $\mathbf{E}_{\text{int}}$.

Также как и с вершинами $\mathbf{E} = \mathbf{E}_{\text{int}} \cup \mathbf{E}_{\text{out}} \cup \mathbf{E}_{\text{in}}$ за исключением случая, когда модель представляет из себя одно ребро, который в этой работе не рассматривается.

Определим теперь понятие состояния модели в момент времени t. Для этого нам понадобится понятие автомобильной группы на ветви (i, j): $\mathbf{A}_k^t = \{\text{Pos}_k, V_k, N_k\}$, обладающей следующими характеристиками:

1. Pos_k – позиция начала группы относительно начала ветви, на которой она расположена.

2. V_k – скорость группы АТС.

3. N_k – размер группы АТС из $\mathbb{R}_{\geq 0} = \mathbb{R}_+$.

Пусть теперь $\mathbf{A}^t_{i,j} = \{\mathbf{A}^t_k\}$ – упорядоченное множество автомобильных групп на ветви (i, j). Причем $\forall l, m : l < m \to \mathrm{Pos}_l > \mathrm{Pos}_m$ – группы не могут обгонять друг друга.

Таким образом, введем состояние системы в момент времени t как $\mathbf{A}^t = \{\mathbf{A}^t_{i,j}\} \cup \{A^t_{\mathrm{out},i,j}\}$, т.е. положение, скорость, размер и тип всех автомобильных групп на всех ветвях дорожно-транспортной сети. Группы АТС $\{A^t_{\mathrm{out},i,j}\}$ представляют собой специальные группы-буферы. Их особые свойства рассматриваются в разделе Общих свойств модели.

Для расчетов нам также понадобится понятие потенциала трансфера $\mathrm{Tr}^t_{(i,j),k}(\mathbf{A}^t, \mathbf{A}^{t-1})$ – число АТС, которые могут съехать с ветви (i, j) на ветвь (j, k) в интервал времени от $t-1$ до t. Данная величина вычисляется заново на каждой временной итерации в зависимости от состояния системы.

2.2. Внешние свойства модели

Определим теперь свойства модели, задаваемые при ее инициализации.

Рассмотрим предварительно три ветви с $d(j) = 3$: $(i, j), (j, k_1), (j, k_2)$. В данной работе мы считаем (j, k_1) продолжением автомагистрали, а (j, k_2) – съездом с нее. Данное распределение полностью задается в момент инициализации модели.

Перечислим все внешние параметры модели для каждой ветви $(i, j) \in \mathbf{E}$ графа $\mathbf{G}$.

1. Длина ветви $l_{i,j}$.

2. Число полос, по которым разрешено движение автомобилей по данной ветви $n_{i,j}$.

3. $I_{i,j} = \{0,1\}$ – идентификатор того, является ветвь съездом или нет. Если является, то $I_{i,j} = 1$.

4. Функция скорости для данной ветви $V = f_{i,j}(\rho)$, $f_{i,j} : \mathbb{Q}_+ \to \mathbb{Q}_+$, где $\rho \in \mathbb{R}_+$ – плотность АТС. В данной работе рассматриваются только ограниченные, непрерывные, монотонно убывающие функции скорости. Процедура получения данной функции из экспериментальных данных детально описана в статье [17].

5. Матрица перемешивания в узле j в момент времени t, задаваемая функцией $M_j(t)$. В случае если $j: \nexists (j,k) \in \mathbf{E}_{\mathrm{out}} \to \forall t: M_j(t) = 0$.

6. Интенсивность источника в узле i в момент времени t, задаваемая функцией $F_i(t)$. Для всех $i: i \notin \mathbf{V}_{\mathrm{in}} \to \forall t: F_i(t) == 0$.

Также у каждой ветви есть буфер АТС $A^t_{\mathrm{out},i,j} = \{\mathrm{Pos}_{\mathrm{out},i,j}, V_{\mathrm{out},i,j}, N_{\mathrm{out},i,j}\}$, представляющий из себя группу АТС с $\mathrm{Pos}_{\mathrm{out},i,j} = l_{i,j}$, $V_{\mathrm{out},i,j} = 0$. Данная группа моделирует очередь на съезд с ребра (i, j) – т.е. группу (j, k) с $I_{i,j} = 1$. Работа с данной группой детально описана в разд. Алгоритмов перемещения и объединения групп АТС 3.

Будем считать, что в модели все автомобили имеют фиксированный размер $L_{\mathrm{car}^{\mathrm{avg}}}$. В дальнейшем, путем изменения этой величины можно также исследовать зависимость поведения автомагистрали от состава потока автомобилей. С его помощью получаем максимальное число автомобилей на ветви (i, j) как $N^{i,j}_{\max} = \dfrac{l_{i,j} \cdot n_{i,j}}{L^{\mathrm{avg}}_{\mathrm{car}}}$.

Введем также понятие динамического размера автомобилей $L_{\mathrm{car}}(V) = L^{\mathrm{avg}}_{\mathrm{car}} + aV$, где $a = 0.504$. Данная величина отражает тот факт, что автомобили в среднем на определенной скорости не сближаются сильнее некоторого расстояния. Сама же константа a, как и данное приближение, взята из книги [19].

Таким образом, получаем, что все характеристики автомобильной группы ограничены сверху.

1. Положение Pos_k группы АТС длиной ветви, на которой группа находится.

2. Скорость V_k ограничена максимальной скоростью на ветви, которую можно определить из функции скорости $f_{i,j}(\rho)$.

3. Максимальный размер N_k ограничен максимальным числом автомобилей на ветви $N^{i,j}_{\max}$.

Однако поскольку в нашей модели все группы АТС движутся так, как будто каждый автомобиль в группе обладает полным знанием обо всех других автомобилях, то это накладывает на группы логическое ограничение на их размер, так как сложно ожидать такого поведения у АТС в огромной группе. Мы в данной работе считаем разумным ограничение в $N_{\max} = 20$ АТС.

Также нам понадобится величина среднего ускорения группы АТС a_{avg} для ограничения увеличения скорости движения групп АТС по ветвям автомагистрали. В данной работе величина ускорения взята за константу и равна 2.2 м/с^2 (см. [20]).

3. АЛГОРИТМЫ ПЕРЕМЕЩЕНИЯ И ОБЪЕДИНЕНИЯ ГРУПП АТС

Определим как группы АТС объединяются в одну, как переезжают с одной ветви на другую, как движутся по ветви и как съезжают на ветви-съезды. Также определим функцию расчета $\text{Tr}^t_{(i,j),k}(\mathbf{A}^t, \mathbf{A}^{t-1})$ на каждом временном шаге.

3.1. Движение групп АТС по ветви

На каждой итерации алгоритма надо рассчитать новое положение групп АТС для каждой ветви. Перерасчет положения групп, а также их скорости производит функция $\text{group\_mover}(\mathbf{A}^t_{i,j}, k, (i,j), \tau, t)$ по алгоритму 1. Расчеты по данному алгоритму сводятся к следующим шагам.

Шаг 1. Для выбранной группы АТС $\mathbf{A}^t_k$ на ветви (i, j) рассчитываем ее скорость V_k' на основании плотности автомобилей на участке автодороги перед ней.

Шаг 2. Рассчитываем новое положение группы АТС.

Шаг 3. Если группа оказалась в конце ветви:

(а) рассчитать, сколько времени она ехала до конца ветви;

(б) в соответствии с матрицей перемешивания часть группы добавляется в буфер-группу $A^t_{\text{out},i,j}$;

(в) оставшаяся часть группы $\mathbf{A}^t_k$ пытается переехать на следующую для нее ветвь с $I_{j,m_1} = 0$;

(г) группа-буфер $A^t_{\text{out},i,j}$ пытается переехать на следующую для нее ветвь с $I_{j,m_1} = 1$.

Заметим, что функция group_transferrer алгоритм 3 вызывается тут с временным шагом τ', что означает, что группа АТС при переезде на новую ветвь будет двигаться меньшее количество времени.

Алгоритм 1. Алгоритм расчета положения и скорости группы АТС

Вход: $\mathbf{A}^t_{i,j}$ – множество характеристик автомобильных групп на ветви;

k – индекс рассматриваемой автомобильной группы;

(i, j) – рассматриваемая ветвь графа;

τ – временной шаг;

t – текущий момент времени;

если $\text{Pos}_k = l_{i,j}$, **то**

Если группа АТС уже в конце ветви, то она просто пытается переехать на следующую ветвь

Пусть j': $(j, j') \in \widetilde{\mathbf{E}}$ – ветвь с $I_{j,j'} = 0$

$\text{group\_transferrer}(\mathbf{A}^t, \mathbf{A}^{t-1}, k, (i,j), (j,j'), \tau')$ из Алгоритма 3

return 0

$\text{Pos}_k += V_k \cdot \tau$

Пусть $\rho = \dfrac{\sum_{m=k+1}^{\text{len}(\mathbf{A}^t_{i,j})} N_m}{l_{i,j} \cdot n_{i,j}}$ – плотность АТС на участке ветви (i, j) перед группой АТС k,

тогда $V_k' = f_{i,j}(\rho)$ – новая скорость группы АТС.

если $V_k' - V_k > a_{\text{avg}} \cdot \tau$, **то**

$V_k' = V_k + a_{\text{avg}} \cdot \tau$

если $\text{Pos}_k \geq l_{i,j}$ и $k = 0$, **то**

$\tau' = \tau \cdot \dfrac{\text{Pos}_k - l_{i,k}}{V_k \cdot \tau}$ – 'оставшееся' время движения автомобильной группы

$\text{Pos}_k = l_{i,k}$

Добавим АТС в группу-буфер $\mathbf{A}^t_{\text{out},i,j}$

$N_{\text{out},i,j} += N_k \cdot M_j(t)$

$N_k = N_k \cdot (1 - M_j(t))$

Оставшиеся АТС должны продолжить движение по магистрали

Пусть $j'\colon (j, j') \in \widetilde{\mathbf{E}}$ – ветвь с $I_{j,j'} = 0$

group_transferrer$(\mathbf{A}^t, \mathbf{A}^{t-1}, k, (i, j), (j, j'), \tau')$ по Алгоритму 3

Группа-буфер пытается съехать

Пусть $j''\colon (j, j'') \in \widetilde{\mathbf{E}}$ – ветвь с $I_{j,j''} = 0$

group_transferrer$(\mathbf{A}^t, \mathbf{A}^{t-1}, k, (i, j), (j, j''), \tau')$ по Алгоритму 3

иначе

если $\text{Pos}_k \geq \text{Pos}_{i,k-1}$, **то**

$\text{Pos}_k = \text{Pos}_{k-1} - L_{\text{car}}(V_{k-1}) \cdot N_{k-1}$

group_union$(\mathbf{A}_{i,j}, k, (i, j), t)$

Если группа АТС $\mathbf{A}^t_k$ все еще существует $V_k = V_k'$

3.2. Объединение двух групп АТС

После изменения положения группы АТС на ветви нужно проверить, не может ли она быть объединена с какой-либо другой группой. Поскольку группы движутся только вперед и в нашей модели не могут обгонять друг друга, то проверка на возможность объединения идет только с группой перед рассматриваемой. То есть для группы АТС j рассматривается возможность ее слияния с группой $j-1$. За слияние групп отвечает функция group_union$(\mathbf{A}^t_{i,j}, k, (i, j), t)$, работающая по алгоритму 2.

Алгоритм 2. Алгоритм объединения групп АТС

Вход: $\mathbf{A}^t_{i,j}$ – множество характеристик автомобильных групп на ветви;

k – индекс рассматриваемой автомобильной группы;

(i, j) – рассматриваемая ветвь графа;

t – текущий момент времени;

если $k = 0$, то

Группу не с чем объединять, так как она самая первая

иначе

если $\text{Pos}_{k-1} - \text{Pos}_k \leq L_{\text{car}}(V_{k-1}) \cdot N_{k-1}$ и $N_{k-1} + N_k \leq N_{\max}$, **то**

Объединяем группы в одну

если $k - 1 = 0$ и $\text{Pos}(k-1) = l_{i,j}$, **то**

$N^j_{\text{exit}} += N_k \cdot M_j(t)$

$N_{k-1} += N_k \cdot (1 - M_j(t))$

иначе

$N_{k-1} += N_k$

del $\mathbf{A}^t_k$ – группа k удаляется

3.3. Перемещение групп АТС между ветвями

Когда группа автомобилей достигает конца ветви, на которой она находится, требуется определить, какая ее часть переедет на следующий сегмент автомагистрали и какое положение и скорость примет на новой ветви. За данные расчеты отвечает функция group_transferrer$(\mathbf{A}^t, \mathbf{A}^{t-1}, k, (i, j), (j, j'), \tau, t)$ по алгоритму 3. Расчеты по данному алгоритму сводятся к следующим шагам.

Шаг 1. Определяем индекс новой группы АТС на ветви (j, j').

Шаг 2. Определяем, может ли группа переехать на новую ветвь полностью. Если да:

(а) создаем новую группу АТС в конце ветви (j, j') с $N_{k'} = N_k$.

(б) Удаляем группу k из $\mathbf{A}^t_{i,j}$.

Шаг 3. Если нет:

(а) создаем новую группу АТС в конце ветви (j, j') с $N_{k'} = N'$, где N' – число АТС, которые могут переехать.

(б) Уменьшаем размер группы k на величину N'.

Шаг 4. Вызываем функцию для перемещения новой группы по ветви (j, j').

Алгоритм 3. Алгоритм перемещения группы АТС на новую ветвь

Вход: $\mathbf{A}^t$ – состояние системы в текущий момент времени;

$\mathbf{A}^{t-1}$ – состояние системы в предыдущий момент времени;

k – индекс рассматриваемой автомобильной группы;

(i, j) – ветвь, с которой хочет съехать группа АТС;

(j, j') – ветвь, на которую хочет съехать группа АТС;

τ – временной шаг;

t – текущий момент времени;

если $N_k \le \text{Tr}^t_{(i,j),j'}(\mathbf{A}^t, \mathbf{A}^{t-1})$, **то**

Группа полностью может переехать на новую ветвь

Пусть $k' = \text{len}(\mathbf{A}^t_{j,j'}) + 1$ – индекс новой группы АТС

Создаем новую группу АТС на ребре (j, j') с индексом k' и характеристиками:

$\text{Pos}_{k'} = 0$, $V_{k'} = f_{j,j'}(\rho)$, $N_{k'} = N_k$

Удаляем группу k из $\mathbf{A}^t_{i,j}$

group_mover$(\mathbf{A}^t_{j,j'}, k', (j, j'), \tau, t)$

иначе

Только часть группы переезжает на новую ветвь

Пусть $k' = \text{len}(\mathbf{A}^t_{j,j'}) + 1$ – индекс новой группы АТС

Создаем новую группу АТС на ребре (j, j') с индексом k' и характеристиками:

$\mathrm{Pos}_{k'} = 0$, $V_{k'} = f_{j,j'}(\rho)$, $N_{k'} = \mathrm{Tr}^t_{(i,j),j'}(\mathbf{A}^t, \mathbf{A}^{t-1})$

$N_k - = \mathrm{Tr}^t_{(i,j),j'}(\mathbf{A}^t, \mathbf{A}^{t-1})$

$\mathrm{group\_mover}(\mathbf{A}^t_{j,j'}, k', (j, j'), \tau, t)$

4. РАСЧЕТНЫЙ ЦИКЛ

4.1. Расчет потенциала трансфера

Для начала определим то, как в конце каждой итерации рассчитывается сколько АТС могут переехать с ветви (i, j) на ветвь (j, k). Особенностью является то, что алгоритм рассчитывает не потенциал трансфера для ветви (i, j) на ветвь (j, k), а все потенциалы трансфера $\forall i : (i, j) \in \mathbf{E} \to \mathrm{Tr}^k_{i,j}$. То есть для ветви (j, k) рассчитываются всевозможные $\mathrm{Tr}^k_{i,j}$.

Данный расчет производит функция $\mathrm{Tr\_calculator}(\mathbf{A}^t, \mathbf{A}^{t-1}, (j, k), \tau)$ по алгоритму 4. В процессе расчета нам также понадобятся величины $Q(i, j) = \max(\rho \cdot f_{i,j}(\rho))$ – максимальный поток АТС на ветви (i, j), $N^{i,j}_{\max}$ – максимальное число АТС на ветви (i, j) и $N^t_{i,j} = \sum_{m=0}^{\mathrm{len}(\mathbf{A}^t_{i,j})} N_m$ – текущее число АТС на ветви (i, j).

Процедура расчета сводится к определению двух величин:

1) P_i – потенциальное количество АТС, которые могут доехать до конца ветви (i, j), предшествующей (j, k);

2) N_{total} – сколько всего АТС может переехать на ветвь (j, k) на основании ее вместимости и максимального потока на ней.

В итоге число АТС, которые могут переехать с ветви (i, j) на ветвь (j, k), определяется формулой $N_{\mathrm{total}} \dfrac{P_i}{\sum P_i}$.

4.2. Процедура расчета

Процедура перехода от состояния системы $\mathbf{A}^{t-1}$ к состоянию $\mathbf{A}^t$ происходит в соответствии со следующим циклом.

1. $\forall (i, j) \in \mathbf{V}_{\mathrm{out}}$: удаляем все группы АТС, находящиеся на этой ветви, так как это ветви – стоки.

2. $\forall (i, j) \in \mathbf{V}_{\mathrm{in}}$: формируем новую группу АТС $k' = \mathrm{len}(\mathbf{A}^t_{i,j}) + 1$ с $\mathrm{Pos}_{k'} = 0$, $V_{k'} = f_{j,j'}(\rho)$, $N_{k'} = F_i(t)$.

3. Пусть $\mathbf{C}$ – некоторое подмножество ветвей графа. Будем выполнять следующие действия пока оно не пусто:

(а) $\mathbf{C} = \{(i, j)\}$, $(i, j) \in \mathbf{V}_{\mathrm{out}}$.

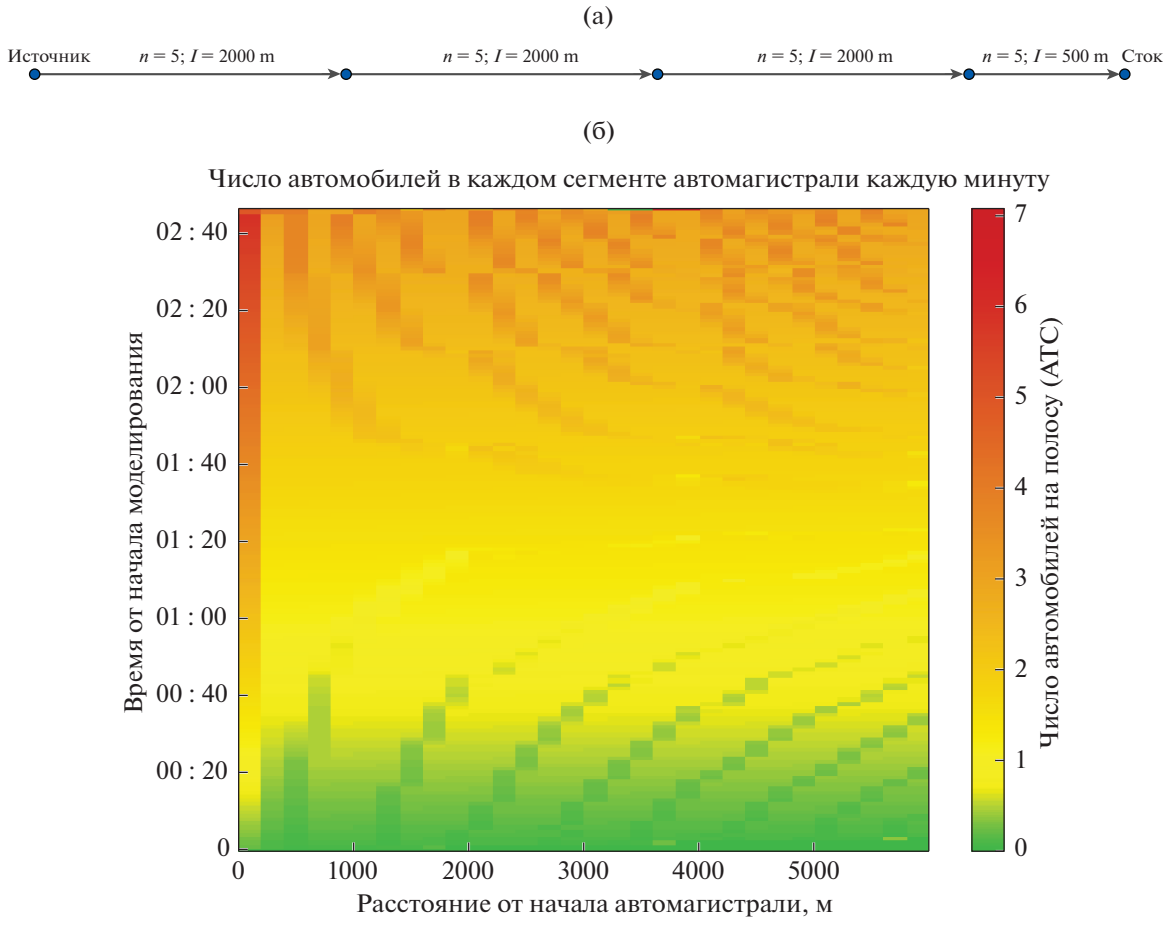
(б) $\forall (i, j) \in \mathbf{C} \to \forall \mathbf{A}^t_k \in \mathbf{A}^t_{i,j} \to \mathrm{group\_mover}(\mathbf{A}^t_{i,j}, k, (i, j), \tau, t)$ – для каждой группы АТС рассчитываем ее новое положение. Причем расчет производится упорядоченно по убыванию величины Pos_k, причем группы-буферы обсчитываются первыми.

(в) $\mathbf{C}' = \{(k, i)\} : \exists j : (i, j) \in \mathbf{C}$ – формируем новое подмножество для расчетов.

(г) $\mathbf{C} = \mathbf{C}'$.

4. $\forall (i, j) \in \mathbf{E} \to \forall \mathbf{A}^t_k \in \mathbf{A}^t_{i,j} \to \mathrm{group\_union}(\mathbf{A}^t_{i,j}, k, (i, j), t)$ – объединяем группы АТС, если это возможно.

Таким образом, получаем состояние системы $\mathbf{A}^t$.



Фиг. 1. (а) Схема простой дороги в модели состоит из 3 сегментов по 2 километра. (б) Тепловая карта автомобилей на простой дороге без перекрестков с линейно нарастающим вплоть до 150 АТС/мин потоком.

Алгоритм 4. Алгоритм расчета $\mathrm{Tr}_{i,j}^{k}$

Вход: $\mathbf{A}^{t}$ – состояние системы в текущий момент времени;

$\mathbf{A}^{t-1}$ – состояние системы в предыдущий момент времени;

(j,k) – ветвь, для которой проводится расчет;

τ – временной шаг;

Пусть $\mathbf{I} = \{i : (i,j) \in \mathbf{E}\}$ – множество ветвей, предшествующих рассматриваемой

для всех $i \in \mathbf{I}$

$P_i = 0$ – число АТС, которые теоретически могут достигнуть конца их ветви на следующей временной итерации

для всех $m = 0; l \le \mathrm{len}(\mathbf{A}_{i,j}^{t}); m{+}{+}$

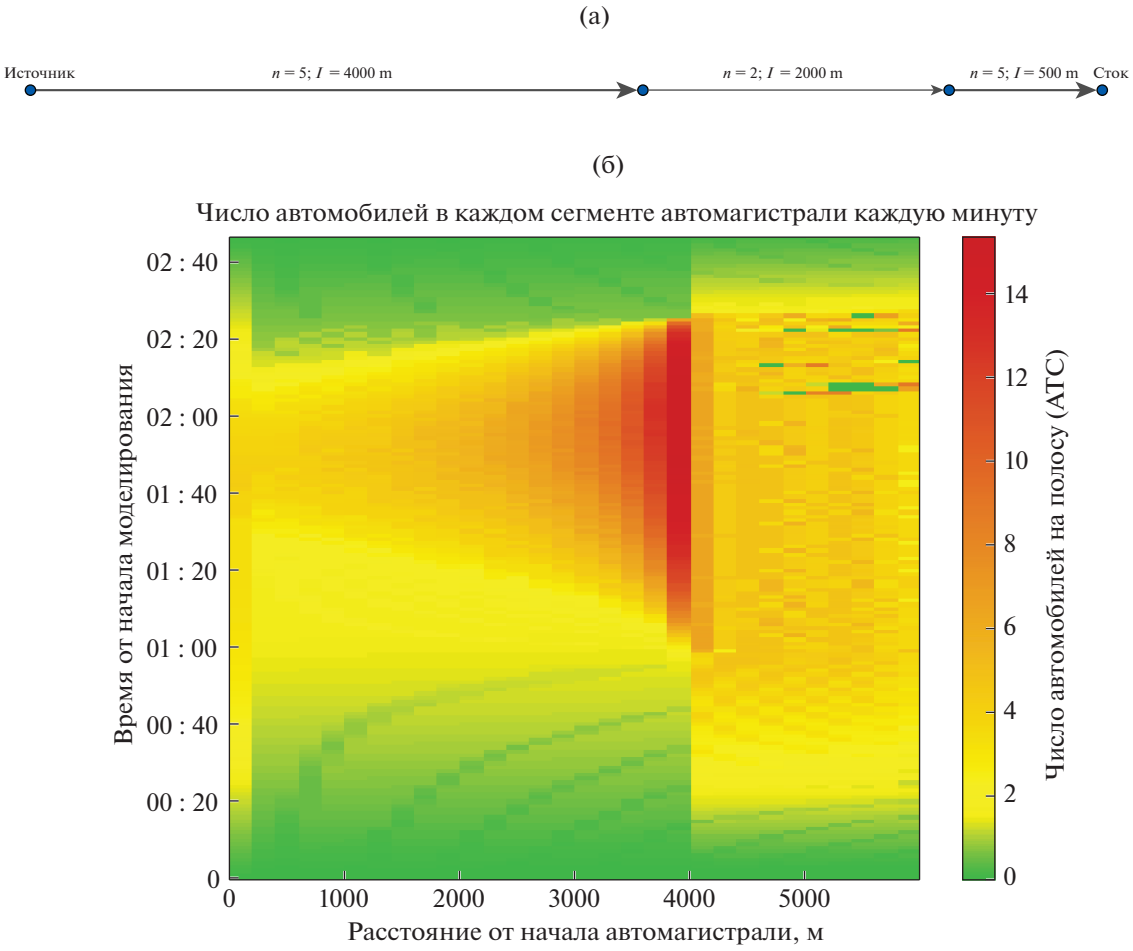
если $\mathrm{Pos}_m + V_m \cdot \tau \ge m_{i,j}$, **то**

если $(j,k) \in \mathbf{E}_{\mathrm{int}}$, **то**

$P_i += N_m \cdot (1 - M_j(t))$

иначе

$P_i += N_m \cdot M_j(t)$



Фиг. 2. (а) Схема дороги. (б) Тепловая карта автомобилей на пятиполосной дороге без перекрестков с сужением до двух полос и синусоидальным потоком на входе.

Определим N_{total} – сколько всего АТС может переехать на ветвь (j,k)

если $N_{\max}^{j,k} - N_{\text{cur}}^{j,k} < Q_(j,k)$, **то**

$$N_{\text{total}} = N_{\max}^{j,k} - N_{\text{cur}}^{j,k}$$

иначе

$$N_{\text{total}} = Q_(j,k)$$

для всех $i \in \mathbf{I}$

$$\text{Tr}_{i,j}^{k} = N_{\text{total}} \cdot \frac{P_i}{\sum P_i}$$

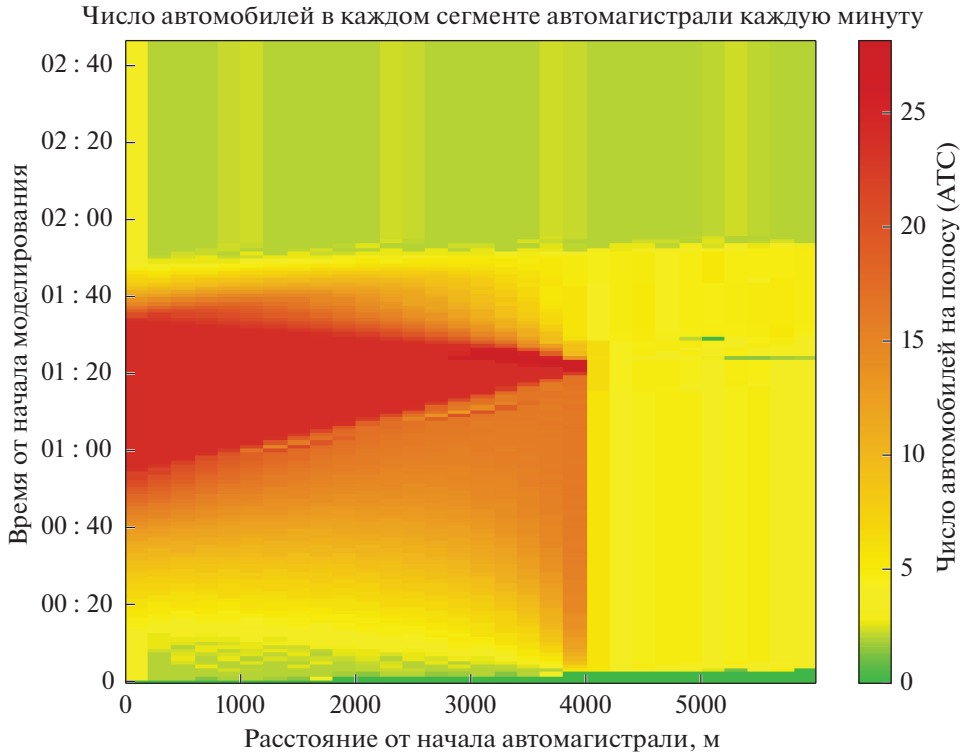
5. ОПИСАНИЕ ДАННЫХ

В данной работе во всех экспериментах использовалась одна фундаментальная диаграмма поток-плотность, полученная анализом реальных данных с дорожных датчиков за 2012 г. Построение фундаментальной диаграммы сводится к следующим шагам.

1. Для каждого надежного датчика на выбранном участке дороги извлечем данные измерений плотности и потока за наблюдаемый период времени. Каждая точка на диаграмме определяется парой значений “плотность–поток” на плоскости $Q(\rho)$.

(а)

(б)



Фиг. 3. (а) Схема дороги. (б) Тепловая карта автомобилей на пятиполосной дороге без перекрестков с сужением до двух полос пропадающим в середине моделирования и постоянным потоком в 100 АТС/мин.

2. Фильтрация выбросов путем построения альфа-оболочек облака точек диаграммы до тех пор, пока разница площадей оболочек для двух смежных итераций не будет мала.

3. Находим опорные точки на границе облака точек диаграммы и строим на их основе функцию-огибающую, которую и принимаем за фундаментальную диаграмму поток-плотность.

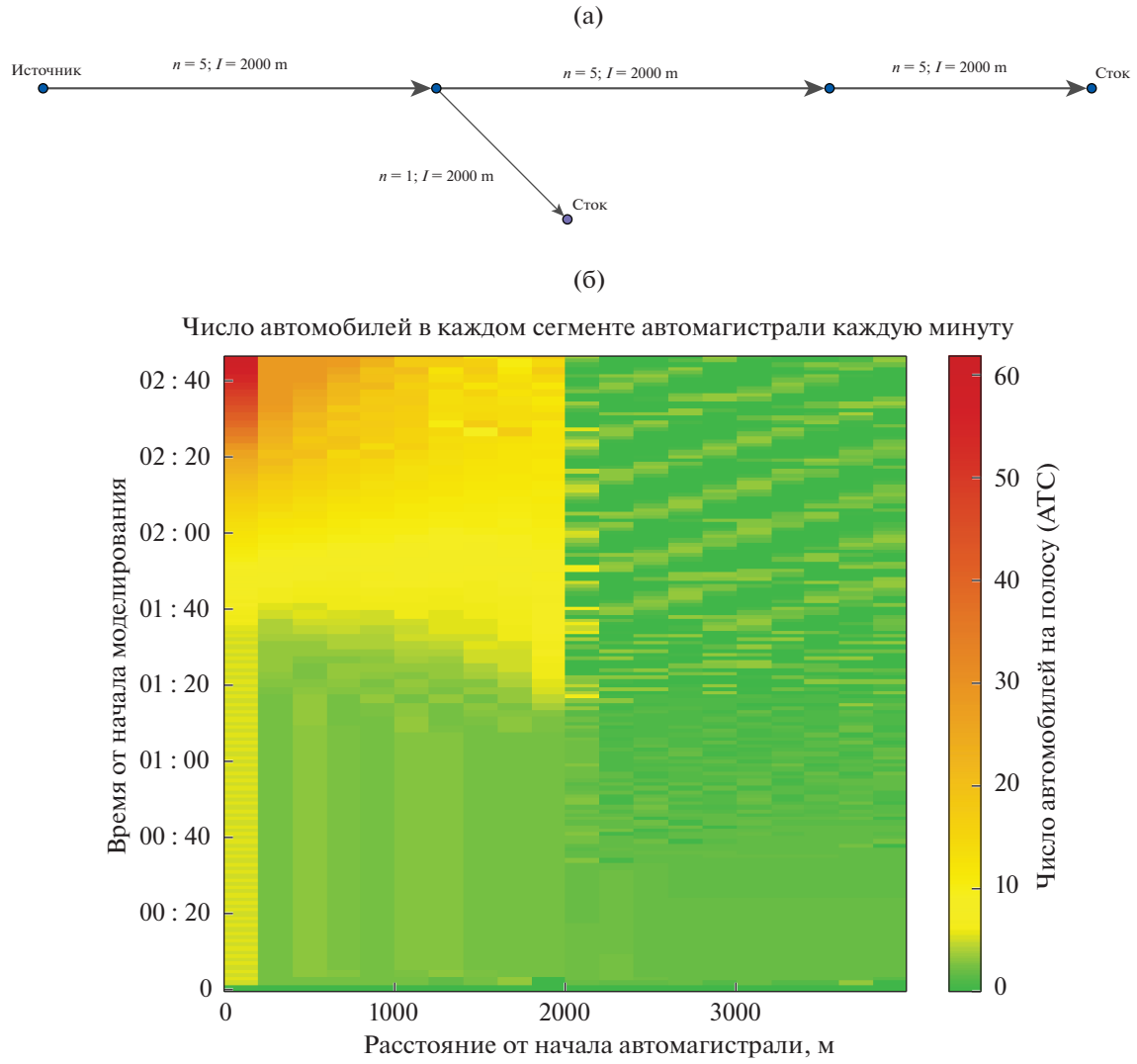
Детально процедура построения фундаментальной диаграммы описана в работе [17].

6. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

В работе проводится серия экспериментов на синтетических данных в различных конфигурациях транспортной сети. Цель экспериментов – проверка адекватности модели на всех режимах работы автомагистрали. Графики представляют из себя тепловые карты, по оси x которых отложено расстояние от начала участка магистрали, по оси y – время. В конце автомагистрали всегда находится небольшая ветвь, представляющая собой сток.

6.1. Прямая дорога

Для начала рассмотрим поведение модели для простой пятиполосной дороги длиной 6 километров без перекрестков с линейно нарастающим вплоть до 150 АТС/мин потоком изображенное на фиг. 1. В модели данная дорога представлена тремя ветвями по 2 километра. Данный эксперимент показывает, что в модели нет существенных краевых эффектов на стыке ветвей.



Фиг. 4. (а) Схема дороги. (б) Тепловая карта автомобилей на пятиполосной дороге со съездом.

6.2. Прямая дорога с сужением и синусоидальным потоком

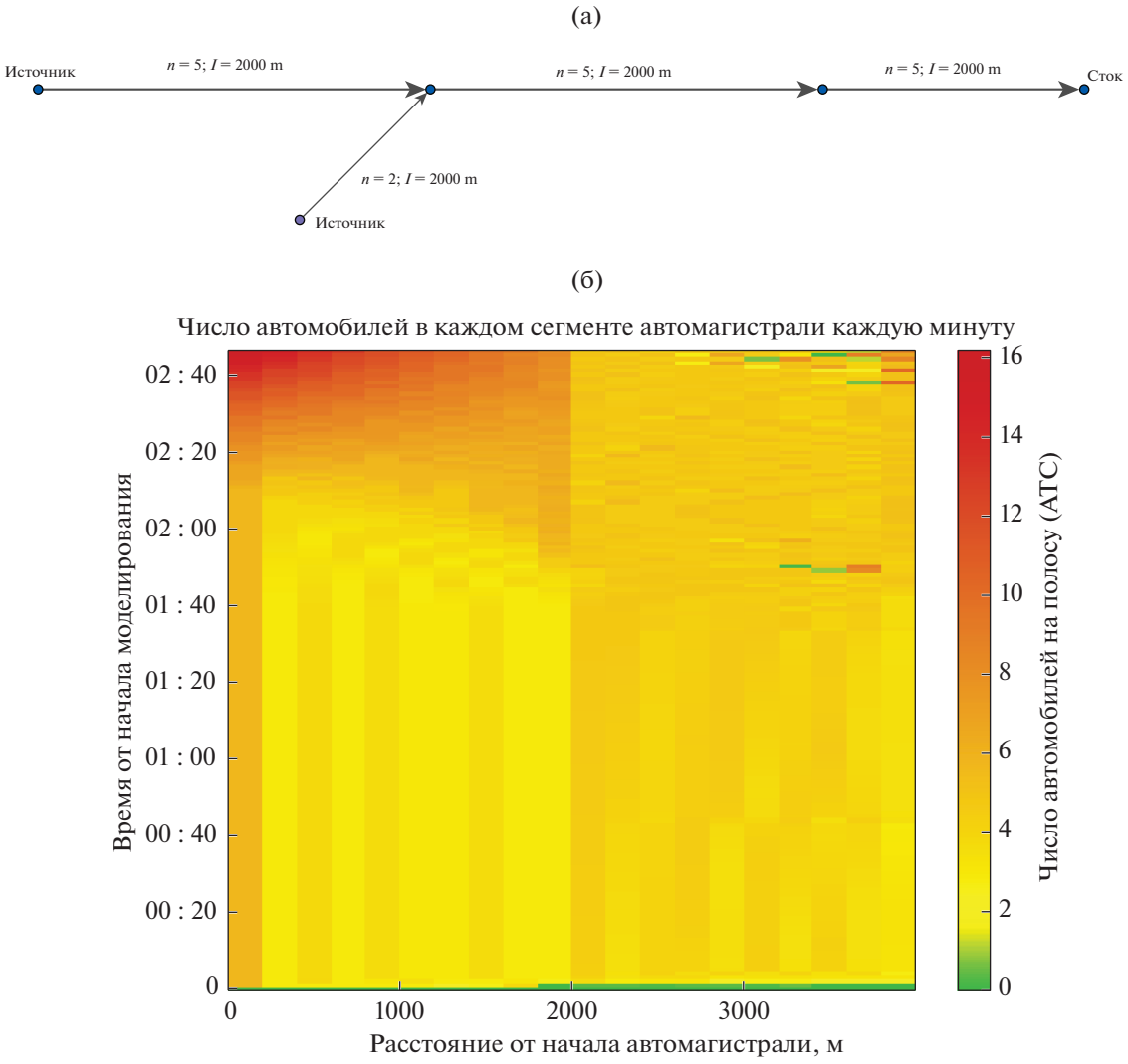
Для следующего эксперимента возьмем прямой участок пятиполосной дороги с сужением до двухполосной. В данном эксперименте с целью рассмотрения как процесса формирования затора, так и его исчезновения пустим на вход синусоидальный поток с периодом, равным времени моделирования, и амплитудой в 85 АТС/мин. Результат моделирования можно наблюдать на фиг. 2. На графике видно, что при уменьшении потока на сегменте, соответствующем двухполосной дороге, наблюдается разрыв потока АТС, который мы связываем с групповыми эффектами модели.

6.3. Прямая дорога с пропадающим сужением

Также рассмотрим ситуацию, когда при постоянном потоке в 100 АТС/мин на пятиполосной дороге с сужением до двухполосной данное сужение в середине моделирования пропадает. Такая ситуация может сложиться, например, при прекращении ремонтных работ или устранении аварии. Результаты моделирования можно наблюдать на фиг. 3.

6.4. Перекресток со съездом

Промоделируем оба варианта перекрестков, возможных в предложенной модели. Перекресток со съездом и перекресток с въездом. В обоих случаях основная автомагистраль – пятиполосная. Въезд или съезд однополосные.



Фиг. 5. (а) Схема дороги. (б) Тепловая карта автомобилей на пятиполосной дороге с въездом с постепенно нарастающим потоком с него.

В эксперименте со съездом входной поток – 65 АТС/мин. Доля съезжающих автомобилей линейно растет с 20% до 60%. На фиг. 4 видно, что из-за недостаточной пропускной способности съезда на основной автомагистрали образуется пробка.

6.5. Перекресток с въездом

В эксперименте со въездом поток на автомагистрали – 140 АТС/мин, поток на въезде линейно растет от 20 до 50 АТС/мин. В данном случае также образуется пробка на основной автомагистрали, что видно на фиг. 5.

7. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В работе изложен новый алгоритм моделирования числа проехавших АТС для задачи моделирования транспортных потоков и проведены пять экспериментов, показывающих его состоятельность.

Эксперименты из разд. 6 показывают работоспособность модели для моделирования всевозможных конфигураций автомагистрали при любом потоке АТС на ней. Показано, что модель адекватно симулирует поведение АТС на автомагистрали как в ситуации достаточной ее про-

пускной способности, так и при ее превышении, а также моделирует различные варианты образования заторных ситуации как при распространении пробки из-за проблем на магистрали на фиг. 2, так и по причине недостаточной пропускной способности прилегающих съездов 4.

Из недостатков стоит отметить отсутствие каких-либо ограничений на генерируемые источниками автомобили в зависимости от уже имеющейся загруженности ветви автомагистрали, а также невозможность в текущем состоянии моделировать перекрестки с несколькими въездами и съездами, инцидентными одному узлу.

С точки зрения развития модели требуется на реальных данных проверить зависимость точности моделирования от размера группы АТС. Провести эксперименты с различными динамическими размерами автомобилей, а также проверить необходимость использования ограничения на ускорение группы АТС в зависимости от ее скорости вместо статического, использованного в данной работе. Также требуются улучшения в работе с группой АТС, моделирующей очередь на съезд с автомагистрали.

СПИСОК ЛИТЕРАТУРЫ

1. *Старожилец В.М., Чехович Ю.В.* Комплексирование данных из разнородных источников в задачах моделирования транспортных потоков // Машинное обучение и анализ данных. 2016. Т. 2. № 3. С. 260–276.
2. *Lighthill M.J., Whitham G.B.* On kinematic waves. II. A theory of traffic flow on long crowded roads // P. Roy. Soc. Lond. A Mat. 1955. V. 229. P. 317–345.
3. *Richards P.I.* Shock waves on the highway // Oper. Res. 1956. V. 4. № 1. P. 42–51.
4. *Whitham J.B.* Linear and nonlinear waves. Hoboken: Wiley, 1974. 656 p.
5. *Daganzo C.F.* Requiem for second-order fluid approximations of traffic flow // Transport. Res. B Meth. 1995. V. 29. № 4. P. 277–286.
6. *Payne H.J.* Models of freeway traffic and control // Math. Models Public Syst. 1998. № 4. P. 51–61.
7. *Papageorgiou M.* Some remarks on macroscopic traffic flow modelling // Transport. Res. A Pol. 1998. V. 32. № 5. P. 323–329.
8. *Aw A., Michel Rascle M.* Resurrection of “second order” models of traffic flow // SIAM J. Appl. Math. 2000. V. 60. № 3. P. 916–938.
9. *Zhang M.* A non-equilibrium traffic model devoid of gas-like behavior // Transport. Res. B Meth. 2002. V. 36. № 3. P. 275–290.
10. *Zhang M.* Anisotropic property revisited – does it hold in multi-lane traffic? // Transport. Res. B Meth. 2003. V. 37. № 6. P. 561–577.
11. *Siebel F., Mauser W.* On the fundamental diagram of traffic flow // SIAM J. Appl. Math. 2006. V. 66. № 4. P. 1150–1162.
12. *Siebel F., Mauser W.* Synchronized flow and wide moving jams from balanced vehicular traffic // Phys. Rev. E. 2006. V. 73. № 6. P. 066108.
13. *Dey P.P., Chandra S., Gangopadhyay S.* Simulation of mixed traffic flow on two-lane roads // J. of Transportation Engng. 2008. V. 134. № 9. P. 361–369.
14. *Guo Hong-Wei, Gao Zi-You, Zhao Xiao-Mei, Xie Dong-Fan*. Dynamics of motorized vehicle flow under mixed traffic circumstance // Communications in Theoretical Physics. 2011. V. 55. № 4. P. 719.
15. *Gundaliya P.J., Tom V. Mathew, Sunder Lall Dhingra.* Heterogeneous traffic flow modelling for an arterial using grid based approach // J. of Advanced Transportation. 2008. V. 42. № 4. P. 467–491.
16. *Lan L.W., Chang C. -W., Gangopadhyay S.* Inhomogeneous cellular automata modeling for mixed traffic with cars and motorcycles // J. of advanced transportation. 2005. V. 39. № 3. P. 323–349.
17. *Алексеенко А.Е., Холодов Я.А., Холодов А.С., Горева А.И., Васильев М.О., Чехович Ю.В., Мишин В.Д., Старожилец В.М.* Разработка, калибровка и верификация модели движения трафика в городских условиях. Ч. I // Компьютерные исследования и моделирование. 2015. Т. 7. № 6. С. 1185–1203.
18. *Kerner B.* The physics of traffic. Berlin: Springer, 2004. 681 p.
19. *Гасников А.В. и др.* Введение в математическое моделирование транспортных потоков. М.: Litres, 2015. С. 89.
20. *Long G.* Acceleration characteristics of starting vehicles // Transportation Research Record. 2000. V. 1737. № 1. P. 58–70.