

ISSN 2712-8687

ПРОБЛЕМЫ УПРАВЛЕНИЯ

5/2021

CONTROL  SCIENCES

РЕДАКЦИОННЫЙ СОВЕТ

С. Н. Васильев, академик РАН,
И. А. Каляев, академик РАН,
В. А. Левин, академик РАН,
Н. А. Махутов, чл.-корр. РАН,
А. Ф. Резчиков, чл.-корр. РАН,
Е. А. Федосов, академик РАН

РЕДКОЛЛЕГИЯ

Ф. Т. Алескеров, д-р техн. наук,
В. Н. Афанасьев, д-р техн. наук,
Н. Н. Бахтадзе, д-р техн. наук,
В. Н. Бурков, д-р техн. наук,
В. М. Вишневский, д-р техн. наук,
В. В. Клочков, д-р экон. наук,
С. А. Краснова, д-р техн. наук,
Н. В. Кузнецов, д-р физ.-мат. наук,
О. П. Кузнецов, д-р техн. наук,
В. В. Кульба, д-р техн. наук,
А. Г. Кушнер, д-р физ.-мат. наук,
А. А. Лазарев, д-р физ.-мат. наук,
В. Г. Лебедев, д-р техн. наук,
В. Е. Лепский, д-р психол. наук,
Н. Е. Максимова, канд. техн. наук
(ответственный секретарь),
А. С. Мандель, д-р техн. наук,
Р. В. Мещеряков, д-р техн. наук,
А. И. Михальский, д-р биол. наук,
Д. А. Новиков, чл.-корр. РАН
(гл. редактор),
Б. В. Павлов, д-р техн. наук,
Ф. Ф. Пашченко, д-р техн. наук
(зам. гл. редактора),
Л. Б. Рапопорт, д-р физ.-мат. наук,
С. В. Ратнер, д-р экон. наук,
Е. Я. Рубинович, д-р техн. наук,
В. Ю. Рутковский, д-р техн. наук,
М. В. Хлебников, д-р физ.-мат. наук,
А. Д. Цвиркун, д-р техн. наук,
П. Ю. Чеботарёв, д-р физ.-мат. наук,
И. Б. Ядыкин, д-р техн. наук

РУКОВОДИТЕЛИ РЕГИОНАЛЬНЫХ РЕДСОВЕТОВ

Владивосток – О. В. Абрамов, д-р техн. наук,
Волгоград – А. А. Воронин, д-р физ.-мат. наук,
Воронеж – С. А. Баркалов, д-р техн. наук,
Курск – С. Г. Емельянов, д-р техн. наук,
Липецк – А. К. Погодаев, д-р техн. наук,
Пермь – В. Ю. Столбов, д-р техн. наук,
Ростов-на-Дону – Г. А. Угольниковский,
д-р техн. наук,
Самара – М. И. Гераськин, д-р экон. наук,
Саратов – В. А. Твердохлебов, д-р техн. наук,
Уфа – Б. Г. Ильясов, д-р техн. наук

ADVISORY BOARD

E. A. Fedosov, Academician of RAS¹,
I. A. Kalyaev, Academician of RAS,
V. A. Levin, Academician of RAS,
N. A. Makhutov, Corr. Member of RAS,
A. F. Rezhnikov, Corr. Member of RAS,
S. N. Vassilyev, Academician of RAS

EDITORIAL BOARD

V. N. Afanas'ev, Dr. Sci. (Tech.),
F. T. Aleskerov, Dr. Sci. (Tech.),
N. N. Bakhtadze, Dr. Sci. (Tech.),
V. N. Burkov, Dr. Sci. (Tech.),
P. Yu. Chebotarev, Dr. Sci. (Phys.-Math.),
V. V. Klochkov, Dr. Sci. (Econ.),
M. V. Khlebnikov, Dr. Sci. (Phys.-Math.),
S. A. Krasnova, Dr. Sci. (Tech.),
V. V. Kulba, Dr. Sci. (Tech.),
A. G. Kushner, Dr. Sci. (Phys.-Math.),
N. V. Kuznetsov, Dr. Sci. (Phys.-Math.),
O. P. Kuznetsov, Dr. Sci. (Tech.),
A. A. Lazarev, Dr. Sci. (Phys.-Math.),
V. G. Lebedev, Dr. Sci. (Tech.),
V. E. Lepskiy, Dr. Sci. (Psych.),
N. E. Maximova, Cand. Sci. (Tech.),
Executive Editor-in-Chief,
A. S. Mandel, Dr. Sci. (Tech.),
R. V. Meshcheryakov, Dr. Sci. (Tech.),
A. I. Michalski, Dr. Sci. (Biol.),
D. A. Novikov, Corr. Member of RAS,
Editor-in-Chief,
F. F. Pashchenko, Dr. Sci. (Tech.),
Deputy Editor-in-Chief,
B. V. Pavlov, Dr. Sci. (Tech.),
L. B. Rapoport, Dr. Sci. (Phys.-Math.),
S. V. Ratner, Dr. Sci. (Econ.),
E. Ya. Rubinovich, Dr. Sci. (Tech.),
V. Yu. Rutkovskii, Dr. Sci. (Tech.),
A. D. Tsvirkun, Dr. Sci. (Tech.),
V. M. Vishnevsky, Dr. Sci. (Tech.),
I. B. Yadykin, Dr. Sci. (Tech.)

LEADERS OF REGIONAL BOARDS

Kursk – S. G. Emelyanov, Dr. Sci. (Tech.),
Lipetsk – A. K. Pogodaev, Dr. Sci. (Tech.),
Perm – V. Yu. Stolbov, Dr. Sci. (Tech.),
Rostov-on-Don – G. A. Ougolnitsky,
Dr. Sci. (Tech.),
Samara – M. I. Geraskin, Dr. Sci. (Econ.),
Saratov – V. A. Tverdokhlebov, Dr. Sci. (Tech.),
Ufa – B. G. Ilyasov, Dr. Sci. (Tech.),
Vladivostok – O. V. Abramov, Dr. Sci. (Tech.),
Volgograd – A. A. Voronin, Dr. Sci. (Phys.-Math.),
Voronezh – S. A. Barkalov, Dr. Sci. (Tech.)

¹Russian Academy of Sciences.



CONTROL SCIENCES
Научно-технический
журнал

6 номеров в год

ISSN 1819-3161 (Print)

ISSN 2712-8687 (Online)

Издается с 2003 года

УЧРЕДИТЕЛЬ и ИЗДАТЕЛЬ

Федеральное государственное
бюджетное учреждение науки
Институт проблем управления
им. В.А. Трапезникова РАН

Главный редактор
чл.-корр. РАН
Д.А. Новиков

Заместитель главного редактора
Ф.Ф. Пашенко

Ответственный секретарь
Н.Е. Максимова

Выпускающий редактор
Л.В. Петракова

Адрес редакции
117997, ГСП-7, Москва,
ул. Профсоюзная, д. 65, к. 410

Тел./факс (495) 198-17-20, доб. 1410

E-mail: pu@ipu.ru

Интернет: <http://pu.mtas.ru>
<http://controlsciences.org>

Опубликовано: 25 октября 2021 г.

Свидетельство о регистрации
ПИ № ФС 77-49203 от 30 марта 2012 г.
выдано Министерством Российской
Федерации по делам печати,
телерадиовещания и средств массовых
коммуникаций

Свидетельство о регистрации
Эл № ФС 77-80482 от 17 февраля 2021 г.
выдано Федеральной службой
по надзору в сфере связи,
информационных технологий и
массовых коммуникаций

Журнал входит в RSCI на платформе
Web of Science и Перечень
рецензируемых научных изданий ВАК

Журнал включен в Российский индекс
научного цитирования (РИНЦ).
На сайте Научной электронной
библиотеки (www.elibrary.ru) доступны
полные тексты статей.

© Федеральное государственное
бюджетное учреждение науки
Институт проблем управления
им. В.А. Трапезникова РАН

ПРОБЛЕМЫ УПРАВЛЕНИЯ

5.2021

СОДЕРЖАНИЕ

Обзоры

Лепский А.Е. Анализ противоречивости информации
в теории функций доверия. Ч. 1. Внешний конфликт 3

Системный анализ

Белов М.В., Новиков Д.А. Структура креативной
деятельности 20

Анализ и синтез систем управления

Глушенко А.И., Петров В.А., Ласточкин К.А.
Адаптивное нейросетевое управление нелинейными объектами
с дефицитом каналов управления на примере двухколесного
балансирующего робота 34

Гулюкина С.И., Уткин В.А. Управление реактором
с непрерывным перемешиванием в условиях
неопределенности и с учетом ограничений
на фазовые переменные и управления 48

Управление в медико-биологических системах

Тасейко О.В., Черных Д.А. Оценка влияния факторов
окружающей среды на показатели смертности населения
старших возрастных групп на примере г. Красноярск 60

Информационные технологии в управлении

Подлазов В.С. Неблокируемые отказоустойчивые
дуальные фотонные коммутаторы
широкой масштабируемости 70

Краткие сообщения

Чернов И.В. Сценарные методы повышения эффективности
реализации жизненного цикла программно-целевого управления
(анализ концепции) 88



CONTROL SCIENCES
Scientific Technical
Journal

6 issues per year

ISSN 1819-3161 (Print)

ISSN 2712-8687 (Online)

Published since 2003

FOUNDER and PUBLISHER

V.A. Trapeznikov

Institute of Control Sciences
of Russian Academy of Sciences

Editor-in-Chief

D.A. Novikov, Corr. Member of RAS

Deputy Editor-in-Chief

F.F. Pashchenko

Executive Editor-in-Chief

N.E. Maximova

Editor

L.V. Petrakova

Editorial address

65 Profsoyuznaya st., office 410,
Moscow 117997, Russia

☎/📠 +7(495) 198-17-20, ext. 1410

✉ pu@ipu.ru

URL: <http://pu.mtas.ru>

<http://controlsciences.org>

Published: October 25, 2021

Registration certificate of

ПИ № ФС 77-49203 от 30 March 2012

issued by the Ministry of Press,
Broadcasting, and Mass Media
of the Russian Federation

Registration certificate of

Эл № ФС 77-80482 of 17 February 2021

issued by the Federal Service
for Supervision of Communications,
Information Technology, and Mass Media

The Journal is indexed in RSCI (Russian
Science Citation Index) on the platform
Web of Science and in the list of peer-
reviewed scientific publications of HAC

On the website of the Scientific electronic
library (www.elibrary.ru) full texts of
articles are available

© V.A. Trapeznikov

Institute of Control Sciences

of Russian Academy of Sciences

CONTROL SCIENCES

5.2021

CONTENTS

Surveys

Lepskiy, A.E. Analysis of Information Inconsistency
in Belief Function Theory. Part I: External Conflict 3

Systems Analysis

Belov, M.V. and Novikov, D.A. The Structure
of Creative Activity 20

Analysis and Design of Control Systems

Glushchenko, A.I., Petrov, V.A., and Lastochkin, K.A.
Adaptive Neural-Network-Based Control
of Nonlinear Underactuated Plants:
An Example of a Two-Wheeled Balancing Robot 34

Gulyukina, S.I. and Utkin, V.A. A Block Approach to CSTR
Control under Uncertainty, State-Space and Control Constraints . . . 48

Control in Medical and Biological Systems

Taseiko, O.V. and Chernykh, D.A. Assessing the Impact
of Environmental Factors on Mortality in Elder Age Groups:
An Example of Krasnoyarsk 60

Information Technologies in Control

Podlazov, V.S. Non-blocking Fault-Tolerant Dual Photon
Switches with High Scalability 70

Brief Communications

Chernov, I.V. Scenario Methods to Improve the Efficiency
of Implementing the Life Cycle of Program-Target Management:
A Conceptual Analysis 88

АНАЛИЗ ПРОТИВОРЕЧИВОСТИ ИНФОРМАЦИИ В ТЕОРИИ ФУНКЦИЙ ДОВЕРИЯ. Ч. 1. ВНЕШНИЙ КОНФЛИКТ¹

А.Е. Лепский

Аннотация. Приведен аналитический обзор результатов по анализу противоречивости информации в рамках теории функций доверия (теории свидетельств Демпстера – Шейфера), который интенсивно развивается в течение последних 10–15 лет. В первой части обзора рассматривается понятие меры внешнего конфликта тел свидетельств. Обсуждаются понятия конфликтных и неконфликтных тел свидетельств, а также основные требования, предъявляемые к мерам внешнего конфликта. Анализируются различные аксиомы меры внешнего конфликта. Приведены результаты по общему виду мер внешнего конфликта, удовлетворяющих системе аксиом. Рассмотрены различные способы построения мер внешнего конфликта: метрический, алгебраический и структурный подходы, оценивание с помощью правил комбинирования. Обсуждаются: проблема робастного оценивания внешнего конфликта, соотношение между мерой конфликта и метрикой на множестве тел свидетельств, согласованность правил комбинирования тел свидетельств и мер конфликта. Материал статьи проиллюстрирован большим количеством примеров.

Ключевые слова: теория функций доверия, правила комбинирования, конфликтность тел свидетельств, мера внешнего конфликта.

ВВЕДЕНИЕ

Во многих задачах анализа данных и принятия решений информацию, поступающую от различных источников (экспертов, аналитиков, мониторинговых данных и пр.), удобно описывать с помощью так называемых тел свидетельств, которые являются основным понятием в теории функций доверия (теории свидетельств Демпстера – Шейфера) [1, 2]. Тело свидетельств представляет собой совокупность множества некоторых подмножеств (так называемых фокальных элементов) из универсального множества и определенную на этих подмножествах функцию масс. В настоящее время теория функций доверия находит многочисленные приложения в анализе данных, обработке экспертной информации, задачах принятия решений и т. д. Это обусловлено как понятной, интерпретируемой, связанной с классическим вероятностным подходом основой самой теории, так и наличием широкого инструментария для моделирования и оценивания таких важных факторов,

как неопределенность и неточность информации, надежность источников информации, противоречивость информации из разных источников, степень незнания и т. д. Кроме того, в теории функций доверия широко развит аппарат комбинирования тел свидетельств (правила комбинирования Демпстера, Ягера, Инагаки и пр.) с учетом перечисленных выше факторов.

В настоящей статье проанализированы основные результаты, полученные за последние 15 лет, которые связаны с анализом противоречивости (конфликтности) источников информации в рамках теории функций доверия. Противоречивость информации, полученной из разных источников, является важной априорной характеристикой этих источников, которая должна учитываться в задаче принятия решения о возможности агрегирования этих источников или выбора источников из некоторого множества для агрегирования. Например, если один аналитик прогнозирует, что стоимость акций некоторой компании будет через месяц в промежутке [40, 50], а другой – в промежутке [70, 85] (все значения указаны в условных единицах), то можно говорить о большом (внешнем) конфликте между соответствующими свидетельствами.

¹Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-11-50077.

Хотя само понятие конфликта (противоречивости) источников информации появилось еще в пионерских работах по теории свидетельств, в последнее время исследование понятия конфликтности в рамках теории функций доверия привлекло внимание многих исследователей и выделилось в отдельное научное направление. В рамках этого направления, в частности, исследовалась аксиоматика мер внешнего и внутреннего конфликта, различные способы оценивания конфликта, свойства мер конфликта, а также рассматривался ряд прикладных задач, связанных с оцениванием конфликта. В частности, оценивание величины конфликтности информации находит применение в задачах управления конфликтами [3], при анализе экспертной информации и рекомендаций [4–6], при анализе результатов голосований [7], в задаче разработке торговых стратегий [8], в задаче регистрации изображений [9] и т. д. Кроме того, связанное с конфликтом понятие метрики между телами свидетельств используется в задачах аппроксимации тел свидетельств [10–13] (которая состоит в том, что «сложное» тело свидетельств с большим числом фокальных элементов заменяется телом свидетельств с меньшим числом фокальных элементов), в задачах классификации [14] и кластеризации [15] данных и т. д. В силу многообразия таких задач, в настоящей статье прикладные аспекты анализа противоречивости в рамках теории функций доверия не рассматриваются.

Основная часть статьи имеет следующую структуру. В § 1 напоминаются основные положения теории функций доверия. В § 2 рассматриваются правила комбинирования тел свидетельств. В § 3 обсуждаются понятия конфликтных и неконфликтных тел свидетельств, а также требования, предъявляемые к мерам внешнего конфликта. В § 4 анализируются различные аксиомы меры конфликта и обсуждается общий вид такой меры, удовлетворяющей некоторой системе аксиом. В § 5 рассматриваются различные способы оценивания внешнего конфликта: метрический (п. 5.1), структурный (п. 5.2), алгебраический (п. 5.3), оценивание конфликта с помощью правил комбинирования (п. 5.4). В § 6 рассматривается робастная процедура оценивания конфликтности. Наконец, в заключении подведены некоторые результаты данного исследования.

Кроме конфликтности между телами свидетельств, рассматривают и конфликтность информации, предоставляемой одним свидетельством (внутренний конфликт). Этому понятию будет посвящена вторая часть обзора.

1. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ДЕМПСТЕРА – ШЕЙФЕРА

Пусть $X = \{x_1, \dots, x_n\}$ – некоторое конечное множество, 2^X – множество всех подмножеств из множества X . Базовым доверительным назначением (БДН, англ. *basic belief assignment*), или функцией масс, в теории Демпстера – Шейфера называют функцию множеств $m: 2^X \rightarrow [0, 1]$, которая удовлетворяет условию $\sum_{A \in 2^X} m(A) = 1$.

Как правило, рассматриваются нормализованные БДН, в которых $m(\emptyset) = 0$. Ненормализованные БДН рассматриваются в рамках так называемой трансферабельной модели доверия (*Transferable Belief Model*), введенной в статье [16]. В этом случае значение $m(\emptyset) > 0$ интерпретируется как мера доверия к тому, что истинное значение $x \notin X$.

Подмножество $A \subseteq X$ называют фокальным элементом БДН m , если $m(A) > 0$. Пару $F = (\mathcal{A}, m)$ из множества всех фокальных элементов $\mathcal{A} = \{A\}$ и соответствующей БДН $m(A)$, $A \in \mathcal{A}$, называют телом свидетельства. Пусть $\mathcal{F}(X)$ – множество всех тел свидетельств на X , а $\mathcal{P}(X)$ – множество всех вероятностных мер на X .

Для случаев, когда тело свидетельств $F = (\mathcal{A}, m)$ известно, в теории Демпстера – Шейфера вводятся в рассмотрение некоторые функции множеств, важнейшими из которых являются функция доверия (*belief function*) $Bel(A) = \sum_{B \subseteq A} m(B)$ и двойственная к ней функция правдоподобия $Pl(A) = 1 - Bel(A^c)$, где A^c – дополнение множества A . Сама функция масс может быть однозначно восстановлена по функции доверия с помощью преобразования Мебиуса $m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B)$. Всегда верно, что $Bel(A) \leq Pl(A) \quad \forall A \subseteq X$, а длина интервала $[Bel(A), Pl(A)]$ определяет степень неопределенности события A [4]. Будем обозначать функции доверия и правдоподобия через Bel_F и Pl_F соответственно, если надо подчеркнуть их зависимость от тела свидетельств $F = (\mathcal{A}, m)$. Функция $Pl(x) = \sum_{A \in \mathcal{A}: x \in A} m(A)$, $x \in X$, называется контурной функцией тела свидетельств.

Тело свидетельств $F = (\mathcal{A}, m)$ и соответствующие ему функции множеств могут быть заданы в векторной форме: $\mathbf{m}$ – $2^{|X|}$ -мерный вектор, координатами которого являются значения функции



масс $m(A)$, $A \in 2^X$, для некоторого упорядочения всех подмножеств из 2^X ; $\mathbf{PI}$ – $2^{|X|}$ -мерный вектор значений $PI(A)$, $A \in 2^X$, и т. д.

Существуют две основные интерпретации БДН и, соответственно, два подхода к построению теории функций доверия. Согласно первой, восходящей к работе [1], $m(A)$ интерпретируется как вероятность случайного события, значением которого является множество A (так называемое случайное множество, англ. *random set* [17]). В статистическом смысле в этом случае БДН рассматривается как относительная частота того, что истинная альтернатива принадлежит множеству A : $m(A) = q(A)/N$, где $q(A)$ – количество наблюдаемых множеств $A \subseteq X$; N – общее количество наблюдений. В альтернативном подходе, развитом в работе [2], сначала вводятся функции доверия и правдоподобия как функции множеств, удовлетворяющие ослабленной аксиоматике вероятностной меры. Затем эти функции используются для оценивания степени неопределенности источников информации (свидетельства). Поэтому теория Демпстера – Шейфера также называется теорией свидетельств или теорией функций доверия.

В статье [18] показано, что

$$Bel(A) = \inf_{P \in \mathcal{P}_{Bel}} P(A), \quad Pl(A) = \sup_{P \in \mathcal{P}_{Bel}} P(A), \quad A \subseteq X,$$

где $\mathcal{P}_{Bel}$ – множество вероятностных мер $P \in \mathcal{P}(X)$, согласованных с функцией Bel , т. е. $Bel(A) \leq P(A) \forall A \subseteq X$. Обратное утверждение неверно, т. е. нижняя грань множества вероятностных мер может и не быть функцией доверия. Множество $\mathcal{P}_{Bel}$ еще называют кредальным множеством функции Bel .

Пример 1. Предположим, что 10 экспертов дают прогноз о перспективности развития трех технологий $X = \{a, b, c\}$: три эксперта высказались в пользу технологий a или b , два – в пользу технологии b , четыре – в пользу технологий b или c и один – в пользу технологии c . Тогда, если не предполагать известным распределение между альтернативами (неопределенность информации), то функции масс равны

$$m(\{a, b\}) = \frac{3}{10}, \quad m(\{b, c\}) = \frac{4}{10}, \\ m(\{b\}) = \frac{2}{10}, \quad m(\{c\}) = \frac{1}{10}.$$

Таким образом, имеем тело свидетельств $F = (\mathcal{A}, m)$ с множеством фокальных элементов $\mathcal{A} = \{\{a, b\}, \{b, c\}, \{b\}, \{c\}\}$. По этому телу свидетельств можно найти значения функций $Bel(A)$ и $Pl(A)$

$\forall A \subseteq X$. В частности, получим, что $Bel(\{a\}) = 0$, $Pl(\{a\}) = 0,3$, $Bel(\{b\}) = 0,2$, $Pl(\{b\}) = 0,9$, $Bel(\{c\}) = 0,1$, $Pl(\{c\}) = 0,5$. Эти числа можно считать нижними и верхними оценками вероятности перспективности развития отдельных технологий: $0 \leq P(\{a\}) \leq 0,3$, $0,2 \leq P(\{b\}) \leq 0,9$, $0,1 \leq P(\{c\}) \leq 0,5$. Причем наибольшую неопределенность имеет прогноз технологии b . ♦

Телу свидетельств $F = (\mathcal{A}, m)$ можно поставить в соответствие так называемую пигнистическую вероятность (*pignistic probability*) $Bet_F \in \mathcal{P}(X)$ [19], значения которой на элементах $x_i \in X$, $i = 1, \dots, n$, равны вероятности этих событий при условии, что в пределах фокальных элементов случайные величины распределены равномерно:

$$Bet_F(x_i) = \sum_{A \in \mathcal{A}, x_i \in A} \frac{m(A)}{|A|}, \quad i = 1, \dots, n.$$

Эти значения совпадают со значениями Шепли [20]

$$Bet_F(x_i) = \sum_{A \in \mathcal{A}, x_i \in A} \frac{(n - |A|)! (|A| - 1)!}{n!} (Bel(A) - Bel(A \setminus \{x_i\}))$$

соответствующей функции доверия Bel . Пигнистическая вероятность произвольного множества $B \subseteq X$ будет равна $Bet_F(B) = \sum_A \frac{|A \cap B|}{|A|} m(A)$. Заметим, что $Bet_F \in \mathcal{P}_{Bel}$.

Теория функций доверия рассматривается не только на конечном множестве X , но и, например, на числовой оси (так называемые непрерывные структуры доверия) [21] или в случае нечетких фокальных элементов [22].

В приложениях рассматривают различные частные случаи функций доверия и соответствующих им тел свидетельств, которые отражают простые и наиболее популярные структуры высказываний о принадлежности истинной альтернативы некоторому множеству. В частности, функция доверия (и тело свидетельств) называется:

– категоричной (*categorical*), если она имеет только один фокальный элемент; соответствующее тело свидетельств является простейшим, и будем обозначать его так: $F_A = (A, 1)$;

– бессодержательной (*vacuous*), если единственным фокальным элементом этой функции является все множество X , $F_X = (X, 1)$; такое тело свидетельств не несет никакой информации о принадлежности истинной альтернативы какому-либо подмножеству из множества X (в этом случае $Bel(A) = 0$ и $Pl(A) = 1 \forall A \neq \emptyset, X$);

– консонантной (*consonant*), если ее фокальные элементы являются вложенными, т. е. $\forall A, B \in \mathcal{A}$: $A \subseteq B$ или $B \subseteq A$; соответствующее тело свидетельств является «уточняющим»;

– простой (*simple*), если БДН имеет не более двух фокальных элементов и в случае, если их два, то X – один из них; это тело свидетельств состоит из простейшего содержательного высказывания « $x \in A$ » с некоторой степенью доверия $m(A)$ и бессодержательного высказывания со степенью доверия $1 - m(A)$;

– догматической (*dogmatic*), если $X \notin \mathcal{A}$ (т. е. тело свидетельств не содержит бессодержательного высказывания: $m(X) = 0$).

Любое тело свидетельств $F = (\mathcal{A}, m)$ можно представить в виде $F = \sum_{A \in \mathcal{A}} m(A) F_A$, т. е. как выпуклую комбинацию категоричных тел свидетельств. Например, тело свидетельств из примера 1 можно записать как $F = 0,2 F_{\{b\}} + 0,1 F_{\{c\}} + 0,3 F_{\{a,b\}} + 0,4 F_{\{b,c\}}$. Простое тело свидетельств можно представить в виде $F_A^\omega = (1 - \omega) F_A + \omega F_X$, где $\omega \in [0, 1]$. В частности, $F_A^0 = F_A$ и $F_A^1 = F_X$. И наоборот, выражение в виде конечной суммы $F = \sum_i \alpha_i F_{A_i}$, $\alpha_i \geq 0 \forall i$, $\sum_i \alpha_i = 1$, задает некоторое тело свидетельств.

Количество незнания в информации, содержащейся в теле свидетельств F , можно оценить с помощью так называемых индексов неточности [23], которые принимают тем большие значения, чем больше фокальных элементов большой мощности и с большей массой содержится в теле свидетельств. Примером такого индекса является обобщенная мера Хартли [24, 25] $H(F) = -\sum_{A \in \mathcal{A}} m(A) \log_2 |A|$. Далее будем использовать нормированную обобщенную меру Хартли $H_0(F) = H(F) / \log_2 |X|$, $H_0: \mathcal{F}(X) \rightarrow [0, 1]$. Заметим, что $H_0(F) = 1 \Leftrightarrow F = F_X$ и $H_0(F) = 0 \Leftrightarrow F = P$, где $P \in \mathcal{P}(X)$.

Несмотря на свою популярность и востребованность, теория функций доверия подверглась многочисленной критике [26–28], в результате которой был инициирован ряд исследований, нацеленных на уточнение границ применения теории, интерпретации результатов и т. д. В частности, в статье [18] было показано, что многие критические замечания связаны со смешением двух точек зрения на функции доверия: первая рассматривает функцию доверия как обобщение вероятностной меры, а вторая – как способ представления свидетельства.

2. ПРАВИЛА КОМБИНИРОВАНИЯ ТЕЛ СВИДЕТЕЛЬСТВ И ПОНЯТИЕ КОНФЛИКТА

Одним из удобных инструментов в теории свидетельств является возможность комбинирования тел свидетельств, т. е. агрегирования информации, полученной из разных источников. Под правилом комбинирования понимают некоторую операцию $\otimes: \mathcal{F}(X) \times \mathcal{F}(X) \rightarrow \mathcal{F}(X)$. Существует несколько общих схем построения правил комбинирования. Наиболее популярной из них является так называемое обобщенное конъюнктивное правило комбинирования $\otimes_\cap$, которое строится по схеме [29]: $F = F_1 \otimes_\cap F_2$, $F = (\mathcal{A}, m_\cap)$, $F_1 = (\mathcal{A}_1, m_1)$ и $F_2 = (\mathcal{A}_2, m_2)$, где

$$m_\cap(A) = \sum_{B \cap C = A} \tilde{m}(B, C),$$

а значения функции множеств $\tilde{m}: 2^X \times 2^X \rightarrow [0, 1]$ удовлетворяют условиям согласования с телами свидетельств F_1 и F_2 :

$$\begin{aligned} \sum_{C \in 2^X} \tilde{m}(B, C) &= m_1(B), \\ \sum_{B \in 2^X} \tilde{m}(B, C) &= m_2(C), \end{aligned} \quad (1)$$

$$B, C \in 2^X.$$

Последняя система может иметь множество решений – в результате получим множество новых тел свидетельств $F = F_1 \otimes_\cap F_2$, которое будем обозначать через $\mathcal{R}_\cap(F_1, F_2)$.

Пример 2. Пусть $X = \{x_1, x_2\}$ и $F_1 = 0,8 F_{\{x_1\}} + 0,2 F_X$, $F_2 = 0,7 F_{\{x_1\}} + 0,3 F_{\{x_2\}}$. Найдем множество $\mathcal{R}_\cap(F_1, F_2)$. Для этого решим систему

$$\begin{aligned} \sum_C \tilde{m}(\{x_1\}, C) &= 0,8, \quad \sum_C \tilde{m}(X, C) = 0,2, \\ \sum_B \tilde{m}(B, \{x_1\}) &= 0,7, \quad \sum_B \tilde{m}(B, \{x_2\}) = 0,3, \\ \tilde{m}(B, C) &\in [0, 1] \quad \forall B, C \in 2^X. \end{aligned}$$

Так как $\tilde{m}(\cdot, \emptyset) = \tilde{m}(\emptyset, \cdot) = \tilde{m}(\cdot, X) = \tilde{m}(\{x_2\}, \cdot) = 0$, то имеем систему с четырьмя переменными из четырех уравнений и неравенств. Решением этой системы будут числа $\tilde{m}(\{x_1\}, \{x_1\}) = 0,5 + t$, $\tilde{m}(\{x_1\}, \{x_2\}) = 0,3 - t$, $\tilde{m}(X, \{x_1\}) = 0,2 - t$, $\tilde{m}(X, \{x_2\}) = t$, $t \in [0; 0,2]$. Тогда $\mathcal{R}_\cap(F_1, F_2) = \{F_1 \otimes_\cap F_2 = (0,3 - t) F_\emptyset + 0,7 F_{\{x_1\}} + t F_{\{x_2\}} : t \in [0; 0,2]\}$. ♦

В частности, если источники информации независимы, то $\tilde{m}(B, C) = m_1(B) m_2(C) \quad \forall B, C \in 2^X$ и мы получим так называемое ненормализованное правило Демпстера $\otimes_{ND}: m_{ND}(A) = \sum_{B \cap C = A} m_1(B) \times m_2(C) \quad \forall A \in 2^X$. Но в этом случае возможно, что $K = K(F_1, F_2) = m_{ND}(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) > 0$.



Величина $K \in [0, 1]$ называется **канонической мерой конфликта** и характеризует степень конфликтности источников информации, описываемых телами свидетельств F_1 и F_2 : чем больше значение этого параметра, тем более противоречивую информацию предоставляют источники. Если равномерно перераспределить величину конфликта по всем фокальным элементам нового тела свидетельств, то получим классическое правило

$$\text{Демпстера [1]} \quad \otimes_D: m_D(A) = \frac{m_{ND}(A)}{1-K} \quad \forall A \in 2^X \setminus \emptyset.$$

Если источники информации являются полностью конфликтными, т. е. $B \cap C = \emptyset$ для всех $B \in \mathcal{A}_1$, $C \in \mathcal{A}_2$, то правило комбинирования Демпстера неприменимо, поскольку в этом случае $K=1$. Если же отнести величину конфликта к массе всего множества $m_{ND}(X)$ (т. е. увеличить массу бессодержательного высказывания $x \in X$ на величину конфликта), то мы получим правило Ягера [30] $\otimes_Y: m_Y(A) = m_{ND}(A) \quad \forall A \in 2^X \setminus \{\emptyset, X\}$, $m_Y(X) = m_{ND}(X) + K$.

Операция комбинирования $\otimes_{ND}$ является ассоциативной и, следовательно, позволяет комбинировать любое конечное число тел свидетельств. Информацию о ряде других правил комбинирования можно найти в работе [31].

В некотором смысле двойственным к правилу Демпстера является дизъюнктивное правило консенсуса $\otimes_{\cup}$ [32, 33] (тело свидетельств, полученное с помощью такого правила, будем обозначать через $F = F_1 \otimes_{\cup} F_2 = (\mathcal{A}, m_{\cup})$):

$$m_{\cup}(A) = \sum_{B \cup C = A} m_1(B)m_2(C), \quad A \in 2^X.$$

Для конъюнктивного и дизъюнктивного правил комбинирования справедлив аналог закона де Моргана [32]:

$$\overline{F_1 \otimes_{ND} F_2} = \overline{F_1} \otimes_{\cup} \overline{F_2},$$

где дополнение $\overline{F} = (\overline{\mathcal{A}}, \overline{m})$ тела свидетельств $F = (\mathcal{A}, m)$ определяется как $\overline{\mathcal{A}} = \{A^c : A \in \mathcal{A}\}$ и $\overline{m}(A) = m(A^c) \quad \forall A \in \overline{\mathcal{A}}$.

Пример 3. Предположим, что две независимые группы экспертов дают прогноз о перспективности развития трех технологий $X = \{a, b, c\}$. Информация первой группы описывается телом свидетельств $F_1 = 0,2F_{\{b\}} + 0,1F_{\{c\}} + 0,3F_{\{a,b\}} + 0,4F_{\{b,c\}}$ (см. пример 1). Информация второй группы описывается телом свидетельств $F_2 = 0,7F_{\{a,c\}} + 0,3F_{\{a,b,c\}}$. Тогда результаты комбинирования этих тел свидетельств с помощью различных конъюнктивных правил и дизъюнктивного правила консенсуса приведены в табл. 1 (указаны только массы фокальных элементов).

Каноническая мера конфликта для этих тел свидетельств равна $K = m_{DN}(\emptyset) = 0,14$. В последней строке табл. 1 приведены значения индекса неточности H_0 для каждого результата комбинирования. ♦

Замечание 1. Конъюнктивное правило является оптимистичным в следующем смысле. Пусть у нас есть два независимых источника информации. Причем первый источник утверждает, что истинная альтернатива принадлежит множеству A (это задается категоричным телом свидетельств F_A), а второй – множеству B (т. е. имеем категоричное тело свидетельств F_B).

Тогда после конъюнктивного комбинирования этих тел свидетельств получим, что истинная

альтернатива принадлежит множеству $A \cap B$. Если же мы применим дизъюнктивное правило консенсуса, то получим, что истинная альтернатива принадлежит множеству $A \cup B$ (см. работу [34]). В этом смысле данное правило является пессимистичным.

Каждому правилу комбинирования и фиксированному телу свидетельств $F = (\mathcal{A}, m)$ можно поставить в соответствие матрицу (так называемую матрицу свидетельств) [35, 36] $R_F^{\otimes} = \{r_F^{\otimes}(A, B)\}_{A, B \in 2^X}$

Таблица 1

Результаты комбинирования тел свидетельств

A	F_1	F_2	$F_1 \otimes_{ND} F_2$	$F_1 \otimes_D F_2$	$F_1 \otimes_Y F_2$	$F_1 \otimes_{\cup} F_2$
$\emptyset$	–	–	0,14	–	–	–
$\{a\}$	–	–	0,21	0,244	0,21	–
$\{b\}$	0,2	–	0,06	0,07	0,06	–
$\{c\}$	0,1	–	0,38	0,442	0,38	–
$\{a, b\}$	0,3	–	0,09	0,104	0,09	–
$\{a, c\}$	–	0,7	–	–	–	0,07
$\{b, c\}$	0,4	–	0,12	0,14	0,12	–
$\{a, b, c\}$	–	0,3	–	–	0,14	0,93
H_0	0,442	0,742	0,132	0,154	0,272	0,974

с элементами $r_F^{\otimes}(A, B) = (F \otimes F_B)(A)$ (здесь F_B – категоричное тело свидетельств). В частности, популярны матричные представления для ненормализованного правила Демпстера $\otimes_{ND}$ и дизъюнктивного правила консенсуса $\otimes_{\cup}$ (обозначаются $R_F^{\cap}$ и $R_F^{\cup}$ соответственно), которые (в силу сделанного выше замечания) задают полностью информацию как о самом теле свидетельств F (поскольку $F \otimes_{ND} F_X = F$ и $F \otimes_{\cup} F_{\emptyset} = F$), так и о результатах его комбинирования с другими категоричными телами свидетельств. Эти результаты для разных категоричных тел свидетельств записаны в матрице $R_F^{\otimes}$ по столбцам. Матрицы $R_F^{\cap}$ и $R_F^{\cup}$ называются матрицами специализации и обобщения тела свидетельств F соответственно, поскольку эти матрицы содержат информацию обо всех телах свидетельств вида $F_B^{\cap} = (\mathcal{A}_B^{\cap}, m)$ и $F_B^{\cup} = (\mathcal{A}_B^{\cup}, m)$ соответственно, где $\mathcal{A}_B^{\cap} = \{A \cap B : A \in \mathcal{A}\}$, $\mathcal{A}_B^{\cup} = \{A \cup B : A \in \mathcal{A}\}$. Нетрудно видеть, что $R_{F_1 \otimes_{ND} F_2}^{\cap} = R_{F_1}^{\cap} \cdot R_{F_2}^{\cap}$. В работе [35] (см. также доклад [37]) рассматривались матричные операторы, соответствующие некоторым параметрическим семействам конъюнктивных и дизъюнктивных правил – так называемые операторы α -слияния (α -junctions), которые при $\alpha=1$ вырождаются в операторы $R_F^{\cap}$ и $R_F^{\cup}$.

Пример 4. Для тел свидетельств $F_1 = 0,8F_{\{x_1\}} + 0,2F_X$ и $F_2 = 0,7F_{\{x_1\}} + 0,3F_{\{x_2\}}$ на множестве $X = \{x_1, x_2\}$ из примера 2 имеем (строки и столбцы соответствуют подмножествам $\emptyset, \{x_1\}, \{x_2\}, X$):

$$R_{F_1}^{\cap} = \begin{pmatrix} 1 & 0 & 0,8 & 0 \\ 0 & 1 & 0 & 0,8 \\ 0 & 0 & 0,2 & 0 \\ 0 & 0 & 0 & 0,2 \end{pmatrix}, R_{F_2}^{\cap} = \begin{pmatrix} 1 & 0,3 & 0,7 & 0 \\ 0 & 0,7 & 0 & 0,7 \\ 0 & 0 & 0,3 & 0,3 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$R_{F_1}^{\cup} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0,8 & 0,8 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0,2 & 0,2 & 1 & 1 \end{pmatrix}, R_{F_2}^{\cup} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0,7 & 0,7 & 0 & 0 \\ 0,3 & 0 & 0,3 & 0 \\ 0 & 0,3 & 0,7 & 1 \end{pmatrix} \cdot \diamond$$

3. ПОНЯТИЕ КОНФЛИКТНЫХ И НЕКОНФЛИКТНЫХ ТЕЛ СВИДЕТЕЛЬСТВ

В общем случае под мерой внешнего конфликта понимают некоторый функционал

$Con_{ext} : \mathcal{F}(X) \times \mathcal{F}(X) \rightarrow [0, 1]$. Поскольку каждое тело свидетельств определяется множеством фокальных элементов и функцией массы на этих множествах, то и конфликт между телами свидетельств зависит от взаимного расположения между множествами фокальных элементов двух тел свидетельств и значений их функций масс.

Мера конфликта в первую очередь должна отражать экстремальные ситуации конфликтности – быть максимальной в случае полного конфликта двух тел свидетельств и минимальной в случае их неконфликтности.

Естественным условием полной конфликтности является непересечение всех пар фокальных элементов двух тел свидетельств. Точнее, тела свидетельств $F_1 = (\mathcal{A}_1, m_1)$ и $F_2 = (\mathcal{A}_2, m_2)$ называются полностью конфликтными, если $A \cap B = \emptyset \forall A \in \mathcal{A}_1, \forall B \in \mathcal{A}_2$.

Однако возможны различные степени неконфликтности. Так, в работе [38] рассматриваются следующие степени неконфликтности тел свидетельств $F_1 = (\mathcal{A}_1, m_1)$ и $F_2 = (\mathcal{A}_2, m_2)$:

- сильная неконфликтность: $\bigcap_{A \in \mathcal{A}_1 \cup \mathcal{A}_2} A \neq \emptyset$,
- (простая) неконфликтность: $A \cap B \neq \emptyset \forall A \in \mathcal{A}_1, \forall B \in \mathcal{A}_2$,
- слабая неконфликтность: $\mathcal{P}_{Bel_1} \cap \mathcal{P}_{Bel_2} \neq \emptyset$, где Bel_i – функция доверия, соответствующая телу свидетельств $F_i, i=1, 2$.

Известно [38], что из сильной неконфликтности следует (простая) неконфликтность, а из неконфликтности следует слабая неконфликтность. Для категоричных тел свидетельств F_A и F_B все эти понятия неконфликтности сводятся к непустоте пересечения их фокальных элементов: $A \cap B \neq \emptyset$. Но, например, если $F_1 = F_2 = \alpha F_A + (1-\alpha)F_B$, где $A \cap B = \emptyset$ и $\alpha \in (0,1)$, то такие одинаковые тела свидетельств будут только слабо неконфликтными, но не будут (просто) неконфликтными, и тем более не будут сильно неконфликтными.

В соответствии с условиями полного конфликта и неконфликтности желательно чтобы мера внешнего конфликта Con_{ext} удовлетворяла условиям [4, 38, 39]:

E1: $Con_{ext}(F_1, F_2) = Con_{ext}(F_2, F_1) \quad \forall F_1, F_2 \in \mathcal{F}(X)$ (симметричность).

E2: $Con_{ext}(F_1, F_2) = 0$, если F_1 и F_2 слабо неконфликтны.

В частности, любое тело свидетельств F будет слабо неконфликтным с самим собой и с бессо-



держательным телом свидетельств F_X . Поэтому при выполнении условия E2 будут верны утверждения (следствия):

а) $Con_{ext}(F, F) = 0 \quad \forall F \in \mathcal{F}(X)$ (нильпотентность);

б) $Con_{ext}(F_X, F) = 0 \quad \forall F \in \mathcal{F}(X)$ (с незнанием нет конфликта).

E3: $Con_{ext}(F_1, F_2) = 1$, если $A \cap B = \emptyset \quad \forall A \in \mathcal{A}_1, \forall B \in \mathcal{A}_2$ (полный конфликт).

Следующее условие связано с понятием специализации свидетельств [32].

Тело свидетельств $F' = (\mathcal{A}', m')$ называется специализацией тела свидетельств $F'' = (\mathcal{A}'', m'')$ (обозначают: $F' \sqsubseteq F''$), если существует разбиение множества $\mathcal{A}' = \mathcal{A}'_1 \cup \dots \cup \mathcal{A}'_k$, $\mathcal{A}'_i \cap \mathcal{A}'_j = \emptyset \quad \forall i \neq j$, $k = |\mathcal{A}''|$: $\bigcup_{A \in \mathcal{A}'_i} A \subseteq B_i$, $\sum_{A \in \mathcal{A}'_i} m'(A) = m''(B_i)$, $\forall B_i \in \mathcal{A}'', i = 1, \dots, k$.

Другими словами, тело свидетельств F' уточняет (специализирует) тело свидетельств F'' . Последнее в этом случае называют обобщением тела свидетельств F' . Заметим, что $F \sqsubseteq F_X \quad \forall F \in \mathcal{F}(X)$.

Если $F' \sqsubseteq F''$, то неточность специализации F' не больше неточности тела свидетельств F'' для любого индекса неточности (например, относительно обобщенной меры Хартли $H(F') \leq H(F'')$).

E4: $Con_{ext}(F', F) \geq Con_{ext}(F'', F) \quad \forall F, F', F'' \in \mathcal{F}(X)$, $F' \sqsubseteq F''$ (антимонотонность по специализации).

Заметим, что каноническая мера конфликта K удовлетворяет всем перечисленным условиям, кроме E2 (в частности, не будет нильпотентной), но будет удовлетворять условию

E2': $Con_{ext}(F_1, F_2) = 0$, если F_1 и F_2 (просто) неконфликтны.

4. АКСИОМАТИКА МЕРЫ ВНЕШНЕГО КОНФЛИКТА

В статье [40] была рассмотрена аксиоматика меры конфликта, основанная на усилении приведенных выше условий E1 – E4:

A1: $Con_{ext}(F_1, F_2) = Con_{ext}(F_2, F_1) \quad \forall F_1, F_2 \in \mathcal{F}(X)$.

A2: $Con_{ext}(F_1, F_2) = 0 \Leftrightarrow F_1$ и F_2 слабо неконфликтны.

A3: $Con_{ext}(F_1, F_2) = 1 \Leftrightarrow A \cap B = \emptyset \quad \forall A \in \mathcal{A}_1, \forall B \in \mathcal{A}_2$.

A4: $Con_{ext}(F', F) \geq Con_{ext}(F'', F) \quad \forall F, F', F'' \in \mathcal{F}(X)$, $F' \sqsubseteq F''$.

Кроме того, рассматривалась аксиома

A5: если $Con_{ext}(F_1, F_2) = a \in [0, 1]$, то $\exists F_i^{(k)} \in \mathcal{F}(X)$, $k = 1, 2$, $i = 1, 2$: $F_i = (1 - a)F_i^{(1)} + aF_i^{(2)}$, $i = 1, 2$, $Con_{ext}(F_1^{(1)}, F_2^{(1)}) = 0$, $Con_{ext}(F_1^{(2)}, F_2^{(2)}) = 1$. Эта аксиома предполагает, что источники информации можно линейно разделить на полностью конфликтные и слабо неконфликтные части в пропорции, равной величине конфликта относительно данной меры конфликта. Это довольно сильное требование. В частности, каноническая мера конфликта этой аксиоме не удовлетворяет. В этом нетрудно убедиться на конкретном примере.

Справедлива следующая теорема о мере конфликта.

Теорема 1 [40]. Если функционал $Con_{ext} : \mathcal{F}(X) \times \mathcal{F}(X) \rightarrow [0, 1]$ удовлетворяет аксиомам A1 – A5, то он имеет вид:

$$Con_{ext}(F_1, F_2) = \inf \{ m_{\cap}(\emptyset) : F = (\mathcal{A}, m_{\cap}) \in \mathcal{R}_{\cap}(F_1, F_2) \}. \quad (2)$$

Пример 5. Для тел свидетельств $F_1 = 0,8F_{\{x_1\}} + 0,2F_X$ и $F_2 = 0,7F_{\{x_1\}} + 0,3F_{\{x_2\}}$ на множестве $X = \{x_1, x_2\}$, применяя обобщенное конъюнктивное правило комбинирования $\otimes_{\cap}$, получим (см. пример 2), что $m_{\cap}(\emptyset) = 0,3 - t$, $t \in [0; 0,2]$. Поэтому $Con_{ext}(F_1, F_2) = \min \{ 0,3 - t : t \in [0; 0,2] \} = 0,1$. Заметим, что каноническая мера конфликта для этих тел свидетельств $K = 0,24$. ♦

Кроме того, в статье [40] было показано, что мера (2) может быть найдена по формуле

$Con_{ext}(F_1, F_2) = \inf \{ Con_{ext}(P_1, P_2) : P_1 \in \mathcal{P}_{Bel_1}, P_2 \in \mathcal{P}_{Bel_2} \}$, где $\mathcal{P}_{Bel_i}$ – кредальные множества, соответствующие телам свидетельств F_i , $i = 1, 2$, а для вероятностных мер

$$Con_{ext}(P_1, P_2) = 1 - \sum_{i=1}^n \min \{ P_1(x_i), P_2(x_i) \}. \quad (3)$$

Пример 6. Для предыдущего примера имеем: $\mathcal{P}_{Bel_1} = \{ \alpha F_{\{x_1\}} + (1 - \alpha)F_X : 0,8 \leq \alpha \leq 1 \}$, $\mathcal{P}_{Bel_2} = \{ F_2 \} = \{ 0,7F_{\{x_1\}} + 0,3F_{\{x_2\}} \}$. Тогда $Con_{ext}(F_1, F_2) = \inf \{ Con_{ext}(P_1, P_2) : P_1 \in \mathcal{P}_{Bel_1}, P_2 \in \mathcal{P}_{Bel_2} \} = \inf_{0,8 \leq \alpha \leq 1} (1 - \min \{ \alpha; 0,7 \} - \min \{ 1 - \alpha; 0,3 \}) = \inf_{0,8 \leq \alpha \leq 1} (1 - 0,7 - (1 - \alpha)) = 0,1$. ♦

Замечание 2. Для вероятностных мер $P_1, P_2 \in \mathcal{P}(X)$ каноническая мера конфликта K будет равна (ср. с формулой (3))

$$K(P_1, P_2) = \sum_{i=1}^n P_1(x_i) \sum_{j=1, j \neq i}^n P_2(x_j) = \sum_{i=1}^n P_1(x_i)(1 - P_2(x_i)) = 1 - \sum_{i=1}^n P_1(x_i)P_2(x_i).$$

В частности, отсюда видно, что каноническая мера конфликта не будет в общем случае нильпотентной, так как $K(P, P) = 1 - \sum_{i=1}^n P^2(x_i)$, $P \in \mathcal{P}(X)$. В частности, для равномерного дискретного распределения P на X имеем $K(P, P) = 1 - \frac{1}{|X|}$.

В статье [4] рассматривалась аксиоматика меры внешнего конфликта Con_{ext} на произвольном конечном множестве тел свидетельств $M = \{F_1, \dots, F_l\}$, $F_i \in \mathcal{F}(X)$, $i = 1, \dots, l$. Пусть 2^M – множество всех подмножеств из множества M . По определению считаем, что $Con_{ext}(B) = 0$, если $|B| = 1$, $B \in 2^M$ и $Con_{ext}(\emptyset) = 0$. Кроме тех аксиом, которые являются непосредственным обобщением аксиом A1 – A4 на множественный случай, рассматривалась также аксиома монотонности

A6: $Con_{ext}(B) \leq Con_{ext}(C)$, если $B \subseteq C$ и $B, C \in 2^M$.

Например, каноническая мера конфликта в силу ассоциативности может быть продолжена на произвольное конечное множество тел свидетельств M . И нетрудно показать, что на этом множестве каноническая мера конфликта будет удовлетворять аксиоме монотонности A6.

5. СПОСОБЫ ОЦЕНИВАНИЯ ВНЕШНЕГО КОНФЛИКТА

Существует несколько подходов к оцениванию конфликтности тел свидетельств. Условно их можно разделить на метрический, структурный, алгебраический, на основе правил комбинирования.

5.1. Метрические способы оценивания конфликта

Метрический подход является одним из наиболее популярных способов оценивания конфликта [41, 42]. Метрики могут вводиться как на самих телах свидетельств, так и на представляющих их функциях множеств или на матричных представлениях тел свидетельств. При этом спе-

цифика задания метрик на множестве тел свидетельств $\mathcal{F}(X)$ состоит в том, что учитываются той или иной степени такие структурные особенности тел свидетельств, как мощность фокальных элементов, степени пересечения фокальных элементов двух тел свидетельств, взаимное расположение фокальных элементов и т. д. Для того чтобы метрику d можно было рассматривать в качестве меры конфликта, ее необходимо нормировать так, чтобы $d: \mathcal{F}(X) \times \mathcal{F}(X) \rightarrow [0, 1]$.

Метрика между телами свидетельств на основе обобщенного евклидова расстояния

$$d_Q(F_1, F_2) = \|\mathbf{m}_1 - \mathbf{m}_2\|_Q = \sqrt{\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T Q (\mathbf{m}_1 - \mathbf{m}_2)}, \quad (4)$$

где $\mathbf{m}_i - 2^{|X|}$ -мерный вектор-столбец, координатами которого являются значения функции масс $m_i(A)$, $A \in 2^X$, $i = 1, 2$, для некоторого упорядочения всех подмножеств из 2^X ; $Q = (q_{A,B})_{A,B \in 2^X}$ – матрица мер сходства между фокальными элементами: $q_{A,B} = q_{B,A} \in [0, 1]$; $q_{A,A} = 1$, $A \in 2^X \setminus \emptyset$; $q_{A,B} = 0$, если $A \cap B = \emptyset$. Тогда $d_Q \in [0, 1]$. Например, одной из популярных метрик такого типа является метрика [41], в которой в качестве меры сходства выступает индекс Жаккарда (*Jaccard index*) [43]: $J = (jac_{A,B})_{A,B \in 2^X}$, $jac_{A,B} = \frac{|A \cap B|}{|A \cup B|}$, если $A, B \neq \emptyset$ и $d_{\emptyset, \emptyset} = 0$. Положительная определенность матрицы (и, соответственно, то, что d_j является метрикой на множестве $\mathcal{F}(X) \times \mathcal{F}(X)$) доказана в статье [44]. В частности, для категоричных тел свидетельств верно, что $d_j(F_A, F_B) = \sqrt{1 - \frac{|A \cap B|}{|A \cup B|}}$. Обобщение метрики (4) с мерой сходства Жаккарда на случай непрерывных структур доверия рассматривалось в работе [45].

Пример 7. Для тел свидетельств $F_1 = 0,8F_{\{x_1\}} + 0,2F_X$ и $F_2 = 0,7F_{\{x_1\}} + 0,3F_{\{x_2\}}$ на множестве $X = \{x_1, x_2\}$ из примера 2 имеем: $J =$

$$= \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{pmatrix}, \quad \mathbf{m}_1 = (0,8; 0; 0,2)^T, \quad \mathbf{m}_2 = (0,7; 0,3; 0)^T$$

$$\text{и } d_j(F_1, F_2) = \frac{1}{2\sqrt{5}} \approx 0,224. \blacklozenge$$

В качестве других мер сходства в формуле (4) используются [46]: коэффициент Соренсена

$$sor_{A,B} = \frac{2|A \cap B|}{|A| + |B|},$$

коэффициент Симпсона

$$sim_{A,B} = \frac{|A \cap B|}{\min\{|A|, |B|\}}$$

и др. Но в этих случаях соответ-



ствующая матрица уже не всегда будет положительно определенной, а функция расстояния не будет (полной) метрикой. В статье [47] была предложена метрика вида (4) в случае, когда X – метрическое пространство, с мерой сходства

$$q_{A,B} = \frac{1}{1 + c \cdot d_H(A,B)}, \text{ где } d_H(A,B) - \text{метрика Хаусдорфа}$$

между множествами A и B метрического пространства X , $c > 0$. Но такая мера сходства не удовлетворяет условию $q_{A,B} = 0$, если $A \cap B = \emptyset^2$. Эта метрика в работе [47], в частности, использовалась в задаче агрегирования информации от разных датчиков в том случае, когда погрешность измерения датчиков и характеристики шума точно не известны.

Пример 8. Найдем расстояние между телами двух свидетельств о прогностической стоимости акций некоторой компании $F_1 = 0,5F_{[40,50]} + 0,3F_{[45,55]} + 0,2F_X$ и $F_2 = 0,3F_{[30,60]} + 0,7F_{[40,50]}$ на множестве $X = [20, 70]$ с помощью метрики (4), где $q_{A,B} = \frac{1}{1 + 0,2d_H(A,B)}$; d_H –

метрика Хаусдорфа. Например, согласно телу свидетельств F_1 стоимость акций прогнозируется в промежутке $[40, 50]$ у. е. со степенью доверия 0,5, в промежутке $[45, 55]$ – со степенью доверия 0,3 в промежутке $X = [20, 70]$ – со степенью доверия 0,2. Будем рассматривать векторные представления этих тел свидетельств только на множествах $\mathcal{A} \cup \mathcal{A}_2 = \{[40, 50], [45, 55], [30, 60], [20, 70]\}$ (с указанным упорядочиванием). Тогда $\mathbf{m}_1 = (0,5; 0,3; 0; 0,2)^T$, $\mathbf{m}_2 = (0,7; 0; 0,3; 0)^T$. Так как мера Хаусдорфа для отрезков равна $d_H([a_1, b_1], [a_2, b_2]) = \max\{|a_1 - a_2|, |b_1 - b_2|\}$, то матрица $Q = (q_{[a_i, b_i], [a_j, b_j]})_{i,j=1}^4$ будет представлять собой

$$Q = \begin{pmatrix} 1 & 1/2 & 1/3 & 1/5 \\ 1/2 & 1 & 1/4 & 1/6 \\ 1/3 & 1/4 & 1 & 1/3 \\ 1/5 & 1/6 & 1/3 & 1 \end{pmatrix}$$

и $d_Q(F_1, F_2) = \sqrt{\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T Q(\mathbf{m}_1 - \mathbf{m}_2)} \approx 0,282$. ♦

Другим примером метрики, вычисляемой непосредственно между телами свидетельств и учитывающей информационную неопределенность, является [11]

$$d(F_1, F_2) = \frac{1}{2|X|-1} \sum_{A \in 2^X} |A| |m_1(A) - m_2(A)|^3$$

² Положительная определенность такой матрицы сходства доказана в статье [47] только для частного случая.

³ В статье [11] рассматривалась ненормированная метрика.

Эта метрика использовалась в статье [11] в задаче аппроксимации сложных тел свидетельств более простыми по структуре фокальными множества телами свидетельств.

Метрика между функциями, взаимно однозначно представляющими тела свидетельств. Среди функций расстояния (метрик, полуметрик, псевдометрик) между взаимно однозначно представляющими тела свидетельств функциями множеств можно выделить метрики Минковского l_p , $1 \leq p \leq \infty$, между функциями доверия или функциями правдоподобия (так как $|Pl_1(A) - Pl_2(A)| = |Bel_1(A^c) - Bel_2(A^c)| \quad \forall A \in 2^X$, то эти метрики равны) (см. работы [11, 48])

$$d_p(F_1, F_2) = c \left(\sum_{A \in 2^X} |Pl_1(A) - Pl_2(A)|^p \right)^{\frac{1}{p}} = c \left(\sum_{A \in 2^X} |Bel_1(A) - Bel_2(A)|^p \right)^{\frac{1}{p}}.$$

При $p = \infty$ вместо суммы вычисляется максимум по всем подмножествам $A \in 2^X$. Здесь (и далее) нормирующий множитель $c > 0$ выбирается из условия $d \in [0, 1]$. Такие метрики часто используют в задачах аппроксимации тел свидетельств.

Метрики между функциями, описывающими (но не представляющими взаимно однозначно) тела свидетельств. Например, метрика Минковского l_p , $1 \leq p \leq \infty$, между пигнистическими вероятностями [49, 50]

$$d_{Bet,p}(F_1, F_2) = c \left(\sum_{A \in 2^X} |Bet_{F_1}(A) - Bet_{F_2}(A)|^p \right)^{\frac{1}{p}}$$

или между вероятностями элементов базового множества

$$d_{Bet,p,x}(F_1, F_2) = c \left(\sum_{x \in X} |Bet_{F_1}(x) - Bet_{F_2}(x)|^p \right)^{\frac{1}{p}},$$

или между так называемыми контурными функциями тел свидетельств [51]

$$d_{Pl,p,x}(F_1, F_2) = c \left(\sum_{x \in X} |Pl_1(x) - Pl_2(x)|^p \right)^{\frac{1}{p}}.$$

Но эти функции будут только псевдометриками на $\mathcal{F}(X) \times \mathcal{F}(X)$, т. е. вместо аксиомы равенства нулю $d(F_1, F_2) = 0 \Leftrightarrow F_1 = F_2$ будет выполняться только аксиома рефлексивности $d(F, F) = 0$.

В частности, псевдометрика $d_{Bet,1,x}$ использовалась в работе [14] в задаче обучения классификатора, в котором классы описывались функциями доверия.

Нормированные полуметрики (т. е. выполняются все аксиомы метрики, кроме неравенства треугольника) **на основе скалярного произведения** векторных представлений тел свидетельств. Примером такой полуметрики служит функция, введенная в работе [42]

$$d(F_1, F_2) = 1 - \frac{\mathbf{Pl}_1^T \cdot \mathbf{Pl}_2}{\|\mathbf{Pl}_1\| \|\mathbf{Pl}_2\|},$$

где $\mathbf{Pl}_i$ – $2^{|X|}$ -мерный вектор-столбец, координатами которого являются значения функции правдоподобия $Pl_i(A)$; $A \in 2^X$; $i=1, 2$; $\|\cdot\|$ – евклидова норма;

Пример 9. Для тел свидетельств $F_1 = 0,8F_{\{x_1\}} + 0,2F_X$ и $F_2 = 0,7F_{\{x_1\}} + 0,3F_{\{x_2\}}$ на множестве $X = \{x_1, x_2\}$ из примера 2 имеем: $\mathbf{Pl}_1 = (1; 0,2; 1)^T$, $\mathbf{Pl}_2 = (0,7; 0,3; 0)^T$ и $d(F_1, F_2) = 1 - \frac{\mathbf{Pl}_1^T \cdot \mathbf{Pl}_2}{\|\mathbf{Pl}_1\| \|\mathbf{Pl}_2\|} \approx 0,301$. ♦

Метрики на основе сравнения интервалов неопределенности $\{[Bel_i(A), Pl_i(A)]\}_{A \in 2^X}$ и $\{[Bel_2(A), Pl_2(A)]\}_{A \in 2^X}$, которые могут быть взаимно однозначно поставлены в соответствие телам свидетельств F_1 и F_2 . Тогда метрика на множестве тел свидетельств может быть задана как продолжение метрики на интервалах:

$$d_{l,p}(F_1, F_2) = c \times$$

$$\times \left(\sum_{A \in 2^X} d_l^p([Bel_1(A), Pl_1(A)], [Bel_2(A), Pl_2(A)]) \right)^{\frac{1}{p}}, \quad (5)$$

$$1 \leq p \leq \infty,$$

где d_l – метрика на интервалах. Например, в качестве d_l может быть использована метрика Хаусдорфа d_H (см. пример 8). В работе [12] исследовалась и применялась в задаче аппроксимации тел свидетельств метрика вида (5), где в качестве метрики d_l рассматривалась p -метрика Вассерштейна $d_{w,p}([a_1, b_1], [a_2, b_2])$ двух равномерных распределений на промежутках $[a_1, b_1]$ и $[a_2, b_2]$. В частно-

сти, нетрудно показать, что $d_{w,2}([a_1, b_1], [a_2, b_2]) = \sqrt{(mean_1 - mean_2)^2 + \frac{1}{3}(rad_1 - rad_2)^2}$, где $mean_i$ и rad_i – середина и полудлина соответственно отрезка $[a_i, b_i]$, $i=1, 2$. В работе [12] показано, что для $p=2$ нормирующий множитель $c = 2^{-\frac{|X|-1}{2}}$. Эта метрика оценивает расхождение между информационными неопределенностями тел свидетельств.

Пример 10. Для тел свидетельств $F_1 = 0,8F_{\{x_1\}} + 0,2F_X$ и $F_2 = 0,7F_{\{x_1\}} + 0,3F_{\{x_2\}}$ на множестве $X = \{x_1, x_2\}$ из примера 2 найдем значения метрик $d_{w,2}(F_1, F_2)$ и $d_{H,2}(F_1, F_2)$. Для $|X|=2$ и $p=2$ нормирующий множитель $c = \frac{1}{\sqrt{2}}$. Вычисления метрик между промежутками неопределенности сведены в табл. 2.

Тогда $d_{w,2}(F_1, F_2) = \frac{1}{10} \sqrt{\frac{13}{3}} \approx 0,208$ и $d_{H,2}(F_1, F_2) = 0,3$. ♦

Метрика Вассерштейна. В статье [52] показано, что задав метрику d на множестве 2^X , ее можно продолжить до метрики на множестве $\mathcal{F}(X)$ как решение задачи Канторовича

$$d_w(F_1, F_2) = \min_{\tilde{m}} \sum_{A \in 2^X} \sum_{B \in 2^X} \tilde{m}(A, B) d(A, B),$$

где минимум берется по всем функциям множеств, удовлетворяющих условиям согласования (1). Полученную метрику можно рассматривать как 1-метрику Вассерштейна между телами свидетельств. Эта метрика хорошо интерпретируется как решение оптимизационной задачи Канторовича.

Пример 11. Для тел свидетельств $F_1 = 0,8F_{\{x_1\}} + 0,2F_X$ и $F_2 = 0,7F_{\{x_1\}} + 0,3F_{\{x_2\}}$ на множестве $X = \{x_1, x_2\}$ из примера 2 имеем: $\tilde{m}(\{x_1\}, \{x_1\}) = 0,5+t$, $\tilde{m}(\{x_1\}, \{x_2\}) = 0,3-t$, $\tilde{m}(X, \{x_1\}) = 0,2-t$, $\tilde{m}(X, \{x_2\}) = t$, $t \in [0; 0,2]$ и $\tilde{m}(A, B) = 0$ для остальных пар $(A, B) \in 2^X \times 2^X$. Рассмотрим на множестве 2^X метрику $d(A, B) = d_J(F_A, F_B) = \sqrt{1 - \frac{|A \cap B|}{|A \cup B|}}$.

Таблица 2

Значения метрик $d_{w,2}$ и $d_{H,2}$ между промежутками $[Bel_i(A), Pl_i(A)]$, $i=1, 2$

A	F ₁				F ₂				d _{w,2} ²	d _{H,2} ²
	Bel ₁	Pl ₁	mean ₁	rad ₁	Bel ₂	Pl ₂	mean ₂	rad ₂		
{x ₁ }	0,8	1	0,9	0,1	0,7	0,7	0,7	0	13/300	0,09
{x ₂ }	0	0,2	0,1	0,1	0,3	0,3	0,3	0	13/300	0,09
X	1	1	1	0	1	1	1	0	0	–



Тогда метрика Вассерштейна между телами свидетельств F_1 и F_2 будет равна

$$d_w(F_1, F_2) = \min_{0 \leq t \leq 0,2} \left\{ (0,3-t) \cdot 1 + (0,2-t) \frac{1}{\sqrt{2}} + t \frac{1}{\sqrt{2}} \right\} = \\ = \min_{0 \leq t \leq 0,2} \left\{ \frac{3+\sqrt{2}}{10} - t \right\} = \frac{1+\sqrt{2}}{10} \approx 0,241. \blacklozenge$$

Метрики между матрицами специализаций (обобщения) тел свидетельств. В статье [53] были введены и исследованы метрики на множестве $\mathcal{F}(X)$ вида

$$d^\cap(F_1, F_2) = c_\cap \|R_{F_1}^\cap - R_{F_2}^\cap\|, \\ d^\cup(F_1, F_2) = c_\cup \|R_{F_1}^\cup - R_{F_2}^\cup\|,$$

где $R_{F_i}^\cap$ и $R_{F_i}^\cup$ – матрицы специализаций и обобщения тела свидетельств F_i соответственно (см. § 3); $i = 1, 2$; $\|\cdot\|$ – матричная норма. Нетрудно видеть,

$$\text{что } c_\cap = \left(\max_{A, B \in 2^X} \|R_{F_A}^\cap - R_{F_B}^\cap\| \right)^{-1}, \quad c_\cup = \left(\max_{A, B \in 2^X} \|R_{F_A}^\cup - R_{F_B}^\cup\| \right)^{-1}.$$

В работе [54] были исследованы обобщения этих метрик для операторных представлений α -слияний, а также было показано, что $d^\cap = d^\cup$. Метрики на основе матриц специализаций представляются более робастными к небольшим изменениям тел свидетельств.

Пример 12. Для тел свидетельств $F_1 = 0,8 \times F_{\{x_1\}} + 0,2 F_X$ и $F_2 = 0,7 F_{\{x_1\}} + 0,3 F_{\{x_2\}}$ на множестве $X = \{x_1, x_2\}$ матрицы специализации $R_{F_1}^\cap$, $R_{F_2}^\cap$ были найдены в примере 4. Тогда, например, для матричной нормы $\|A\|_1 = \max_j \sum_i |a_{ij}|$ имеем $c_\cap = \frac{1}{2}$ и

$$d_1^\cap(F_1, F_2) = \frac{1}{2} \|R_{F_1}^\cap - R_{F_2}^\cap\|_1 = \\ = \frac{1}{2} \left\| \begin{pmatrix} 0 & -0,3 & 0,1 & 0 \\ 0 & 0,3 & 0 & 0,1 \\ 0 & 0 & -0,1 & -0,3 \\ 0 & 0 & 0 & 0,2 \end{pmatrix} \right\|_1 = 0,3. \blacklozenge$$

В ряде работ отмечалось, что использование только одной метрики или функционала для характеристики внешнего конфликта недостаточно. В частности, каноническая мера конфликта будет принимать большие значения для пары одинаковых (т. е. слабо неконфликтных) тел свидетельств, «близких» к равновероятному распределению (см. замечание 2), для которых метрическая составляющая будет равна нулю. С другой стороны, конфликт между телами свидетельств зачастую не сводится к вычислению только метрической составляющей. Действительно, категоричные консонантные тела свидетельств F_A и F_B , где $A \subseteq B$, являются сильно неконфликтными (и $K(F_A, F_B) = 0$), однако метрика (4) с индексом Жаккарда для этих тел свидетельств, равная

$$d_J(F_A, F_B) = \sqrt{1 - \frac{|A|}{|B|}}, \text{ будет принимать большие значения, если } |A| \ll |B|.$$

Поэтому в некоторых работах предлагалось использовать не одну, а совокупность меры конфликта и метрики на множестве тел свидетельств $\mathcal{F}(X)$. Так, в статье [50] была рассмотрена пара $(K(F_1, F_2), d_{Bet, \infty}(F_1, F_2))$, где $d_{Bet, \infty}(F_1, F_2) = \max_{A \subseteq X} |Bet_{F_1}(A) - Bet_{F_2}(A)|$. Большие значения каждой составляющей этой пары гарантируют большой конфликт между телами свидетельств. В частности, большое значение метрики $d_{Bet, \infty}(F_1, F_2)$ означает, что тела свидетельств далеки от вероятностных распределений.

В общем случае каноническую меру конфликта можно представить в виде [55]

$$K(F_1, F_2) = E_Q(\mathbf{m}_1) + E_Q(\mathbf{m}_2) + \|\mathbf{m}_1 - \mathbf{m}_2\|_Q^2 - \sum_{B, C} r_{B, C} m_1(B) m_2(C), \quad (6)$$

где

$$E_Q(\mathbf{m}_i) = \frac{1}{2} - \|\mathbf{m}_i\|_Q^2 = \frac{1}{2} \sum_B m_i(B) \left(1 - \sum_C q_{B, C} m_i(C) \right); \\ i = 1, 2; \quad r_{B, C} = \begin{cases} 0, & B \cap C = \emptyset \vee B = C, \\ 1 - q_{B, C}, & \text{иначе.} \end{cases}$$

Последнее слагаемое в формуле (6) характеризует взаимодействие слабо пересекающихся фокальных элементов. Функционал $E_Q: \mathcal{F}(X) \rightarrow [0, \frac{1}{2}]$ по своим свойствам близок к энтропийным функционалам и характеризует величину внутреннего конфликта тела свидетельств. В частности, для полностью конфликтных тел свидетельств $K(F_1, F_2) = E_I(\mathbf{m}_1) + E_I(\mathbf{m}_2) + \|\mathbf{m}_1 - \mathbf{m}_2\|_I^2$, где I – единичная матрица. Применение энтропий и дивергенций для оценивания неопределенности тел свидетельств рассматривалось в работе [56].

5.2. Структурные способы оценивания конфликта

В статье [39] рассматривался структурный подход к оцениванию конфликта, в котором учитывалась степень включения фокальных элементов одного тела свидетельств в фокальные элементы другого тела свидетельств. Для этого использовался

индекс включения множеств $Inc(A, B) = \begin{cases} 1, A \subseteq B, \\ 0, A \not\subseteq B, \end{cases}$

$A, B \in 2^X$, с помощью которого вычислялась степень включения одного множества фокальных элементов в другое:

$$\delta(\mathcal{A}_1, \mathcal{A}_2) = \max\{d(\mathcal{A}_1, \mathcal{A}_2), d(\mathcal{A}_2, \mathcal{A}_1)\},$$

$$d(\mathcal{A}_1, \mathcal{A}_2) = \frac{1}{|\mathcal{A}_1||\mathcal{A}_2|} \sum_{A_1 \in \mathcal{A}_1} \sum_{A_2 \in \mathcal{A}_2} Inc(A_1, A_2).$$

Тогда мера конфликта определялась по формуле

$$Con_{ext}(F_1, F_2) = (1 - \delta(\mathcal{A}_1, \mathcal{A}_2)) d_J(F_1, F_2),$$

где d_J – расстояние (4) с индексом Жаккарда. Эта мера будет удовлетворять условиям E1, E2', E3, а также следствиям а) и б).

Замечание 3. Индекс включения множеств $Inc(A, B)$ можно рассматривать как значение меры доверия η_A , соответствующей категоричному телу свидетельств F_A , на множестве B . Тогда произвольную функцию доверия Bel , соответствующую телу свидетельств $F = (\mathcal{A}, m)$, можно представить в виде $Bel = \sum_{A \in \mathcal{A}} m(A) \eta_A$.

5.3. Алгебраические способы оценивания конфликта

В докладе [57] был предложен алгебраический подход к построению меры конфликта на основе мер сходства множеств: $Q = (q_{A,B})$, $q_{A,B} = q_{A,B} \in [0, 1]$, $q_{A,A} = 1$, $A \in 2^X \setminus \emptyset$, $q_{A,B} = 0$, если $A \cap B = \emptyset$.

Назовем функционал $Con_Q: \mathcal{F}(X) \times \mathcal{F}(X) \rightarrow [0, 1]$ билинейным, если $Con_Q(\alpha F_1 + \beta F_2, F) = \alpha Con_Q(F_1, F) + \beta Con_Q(F_2, F)$ для всех $\alpha, \beta \in [0, 1]$, $\alpha + \beta = 1$, $F, F_1, F_2 \in \mathcal{F}(X)$. Вместо условия E4 (антимонотонность по специализации) будем рассматривать его ослабленный вариант

E4': если $F = (\mathcal{A}_0, m)$, $F' = (\mathcal{A} \cup \{A'\}, m')$, $F'' = (\mathcal{A} \cup \{A''\}, m'')$, $m'(A) = m''(A) \quad \forall A \in \mathcal{A}$, $m'(A') = m''(A'')$ и $q_{A',B} \leq q_{A'',B} \quad \forall B \in \mathcal{A}_0$, то $Con_Q(F', F) \geq Con_Q(F'', F)$.

Теорема 2 [57]. Функционал Con_Q будет билинейной мерой конфликта на множестве $\mathcal{F}(X) \times \mathcal{F}(X)$, удовлетворяющей условиям E1, E3 и E4', тогда и только тогда, когда

$$Con_Q(F_1, F_2) = \sum_{A \in \mathcal{A}_1, B \in \mathcal{A}_2} \gamma(A, B) m_1(A) m_2(B) = K(F_1, F_2) + \sum_{A \cap B \neq \emptyset} \gamma(A, B) m_1(A) m_2(B), \quad (7)$$

где коэффициенты $\gamma(A, B) = Con_Q(F_A, F_B) \in [0, 1]$ удовлетворяют условиям: $\gamma(A, B) = \gamma(B, A)$ и $\gamma(A', B) \geq \gamma(A'', B)$, если $q_{A',B} \leq q_{A'',B}$; $\gamma(A, B) = 1$, если $A \cap B = \emptyset$.

Из формулы (7) следует, что каноническая мера конфликта $K(F_1, F_2)$ является наименьшей из всех билинейных мер конфликта: $Con_Q(F_1, F_2) \geq K(F_1, F_2)$.

Замечание 4. Коэффициенты $\gamma(A, B) = Con_Q(F_A, F_B) = \psi(q_{A,B})$, $A, B \neq \emptyset$, будут удовлетворять условиям а) – в) теоремы 2, если ψ – невозрастающая функция, $\psi(1) = 0$, $\psi(0) = 1$ и $q_{A,B} = |A \cap B| / \min\{|A|, |B|\}$. В частности, если $q_{A,B} = \begin{cases} 1, A \cap B \neq \emptyset, \\ 0, A \cap B = \emptyset, \end{cases}$ – примитивная мера пересечения, то $Con_Q(F_1, F_2) = K(F_1, F_2)$.

5.4. Оценивание конфликта с помощью правил комбинирования

Каноническая мера конфликта $K(F_1, F_2)$ двух тел свидетельств $F_1 = (\mathcal{A}_1, m_1)$ и $F_2 = (\mathcal{A}_2, m_2)$ численно равна массе пустого множества их комбинирования $F_1 \otimes_{ND} F_2 = (\mathcal{A} \wedge \mathcal{A}_2, m_{ND})$, $\mathcal{A} \wedge \mathcal{A}_2 = \{A \cap B : A \in \mathcal{A}_1, B \in \mathcal{A}_2\}$, с помощью ненормализованного правила Демпстера: $K(F_1, F_2) = m_{DN}(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$.

Мера конфликта из теоремы 1 равна инфимуму масс пустых множеств по всем результатам комбинирования с помощью обобщенного конъюнктивного правила $\otimes_{\cap}$.

Эти примеры показывают, что мера конфликта должна быть согласована с правилами комбинирования тел свидетельств. В частности, этот вопрос обсуждался в работе [55].

Непосредственным обобщением правила комбинирования Демпстера является его весовой аналог: если $F_1 = (\mathcal{A}_1, m_1)$, $F_2 = (\mathcal{A}_2, m_2)$, то $F_Q = (\mathcal{A}, m_Q) = F_1 \otimes_Q F_2$, где $\mathcal{A} = \mathcal{A}_1 \wedge \mathcal{A}_2$ и

$$m_Q(C) = \frac{1}{K_Q} \sum_{A \cap B = C} q_{A,B} m_1(A) m_2(B),$$

а $Q = (q_{A,B})$ – мера сходства множеств: $q_{A,B} = q_{A,B} \in [0, 1]$, $q_{A,A} = 1$, $A \in 2^X \setminus \emptyset$, $q_{A,B} = 0$, если $A \cap B = \emptyset$. Коэффициент нормировки $K_Q = \sum_C \sum_{A \cap B = C} q_{A,B} m_1(A) m_2(B) = \sum_{A,B} q_{A,B} m_1(A) m_2(B)$.



Такое правило рассматривалось, например, в работе [58].

Тогда мерой конфликта, согласованной с правилом комбинирования $\otimes_Q$, будет мера $Con_Q(F_1, F_2) = 1 - K_Q$ вида (7), в которой $\gamma(A, B) = 1 - q_{A,B}$, $A, B \in \mathcal{A}$.

Мера конфликта (7) учитывает пары не только непересекающихся фокальных элементов, но и пары «слабопересекающихся» фокальных элементов, т. е. такие пары, для которых мощность пересечения мала по сравнению с мощностью каждого из этих множеств.

Пример 13. Для тел свидетельств $F_1 = 0,2F_{\{b\}} + 0,1F_{\{c\}} + 0,3F_{\{a,b\}} + 0,4F_{\{b,c\}}$ и $F_2 = 0,7F_{\{a,c\}} + 0,3F_{\{a,b,c\}}$ на множестве $X = \{a, b, c\}$ из примера 3 для $\gamma(A, B) = 1 - jac_{A,B} = 1 - \frac{|A \cap B|}{|A \cup B|}$ имеем: $Con_j(F_1, F_2) = 59/120$, но $K(F_1, F_2) = 0,14$. Видно, что учет «слабопересекающихся» пар фокальных элементов значительно увеличил величину конфликта. ♦

Меру конфликта (7) удобно использовать и в том случае, когда фокальные элементы являются подмножествами числовой оси $\mathbb{R}$.

Пример 14. Для двух тел свидетельств о прогностической стоимости акций некоторой компании $F_1 = 0,5F_{[40,50]} + 0,3F_{[45,55]} + 0,2F_X$ и $F_2 = 0,3F_{[30,60]} + 0,7F_{[40,50]}$ на множестве $X = [20, 70]$ (см. пример 8) и $\gamma(A, B) = 1 - jac_{A,B} = 1 - \frac{|A \cap B|}{|A \cup B|}$, где $|A|$ – мера Лебега множества A на числовой оси $\mathbb{R}$. Тогда $Con_j(F_1, F_2) = 0,163$, но $K(F_1, F_2) = 0$. ♦

Также можно меру конфликта (7) применять в случае нечетких фокальных элементов.

Пример 15. Пусть в примере 14 $F_1 = 0,5F_{(35,40,50,55)} + 0,3F_{(40,45,55,60)} + 0,2F_X$ и $F_2 = 0,3F_{(20,30,60,70)} + 0,7F_{(35,40,50,55)}$, $\tilde{A} = (a_1, a_2, a_3, a_4)$, $a_1 \leq a_2 \leq a_3 \leq a_4$ – трапецевидное нечеткое число на множестве $X = [20, 70]$ с функцией принадлежности

$$\mu_{\tilde{A}}(x) = \max \left\{ 0, \min \left\{ \frac{x - a_1}{a_2 - a_1}, 1, \frac{a_4 - x}{a_4 - a_3} \right\} \right\}, \text{ если } a_1 <$$

$a_2 \leq a_3 < a_4$. Заметим, что в этом примере ядра всех нечетких фокальных элементов совпадают с соответствующими неразмытыми фокальными элементами из предыдущего примера: $\text{Ker} \tilde{A} = \text{Ker}(a_1, a_2, a_3, a_4) =$

$$= [a_2, a_3]. \text{ Учитывая, что } \gamma(\tilde{A}, \tilde{B}) = 1 - \frac{|\tilde{A} \cap \tilde{B}|}{|\tilde{A} \cup \tilde{B}|} =$$

$$= 1 - \int_X \min \{ \mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x) \} dx / \int_X \max \{ \mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x) \} dx, \text{ получим: } Con_j(F_1, F_2) = 0,365. \diamond$$

6. РОБАСТНЫЕ ОЦЕНКИ ВНЕШНЕГО КОНФЛИКТА

Некоторые способы вычисления меры конфликта (в частности, каноническая мера конфликта) неустойчивы к «малым» изменениям тел свидетельств. При этом сами тела свидетельств могут формироваться субъективно и зависеть от особенностей источников информации. Например, если один эксперт дает «осторожный» прогноз о стоимости акций некоторой компании в промежутке (30, 40) у. е., а другой эксперт делает «оптимистичный» прогноз – (38, 45) у. е., то можно считать, что у нас есть два категоричных тела свидетельства: F_A и F_B , где $A = (30, 40)$ и $B = (38, 43)$. В этом случае каноническая мера конфликта $K(F_A, F_B) = 0$. На самом деле «осторожный» эксперт при уточнении своего прогноза имел в виду тело свидетельств $F_1 = 0,8F_{(30,38)} + 0,2F_{(38,40)}$, а «оптимистичный» эксперт тоже мог бы уточнить свой прогноз как тело свидетельств $F_2 = 0,2F_{(35,38)} + 0,7F_{(38,42)} + 0,1F_{(42,44)}$. В этом случае получили бы $K(F_1, F_2) = 0,7$.

Одним из подходов к повышению робастности оценивания конфликта является использование процедур специализации-обобщения. Процедура специализации представляет собой разбиение фокальных элементов на более «мелкие» множества с одновременным распределением функции масс по этим множествам. Без ограничения общности можно считать, что для тела свидетельств $F_1 = (\{A_i\}, m_1)$ его специализацией будет тело свидетельств $F_2 = (\{B_{ij}\}, m_2) = S(F_1)$ (обозначение: $F_2 \sqsubseteq F_1$), где $\cup_j B_{ij} = A_i$ и $\sum_j m_2(B_{ij}) = m_1(A_i) \quad \forall i$. А обобщением тела свидетельств $F_1 = (\{A_i\}, m_1)$ будет тело свидетельств $F_3 = (\{C_{ij}\}, m_3) = G(F_1)$ ($F_1 \sqsubseteq F_3$): $\cap_j C_{ij} = A_i$ и $\sum_j m_3(C_{ij}) = m_1(A_i) \quad \forall i$).

Будем рассматривать только те специализации-обобщения тела свидетельств $F = (\mathcal{A}, m)$, которые достаточно близки к телу свидетельств F . Степень близости будем измерять с помощью индекса неточности $f: \mathcal{F}(X) \rightarrow [0, 1]$ [23], примером которого является нормированная обобщенная мера Хартли: $H_0(F) = \frac{1}{\ln|X|} \sum_{A \in \mathcal{A}} m(A) \ln|A|$. При специализации индекс неточности не увеличивается: $f(F_2) \leq f(F_1)$, если $F_2 \sqsubseteq F_1$. Аналогично, при обобщении индекс неточности не уменьшается.

Пусть $S_\varepsilon(F) = \{F' \sqsubseteq F : f(F) - f(F') < \varepsilon\}$, $G_\varepsilon(F) = \{F \sqsubseteq F' : f(F') - f(F) < \varepsilon\}$ – множества

всех специализаций и обобщений тела свидетельств F соответственно, которые находятся в ε -окрестности тела свидетельств F относительно индекса неточности f , $SG_\varepsilon(F) = S_\varepsilon(F) \cup G_\varepsilon(F)$.

Тогда в качестве меры конфликта можно использовать величину $K_\varepsilon(F_1, F_2) = \text{MEAN}_{\tilde{F}_i \in SG_\varepsilon(F_i), i=1,2} (K(\tilde{F}_1, \tilde{F}_2))$,

где MEAN – некоторый оператор усреднения. Например, пусть $SG_\varepsilon(F_1) = \{F_{1,j}\}$, $SG_\varepsilon(F_2) = \{F_{2,k}\}$,

где $F_{i,s} = F_{i,s}(\theta)$, $i=1, 2$, $\theta \in [0, 1]^N$ – вектор параметров (значений функции масс). Тогда $K(F_{1,j}, F_{2,k}) = \varphi_{j,k}(\theta)$, причем $\theta \in D_{j,k} = \{\theta : |f(F_{1,j}(\theta)) - f(F_1)| < \varepsilon, |f(F_{2,k}(\theta)) - f(F_2)| < \varepsilon\}$.

В качестве оператора MEAN можно использовать усреднение средних интегральных значений

$$K_\varepsilon(F_1, F_2) = \frac{1}{|\{F_{1,j}\}| \cdot |\{F_{2,k}\}|} \sum_{j,k} I_{j,k},$$

где $I_{j,k} = \frac{1}{V(D_{j,k})} \int_{D_{j,k}} \varphi_{j,k}(\theta) d\theta$; V – мера Лебега на пространстве параметров. Возможны и другие способы формирования множества $SG_\varepsilon(F)$ специализаций-обобщений тела свидетельств F .

Пример 16 [4]. Пусть $X = \{x_1, x_2, x_3\}$, $F_1 = F_{\{x_1, x_2\}}$, $F_2 = F_{\{x_3\}}$. Тогда $K_0(F_1, F_2) = 1$. Рассмотрим теперь специализации-обобщения тел свидетельств F_1 и F_2 . Для оценки неточности будем использовать нормированную обобщенную меру Хартли H_0 . Пусть $c = \frac{\ln 2}{\ln 3}$. Тогда $H_0(F_1) = c$, $H_0(F_2) = 0$. Для специализаций имеем $S(F_1) = \{F_{1,0}, F_{1,1}\}$, где $F_{1,0} = F_1$, $F_{1,1} = \theta F_{\{x_1\}} + (1-\theta)F_{\{x_2\}}$, $H_0(F_{1,1}) = 0$, $S(F_2) = \{F_{2,0}\}$, $F_{2,0} = F_2$. Обобщениями тел свидетельств F_1 и F_2 будут множества

$$G(F_1) = \{F_{1,0}, F_{1,2}\},$$

$$G(F_2) = \{F_{2,0}, F_{2,1}, \dots, F_{2,4}\},$$

где $F_{1,2} = \theta F_{\{x_1, x_2\}} + (1-\theta)F_{\{x_1, x_2, x_3\}}$ и $H_0(F_{1,2}) = c\theta + 1 - \theta$;

$$F_{2,1} = \theta F_{\{x_1, x_3\}} + (1-\theta)F_{\{x_3\}} \text{ и } H_0(F_{2,1}) = c\theta,$$

$$F_{2,2} = \theta F_{\{x_2, x_3\}} + (1-\theta)F_{\{x_3\}} \text{ и } H_0(F_{2,2}) = c\theta,$$

$$F_{2,3} = \theta F_{\{x_1, x_3\}} + (1-\theta)F_{\{x_2, x_3\}} \text{ и } H_0(F_{2,3}) = c,$$

$$F_{2,4} = \theta F_{\{x_1, x_2, x_3\}} + (1-\theta)F_{\{x_3\}} \text{ и } H_0(F_{2,4}) = \theta.$$

Тогда $SG(F_1) = \{F_{1,0}, F_{1,1}, F_{1,2}\}$, $SG(F_2) = \{F_{2,0}, F_{2,1}, \dots, F_{2,4}\}$.

Пусть $\varepsilon < 1 - c$. Тогда $SG_\varepsilon(F_1) = \{F_{1,0}, F_{1,2}\}$ и $SG_\varepsilon(F_2) = \{F_{2,0}, F_{2,1}, F_{2,2}, F_{2,4}\}$. Теперь имеем $K(F_{1,0}, F_{2,0}) = 1$, $K(F_{1,0}, F_{2,1}) = K(F_{1,0}, F_{2,2}) = 1 - \theta$ при $0 < c\theta < \varepsilon$, $K(F_{1,0}, F_{2,4}) = 1 - \theta$ при $0 < \theta < \varepsilon$,

$K(F_{1,2}, F_{2,0}) = \theta$ при $0 < (1-\theta)(1-c) < \varepsilon$, $K(F_{1,2}, F_{2,1}) = K(F_{1,2}, F_{2,2}) = \theta_1(1-\theta_2)$ при $0 < (1-\theta_1)(1-c) < \varepsilon$ и $0 < c\theta_2 < \varepsilon$, $K(F_{1,2}, F_{2,4}) = \theta_1(1-\theta_2)$ при $0 < (1-\theta_1) \times (1-c) < \varepsilon$ и $0 < \theta_2 < \varepsilon$.

Теперь $K_\varepsilon(F_1, F_2) = \frac{1}{2.4} \sum_{j,k} I_{j,k}$, где $I_{j,k} = \frac{1}{V(D_{j,k})} \times \int_{D_{j,k}} \varphi_{j,k}(\theta) d\theta$, т. е. $I_{0,0} = 1$, $I_{0,1} = I_{0,2} = \frac{c}{\varepsilon} \int_0^\varepsilon (1-\theta) d\theta = 1 - \frac{1}{2c} \varepsilon$, $I_{0,4} = \frac{1}{\varepsilon} \int_0^\varepsilon (1-\theta) d\theta = 1 - \frac{1}{2} \varepsilon$, $I_{2,0} = \frac{1-c}{\varepsilon} \int_{1-\frac{c}{\varepsilon}}^1 \theta d\theta = 1 - \frac{1}{2(1-c)} \varepsilon$, $I_{2,1} = I_{2,2} = \frac{1-c}{\varepsilon} \cdot \frac{c}{\varepsilon} \int_{1-\frac{c}{\varepsilon}}^1 \int_0^\varepsilon \theta_1(1-\theta_2) d\theta_1 d\theta_2 = (1 - \frac{1}{2(1-c)} \varepsilon)(1 - \frac{1}{2c} \varepsilon)$, $I_{2,4} = \frac{1-c}{\varepsilon} \cdot \frac{1}{\varepsilon} \int_{1-\frac{c}{\varepsilon}}^1 \int_0^\varepsilon \theta_1(1-\theta_2) d\theta_1 d\theta_2 = (1 - \frac{1}{2(1-c)} \varepsilon)(1 - \frac{1}{2} \varepsilon)$.

Таким образом, $K_\varepsilon(F_1, F_2) = 1 - \varepsilon \frac{(2-c)(c+1)}{8c(1-c)} + \varepsilon^2 \frac{(2+c)}{4c(1-c)}$ при $\varepsilon < 1 - c$. ♦

ЗАКЛЮЧЕНИЕ

В работе дан аналитический обзор современного состояния исследований по анализу противоречивости (конфликтности) информации, полученной из нескольких источников в рамках теории функций доверия. В частности, можно выделить следующие моменты текущего состояния исследований по анализу противоречивости информации:

- в настоящее время в научном сообществе сформировались определенные требования к мере внешнего конфликта, важнейшими из которых являются:

- мера конфликта должна отражать различные степени градации конфликтности тел свидетельств: от полной неконфликтности до той или иной степени конфликтности (слабая, простая, сильная);
- мера конфликта должна быть антимонотонной по специализации;

- эти требования положены в основу аксиоматики меры внешнего конфликта; найден общий вид меры внешнего конфликта, удовлетворяющий некоторой системе аксиом;

- существует несколько способов оценивания внешнего конфликта, среди которых условно можно выделить метрический, структурный, алгебраический, на основе правил комбинирования;

- вместе с мерой конфликта в теории функций доверия рассматривается и понятие расстояния между телами свидетельств, но мера конфликта не сводится к вычислению расстояния;

- робастное вычисление меры конфликта достигается с помощью процедуры обобщения специализации.



Вместе с тем, в исследовании анализа противоречивости информации в рамках теории функций доверия остается и ряд открытых проблем. Среди них можно выделить:

- построение мер внешнего конфликта, полнее отражающих различные степени конфликтности тел свидетельств и их структурные особенности;
- исследование согласованности правил комбинирования и мер конфликта в задаче принятия решения о выборе тел свидетельств для комбинирования;
- исследование взаимосвязи логической цепочки понятий «противоречивость» – «согласованность» – «взаимовлияние» источников информации;
- нахождение общего вида меры конфликта для других систем аксиом;
- исследование мер конфликта для тел свидетельств, определенных на пространстве с мерой.

И конечно, актуальными являются прикладные задачи, связанные с анализом конфликтности источников информации, управления конфликтом и т. д.

ЛИТЕРАТУРА

1. *Dempster, A.P.* Upper and Lower Probabilities Induced by a Multivalued Mapping // *Annals of Mathematical Statistics.* – 1967. – Vol. 38. – P. 325–339.
2. *Shafer, G.* A Mathematical Theory of Evidence. – Princeton, N.J.: Princeton University Press, 1976.
3. *Lefevre, E., Colot, O., Vannoorenbergh, P.* Belief Function Combination and Conflict Management // *Inf. Fusion.* – 2002. – Vol. 3, no. 2. – P. 149–162.
4. *Bronevich, A., Lepskiy, A., Penikas, H.* The Application of Conflict Measure to Estimating Incoherence of Analyst's Forecasts about the Cost of Shares of Russian Companies // *Procedia Computer Science.* – 2015. – Vol. 55. – P. 1113–1122.
5. *Kutynina, E., Lepskiy, A.* Aggregation of Forecasts and Recommendations of Financial Analysts in the Framework of Evidence Theory // In: *Kacprzyk, J., Szmidt, E., Zadrożny, S., Atanassov, K., Krawczak, M. (eds.) Advances in Intelligent Systems and Computing.* – Vol. 642. – Springer, Cham, 2018. – P. 370–381.
6. *Bronevich, A.G., Spiridenkova, N.S.* Measuring Uncertainty for Interval Belief Structures and Its Application for Analyzing Weather Forecasts // In: *Kacprzyk, J., Szmidt, E., Zadrożny, S., Atanassov, K., Krawczak, M. (Eds.) Advances in Fuzzy Logic and Technology 2017, Advances in Intelligent Systems and Computing.* – Vol. 641. – Springer, Cham, 2018. – P. 273–285.
7. *Lepskiy, A., Smolev, V.* Application of Non-additive Measures and Integrals for Analysis of the Importance of Party Positions for Voting // In: *Atlantis Studies in Uncertainty Modelling: Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (EUSFLAT 2019).* – Atlantis Press, 2019. – Vol. 1. – P. 321–327.
8. *Lepskiy, A., Suevalov, A.* Application of the Belief Function Theory to the Development of Trading Strategies // *Procedia Computer Science.* – 2019. – Vol. 162. – P. 235–242.
9. *Rominger, C., Martin, A.* Using the Conflict: An Application to Sonar Image Registration // In: *Workshop on the Theory of Belief Functions.* – Brest, France, 2010. – P. 1–6.
10. *Harmanec, D.* Faithful Approximations of Belief Functions // In: *Laskey, K.B., Prade, H. (Eds.), Uncertainty in Artificial Intelligence 15 (UAI99).* – Stockholm, Sweden, 1999.
11. *Denœux, T.* Inner and Outer Approximation of Belief Structures Using a Hierarchical Clustering Approach // *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.* – 2001. – Vol. 9. – P. 437–460.
12. *Han, D., Dezert, J., Yang, Y.* Belief Interval-Based Distance Measures in the Theory of Belief Functions // *IEEE Transactions on Systems, Man and Cybernetics: Systems, IEEE.* – 2016. – Vol. 48, no. 6. – P. 833–850.
13. *Bronevich, A., Lepskiy, A.* Measures of Conflict, Basic Axioms and Their Application to the Clusterization of a Body of Evidence // *Fuzzy Sets and Systems.* – 2021. – <https://doi.org/10.1016/j.fss.2021.04.016>
14. *Denœux, T.* A Neural Network Classifier Based on Dempster-Shafer Theory // *IEEE Transactions on Systems, Man and Cybernetics.* – 2000. – Vol. 30. – P. 131–150.
15. *Ye, Q., Wu, X., Chen, Z.* An Approach for Evidence Clustering Using Generalized Distance // *Journal of Electronics.* – 2009. – Vol. 26. – P. 18–23.
16. *Smets, P., Kennes, R.* The Transferable Belief Model // *Artificial Intelligence.* – 1994. – Vol. 66. – P. 191–234.
17. *Nguyen, H.T.* On Random Sets and Belief Functions // *J. Math. Anal. Appl.* – 1978. – Vol. 65. – P. 531–542.
18. *Halpern, J.Y., Fagin, R.* Two Views of Belief: Belief as Generalized Probability and Belief as Evidence // *Artificial Intelligence.* – 1992. – Vol. 54, no. 3. – P. 275–317.
19. *Smets, P.* Decision Making in TBM: The Necessity of the Pignistic Transformation // *International Journal of Approximate Reasoning.* – 2005. – Vol. 38. – P. 133–147.
20. *Shapley, L.* A Value for N-person Games // *Contributions to the Theory of Games. II (28) in Annals of Mathematics Studies,* Princeton University Press, 1953. – P. 307–317.
21. *Smets, P.* Belief Functions on Real Numbers // *International Journal of Approximate Reasoning.* – 2005. – Vol. 40, no. 3. – P. 181–223.
22. *Yen, J.* Generalizing the Dempster-Shafer Theory to Fuzzy Sets // *IEEE Transactions on Systems, Man, and Cybernetics.* – 1990. – Vol. 20, no. 3. – P. 559–570.
23. *Bronevich, A., Lepskiy, A.* Imprecision Indices: Axiomatic, Properties and Applications // *Int. J. of General Systems.* – 2015. – Vol. 44, no. 7-8. – P. 812–832.
24. *Higashi, M., Klir, G.J.* Measures of Uncertainty and Information Based on Possibility Distributions // *Int. J. General Systems.* – 1983. – No. 9. – P. 43–58.
25. *Dubois, D., Prade, H.* A Note on Measures of Specificity for Fuzzy Sets // *Int. J. of General Systems.* – 1985. – No. 10. – P. 279–283.
26. *Diaconis, P., Zabell, S.L.* Updating Subjective Probability // *Journal of the American Statistical Society.* – 1982. – Vol. 77, no. 380. – P. 822–830.
27. *Hunter, D.* Dempster-Shafer vs. Probabilistic Logic // In *PTOC. Third AAAI Uncertainty in Artificial Intelligence Workshop,* 1987. – P. 22–29.
28. *Pearl, J.* Reasoning with Belief Functions: An Analysis of Compatibility // *International Journal of Approximate Reasoning.* – 1990. – Vol. 4, no. 5. – P. 363–389.

29. Cattaneo, M. Combining Belief Functions Issued from Dependent Sources // In: Proc. Third International Symposium on Imprecise Probabilities and Their Application (ISIPTA'03). – Lugano, Switzerland, 2003. – P. 133–147.
30. Yager, R.R. On the Dempster-Shafer Framework and New Combination Rules // Information Sciences. – 1987. – Vol. 41. – P. 93–138.
31. Sentz, K., Ferson, S. Combination of Evidence in Dempster-Shafer Theory // In: Report SAND 2002-0835, Sandia National Laboratories, 2002.
32. Dubois, D., Prade, H. A Set-Theoretic View on Belief Functions: Logical Operations and Approximations by Fuzzy Sets // Int. J. of General Systems. – 1986. – No. 12. – P. 193–226.
33. Dubois, D. and Prade, H., On the Combination of Evidence in Various Mathematical Frameworks, in *Reliability Data Collection and Analysis. Eurocourses (Reliability and Risk Analysis)*, Flamm, J., Luisi, T., vol. 3, Dordrecht: Springer, 1992, pp. 213–241.
34. Lepskiy, A. General Schemes of Combining Rules and the Quality Characteristics of Combining // F. Cuzzolin (Ed.): BELIEF 2014, LNAI 8764. – Springer-Verlag, 2014. – P. 29–38.
35. Smets, P. The Alpha-Junctions: Combination Operators Applicable to Belief Functions // In First Int. Joint Conference on Qualitative and Quantitative Practical Reasoning (ECSQUARU-FAPR'97), 1997. – Lecture Notes in Computer Sciences, Vol. 1244. – Bad Honef, Germany: Springer International Publishing, 1997. – P. 131–153.
36. Smets, P. The Application of the Matrix Calculus to Belief Functions // International Journal of Approximate Reasoning. – 2002. – Vol. 31, no. 1-2. – P. 1–30.
37. Pichon, F., Denœux, T. Interpretation and Computation of Alpha-Junctions for Combining Belief Functions // In 6th Int. Symposium on Imprecise Probability: Theories and Applications (ISIPTA '09). – Durham, U.K., 2009.
38. Destercke, S., Burger, T. Toward an Axiomatic Definition of Conflict between Belief Functions // IEEE Transactions on Cybernetics. – 2013. – Vol. 43, no. 2. – P. 585–596.
39. Martin, A. About Conflict in the Theory of Belief Functions // Belief Functions: Theory and Applications, Advances in Intelligent and Soft Computing. – 2012. – Vol. 164. – P. 161–168.
40. Bronevich, A., Rozenberg, I. The Contradiction Between Belief Functions: Its Description, Measurement, and Correction Based on Generalized Credal Sets // International Journal of Approximate Reasoning. – 2019. – Vol. 112. – P. 119–139.
41. Jousselme, A.-L., Grenier, D., Bossé, E. A New Distance between Two Bodies of Evidence // Information Fusion. – 2001. – No. 2. – P. 91–101.
42. Jousselme, A.-L., Maupin, P. Distances in Evidence Theory: Comprehensive Survey and Generalizations // International Journal of Approximate Reasoning. – 2012. – Vol. 53. – P. 118–145.
43. Deza, M.M., Deza, E. Encyclopedia of Distances // Springer, Berlin Heidelberg, 2009.
44. Bouchard, M., Jousselme, A.-L., Doré, P.-E. A Proof for the Positive Definiteness of the Jaccard Index Matrix // International Journal of Approximate Reasoning. – 2013. – Vol. 54. – P. 615–626.
45. Attiaoui, D., Doré, P.-E., Martin, A., Ben Yaghlane, B. A Distance between Continuous Belief Functions // In Proc. of the Scalable Uncertainty Management (SUM) Conference, LNAI, Vol. 7520. – Springer-Verlag, Berlin Heidelberg, 2012. – P. 194–205.
46. Diaz, J., Rifqi, M., Bouchon-Meunier, B. A Similarity Measure between Basic Belief Assignments // In: Proceedings of the 9th International Conference Information Fusion. – Firenze, Italy, 2006.
47. Sunberg, Z., Rogers, J. A Belief Function Distance Metric for Orderable Sets // Information Fusion. – 2013. – Vol. 14. – P. 361–373.
48. Cuzzolin, F. Consistent Approximations of Belief Functions // In: 6th International Symposium on Imprecise Probability: Theories and Applications. – Durham, United Kingdom, 2009.
49. Tessem, B. Approximations for Efficient Computation in the Theory of Evidence // Artificial Intelligence. – 1993. – Vol. 61. – P. 315–329.
50. Liu, W. Analysing the Degree of Conflict among Belief Functions // Artificial Intelligence. – 2006. – Vol. 170. – P. 909–924.
51. Mercier, D., Quost, B., Denœux, T. Refined Modeling of Sensor Reliability in the Belief Function Framework Using Contextual Discounting // Information Fusion. – 2008. – No. 9. – P. 246–258.
52. Bronevich, A., Rozenberg, I. The Measurement of Relations on Belief Functions Based on the Kantorovich Problem and the Wasserstein Metric // International Journal of Approximate Reasoning. – 2021. – Vol. 131. – P. 108–135.
53. Loudahi, M., Klein, J., Vannobel, J.-M., Colot, O. New Distances between Bodies of Evidence Based on Dempsterian Specialization Matrices and Their Consistency with the Conjunctive Combination Rule // International Journal of Approximate Reasoning. – 2014. – Vol. 55, no. 5. – P. 1093–1112.
54. Loudahi, M., Klein, J., Vannobel, J.-M., Colot, O. Evidential Matrix Metrics as Distances between Meta-data Dependent Bodies of Evidence // IEEE Transactions on Systems, Man, and Cybernetics. – 2016. – Vol. 46, no. 1. – P. 109–122.
55. Lepskiy, A. On the Conflict Measures Agreed with the Combining Rules // In: Destercke, S., Denœux, T., Cuzzolin, F., Martin, A. (eds) Belief Functions: Theory and Applications. BELIEF 2018. Lecture Notes in Computer Science. – Vol. 11069. – Springer, Cham, 2018 – P. 172–180.
56. Bronevich, A.G., Rozenberg, I.N. Metrical Approach to Measuring Uncertainty // In: Lesot, M.J. et al. (eds) Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU 2020. Communications in Computer and Information Science, Vol. 1238. – Springer, Cham, 2020.
57. Lepskiy, A. About Relation between the Measure of Conflict and Decreasing of Ignorance in Theory of Evidence // Proceedings of the 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13). – Amsterdam – Beijing – Paris: Atlantis Press, 2013. – P. 355–362.
58. Zhang, L. Representation, Independence and Combination of Evidence in the Dempster-Shafer Theory // In: Yager, R.R., Kacprzyk, J., Fedrizzi, M. (eds.) Advances in the Dempster-Shafer Theory of Evidence. – John Wiley & Sons, New York, 1994. – P. 51–69.

Статья представлена к публикации членом редколлегии
П.Ю. Чеботаревым.

Поступила в редакцию 19.07.2021,
после доработки 27.08.2021.
Принята к публикации 31.08.2021

Лепский Александр Евгеньевич – д-р физ.-мат. наук, Национальный исследовательский университет «Высшая школа экономики», Москва, ✉ alex.lepskiy@gmail.com.

ANALYSIS OF INFORMATION INCONSISTENCY IN BELIEF FUNCTION THEORY. PART I: EXTERNAL CONFLICT

A.E. Lepskiy

National Research University Higher School of Economics, Moscow, Russia

✉ alex.lepskiy@gmail.com

Abstract. The analysis results of information inconsistency within belief function theory (the Dempster–Shafer theory of evidence) are reviewed. This theory has been intensively developing over the past 10–15 years. Part I of the survey considers the measure of external conflict between bodies of evidence. The concepts of conflict and non-conflict bodies of evidence and the basic requirements applied to measures of external conflict are discussed. Different axioms of the measure of external conflict are analyzed. The general forms of measures of external conflict that satisfy the system of axioms are given. Different methods for constructing measures of external conflict (metric, algebraic, and structural approaches; evaluation by combining rules) are presented. The robust estimation of external conflict, the relationship between its measure and the metric on the set of bodies of evidence, and the consistency of combining rules and measures of external conflict are discussed. Many illustrative examples are provided.

Keywords: theory of belief functions, combining rules, inconsistency of bodies of evidence, measure of external conflict.

Funding. This work was supported by the Russian Foundation for Basic Research, project no. 20-11-50077.

СТРУКТУРА КРЕАТИВНОЙ ДЕЯТЕЛЬНОСТИ

М.В. Белов, Д.А. Новиков

Аннотация. Рассмотрена специфика креативной/творческой деятельности. Выделены три фазы такой деятельности: открытие новой предметной области и накопление «базовых знаний», освоение предметной области, массовое продуктивное использование. На основании анализа жизненного цикла креативной деятельности показано, что креативность сосредоточена в этапе целеполагания и только в нем. Предложены качественная модель развития знаний/опыта и теоретико-графовая модель структуры предметной области. Для описания и исследования каждой из фаз креативной деятельности могут как развиваться новые, так и использоваться известные модели, в том числе разработанные и изложенные в предшествующих работах авторов: для первой фазы – модели оптимального распределения усилий исследователя между проверяемыми гипотезами и выбора оптимальной последовательности проверяемых гипотез; для второй фазы – математические модели опыта; для третьей фазы могут быть использованы как структурные и алгоритмические модели, так и оптимизационные модели.

Ключевые слова: креативная деятельность, опыт, творчество, предметная область, выдвижение и проверка гипотез.

ВВЕДЕНИЕ

Деятельность (activity) – активное взаимодействие человека с окружающей действительностью, в ходе которого человек выступает как *субъект*, целенаправленно воздействующий на *предмет* [1]. Деятельность – форма активности человека, направленная на познание, преобразование окружающего мира, себя и условий своего существования.

Под *элементарной* понимают такую деятельность, цели, технологии и результат которой не имеют собственной внутренней структуры¹.

В противоположность этому, деятельность, не являющаяся элементарной, в монографии [2] было предложено называть комплексной. То есть *комплексная деятельность* (КД) – деятельность, обладающая нетривиальной внутренней структурой, с множественными и/или изменяющимися целями, субъектом, технологией, ролью предмета в его целевом контексте.

В монографии [2] предложена классификация видов деятельности. В том числе выделены регулярная и креативная деятельность.

Регулярная КД – деятельность, реализуемая по известной технологии для получения априори специфицированных результатов. Структура и технология регулярной КД являются детерминированными.

Креативная КД – деятельность, технология которой не полностью определена (не полностью известна) на момент начала деятельности и поэтому создается в ходе реализации деятельности. Неизвестность технологии вызвана неопределенностью спроса и/или априорной неопределенностью в спецификации результата деятельности.

«Синонимом» креативной деятельности является *творческая деятельность* (дословный перевод на английский язык прилагательного «творческий» – “creative”). На сегодняшний день исторически сложились две парадигмы определения и исследования *творчества*. В рамках первой парадигмы – «деятельностной» – творчество рассматривается как деятельность (наиболее яркими примерами являются *научная деятельность* [3] и *художественная деятельность* [4]):

- творчество – деятельность человека, создающая новые материальные и духовные ценности, обладающие общественной значимостью [55];

¹ В случае элементарной деятельности нет необходимости рассматривать субъект и предмет вместе с собственно деятельностью – они играют роль понятного контекста (в течение периода деятельности эволюционирует только её предмет в соответствии с используемой субъектом технологией).



- творчество – это всякая практическая или теоретическая деятельность человека, в которой возникают новые (по крайней мере, для субъекта деятельности) результаты (*знания*, решения, способы действия, материальные продукты) [6];

- творчество – деятельность, результатом которой является создание новых материальных или духовных ценностей [7].

В рамках второй парадигмы – *психологии творчества* – творчество рассматривается как «взаимодействие, ведущее к развитию» [8, 9]. То есть в случае творчества, по мнению Я.А. Пономарева [8] и представителей его научной школы, субъект должен как бы допустить в себе активность объекта (внешней среды), поэтому взаимодействие, в отличие от деятельности, предполагает встречное действие объекта. Механизм этого встречного действия связан с такими категориями, как интуиция, озарение, когнитивное бессознательное, дефокусировка внимания, побочный продукт действия и т. д.

На самом деле, эти два исследовательских подхода не противоречат, а взаимно дополняют друг друга, так как, независимо от принятой позиции, для ответа на вопрос, откуда в сознании субъекта возникает образ «продукта творчества» (замысел у художника, гипотеза у ученого и т. д.), необходимо вводить те или иные предположения. В рамках психологии творчества в том числе исследуются механизмы отражения внешнего мира в сознании субъекта с учетом опыта последнего. В рамках деятельностного подхода, которому следует настоящая работа, вводится предположение о существовании графа, объективно и адекватно описывающего всю структуру предметной области (см. рис. 4 далее).

Креативность стала очень популярным предметом исследований в менеджменте и психологии управления с 1980–1990-х годов – см., например, обзоры в работах [10–13]. С тех пор наблюдается значительный поток публикаций по этой проблематике (см. наукометрический анализ в статье [14]), однако в упомянутых областях результаты исследований креативности носят качественный характер, как максимум – на уровне структурных моделей [15].

«Общая модель» деятельности, предложенная в книге [16], описывает деятельность субъектов: учитывает их активный выбор и деятельность во внешней среде, позволяет оценить получаемый результат деятельности и освоение технологий. «Общая модель» также описывает и позволяет исследовать динамику знаний, опыта, технологий, используя множества допустимых *структурных*

элементов деятельности (СЭДов) как функции времени и предыдущих значений действий, опыта, результата деятельности, состояния субъекта и значений *факторов неопределенности* (ФН). Однако соотношения «общей модели», используемые непосредственно, не позволяют получить каких-либо конструктивных результатов в силу своей аналитической сложности.

В настоящей работе рассматривается специфика креативной деятельности и, в терминах методологии комплексной деятельности [2], формулируется ее общая структура. В рамках модели жизненного цикла предметной области выделены и проанализированы его ключевые фазы, в том числе показано, что креативность «сосредоточена» в этапе целеполагания.

В § 1 вводятся базовые определения, в § 2 рассматриваются жизненный цикл предметной области, в § 3 локализуются креативные аспекты в жизненном цикле КД, § 4 содержит качественную модель развития знаний/опыта, § 5 – модель структуры предметной области.

1. ОПРЕДЕЛЕНИЯ

На основании определений понятий «знание», «опыт», «познание», «умение», данных в словаре [17], введем определения знаний и опыта. Причем в рамках настоящей работы будем считать понятия «опыт» и «знания» (индивида или группы людей) эквивалентными. *Опыт/знание* определим как результат процесса познания действительности, отраженный в сознании индивида или группы индивидов, а также в доступных им вещественных формах (документах и пр.) в виде представлений, понятий, суждений, умозаключений, теорий, навыков, обеспечивающих возможность выполнения людьми определенной деятельности в определенных условиях. Под *сознанием* будем понимать процесс и результат создания модели мира в определенных целях [18]; в этом смысле креативная деятельность тесно связана с сознанием, так как обе эти категории предполагают создание (дословный перевод – «creation») новых знаний.

Определим *элемент знаний* как утверждение о свойствах внешнего мира, справедливость которого в определенный момент (период) времени подтверждена наблюдениями в процессе выполнения СЭДа или может быть проверена посредством выполнения СЭДа (системы СЭДов).

Например, элемент знаний может быть задан как подмножество декартового произведения множеств возможных значений параметров внешней

среды (включая предметную часть технологий) – он может определять множество возможных значений параметров, в том числе в разные моменты времени. То есть описывать последовательность смены состояний элементов внешней среды или отношения между их параметрами, в том числе под влиянием реализуемой деятельности субъектов.

Активность, в том числе способность к самостоятельному целеполаганию, выбору состояний/действий и рефлексии, является базовой характеристикой человека (*активного элемента* – АЭ).

Будем говорить, что элемент знаний известен *активной системе* (АС – системе, включающей АЭ, т. е. человека), если гипотеза о справедливости соответствующего утверждения подтверждена в ходе выполнения одного или нескольких СЭДов.

Для каждого элемента знаний существуют *предусловия* – множество элементов знаний, которые должны быть известны для того, чтобы гипотеза о справедливости соответствующего утверждения могла быть проверена. В каждый момент времени может быть определено множество элементов знаний, гипотезы которых могут быть проверены, т. е. элементов знаний, *предусловия* (текущий «фронт познания» – см. Рис. 4 ниже) которых удовлетворяют текущему состоянию опыта/знаний.

Тогда каждый элемент знаний в любой момент времени может находиться в одном из следующих состояний относительно АС:

- недоступен для проверки гипотезы – *предусловия* не удовлетворены,
- доступен для проверки гипотезы, но гипотеза не проверена,
- известен – гипотеза проверена.

Формализуем накопленные к текущему моменту опыт/знания АС в виде алгебры множеств множества элементов знаний, известных АС на текущий момент.

Следуя определению [2] *технологии* как системы условий, критериев, форм, методов и средств последовательного достижения поставленной цели, будем рассматривать технологию состоящей из двух компонентов: технологических знаний и предметов, являющихся *средствами* деятельности. При этом технологические знания являются, очевидно, подмножеством опыта/знаний в целом, а предметная часть технологии (средства) может рассматриваться как составляющая часть внешней среды (ресурсы деятельности).

Историческая практика показывает, что процесс развития знаний/опыта человечества носит «скачкообразный» характер: кратковременные периоды

(почти моменты – *научные революции* по Т. Куну [19]) формирования новых парадигм – новых областей знаний – сменяются относительно продолжительными периодами так называемого *нормального развития* – освоения и продуктивного использования знаний. Этот процесс естественным образом «выделяет» *предметные области* – подмножества множеств элементов знаний, возможно, содержательно близких, взаимосвязанных друг с другом. Рассмотрим их жизненные циклы.

2. ЖИЗНЕННЫЙ ЦИКЛ ПРЕДМЕТНОЙ ОБЛАСТИ

На основании анализа общественно-исторической практики выделим фазы² (см. рис. 1 и рис. 2) жизненного цикла (ЖЦ) области знаний.

Фаза I (*открытие новой предметной области и накопление «базовых знаний»*). В рамках этой фазы реализуется (последовательно и/или параллельно) множество СЭДов, целями которых является получение знаний: проверка гипотез о справедливости утверждений, составляющих элементы знаний/опыта. При реализации СЭДов проверки гипотез значимые условия утверждений задаются (выбираются) субъектом (индивидуальным или коллективным). Соответствующие математические модели рассматриваются в подразделе 5.2 книги [16].

Возможный результат каждой такой проверки бинарен и неизвестен априори (!), поэтому в каждом СЭДе реализуется истинная неопределенность [20] и, в зависимости от полученного результата, гипотеза или отвергается, или подтверждается. Эти СЭДы могут реализовываться в форме деятельности и над вещественными предметами (например, физический эксперимент), и над информационными (например, математическое моделирование) объектами, и над мыслимыми (мысленный эксперимент).

Будем считать, что первая фаза продолжается до тех пор, пока неизвестно, какое практически полезное применение могут иметь знания из данной области, пока не сформулированы полезные цели и не разрабатываются технологии их достижения.

Цель каждого из СЭДов и содержание фазы в целом – получение знаний о внешней среде.

² Границы фаз и начало жизненного цикла в целом являются условными. В большинстве случаев невозможно указать единственное событие (момент его наступления), указывающее начало фазы или жизненного цикла. Это и не требуется для построения формальных моделей.

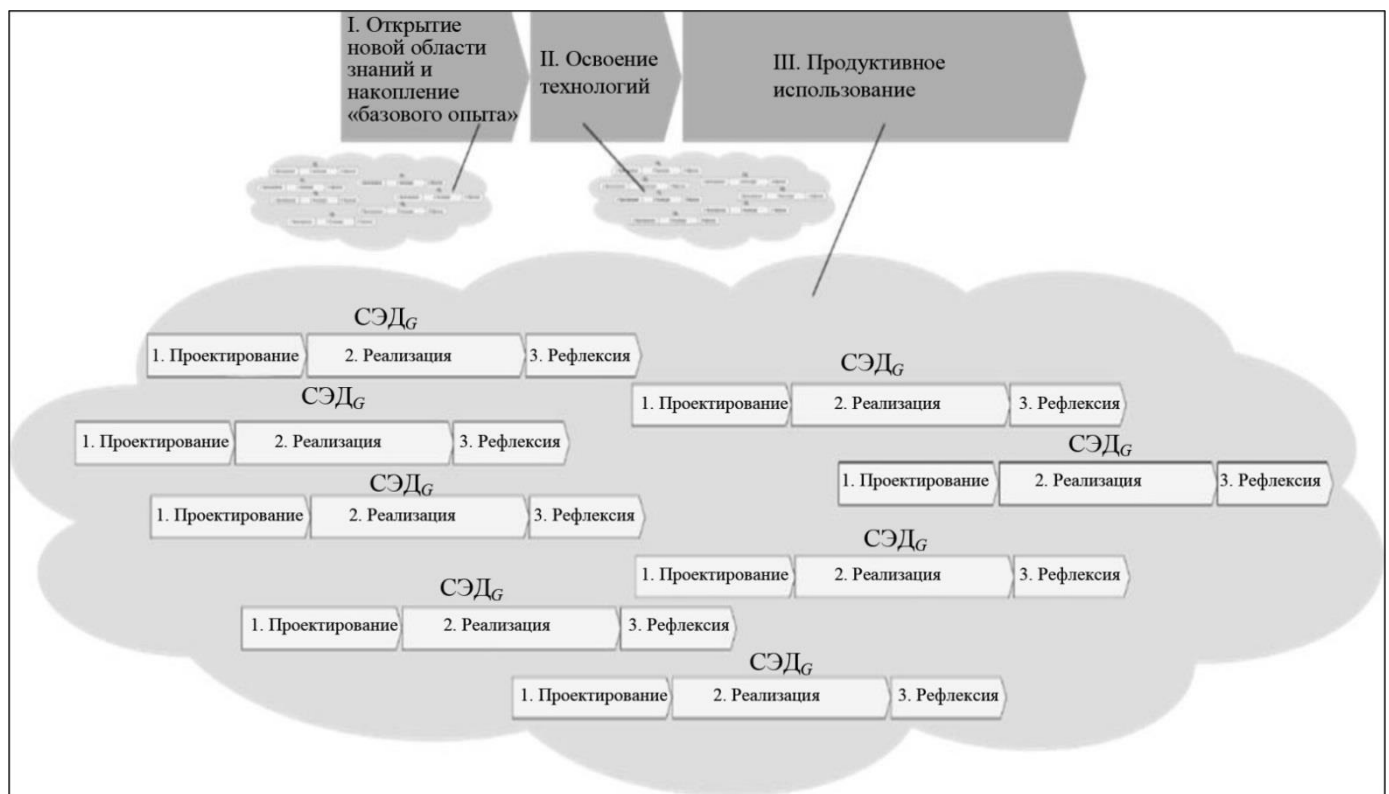


Рис. 1. Модель жизненного цикла предметной области

Таким образом, происходит накопление «базовых знаний/опыта» – получение знаний о свойствах (значениях и законах их динамики) ФН, а также о технологиях КД, выполняемых при определенном множестве значений ФН. На первой фазе технологии КД нацелены на получение новых знаний – построение модели окружающей действительности, а не на получение полезного результата.

Фаза II (освоение предметной области) – разработка технологии и единичное продуктивно-экспериментальное использование. Признаком перехода от фазы I к фазе II является появление гипотез о возможном полезном применении знаний – формулирование новых полезных целей. Цели СЭДов на фазе II – формирование технологий получения полезных результатов на основе модели ФН, построенной на фазе I. При этом значения ФН не выбираются субъектом, а реализуются «естественным выбором» внешней среды. То есть не все значимые условия утверждений задаются субъектом (как это было на фазе I), часть из них определяются значениями ФН. При этом могут реализовываться и уже известные значения ФН (тогда новые элементы знаний/опыта не формируются) или априори не встречавшиеся значения ФН (тогда

формируются новые элементы знаний/опыта). Целью такого многократного повторения известной технологии является подтверждение того, что технология позволяет получить желаемый продуктивный результат КД при (всех или) большинстве значений ФН, или выявление новых значений ФН. Для описания этого процесса адекватны модели, представленные в работах [21, 22].

В каждой предметной области и их комбинации может быть создано конечное количество «разумных», «рациональных», «оптимальных» технологий (например, наилучшая с учетом имеющихся в данный момент материалов конструкция электродвигателя, или паровой машины, или самолета).

Фаза III (массовое продуктивное использование). Признаком перехода от фазы II к фазе III является массовое / не единичное использование известных технологий. Реализуется множество СЭДов с уже известной (из фазы II) технологией с целью получения продуктивного результата, ради которого технология создавалась. При этом могут реализовываться и уже известные значения ФН (тогда получается желаемый продуктивный результат) или априори не встречавшиеся значения ФН (тогда формируется новый опыт, а продуктивный резуль-

тат может быть потерян). При встрече с новым значением ФН может происходить возвращение к фазе I или II в рамках данной предметной области или открытие новой предметной области.

Формированию новых СЭДов на фазах I и II соответствует формирование гипотез об элементах знаний, формированию новых СЭДов на фазе III – порождение потребностей в результатах продуктивной КД, направленной на получение полезного результата.

3. КРЕАТИВНЫЕ АСПЕКТЫ В ЖИЗНЕННОМ ЦИКЛЕ КОМПЛЕКСНОЙ ДЕЯТЕЛЬНОСТИ

Зададимся вопросом, какие из элементов (процессуальных, внутренних или внешних) «отвечают» за то, что деятельность является креативной? Другими словами, где «сосредоточена» креативность? Для этого проанализируем *жизненный цикл (ЖЦ) КД*, приведенный в табл. 10 монографии [2], а также в табл. 1 настоящей статьи.

Таблица 1

**Фазы, стадии и этапы жизненного цикла
элемента комплексной деятельности (А-СЭДа) и их содержание**

Фаза	Стадия	№	Название этапа	Содержание этапа
Проектирование	I. Фиксация спроса и осознание потребности	1	Фиксация спроса и осознание потребности	Вышестоящий В-СЭД или внешняя среда формирует спрос на результаты элемента КД. Субъект фиксирует спрос, осознает потребности и принимает решение осуществлять деятельность
	II. Целеполагание и структурирование целей и задач	2	Формирование логической модели	Потребность структурируется, проверяется, является ли она известной, и соответственно, деятельность – регулярной. Если КД регулярна, этап сводится к извлечению информации о логической модели из информационного хранилища. Иначе формируется структура целей. Цели формулируются в терминах ожидаемых характеристик результатов элементов КД. Проверяется непротиворечивость / модифицируется структура целей. Каждой цели А-СЭДа ставится в соответствие роль субъекта и технология (результат уже поставлен в соответствие ранее) – специфицируются характеристики субъектов и технологий. Результат этапа: логическая модель – структура А-СЭДа в виде перечня нижестоящих СЭДов (Н-СЭДов) и элементарных операций (Н-Оп)
	III. Формирование технологии	3	Проверка готовности технологии и достаточности ресурсов	Проверяется наличие уже известных компонентов технологии А-СЭДа: причинно-следственной модели А-СЭДа, технологий всех Н-СЭДов и технологий всех Н-Оп. Проверяется логическая согласованность А-СЭДа и пулов ресурсов – наличие и достаточность ресурсов для назначения субъектов Н-СЭДов и обеспечения технологий Н-Оп с учетом использования данных ресурсов параллельно при реализации других СЭДов. Результат этапа: подтверждение готовности технологии, подтверждение наличия необходимых ресурсов и переход к этапу 7 или выполнение этапов 4, 5 или 6 соответственно
		4	Создание причинно-следственной модели	Определяются и описываются причинно-следственные связи между целями/результатами нижестоящих элементов (Н-СЭДов и Н-Оп). Описываются возможные события неопределенности и правила реагирования на них (выполняемые СЭДы или эскалация на вышестоящий уровень). Результат этапа: причинно-следственная модель А-СЭДа

См. окончание табл. 1



Окончание табл. 1

Фаза	Стадия	№	Название этапа	Содержание этапа
Проектирование	III. Формирование технологии	5	Создание технологии нижестоящих элементов	Для элементарной операции, в силу ее специфичности и отсутствия внутренней структуры, процесс проектирования и описания элементов технологии специфичен и поэтому не допускает общего описания. Для всех нижестоящих Н-СЭДов, для которых отсутствуют готовые технологии, рекурсивно выполняются этапы с первого по шестой их жизненных циклов. Результат этапа: технологии нижестоящих элементарных операций Н-Оп и технологии нижестоящих Н-СЭДов
		6	Формирование/ модернизация ресурсов	При отсутствии необходимых ресурсов порождаются цели, отвечающие за их генерацию, выполняются СЭДы, обеспечивающие создание или модернизацию пулов ресурсов. Результат этапа: необходимые пулы ресурсов
		7	Календарное и ресурсное планирование	Формируется календарно-сетевой график. Проверяется согласованность ключевых сроков потребности. Проверяется временная согласованность календарно-сетевого графика и пула ресурсов с учетом использования ресурсов другими элементами деятельности. При несогласованности выполняется возврат к этапам 2–4 или осуществляется эскалирование невозможности удовлетворить сроки субъекту вышестоящего СЭДа. Результат этапа: календарно-сетевой план и график использования ресурсов
		8	Оптимизация	Осуществляется оптимизация динамики использования ресурсов (с учетом возможности реализации с использованием данных ресурсов параллельно других КД). Результат этапа: оптимальный календарно-сетевой план и график использования ресурсов
		9	Назначение субъектов и определение ответственности	Фиксируется матрица ответственности (соответствие между субъектами СЭДов и персоналом). Фактически назначение субъектов означает формирование спроса на результаты нижестоящих СЭДов и таким образом – рекуррентное обращение к описываемому здесь процессу выполнения ЖЦ Н-СЭДов: выполняются все стадии фазы Проектирования. Результат этапа: матрица ответственности, которая совместно со структурой А-СЭДа детерминирует его организацию
		10	Назначение ресурсов	Соответственно технологиям элементарных операций производится запрос и назначение требуемых для выполнения технологий ресурсов. Результат этапа: матрица назначений ресурсов элементарным операциям
Реализация	IV. Выполнение действий и получение результата	11	Выполнение действий и получение результата	Соответственно причинно-следственной модели многократно и постоянно выполняется проверка предусловий начала действий элементарных операций Н-Оп и Н-СЭДов и их запуск. Выполнение элементарных операций Н-Оп. Запуск выполнения нижестоящих Н-СЭДов. Результат этапа: выполнение действий А-СЭДа и получение результата
Рефлексия	V. Оценивание результата и рефлексия	12	Оценивание результата и рефлексия	Сопоставление характеристик результата с требуемыми. Сопоставление объемов ресурсов с заданными. Проектирование требований к корректировкам целей, технологии и т. д.

Стадия II жизненного цикла КД (см. табл. 1) – это *целеполагание* (формирование структуры целей) и проверка, существуют ли технологии под все подцели. Если нет, то формулируется цель – разработка соответствующей технологии. Следует отметить, что иногда цели – как предвосхищаемый образ будущего результата – могут (и, в случае творчества, зачастую так и происходит) формулироваться субъектом неосознанно (с точки зрения самого субъекта, целеполагание при этом как бы «отсутствует»), приводя затем к «неожиданным» результатам.

Из табл. 1 видно, что *креативность сосредоточена в стадии целеполагания* и только в ней! Действительно, формирование технологии и реализация – стадии III и IV и последующие (см. табл. 1) – всегда (!) выполняются известными средствами и методами, так как нельзя быть уверенным в результате, если не уверен в средствах, методах способах его получения – в технологии. Именно на стадии целеполагания субъект ставит цель, которая не достигалась до этого и про которую неизвестно, может ли она быть достигнута. При этом известная технология КД отсутствует, поэтому субъект производит декомпозицию поставленной цели на подцели (это эвристическая, креативная активность) и проверяет наличие известной технологии для каждой из подцелей. При отсутствии технологии – де-

композирует подцель. И так далее до тех пор, пока для всех подцелей детализированной таким образом цели не будут обнаружены известные технологии. Реализация набора известных технологий для достижения полученной структуры (под)целей является регулярной КД. Впрочем, как и последующая рефлексия – проверка соответствия результата исходной цели.

В общую схему жизненного цикла КД и жизненного цикла предметной области укладывается и научная КД, и практическая КД (включая инженерную КД), и художественная КД.

Научная КД. Сейчас во многих отраслях знаний уже сложились свои парадигмы – устоявшиеся исследовательские подходы – методы исследований и представления результатов. Альтернативные подходы воспринимаются с априорным подозрением, т. е. исследователи стремятся следовать известной – регулярной – технологии, в которую «вкраплены» СЭДы «проверки гипотез», которые тоже регулярны! Исключением являются моменты научных революций. Приведем примеры (табл. 2).

Художественная КД [4] «устроена» аналогично; примерами проектов, реализуемых множеством людей, являются:

- съемки фильма,
- постановка спектакля,
- создание монументальных объектов.

Таблица 2

Примеры фаз ЖЦ областей знаний

Фаза ЖЦ области знаний	Значения ФН	Примеры, связанные с электричеством	Примеры, связанные с атомной энергией
Фаза I. Открытие новой области знаний и накопление «базового опыта»	Генерируются/выбираются субъектом и параллельно реализуются	– опыты древних с янтарем и шерстью, наблюдения за электрическими угрями, – работы и прибор Гилберта, – опыты Франклина с воздушным змеем, – опыты и приборы Гальвани и Вольта, – генератор Фарадея, – и др.	– идеи-гипотезы древних о строении вещества, – теория Дальтона, открытие Менделеева, – рентген, – модели атома Бора и Резерфорда, – наблюдение Беккереля, работы Кюри, Флерова и Петряка, – и др.
Фаза II. Освоение области знаний. Цель – разработка и освоение технологий	Реализуются, а субъект фиксирует новые значения ФН	– применение электричества для освещения, промышленного привода, транспорта и т. д., – промышленные устройства Теслы, Эдисона, Дюво-Добровольского и др.	– «Манхэттенский проект» – первый реактор, бомба, технологии обогащения, – ядерные боеприпасы, – первые АЭС
Фаза III. Массовое продуктивное использование. Цель – использование разработанных технологий для получения новых результатов	Реализуются	Серийное производство и массовое применение	



Формирование замысла (сюжета, смысла, ...) – целостного образа – книги/картины (или их элементов, если субъект делит предмет творчества на элементы) – является элементарной, но креативной деятельностью. После этого субъект выражает замысел до тех пор, пока ему не понравится результат, т. е. тоже имеет место проверка гипотез (или переосмысление замысла). При этом технология накладывания краски, обработки мрамора или набора текста/формул – регулярна. Такая технология тоже может быть предметом первой фазы (см. п. с) ниже). Но, будучи выработанной, она становится регулярной.

Таким образом, для художественной деятельности имеем:

- формирование замысла художественного произведения – формирование гипотезы/гипотез;
- попытки выразить замысел и проверка «нравится или не нравится» – проверка гипотез;
- художественная техника (гуашь, масло, глина, бронза, мазок) – с точностью до авторской индивидуальности, отражающей замысел, регулярна и «импортируется» из промышленности.

Стадии I и II (и V) КД (см. табл. 1) всегда реализуются применительно к информационной модели предметной области и в ее терминах. Остальные стадии, III и IV, могут требовать КД, связанной с вещественным предметом.

Цели/гипотезы (стадия II) КД (см. табл. 1) могут описывать любые подмножества элементов (и связей между ними) корпуса знаний независимо от модели их представления. В том числе гипотезы могут описывать новые компоненты технологий.

И обобщенная схема одиночного креативного СЭДа, и ЖЦ области знаний совпадают со *схемой научной деятельности*:

А. Осмысление существующих предметных областей;

Б. Формирование целей – в случае креативной КД это выдвижение гипотез о возможности достижения целей, так как на момент целеполагания технология неизвестна, следовательно, и возможность достижения цели также неизвестна. В случае научной КД целью КД является получение новых знаний, т. е. непосредственно проверка гипотез о закономерностях внешней «относительно исследователя» среды;

В. Проверка гипотез;

Г. Обобщение и формирование новых закономерностей/технологий;

Д. Переход к п. А.

Пункты А – Б – это целеполагание (формирование структуры целей) и проверка, существуют ли технологии под все подцели. Повторим, что креативность сосредоточена в этом.

Проверка гипотез (п. В) всегда (!) выполняется посредством использования известной технологии, так как нельзя быть уверенным в результате, если не уверен в технологии – средствах, методах, способах его получения. Если технология проверки неизвестна, производится декомпозиция гипотезы и определяется причинно-следственная структура проверки гипотез нижестоящего уровня с последующим агрегированием результатов. Операции декомпозиции, формирования причинно-следственной структуры и агрегирования являются известными – проверенными компонентами общесистемной технологии достижения комплексных целей / проверки комплексных гипотез. Декомпозиция производится до тех пор, пока для всех целей/гипотез не будет найдено известной технологии. Получается фрактальная совокупность СЭДов. После этого реализуются действия СЭДов, следуя причинно-следственным структурам. После этого исходная гипотеза или подтверждается, или отвергается.

Гипотезы (п. Б) могут описывать любые подмножества элементов (и связей между ними) корпуса знаний независимо от модели представления. Как отмечалось выше, гипотезы могут описывать новые компоненты технологий.

Формирование целей всегда осуществляется в ответ на потребность некоторых заинтересованных лиц и/или решение проблем (что эквивалентно потребностям) заинтересованных лиц, в частном случае таким заинтересованным лицом является сам субъект (в научной и художественной деятельности таким «заинтересованным лицом» в большинстве случаев является сам ученый или художник).

Итак, можно говорить, что исполнение комплексной деятельности всегда регулярно: все этапы (см. табл. 1), следующие за целеполаганием и структурированием целей, реализуются, следуя известной на момент начала действия технологии. То есть, как только цель структурирована и создана структура СЭДов (с верифицированной, известной технологией, приводящей к требуемому результату), выполнение КД становится регулярным.

4. КАЧЕСТВЕННАЯ МОДЕЛЬ РАЗВИТИЯ ЗНАНИЙ/ОПЫТА

В рамках любой АС существование/функционирование активных элементов состоит в том, что:

а) активная система как комплексный субъект «постоянно» реализует множество регулярных СЭДов;

б) наступают события истинной неопределенности;

с) активные элементы, составляющие АС, рефлексируют – осмысливают факты наступления этих событий;

д) (пере)формируется структура целей, в результате появляется структура СЭДов и структура комплексных субъектов;

е) (для новой структуры целей) разрабатывается новая технология или сводится путем декомпозиции к известным;

ф) продолжается реализация регулярной, но уже другой КД, происходит возврат к п. а).

Цепочки а) – ф) реализуются для всех этапов ЖЦ «областей знаний» / предметных областей (см. Рис. 2).

События истинной неопределенности (b) и события переформирования структур целей (d) происходят асинхронно, они «связаны» процессом *рефлексии*, продолжительность и результат которого также неопределенны. Само переформирование целей – проявление истинной неопределенности субъекта.

Таким образом, неопределенность субъекта проявляется в двух формах:

– решение о выполнении деятельности (или отказе),

– формирование структуры целей.

Не для всех предметных областей жизненный цикл включает все три фазы: в некоторых случаях может оказываться, что развитие предметной области не приводит (или пока еще не приводит) к ее продуктивному использованию, и жизненный цикл прерывается на первой или второй фазе (соответствующие графические образы представлены в верхней части рис. 2).

Рефлексия является оцениванием имеющегося опыта и внешней среды, в том числе событий истинной неопределенности. Рефлексия, с одной стороны, предшествует целеполаганию и является его источником – так осуществляется генерирование гипотез. С другой стороны, рефлексия фиксирует опыт – так осуществляется подтверждение / отклонение гипотез.

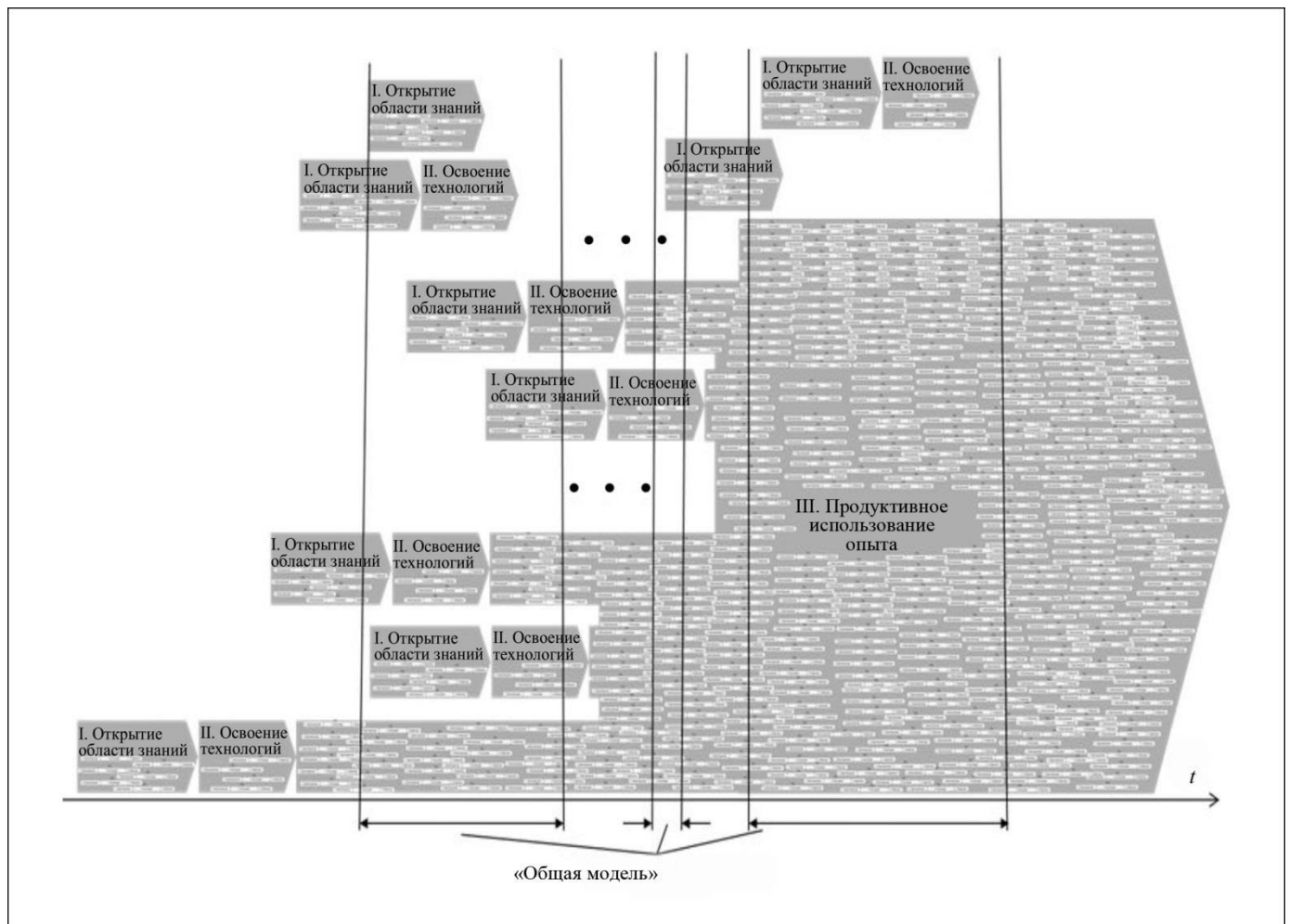


Рис. 2. Модель развития знаний/опыта активной системы

«Общая модель» (см. введение и книгу [16]) описывает в том числе и эволюцию знаний/опыта, используя динамику множеств допустимых СЭДов и зависимость их от предыстории, однако, как уже отмечалось выше, соотношения «общей модели», используемые непосредственно, не позволяют получить каких-либо конструктивных результатов.

Поэтому конкретизируем «общую модель АС» [16] таким образом, чтобы аналитически исследовать развитие знаний/опыта – процесс открытия новых предметных областей и накопления «базовых знаний», для этого:

- i. абстрагируемся от множественности АЭ;
- ii. «избавимся» от множеств-функций, описывающих СЭДы, и представим процесс эволюции иным образом.

Рассмотрим особенности реализации жизненного цикла опыта/знаний, учитывая цели i-ii (см. табл. 3). Множество допустимых действий каждого АЭ состоит из СЭДов, каждый из которых может быть отнесен к одному из подмножеств, характерных для различных фаз ЖЦ предметных областей:

- фаза I – СЭДы проверки гипотез, доступных при текущем уровне опыта,
- фаза II – СЭДы освоения технологий,
- фаза III – СЭДы продуктивного использования освоенных технологий.

При реализации СЭДов подмножества I АЭ выбирает значения параметров и технологии, и внешней среды, проверяя справедливость гипотезы именно при таком сочетании параметров. Истинная неопределенность проявляется в виде результата КД, который априори неизвестен АЭ. Если при этом истинная неопределенность внешней среды проявляется таким образом, что ФН принимают отличные от требуемого значения, то полученный результат КД характеризует проверку гипотезы, отличной от исходной. При завершении проверки гипотезы множества значений ФН и доступных технологий могут трансформироваться (см. табл. 3, Рис. 3 и Рис. 4).

При реализации СЭДов подмножеств II и III АЭ выбирают номер выполняемого элемента КД – параметры технологии, при этом параметры внешней

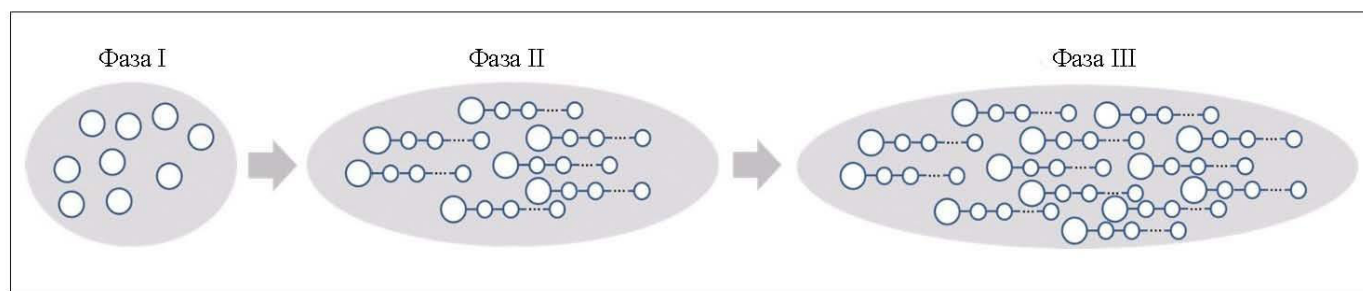


Рис. 3. Жизненный цикл предметной области

Таблица 3

Фазы креативной КД

Выбор АЭ	Значение ФН	Итог	Последствие
Проверка гипотезы, фаза I	Совпадает с требуемым	Гипотеза проверена	Множество доступных технологий трансформировано.
	Не совпадает с требуемым	Проверена иная гипотеза	Возможно, множество значений ФН трансформировано
Освоение технологии, фаза II	Известное	Гипотеза проверена	Уровень освоения неизменен
	Неизвестное		Уровень освоения повысился
Использование технологии, фаза III	Известное	Продуктивный СЭД выполнен	Получен ожидаемый полезный результат
	Неизвестное		Уровень освоения повысился, ожидаемый полезный результат может не быть получен

среды – номер состояния ФН – реализуются независимо от АЭ и являются априори неопределенными для АЭ. При этом результат КД зависит от значений параметров и технологии, и внешней среды.

5. МОДЕЛЬ СТРУКТУРЫ ПРЕДМЕТНОЙ ОБЛАСТИ

Пусть задан связный ориентированный граф без циклов – *сеть* $G = (N, E)$ с правильной нумерацией (отсутствуют дуги, идущие из вершины с большим номером в вершину с меньшим номером), вершины которой соответствуют предметным областям (множествам гипотез)/гипотезам/утверждениям, а множество дуг $E \subseteq N \times N$ отражает логические связи между вершинами (см. Рис. 4).

Через $N_i = \{j \in N \mid (j; i) \in E\}$ обозначим множество непосредственных *предшественников* i -й вершины в сети G , $i \in N$. Считаем, что сеть G имеет множество $N_0 \subseteq N$ *входов* (вершин, не имеющих предшественников – аксиом и/или фактов, являющихся признанным общим знанием).

Будем считать, что для каждой вершины сети G определено *предусловие* – булев предикат $\pi_i(\cdot)$, определенный на множестве $|N_i| + 1$ *входов* двух типов – исходных фактов $z_{N_i} = \{z_j, j \in N_i\}$ и *внешних условий* $\omega_i(\cdot) \in \Omega_i$ – и вычисляющий бинарный *выход* (новый факт) – логическую переменную $z_i = \pi_i(z_{N_i}, \Omega_i)$, которую будем считать определенной, если значение выхода одинаково при любых допустимых значениях внешних условий; в противном случае выход будем считать неопределенным.

Следовательно, проверка гипотезы при известных исходных фактах заключается в нахождении значения выхода при некотором (или некоторых) значении внешних условий – первая фаза креативной КД. Затем, на второй фазе креативной КД, неизменность значения выхода проверяется при различных (всех допустимых) значениях внешних условий (ФН).

Таким образом, i -я вершина графа G задается кортежем $(N_i, \Omega_i, \pi_i(\cdot))$, включающим исходные факты, внешние условия и логический предикат.

Обозначим через G_t достоверно известный исследователю/лям (неважно, сколько субъектов осуществляют параллельно проверку гипотез, обмениваясь результатами) в момент времени t подграф графа G . В примере на рис. 4 граф G_t затенен.

В рамках рассматриваемой модели структуры предметной области *гипотеза* – предположение о

том, что справедливо некоторое утверждение или их совокупность. Гипотеза – подграф/вершина, в котором входные дуги всех вершин либо содержатся в нем, либо исходят из графа G_t .

Обоснование или *опровержение гипотезы* – проверка определенности результата утверждения при всех фигурирующих в нем внешних условиях. Частный случай – когда предикат известен и надо найти максимальное множество внешних условий, когда его значение является определенным.

Модель проверки гипотез рассматривается в подразделе 5.2 книги [16], где в том числе ставятся и решаются задачи оптимального распределения усилий исследователя между проверкой различных выдвинутых гипотез и задачи выбора оптимальной последовательности проверки гипотез.

На первой фазе ЖЦ предметной области «элемент знаний» (гипотеза) – вершина графа на рис. 4 – может быть, как отмечалось в настоящем разделе выше, следующих типов:

- недоступен для проверки гипотезы – предусловия не удовлетворены (обозначены пунктиром),
- доступен для проверки гипотезы, но гипотеза не проверена (обозначен тонкой линией),
- известен – гипотеза проверена (обозначен жирной линией).

Множество известных элементов знаний (вершин, обозначенных на рис. 4 жирными линиями) – *текущий объем известных знаний*. Множество доступных для проверки гипотез (вершин, обозначенных на рис. 4 тонкими линиями) – *текущий горизонт познания*. Текущий объем известных знаний совместно с текущим горизонтом познания представляют собой известный исследователю в текущий момент подграф G_t графа G . Каждая вершина, обозначенная тонкой линией, представляет собой одну гипотезу или множество независимо проверяемых (параллельно или последовательно) гипотез.

Возможны два типа действий АЭ на каждом шаге – *алгоритмические* и *креативные* действия. Первые – полностью автоматические генерация и содержательный анализ всех логически возможных следствий из имеющегося корпуса знаний G_t . Результатом является граф $\hat{G}_t$ (условно его можно называть «логическим замыканием» графа G_t). Вторые – выдвижение и обоснование или опровержение гипотез – новых подграфов G_t^h графа $G \setminus \hat{G}_t$. Обозначим через G_t^{h+} множество подтвержденных гипотез, тогда $G_{t+1} = \hat{G}_t \cup G_t^{h+}$. Выдви-

жение/генерация гипотез является существенно креативным и неформализуемым этапом, поэтому при моделировании, наверное, целесообразно использовать стохастические описания моментов возникновения гипотез и продолжительностей их проверки.

Таким образом, новые гипотезы могут выдвигаться либо автоматически (алгоритмически), либо креативно на первой фазе, а также алгоритмически или креативно в качестве реакции на события истинной неопределенности на второй (см. стрелку 6 на рис. 4) или третьей (см. стрелку 7 на рис. 4) фазах креативной КД.

Для формализации процесса проверки (обоснования или опровержения) выдвинутых гипотез можно использовать модели, рассматриваемые ниже в настоящем разделе:

- на фазе I:
 - переход к новому горизонту познания происходит в момент научной революции [19]; как и почему это происходит, мы описывать не беремся, считая, что граф G задан; и это следует признать наиболее существенным допущением рассматриваемых моделей креативной КД;
 - проверка гипотез описывается моделью, приведенной в подразделе 5.2 книги [16];
 - содержанием стрелок типа 1 и 2 может быть детализация или увеличение/снижение размерности (анализ/синтез, декомпозиция/обобщение);
- на фазе II отработка технологий описывается «моделями опыта» [21, 22];
- на фазе III практическая деятельность описывается общими схемами КД, приведенными в монографии [2].

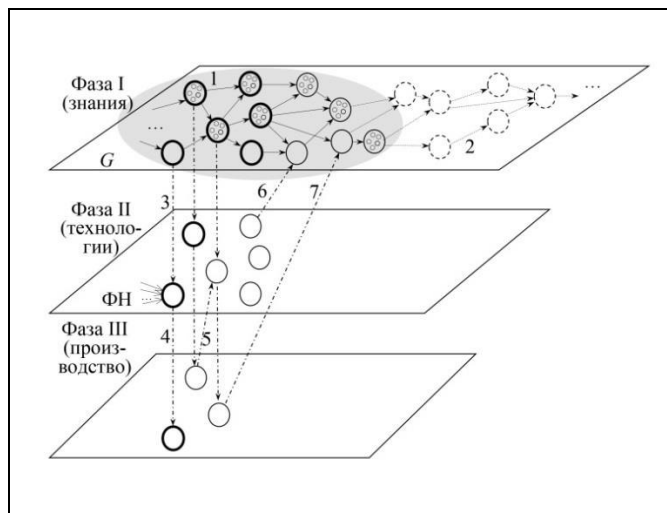


Рис. 4. Фазы креативной деятельности (ЖЦ предметной области)

Цепочка 3–4 описывает ЖЦ «фундаментальные исследования – разработка технологий – производство».

Связи 5, 6 и 7 иллюстрируют, что проблемы, возникшие на второй или третьей фазе (ситуации истинной неопределенности для этой фазы), могут потребовать возврата к предыдущей или предыдущим фазам с выдвижением новых гипотез, их проверкой и/или с разработкой соответствующих технологий.

ЗАКЛЮЧЕНИЕ

Выделены три фазы жизненного цикла креативной деятельности (предметной области):

- фаза I – открытие новой предметной области и накопление базовых знаний (генерация и проверка гипотез);
- фаза II – освоение предметной области;
- фаза III – массовое продуктивное использование.

Показано, что креативность «сосредоточена» в этапе целеполагания (применительно к научной или художественной деятельности – в генерации гипотез). Для описания и исследования каждой из этих фаз могут как развиваться новые, так и использоваться известные модели, в том числе:

- для первой фазы в книге [16] предложены математические модели оптимального распределения усилий исследователя между проверяемыми гипотезами и выбора оптимальной последовательности проверяемых гипотез;
- для второй фазы целесообразно использовать математические модели опыта, приведенные в работах [21, 22];
- для третьей фазы могут быть использованы как структурные и алгоритмические модели, приведенные в монографии [2], так и оптимизационные модели, содержащиеся в книге [23].

ЛИТЕРАТУРА

1. Новиков А.М., Новиков Д.А. Методология. – М.: Синтег, 2007. – 668 с. [Novikov, A.M., Novikov, D.A. Metodologiya. – Moscow: Sinteg, 2007. – 668 s. (In Russian)]
2. Белов М.В., Новиков Д.А. Методология комплексной деятельности. – М.: Ленанд, 2018. – 320 с. [Belov, M.V., Novikov, D.A. Methodology of Complex Activity: Foundations of Understanding and Modelling. – Cham: Springer, 2020. – 223 p.]
3. Новиков А.М., Новиков Д.А. Методология научного исследования. – М.: Либроком, 2009. – 280 с. [Novikov, A.M., Novikov, D.A. Metodologiya nauchnogo issledovaniya. – Moscow: Librokom, 2009. – 280 s. (In Russian)]

4. Новиков А.М. Методология художественной деятельности. – М.: Эгвес, 2008. – 72 с. [Novikov, A.M. Metodologiya khudozhestvennoi deyatel'nosti. – Moscow: Egves, 2008. – 72 s. (In Russian)]
5. Рубинштейн С.Л. Основы общей психологии. – СПб: Издательство «Питер», 2000. – 712 с. [Rubinshtein, S.L. Osnovy obshchei psikhologii. – Saint-Petersburg: Izdatel'stvo «Piter», 2000. – 712 s. (In Russian)]
6. Психологический словарь / Под ред. В.П. Зинченко, Б.Г. Мещерякова. 2-е изд. – М.: Педагогика-Пресс, 1999. – 440 с. [Psikhologicheskii slovar' / Pod red. V.P. Zinchenko, B.G. Meshcheryakova. 2-e izd. – Moscow: Pedagogika-Press, 1999. – 440 s. (In Russian)]
7. Краткий психологический словарь / Сост. Л.А. Карпенко. Под общ. ред. А.В. Петровского, М.Г. Ярошевского. – М.: Феникс, 1998. – 512 с. [Kratkii psikhologicheskii slovar' / Sost. L.A. Karpenko. Pod obshch. red. A.V. Petrovskogo, M.G. Yaroshevskogo. – Moscow: Feniks, 1998. – 512 s. (In Russian)]
8. Пономарев Я.А. Психология творчества. – М.: ИП АН СССР, 1976. – 304 с. [Ponomarev, Ya.A. Psikhologiya tvorchestva. – Moscow: IP AN SSSR, 1976. – 304 s. (In Russian)]
9. Психология творчества: школа Я.А. Пономарева / Под ред. Д.В. Ушакова. – М.: ИП РАН, 2006. – 624 с. [Psikhologiya tvorchestva: shkola Ya.A. Ponomareva / Pod red. D.V. Ushakova. – Moscow: IP RAN, 2006. – 624 s. (In Russian)]
10. Amabile, T., Pratt, M. The Dynamic Componential Model of Creativity and Innovation in Organizations // Research in Organizational Behavior. – 2016. – Vol. 36. – P. 157–183.
11. Handbook of Organizational Creativity / Ed. By M. Mumford. – N.Y.: Academic Press, 2011. – 753 p.
12. Handbook of the Management of Creativity and Innovation / Ed. by M. Tangand C. Werner. – Singapoure: Worldscientific, 2017. – 420 p.
13. The Cambridge Handbook of Creativity / Ed. by J. Kaufman and R. Sternberg. – Cambridge: Cambridge University Press, 2010. – 508 p.
14. Cai, W., Khapovs, S.N., Bosnik, B., Lysova, E.I. Optimizing Employee Creativity in the Digital Era: Uncovering the Interactional Effects of Abilities, Motivations, and Opportunities // Int. J. Environ. Res. Public Health. – 2020. – Vol. 17. – P. 1038–1057.
15. Amabile, T. Componential Theory of Creativity / Working Paper No 12-096. – Harvard: Harvard Business School, 2012. – 10 p.
16. Белов М.В., Новиков Д.А. Модели деятельности. – М.: Ленанд, 2021. – 220 с. [Belov, M.V., Novikov, D.A. Modeli deyatel'nosti. – Moscow: Lenand, 2021. – 220 s. (In Russian)]
17. Платонов К.К. Краткий словарь системы психологических понятий. – М.: Высшая школа, 1984. – 174 с. [Platonov, K.K. Kratkii slovar' sistemy psikhologicheskikh ponyatii. – Moscow: Vysshaya shkola, 1984. – 174 s. (In Russian)]
18. Митио К. Будущее разума. – М.: Альпина нон-фикшн, 2018. – 504 с. [Michio, K. The Future of the Mind: The Scientific Quest to Understand, Enhance, and Empower the Mind. – N.-Y.: Doubleday, 2014. – 400 p.]
19. Кун Т. Структура научных революций. – М.: Прогресс, 1975. – 288 с. [Kuhn, T.S. The Structure of Scientific Revolutions. – Chicago: University of Chicago Press, 1962. – 210 p.]
20. Knight, F. Risk, Uncertainty and Profit / Hart Schaffner and Marx Prize Essays. No. 31. – Boston and New York: Houghton Mifflin, 1921. – 381 p.
21. Белов М.В., Новиков Д.А. Модели опыта // Проблемы управления. 2021. – № 1. – С. 43–60. [Belov, M.V., Novikov, D.A. Models of Experience. – 2021. – Control Sciences. – No. 1. – P. 37–52.]
22. Белов М.В., Новиков Д.А. Модели технологий. – М.: Ленанд, 2018. – 157 с. [Belov, M.V., Novikov, D.A. Models of Technologies. – Cham: Springer Nature, 2020. – 113 p.]
23. Белов М.В., Новиков Д.А. Управление жизненными циклами организационно-технических систем. – М.: Ленанд, 2020. – 384 с. [Belov, M.V., Novikov, D.A. Upravlenie zhiznennymi tsiklami organizatsionno-tekhnicheskikh sistem. – Moscow: Lenand, 2020. – 384 s. (In Russian)]

Статья представлена к публикации руководителем
регионального редсовета А.А. Ворониным.

Поступила в редакцию 16.05.2021,
после доработки 31.07.2021.
Принята к публикации 2.08.2021.

Белов Михаил Валентинович – д-р техн. наук, Сколковский институт науки и технологий, г. Москва,
✉ mbelov59@mail.ru,

Новиков Дмитрий Александрович – чл.-корр. РАН, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, ✉ novikov@ipu.ru.



THE STRUCTURE OF CREATIVE ACTIVITY

M.V. Belov¹ and D.A. Novikov²

¹ Skolkovo Institute of Science and Technology, Moscow, Russia

² Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

¹✉ mbelov59@mail.ru, ²✉ novikov@ipu.ru

Abstract. The specifics of creative activity are considered. There are three phases of such activity: discovering a new knowledge domain (subject matter) and accumulating basic knowledge, mastering the knowledge domain, and mass productive use. The life cycle of creative activity is analyzed. As shown by the analysis, creativity is concentrated in the stage of goal-setting only. A qualitative model for mastering knowledge (experience) and a graph-theoretic structural model of a knowledge domain are proposed. New models can be developed, and well-known models can be used to describe and study each phase of creative activity, including those introduced by the authors earlier: in the first phase, optimal distribution models for the researcher's efforts between the tested hypotheses and optimal scheduling models for tested hypotheses; in the second phase, mathematical models of experience; in the third phase, structural and algorithmic models and optimization models.

Keywords: creative activity, experience, creativity, knowledge domain, making and testing hypotheses.

АДАПТИВНОЕ НЕЙРОСЕТЕВОЕ УПРАВЛЕНИЕ НЕЛИНЕЙНЫМИ ОБЪЕКТАМИ С ДЕФИЦИТОМ КАНАЛОВ УПРАВЛЕНИЯ НА ПРИМЕРЕ ДВУХКОЛЕСНОГО БАЛАНСИРУЮЩЕГО РОБОТА¹

А.И. Глущенко, В.А. Петров, К.А. Ласточкин

Аннотация. Для управления нелинейными объектами с целью компенсации несогласованного параметрического возмущения в условиях дефицита каналов управления предложен новый метод, основанный на беспойсковом адаптивном управлении с эталонной моделью. Регулятор в данном случае состоит из базовой LQ-части и адаптивного компенсатора, при определенных допущениях позволяющего уменьшить по норме указанное нелинейное возмущение. В силу своих универсальных аппроксимационных свойств, в качестве компенсатора выступает оперативно обучаемая многослойная нейронная сеть. Синтезированы законы настройки параметров нейронной сети компенсатора на основе второго метода Ляпунова и метода обратного распространения ошибки. Строго аналитически доказана асимптотическая сходимость к заданной области ошибки слежения нелинейным объектом с дефицитом каналов управления за эталонной моделью. Численные эксперименты по применению разработанной системы управления на математической модели балансирующего робота LEGO EV3 в системе Matlab/Simulink подтвердили полученные теоретические результаты.

Ключевые слова: адаптивное управление с эталонной моделью, балансирующий робот, компенсация несогласованных параметрических возмущений, нейронные сети, оперативное обучение, устойчивость.

ВВЕДЕНИЕ

В современной практике теории автоматического управления все большую актуальность приобретают задачи управления техническими объектами с существенным влиянием параметрических неопределенностей. Классическими примерами таких объектов управления являются манипуляторы [1, 2], беспилотные и пилотируемые летательные аппараты в особых режимах эксплуатации [3, 4], промышленные электроприводы [5], а также технологические процессы в химической [6] и металлургической [7, 8] отраслях промышленности.

Большинство современных методов построения систем управления для объектов с параметрическими неопределенностями можно разделить на робастные подходы и методы беспойскового адаптивного управления. Робастные системы проекти-

руются таким образом, чтобы показатели качества замкнутой системы управления (обычно запасы устойчивости по фазе и амплитуде) в наихудших условиях работы объекта управления удовлетворяли априорно сформированным условиям [9]. Адаптивные же системы выполняют оперативную оценку неопределенности, а затем на ее основе формируют управляющее воздействие таким образом, чтобы минимизировать отклонение поведения объекта управления от эталонного [3]. По сравнению с робастными подходами, адаптивные беспойсковые системы не требуют априорного знания диапазона вариаций параметров объекта управления (максимальной величины параметрической неопределенности), а также, при достаточном запасе по мощности управляющего воздействия, позволяют получить не компромиссное качество управления, а эталонное.

Все методы построения адаптивных беспойсковых систем управления с эталонной моделью можно разделить на прямые, не прямые и композитные [3, 10, 11]. В первом случае выполняется непосред-

¹Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований (проект 18-47-310003-р_а).



ственно настройка параметров регулятора заранее выбранной структуры, во втором случае сначала оцениваются параметры объекта управления и/или параметрической неопределенности, а затем, исходя из полученной информации, рассчитываются параметры регулятора. Системы композитного адаптивного управления особым образом сочетают подходы прямого и непрямого адаптивного управления для повышения качества процесса адаптации.

Основными проблемами всех трех групп методов беспоиcкового адаптивного управления являются [3, 10]:

- сходимость настраиваемых параметров к идеальным значениям только при выполнении ограничительного условия незатухающего возбуждения регрессора,
- неудовлетворительное с точки зрения практики качество переходных процессов по настраиваемым параметрам, управлению и ошибке слежения (особенно при увеличении количества настраиваемых параметров) при использовании стандартного контура настройки,
- необходимость экспериментального, ручного выбора матрицы коэффициента усиления контура адаптации,
- необходимость знания знака матрицы коэффициентов усиления объекта управления.

В последние годы к решению данных проблем было приложено немало усилий как отечественных, так и зарубежных исследователей. Из последних значительных результатов отметим работы [12–15].

Однако существует и пятая, более фундаментальная и менее проработанная проблема теории адаптивного беспоиcкового управления – наличие допущения о выполнении условия адаптируемости объекта управления [3, 10]. Суть ее заключается в том, что методы беспоиcкового адаптивного управления в общей постановке могут быть применены только в том случае, если параметрическая неопределенность объекта согласована с управляющим сигналом, что означает теоретическую возможность полной компенсации неопределенности путем прямого вычитания сформированного управления. Если условие адаптируемости не выполняется, то прямая компенсация невозможна, а неопределенность (параметрическое возмущение) является несогласованной (*unmatched uncertainties*) и требуется применение специальных средств и способов синтеза закона управления и его настройки.

В целом, можно выделить два основных класса возмущений, несогласованных с сигналом управления. К первому классу относятся возмущения,

находящиеся в описании объекта управления в автономных дифференциальных уравнениях. Ко второму же классу относятся возмущения, возникающие в неавтономных уравнениях объекта управления в условиях дефицита каналов управления (в зарубежной литературе такие системы известны как *underactuated systems with unmatched uncertainties*).

Для компенсации несогласованных возмущений, принадлежащих к первому классу, уже достаточно долгое время существуют подходы, относящиеся к методам адаптивного бэкстеппинга [16] и не прямой компенсации на основе функций настройки [17]. К недостаткам этих подходов можно отнести высокий динамический порядок закона управления и его настройки, а также увеличение сложности процесса их синтеза с ростом порядка объекта управления. Для решения этих проблем недавно были предложены новые методы [18–20], позволяющие другим образом учитывать и компенсировать влияние несогласованных параметрических возмущений. Упомянутые решения основаны на непосредственном объединении теории адаптивного [10] и робастного [9] управления. В частности, в работах [18, 20] с помощью методов непрямого беспоиcкового адаптивного управления предлагается выполнять оценку несогласованного возмущения, а затем, по полученной информации, с помощью LMI-синтеза перевычислять параметры регулятора и эталонной модели, что в совокупности позволяет адаптивно поддерживать робастность замкнутой системы к произвольным несогласованным возмущениям, компенсируя при этом влияние согласованных возмущений.

Для компенсации влияния несогласованных возмущений из второй группы в литературе известно не так много методов, относящихся к теории беспоиcкового адаптивного управления с эталонной моделью. Основная трудность здесь заключается в том, что дефицит каналов управления ведет к присутствию одного сигнала управления сразу в нескольких уравнениях, что, в свою очередь, в общем случае ведет к нетривиальной задаче синтеза компенсирующего управления. Для решения этой задачи в геометрической и нелинейной теории управления хорошо известны [21] различные методы замены координат [21–25], позволяющие перейти от модели объекта с дефицитом управляющих воздействий к эквивалентной ей модели в нормальном виде, позволяя тем самым выбрать закон управления по методу линеаризации обратной связью [21]. Недостатками методов замены координат являются сложность, а иногда и невозможность вычисления точного преобразования для нелинейных объектов большой размерности, а

также зависимость самого преобразования от параметров объекта управления, что приводит к необходимости его адаптации. Поэтому указанная задача до сих пор остается в полной мере нерешенной.

В настоящей же работе для компенсации несогласованного параметрического возмущения в условиях дефицита каналов управления предлагается новый подход. Его основой выступает предположение о существовании идеального компенсирующего управления, позволяющего уменьшить несогласованное возмущение по норме. На основе принятого предположения представляется возможным получить компенсирующее управление, решив оптимизационную задачу. Так как большинство реальных объектов с дефицитом управляющих воздействий описывается нелинейными дифференциальными уравнениями, то и несогласованное возмущение является нелинейным, а задача оптимизации – нелинейно параметризованной и в общей постановке достаточно трудно разрешимой. Поэтому для ее решения именно в общей постановке в данной работе предлагается для формирования компенсирующего управляющего воздействия использовать искусственные нейронные сети, обладающие универсальными аппроксимирующими свойствами [26]. Законы настройки параметров нейронной сети компенсатора при этом предлагается синтезировать путем сочетания второго метода Ляпунова и метода обратного распространения ошибки.

Результатом работы является полученный закон формирования компенсирующего нейросетевого управления и алгоритм оперативной настройки его параметров, обеспечивающие асимптотическую сходимость к заданной области ошибки слежения нелинейным объектом с дефицитом каналов управления за выбранной эталонной моделью.

Двухколесный балансирующий робот представляет собой классический пример нелинейного объекта с дефицитом каналов управления, поэтому в данной работе, не ограничивая общности, синтез закона формирования адаптивного нейросетевого управления был осуществлен именно на его примере.

В работе используются следующие обозначения: $(\cdot)(i, j)$ или $(\cdot)_{i,j}$ – элемент матрицы, стоящий на пересечении i -й строки и j -го столбца; $\det(\cdot)$ – определитель матрицы; $\text{tr}(\cdot)$ – след матрицы (сумма элементов главной диагонали); $\text{vec}(\cdot)$ – операция векторизации матрицы (запись столбцов друг под другом); $\text{diag}\{a \ b \ \dots \ c\}$ – диагональная матрица с элементами диагонали $a, b, \dots, c$; $\|\cdot\|_\infty$ – L_∞

норма матрицы; $\|\cdot\|$ – Евклидова норма вектора или Фробениусова норма матрицы (будет ясно из контекста); $(\cdot)_{n \times n}$ – матрица размерностью $n \times n$.

1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДВУХКОЛЕСНОГО БАЛАНСИРУЮЩЕГО РОБОТА

Дифференциальные уравнения, описывающие динамику двухколесного балансирующего робота, могут быть получены с помощью формализма второго метода Эйлера – Лагранжа [27] и после приведения в форму Коши имеют вид:

$$\begin{cases} \dot{x}_1 = x_3, \\ \dot{x}_2 = x_4, \\ \dot{x}_3 = -2(\beta + f_w)[E^{-1}(1, 1) - E^{-1}(1, 2)]x_3 - \\ - 2\beta[E^{-1}(1, 2) - E^{-1}(1, 1)]x_4 + \\ + [E^{-1}(1, 1)MLRx_4^2 + E^{-1}(1, 2)MgL]\sin(x_2) + \\ + \alpha[E^{-1}(1, 1) - E^{-1}(1, 2)](u_1 + u_2), \\ \dot{x}_4 = -2(\beta + f_w)[E^{-1}(2, 1) - E^{-1}(2, 2)]x_3 - \\ - 2\beta[E^{-1}(2, 2) - E^{-1}(2, 1)]x_4 + \\ + [E^{-1}(2, 1)MLRx_4^2 + E^{-1}(2, 2)MgL]\sin(x_2) + \\ + \alpha[E^{-1}(2, 1) - E^{-1}(2, 2)](u_1 + u_2), \end{cases} \quad (1.1)$$

$$E^{-1} = \begin{bmatrix} \frac{J + ML^2 + 2n^2J_m}{\det(E)} & -\frac{MLR \cdot \cos(x_2) - 2n^2J_m}{\det(E)} \\ -\frac{MLR \cdot \cos(x_2) - 2n^2J_m}{\det(E)} & \frac{2J_w + (2m_w + M)R^2 + 2n^2J_m}{\det(E)} \end{bmatrix},$$

$$\det(E) = (2J_w + (2m_w + M)R^2 + 2n^2J_m) \times$$

$$\times (J + ML^2 + 2n^2J_m) - (MLR \cdot \cos(x_2) - 2n^2J_m)^2,$$

$$\beta = \frac{nK_t K_b}{R_m} + f_m, \quad \alpha = \frac{nK_t}{R_m}.$$

Модель (1.1) получена в допущении структурной и параметрической идентичности исполнительных двигателей. Параметры модели (1.1) имеют следующий физический смысл: J_w – момент инерции колеса; m_w – масса колеса; M – масса робота; R – радиус колеса; n – передаточное число редуктора; J_m – момент инерции двигателя; L – расстояние от центра масс до оси колеса; K_t – постоянная момента двигателя; R_m – сопротивление якорной цепи двигателя; K_b – постоянная противо-ЭДС; f_m – коэффициент трения между телом робота и валом двигателя; f_w – коэффициент трения между колесом и поверхностью движения; J – момент инерции тела робота; g – ускорение свободного



падения. Координатами состояния робота являются: x_1 – средний угол поворота колес; x_2 – угол отклонения тела робота от нормали; x_3 – скорость поворота колес; x_4 – скорость отклонения тела робота от нормали. В качестве управляющего воздействия выступают напряжения u_1 и u_2 на левый и правый двигатели соответственно.

Для удобства дальнейших рассуждений введем в модель (1.1) обозначения:

$$\begin{aligned} f_3(x_2, x_3, x_4) = & \\ = -2(\beta + f_w) & [E^{-1}(1, 1) - E^{-1}(1, 2)]x_3 - \\ -2\beta & [E^{-1}(1, 2) - E^{-1}(1, 1)]x_4 + \\ + [E^{-1}(1, 1)MLR & x_4^2 + E^{-1}(1, 2)MgL] \sin(x_2), \\ g_{31}(x_2) = g_{32}(x_2) = & \alpha [E^{-1}(1, 1) - E^{-1}(1, 2)], \\ f_4(x_2, x_3, x_4) = & \\ = -2(\beta + f_w) & [E^{-1}(2, 1) - E^{-1}(2, 2)]x_3 - \\ -2\beta & [E^{-1}(2, 2) - E^{-1}(2, 1)]x_4 + \\ + [E^{-1}(2, 1)MLR & x_4^2 + E^{-1}(2, 2)MgL] \sin(x_2), \\ g_{41}(x_2) = g_{42}(x_2) = & \alpha [E^{-1}(2, 1) - E^{-1}(2, 2)]. \end{aligned} \quad (1.2)$$

2. ПОСТАНОВКА ЗАДАЧИ

С учетом обозначений (1.2) представим модель (1.1) в пространстве координат состояний:

$$\begin{aligned} \dot{x} = A_0 x + B_3 f_3(x_2, x_3, x_4) + & \\ + B_4 f_4(x_2, x_3, x_4) + g(x)u, & \\ A_0 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}; B_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}; B_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}; & \\ g(x_2) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ g_{31}(x_2) & g_{32}(x_2) \\ g_{41}(x_2) & g_{42}(x_2) \end{pmatrix}. & \end{aligned} \quad (2.1)$$

Здесь $x \in R^4$ – доступный для измерения вектор координат состояния робота; $u = [u_1; u_2] \in R^2$ – вектор напряжений на левый и правый двигатели; $f_3(x_2, x_3, x_4)$, $f_4(x_2, x_3, x_4)$ и $g(x_2)$ – нелинейные достаточно гладкие по Липшицу функции, определенные в выражениях (1.2).

Допущение 1. Управляющее воздействие u является таким, что $u_1 = u_2$.

Принятое допущение является классическим для случая, когда не требуется управление вращением робота относительно центра масс в горизонтальной плоскости.

В силу допущения 1 модель (1.1) с двумя управляющими воздействиями может быть приведена к эквивалентной модели с одним управляющим воздействием $v = u_1$:

$$\begin{aligned} \dot{x} = A_0 x + B_3 f_3(x_2, x_3, x_4) + & \\ + B_4 f_4(x_2, x_3, x_4) + Bv, & \\ B = \begin{bmatrix} 0 & 0 & g_3(x_2) & g_4(x_2) \end{bmatrix}^T; & \\ \begin{cases} g_3(x_2) = g_{31}(x_2) + g_{32}(x_2), \\ g_4(x_2) = g_{41}(x_2) + g_{42}(x_2). \end{cases} & \end{aligned} \quad (2.2)$$

Дополнительно будем рассматривать вспомогательную линейную модель, полученную из модели (2.2) путем линеаризации функций $f_3(x_2, x_3, x_4)$, $f_4(x_2, x_3, x_4)$ и нелинейных элементов вектора B в точке неустойчивого равновесия робота ($x_2 = 0$):

$$\begin{aligned} \dot{x} = A_0 x + B_3 \bar{f}_3(x_2, x_3, x_4) + & \\ + B_4 \bar{f}_4(x_2, x_3, x_4) + \bar{B}v = \bar{A}x + \bar{B}v. & \end{aligned} \quad (2.3)$$

Замечание 1. Для получения из модели (1.1) модели (2.3) необходимо задаться некоторыми значениями физических и геометрических параметров модели робота (1.1) и воспользоваться справедливыми в малой окрестности точки линеаризации $x_2 = 0$ равенствами $\lim_{x_2 \rightarrow 0} \sin(x_2) = x_2$; $\lim_{x_2 \rightarrow 0} \cos(x_2) = 1$;

$x_4^2 = 0$. Параметры робота при получении линеаризованной модели (2.3) могут быть известны не точно.

На основе линейной модели (2.3) может быть рассчитан LQ-оптимальный закон управления

$$v_{LQ} = K_{LQ}(r - x) = K_{LQ}e, \quad (2.4)$$

где $r \in R^4$ – вектор заданий по координатам состояния робота, а матрица K_{LQ} получена путем решения задачи минимизации критерия

$$J = \frac{1}{2} \int_0^\infty x^T Q_{LQ} x + R_{LQ} v^2 d\tau, \quad (2.5)$$

а $Q_{LQ} \in R^{4 \times 4}$ и $R_{LQ} \in R$ – положительно определенные диагональные матрицы.

Желаемое качество управления нелинейным объектом (2.2) задается системой (2.3), замкнутой регулятором (2.4):

$$\begin{aligned} \dot{x}_{ref} = A_{ref} x_{ref} + B_{ref} r, & \\ A_{ref} = \bar{A} - \bar{B} K_{LQ}; B_{ref} = \bar{B} K_{LQ}. & \end{aligned} \quad (2.6)$$

Чтобы получить уравнение в отклонениях между уравнениями нелинейного объекта (2.2) и его

линейной эталонной моделью (2.6), введем соотношения:

$$\begin{aligned} f_3(x_2, x_3, x_4) &= \bar{f}_3(x_2, x_3, x_4) + \Delta_{f3}, \\ f_4(x_2, x_3, x_4) &= \bar{f}_4(x_2, x_3, x_4) + \Delta_{f4}, \\ B &= \bar{B} + \Delta_B, \end{aligned} \quad (2.7)$$

где Δ_{f3} , Δ_{f4} , Δ_B – неизвестные достаточно гладкие по Липшицу функции, вызванные параметрическими возмущениями и проявлением нелинейностей при $x_2 \gg 0$.

Подставляя соотношения (2.7) в модель (2.2), имеем:

$$\begin{aligned} \dot{x} &= A_0 x + B_3 [\bar{f}_3(x_2, x_3, x_4) + \Delta_{f3}] + \\ &+ B_4 [\bar{f}_4(x_2, x_3, x_4) + \Delta_{f4}] + (\bar{B} + \Delta_B) v. \end{aligned} \quad (2.8)$$

На основании выражения (2.8) закон управления v выберем в виде

$$v = v_{LQ} - v_{ad}. \quad (2.9)$$

Тогда, учитывая выражения (2.3), (2.6), (2.7) и (2.9), уравнение (2.8) имеет вид:

$$\begin{aligned} \dot{x} &= A_{ref} x + B_{ref} r + \\ &+ \underbrace{B_3 \Delta_{f3} + B_4 \Delta_{f4} + \Delta_B v - \bar{B} v_{ad}}_{\Lambda(z)}, \end{aligned} \quad (2.10)$$

где $\Lambda(z)$ – неизвестная достаточно гладкая по Липшицу функция, описывающая влияние параметрических возмущений и нелинейностей на качество управления балансирующим роботом, а $z = [x_2 \ x_3 \ x_4 \ v] \in D \subset R^4$ – аргумент функции $\Lambda(z)$, определенный в компактной области D пространства R^4 .

Уравнение в отклонениях между формулами (2.10) и (2.6) имеет вид:

$$\dot{e}_{ref} = A_{ref} e_{ref} + \Lambda(z) - \bar{B} v_{ad}, \quad (2.11)$$

где $e_{ref} = x - x_{ref}$ – ошибка слежения объектом управления (2.10) за эталонной моделью (2.6).

Нетрудно проверить, что для вектора $\bar{B}$ не существует псевдообратной матрицы Мура – Пенроуза такой, что $\bar{B}^+ \bar{B} = I$. Следовательно, $\Lambda(z)$ является несогласованным с сигналом управления возмущением. Поэтому для разрешения задачи синтеза управления v_{ad} в адаптивной постановке выдвинем следующее предположение относительно сигнала компенсации.

Предположение 1. Существует сигнал компенсации v_{ad}^* аргумента z такой, что верно

$$\begin{aligned} \|\Lambda(z) - \bar{B} v_{ad}^*\| &\leq \varepsilon_\Lambda < \Lambda_{\max}, \\ v_{ad}^* &= \arg \left[\min_{v_{ad}^*} \left\{ \sup \|\Lambda(z) - \bar{B} v_{ad}^*\| \right\} \right], \end{aligned} \quad (2.12)$$

где $\Lambda_{\max} = \sup_{z \in L_\infty} \|\Lambda(z)\|$, ε_Λ – погрешность аппроксимации $\Lambda(z)$ с помощью сигнала компенсации v_{ad}^* .

Замечание 2. Если предположение 1 не выполняется, то задача адаптивной компенсации возмущения $\Lambda(z)$ является принципиально неразрешимой в классе гладких функций.

Для большей наглядности проиллюстрируем на примере случай, при котором предположение 1 справедливо. Пусть, например, разность $\Lambda(z) - \bar{B} v_{ad}^*$ имеет вид

$$\Lambda(z) - \bar{B} v_{ad}^* = \begin{bmatrix} 0 \\ 0 \\ a_{32}x_2 + a_{33}x_3 + a_{34}x_4 \\ a_{42}x_2 + a_{43}x_3 + a_{44}x_4 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ b_3 \\ b_4 \end{bmatrix} v_{ad}^*.$$

Тогда, если выполняется хотя бы одно из соотношений

$$\frac{a_{32}}{a_{42}} \propto \frac{b_3}{b_4}; \quad \frac{a_{33}}{a_{43}} \propto \frac{b_3}{b_4}; \quad \frac{a_{34}}{a_{44}} \propto \frac{b_3}{b_4},$$

то предположение 1 оказывается верным, а часть несогласованной с сигналом управления неопределенности $\Lambda(z)$ может быть компенсирована.

Таким образом, в условиях отсутствия информации о структуре функции $\Lambda(z)$, а также при невозможности ее полной прямой компенсации в работе ставится задача минимизации ошибки (2.12) и обеспечения асимптотической сходимости ошибки слежения (2.11) к ограниченному множеству

$$\lim_{t \rightarrow \infty} \|e_{ref}(t)\| \leq \bar{\varepsilon}_{e_{ref}}, \quad (2.13)$$

где $\bar{\varepsilon}_{e_{ref}}$ – предельное значение ошибки слежения.

Прежде чем переходить к основному результату работы, введем и обоснуем ограничение на значение ошибки $\bar{\varepsilon}_{e_{ref}}$. Для этого оценим миноранту и мажоранту ошибки слежения e_{ref} . Положив $v_{ad} = v_{ad}^*$ и $v_{ad} = 0$, из уравнения (2.11) получаем соответственно оценку снизу и сверху на производную $\dot{e}_{ref}$:

$$A_{ref} e_{ref} + \Lambda(z) - \bar{B} v_{ad}^* \leq \dot{e}_{ref} < A_{ref} e_{ref} + \Lambda(z). \quad (2.14)$$

Для вычисления миноранты и мажоранты ошибки e_{ref} рассмотрим квадратичную форму

$$L = e_{ref}^T P e_{ref}, \quad (2.15)$$

$$\lambda_{\min}(P) \|e_{ref}\|^2 \leq L \leq \lambda_{\max}(P) \|e_{ref}\|^2,$$

где P – решение уравнения Ляпунова $A_{ref}^T P + PA_{ref} = -Q$, $Q > 0$.



Производная квадратичной формы (2.15) в силу неравенства (2.14) приобретает вид:

$$\begin{aligned} & e_{ref}^T (A_{ref}^T P + P A_{ref}) e_{ref} + \\ & + 2e_{ref}^T P [\Lambda(z) - \bar{B}v_{ad}^*] \leq \dot{L} < \\ & < e_{ref}^T (A_{ref}^T P + P A_{ref}) e_{ref} + 2e_{ref}^T P \Lambda(z). \end{aligned} \quad (2.16)$$

С учетом предположения 1 из неравенства (2.16) получим:

$$\begin{aligned} & -\lambda_{\min}(Q) \|e_{ref}\|^2 + 2\varepsilon_{\Lambda} \lambda_{\max}(P) \|e_{ref}\| \leq \\ & \leq \dot{L} < -\lambda_{\min}(Q) \|e_{ref}\|^2 + 2\lambda_{\max}(P) \Lambda_{\max} \|e_{ref}\|. \end{aligned} \quad (2.17)$$

Пользуясь справедливым для любых $a > 0$ и $b > 0$ неравенством

$$\begin{aligned} -a^2 + ab &= \frac{1}{2} [-a^2 - (a-b)^2 + b^2] \leq \\ & \leq -\frac{1}{2} a^2 + \frac{1}{2} b^2, \end{aligned} \quad (2.18)$$

из (2.17) имеем

$$\begin{aligned} & -\frac{\lambda_{\min}(Q)}{2\lambda_{\max}(P)} L + \frac{2\varepsilon_{\Lambda}^2 \lambda_{\max}^2(P)}{\lambda_{\min}(Q)} \leq \dot{L} < \\ & < -\frac{\lambda_{\min}(Q)}{2\lambda_{\max}(P)} L + \frac{2\lambda_{\max}^2(P) \Lambda_{\max}^2}{\lambda_{\min}(Q)}. \end{aligned}$$

С учетом определения (2.15) имеем миноранту и мажоранту ошибки слежения:

$$\begin{aligned} & \frac{1}{\lambda_{\min}(P)} e^{-\frac{\lambda_{\min}(Q)}{2\lambda_{\max}(P)} t} \|e_{ref}(0)\|^2 + \\ & + \frac{4\varepsilon_{\Lambda}^2 \lambda_{\max}^3(P)}{\lambda_{\min}(P) \lambda_{\min}^2(Q)} \leq \|e_{ref}\|^2 < \\ & < \frac{1}{\lambda_{\min}(P)} e^{-\frac{\lambda_{\min}(Q)}{2\lambda_{\max}(P)} t} \|e_{ref}(0)\|^2 + \\ & + \frac{4\lambda_{\max}^3(P) \Lambda_{\max}^2}{\lambda_{\min}(P) \lambda_{\min}^2(Q)}. \end{aligned} \quad (2.19)$$

Приняв $t \rightarrow \infty$, из выражения (2.19) имеем оценку предельной ошибки слежения:

$$\begin{aligned} & 2\varepsilon_{\Lambda} \frac{\lambda_{\max}(P)}{\lambda_{\min}(Q)} \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} \leq \bar{\varepsilon}_{e_{ref}} < \\ & < 2\Lambda_{\max} \frac{\lambda_{\max}(P)}{\lambda_{\min}(Q)} \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}. \end{aligned} \quad (2.20)$$

Таким образом, в соответствии с выражениями (2.13) и (2.20) требуется построить закон компенсации v_{ad} , обеспечивающий асимптотическую схо-

димость ошибки слежения e_{ref} к заданному множеству с границей $\bar{\varepsilon}_{e_{ref}}$.

3. ВСПОМОГАТЕЛЬНЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ НЕЙРОСЕТЕВОГО УПРАВЛЕНИЯ

Для достижения поставленной цели в условиях неизвестной структуры нелинейной функции $\Lambda(z)$, а также нелинейной параметризации задачи оптимизации (2.12) в данной работе для формирования компенсационного управления v_{ad} предлагается использовать нейронные сети, обладающие универсальной аппроксимационной способностью [26]. В этом разделе будут приведены основные сведения из теории нейросетевого управления, которые будут использованы в дальнейшем.

Утверждение 1 [26]. Любая непрерывная функция $f(z): R^n \rightarrow R$ может быть равномерно аппроксимирована в компактной области $D \subset R^n$ с помощью нейронной сети с одним скрытым слоем с сигмоидальной активационной функцией, то есть для всех $\bar{\varepsilon}_{NN} > 0$ и $z \in D$ существуют V , W и b^1 , b^2 такие, что

$$\|f(z) - f_{NN}(z)\|_{\infty} = \|f(z) - W^T \sigma(V^T \bar{z})\|_{\infty} \leq \bar{\varepsilon}_{NN},$$

$$\bar{z} = [b^1 \quad z]^T; \sigma(V \cdot \bar{z}) = [b^2 \quad \sigma_1 \quad \sigma_2 \quad \dots \quad \sigma_{N_2}]^T;$$

$$V = \begin{bmatrix} \theta_1^V & \dots & \theta_{N_2}^V \\ v_{1,1} & \dots & v_{1,N_2} \\ \vdots & \ddots & \vdots \\ v_{N_1,1} & \dots & v_{N_1,N_2} \end{bmatrix}^T; W = \begin{bmatrix} \theta_1^W & \dots & \theta_{N_3}^W \\ w_{1,1} & \dots & w_{1,N_3} \\ \vdots & \ddots & \vdots \\ w_{N_2,1} & \dots & w_{N_2,N_3} \end{bmatrix}^T,$$

где $V \in R^{N_1 \times N_2}$ – матрица весов скрытого слоя; $W \in R^{N_2 \times N_3}$ – матрица весов выходного слоя; b^1 – смещение скрытого слоя; b^2 – смещение выходного слоя; σ – сигмоидальная функция активации скрытого слоя.

Другими словами, нейронная сеть с сигмоидальной функцией активации способна аппроксимировать в компактной области $D \in R^n$ любую непрерывную функцию аргумента z с ошибкой $\bar{\varepsilon}_{NN} = \sup_{z \in D} \|\varepsilon(z)\|$:

$$f(z) = W^T \sigma(V^T \bar{z}) + \varepsilon(z). \quad (3.1)$$

При этом ошибка $\bar{\varepsilon}_{NN}$ может быть сделана произвольно малой путем увеличения числа нейронов N_2 скрытого слоя.

Поскольку утверждение 1 устанавливает существование идеальных параметров нейронной сети,

но не определяет их значение, то уравнение (3.1) понимается как определяющее идеальный выход нейронной сети, а на его основе вводится понятие нейронной сети с параметрами, настраиваемыми в процессе обучения:

$$\hat{f} = \hat{W}^T \sigma(\hat{V}^T \bar{z}). \quad (3.2)$$

Получим ошибку между текущим (3.2) и идеальным (3.1) выходом нейронной сети:

$$e = W^T \sigma(V^T \bar{z}) - \hat{W}^T \sigma(\hat{V}^T \bar{z}) + \varepsilon(z). \quad (3.3)$$

Тогда идеальные параметры нейронной сети V и W могут быть определены путем решения задачи оптимизации функции ошибки (3.3) по настраиваемым параметрам

$$(V, W) = \arg \left[\min_{(\tilde{V}, \tilde{W})} \left\{ \sup_{z \in D} |e| \right\} \right]. \quad (3.4)$$

Допущение 2. В компактном множестве D идеальные веса нейронной сети ограничены:

$$\|V\| \leq V_M; \|W\| \leq W_M.$$

Так как задача оптимизации (3.4) нелинейно параметризована вследствие нелинейности функции активации скрытого слоя нейронной сети, то ошибку (3.3) переписывают в приближенном линейно параметризованном виде с помощью разложения в ряд Тейлора функции активации скрытого слоя.

Утверждение 2 [28]. Линейно параметризованная ошибка e_{lin} нейросетевого аппроксимации описывается выражением

$$e_{lin} = e = \tilde{W}^T \left(\sigma(\hat{V}^T \bar{z}) - \sigma'(\hat{V}^T \bar{z}) \hat{V}^T \bar{z} \right) + \hat{W}^T \sigma'(\hat{V}^T \bar{z}) \tilde{V}^T \bar{z} + \varepsilon(z) - d, \quad (3.5)$$

где $\sigma'(\hat{V}^T \bar{z}) = \text{diag}\{0 \quad \sigma'_1 \quad \sigma'_2 \quad \dots \quad \sigma'_{N_2}\}$ – производная функции активации скрытого слоя; $\tilde{V} = V - \hat{V}$ – параметрическая ошибка скрытого слоя нейронной сети; $\tilde{W} = W - \hat{W}$ – параметрическая ошибка выходного слоя нейронной сети; d – остаточный член. Разница $\varepsilon(z) - d$ в уравнении (3.5) ограничена [28, 29] в соответствии с условием

$$\|\varepsilon(z) - d\| \leq \alpha_1 \|\tilde{Z}\| + \alpha_2, \quad \alpha_1 > 0, \alpha_2 > 0, \tilde{Z} = \begin{bmatrix} \tilde{W} & 0 \\ 0 & \tilde{V} \end{bmatrix}. \quad (3.6)$$

Тогда задача (3.4) эквивалентна линейно параметризованной задаче вида

$$(V, W) = \arg \left[\min_{(\tilde{V}, \tilde{W})} \left\{ \sup_{z \in D} |e_{lin}| \right\} \right]. \quad (3.7)$$

Решением задачи оптимизации (3.7) с помощью градиентного метода получают законы настройки текущих параметров нейронной сети [28]

$$\begin{cases} \dot{\hat{W}} = \Gamma_W \left(\sigma(\hat{V}^T \bar{z}) - \sigma'(\hat{V}^T \bar{z}) \hat{V}^T \bar{z} \right) e, \\ \hat{W}(0) = 0_{N_2 \times N_3}, \\ \dot{\hat{V}} = \Gamma_V \bar{z} e \hat{W}^T \sigma'(\hat{V}^T \bar{z}), \quad \hat{V}(0) = 0_{N_1 \times N_2}. \end{cases} \quad (3.8)$$

Таким образом, в этом разделе статьи были кратко изложены основные сведения из теории нейросетевого управления, которые будут использованы в следующем разделе для построения компенсирующего нейроуправления v_{ad} .

4. ОСНОВНОЙ РЕЗУЛЬТАТ

Допуская выполнение предположения 1, прибавим и вычтем из уравнения (2.11) величину $\bar{B}v_{ad}^*$:

$$\begin{aligned} \dot{e}_{ref} &= A_{ref} e_{ref} + \Lambda(z) - \bar{B}v_{ad} \pm \bar{B}v_{ad}^* = \\ &= A_{ref} e_{ref} + \bar{B} [v_{ad}^* - v_{ad}] + \Lambda(z) - \bar{B}v_{ad}^*. \end{aligned} \quad (4.1)$$

Так как $z \in D \subset R^4$ согласно постановке задачи, а v_{ad}^* является функцией аргумента z по предположению 1, то функция v_{ad}^* , согласно утверждению 1, может быть аппроксимирована с помощью искусственной нейронной сети:

$$v_{ad}^* = W^T \sigma(V^T \bar{z}) + \varepsilon(z), \quad (4.2)$$

$$\bar{z} = \begin{bmatrix} b^1 & z \end{bmatrix}^T = \begin{bmatrix} b^1 & x_2 & x_3 & x_4 & v \end{bmatrix}^T.$$

Поэтому выберем управление v_{ad} в виде

$$v_{ad} = \hat{W}^T \sigma(\hat{V}^T \bar{z}). \quad (4.3)$$

Тогда с учетом выражений (4.3), (4.2) и (3.5) уравнение (4.1) может быть переписано в виде

$$\begin{aligned} \dot{e}_{ref} &= A_{ref} e_{ref} + \bar{B} \left[\tilde{W}^T \left(\sigma(\hat{V}^T \bar{z}) - \sigma'(\hat{V}^T \bar{z}) \hat{V}^T \bar{z} \right) + \right. \\ &\quad \left. + \hat{W}^T \sigma'(\hat{V}^T \bar{z}) \tilde{V}^T \bar{z} + \varepsilon(z) - d \right] + \Lambda(z) - \bar{B}v_{ad}^*. \end{aligned} \quad (4.4)$$

На основании законов (3.8) введем законы настройки весов скрытого и выходного слоя нейронной сети компенсатора (4.3):

$$\begin{cases} \dot{\hat{W}} = \Gamma_W \left[\left(\sigma(\hat{V}^T \bar{z}) - \sigma'(\hat{V}^T \bar{z}) \hat{V}^T \bar{z} \right) \times \right. \\ \quad \left. \times e_{ref}^T P \bar{B} - \sigma_W \hat{W} \right], \quad \hat{W}(0) = 0_{N_2 \times N_3}, \\ \dot{\hat{V}} = \Gamma_V \left[\bar{z} e_{ref}^T P \bar{B} \hat{W}^T \sigma'(\hat{V}^T \bar{z}) - \sigma_V \hat{V} \right], \\ \hat{V}(0) = 0_{N_1 \times N_2}, \end{cases} \quad (4.5)$$

где $\sigma_W > 0$, $\sigma_V > 0$ – коэффициенты сигма-модификаций [10].



Замечание 3. Вопреки устоявшемуся мнению, нейронная сеть-компенсатор (4.3) не требует какого-либо предварительного автономного обучения, а способна обучаться, начиная с нулевых параметров слоев, по формулам (4.5) непосредственно в процессе функционирования объекта управления.

На основании уравнения (4.4) введем понятие обобщенного вектора ошибок $\zeta = [e_{ref}^T \text{vec}^T(\tilde{W}) \text{vec}^T(\tilde{V})]^T$ и проанализируем его свойства в следующей теореме.

Теорема 1. Если закон компенсации v_{ad} выбран в виде (4.3), а настройка его параметров осуществляется по формулам (4.5), то обобщенная ошибка ζ равномерно предельно ограничена, а с увеличением числа нейронов скрытого слоя N_2 и уменьшением значений коэффициентов σ_w, σ_v значение установившейся ошибки слежения e_{ref} может быть уменьшено до заданного неравенствами (2.13) и (2.20).

Доказательство теоремы 1 приведено в приложении.

Таким образом, при использовании нейросетевого закона компенсации (4.3) и настройки его параметров по формулам (4.5) может быть обеспечена асимптотическая сходимость к заданной области ошибки слежения e_{ref} .

Замечание 4. Сформулированные в теореме 1 рекомендации увеличения числа нейронов скрытого слоя N_2 и уменьшения значений коэффициентов σ_v, σ_w имеют довольно простые интерпретации:

- увеличением в уравнении (4.4) числа нейронов N_2 обеспечивается выполнение неравенства $\Lambda(z) + \bar{B}[-v_{ad}^* + \varepsilon(z) - d] \leq \|\varepsilon_\Lambda + \bar{B}[\varepsilon(z) - d]\| < \Lambda_{\max}$, которое означает, что привносимая нейронной сетью в замкнутый контур неопределенность $\bar{B}[\varepsilon(z) - d]$ не увеличивает неопределенность системы после компенсации ε_Λ до исходного значения $\Lambda_{\max}$.

- выбор малых значений коэффициентов σ_v, σ_w позволяет настраиваемой нейронной сети точнее аппроксимировать идеальный сигнал компенсации $W^T \sigma(V^T \bar{z})$ и тем самым сильнее уменьшать ошибку $\tilde{W}^T (\sigma(\hat{V}^T \bar{z}) - \sigma'(\hat{V}^T \bar{z}) \hat{V}^T \bar{z}) + \hat{W}^T \times \times \sigma'(\hat{V}^T \bar{z}) \tilde{V}^T \bar{z}$. Однако уменьшение значений коэффициентов σ_w, σ_v также снижает робастность законов настройки (4.5) к некомпенсируемой неопределенности $\|\varepsilon_\Lambda + \bar{B}[\varepsilon(z) - d]\|$ (величина γ_1 в формуле (A.7) при близких к нулю значениях ко-

эффициентов σ_w, σ_v может быть отрицательной), что требует выбора значений коэффициентов σ_w, σ_v исходя из классического компромисса между качеством отслеживания идеальной траектории x_{ref} и робастностью системы к некомпенсируемой неопределенности $\|\varepsilon_\Lambda + \bar{B}[\varepsilon(z) - d]\|$.

5. ЭКСПЕРИМЕНТАЛЬНОЕ ПОДТВЕРЖДЕНИЕ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Эксперименты по применению разработанной системы управления были проведены на математической модели балансирующего робота LEGO EV3 в системе Matlab/Simulink. Значения номинальных параметров модели (1.1) приведены в таблице.

Номинальные параметры робота

Параметр	Значение	Параметр	Значение
$J_w, \text{кг} \cdot \text{м}^2$	$8,75 \cdot 10^{-6}$	$K_b, \text{В} \cdot \text{с/рад}$	0,468
$m_w, \text{кг}$	0,024	$R_m, \text{Ом}$	6,69
$R, \text{м}$	0,027	f_m	0,0022
$n, \text{ед.}$	1	f_w	0
$J_m, \text{кг} \cdot \text{м}^2$	10^{-5}	$g, \text{м/с}^2$	9,81
$L, \text{м}$	0,105	$M, \text{кг}$	0,8
$K_t, \text{Н} \cdot \text{м/А}$	0,317		

Момент инерции тела робота J был рассчитан по формуле $J = ML^2/3 = 0,0029 \text{ кг} \cdot \text{м}^2$. Матрица коэффициентов усиления LQ-регулятора K_{LQ} была получена по линеаризованной модели робота (2.3) с параметрами, приведенными в таблице, путем решения задачи оптимизации критерия (2.5) с матрицами $Q = I$ и $R = 1$:

$$K_{LQ} = [-0,7071 \quad -77,0619 \quad -1,5816 \quad -9,3949].$$

В экспериментах использовалась нейронная сеть с сигмоидальной функцией активации с четырьмя нейронами на входном слое ($N_1 = 5$), сорока нейронами на скрытом слое ($N_2 = 41$) и одним нейроном на выходном слое ($N_3 = 1$). Варьируемые параметры контура настройки параметров нейросети во всех экспериментах были такими:

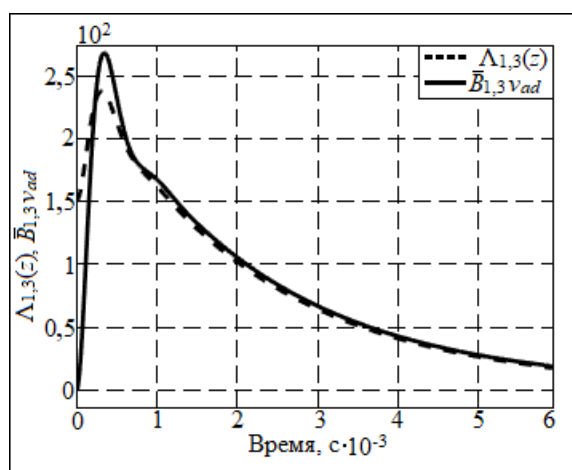
$$\Gamma_w = 10^5 I_{N_2 \times N_2}; \Gamma_v = 10^{-3} I_{N_2 \times N_2}; \sigma_w = 0,1; \sigma_v = 0,001.$$

Чтобы перейти от управления v обратно к управлению u , в экспериментах, в соответствии с допущением 1, было использовано соотношение $v = u_1 = u_2$.

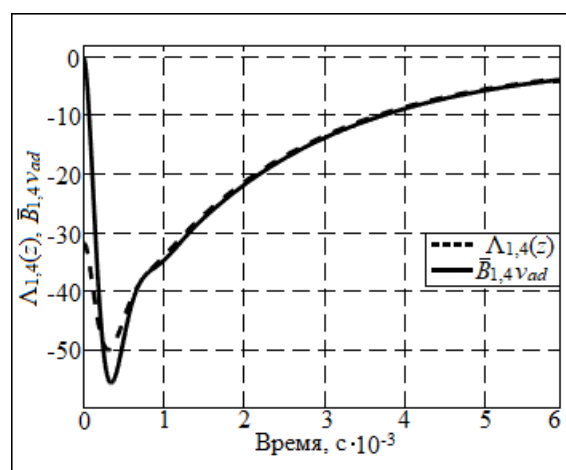
Всего в работе было проведено два эксперимента. В первом эксперименте проверялась способность компенсатора (4.3) аппроксимировать возмущения, вызванные изменением параметров робота при его работе в окрестности точки линеари-

зации $x_2 = 0$. Во втором эксперименте проверялась способность нейросетевого компенсатора (4.3) аппроксимировать возмущения, вызванные изменениями параметров робота и нелинейностями, возникающими при работе в области пространства, отличной от точки линеаризации. Начальные условия объекта управления (2.1) и эталонной модели (2.6) во всех экспериментах совпадали, а в качестве задающего воздействия r использовался нулевой вектор, что соответствует режиму стабилизации балансирующего робота.

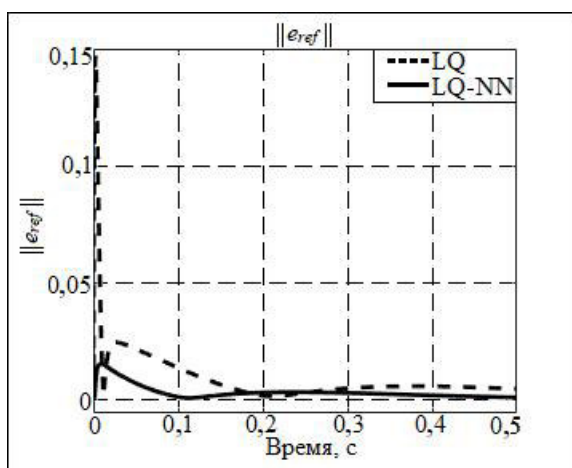
В первом эксперименте модели (2.1) и (2.6) начинали движение из точки фазового пространства с координатами $x(0) = [0 \ 0,01 \ 0 \ 0]^T$, а функция $\Lambda(z)$ была вызвана увеличением массы робота M в два раза относительно номинального значения, приведенного в таблице. На рис. 1 представлено сравнение элементов вектора v_{ad} и элементов функции $\Lambda(z)$ в этом эксперименте.



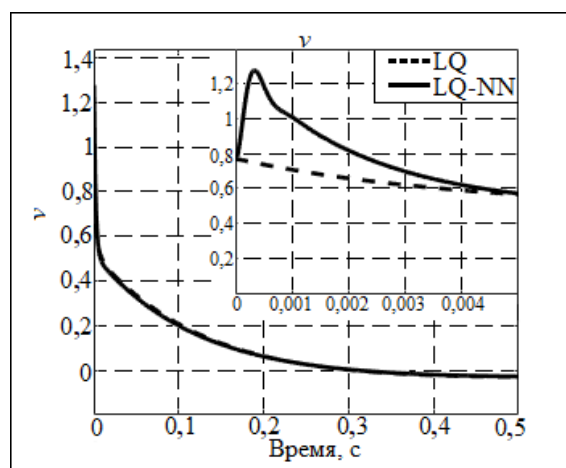
а



б

Рис. 1. Результаты нейросетевой аппроксимации: а – для $\Lambda_{1,3}(z)$, б – для $\Lambda_{1,4}(z)$


а



б

Рис. 2. Сравнение: а – норм ошибок слежения e_{ref} систем LQ-NN и LQ; б – управляющих воздействий v систем LQ-NN и LQ

Переходные процессы, приведенные на рис. 1, демонстрируют высокую точность и достаточное быстродействие аппроксимации нейросетью возмущающего воздействия $\Lambda(z)$.

На рис. 2, а и б представлены соответственно сравнения норм ошибок e_{ref} и управляющих воздействий v , полученных в первом эксперименте при применении системы управления с нейросетью (LQ-NN) и без нее (LQ).

Поскольку верхняя граница целевого множества (2.13), (2.20) задается верхней оценкой на траектории замкнутой системы с LQ-регулятором (при $v_{ad} = 0$), то из рис. 2, а следует равномерная предельная ограниченность ошибки слежения e_{ref} целевым множеством (2.20), что валидирует выводы, сделанные в теореме 1. В соответствии с рис. 2, б затраты суммарного управляющего воздействия v на компенсацию неопределенности $\Lambda(z)$ незначительны.

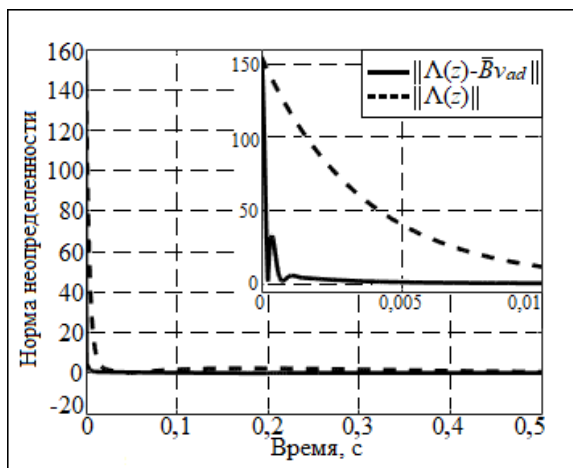
На рис. 3, а представлено сравнение неопределенности $\|\Lambda(z)\|$ в системе управления с LQ-регулятором и неопределенности после компенсации $\|\Lambda(z) - \bar{B}v_{ad}\|$ в системе управления с нейросетевым компенсатором LQ-NN. Так как в соответствии с рис. 3, а площадь под кривой неопределенности системы после компенсации $\|\Lambda(z) - \bar{B}v_{ad}\|$ в течение всего времени эксперимента меньше площади под кривой неопределенности системы без компенсации $\|\Lambda(z)\|$, то из рис. 3, а следует выполнение неравенства

$$\forall t > 0: \dot{e}_{ref} \leq A_{ref} e_{ref} + \Lambda(z) - \bar{B}v_{ad} < A_{ref} e_{ref} + \Lambda(z).$$

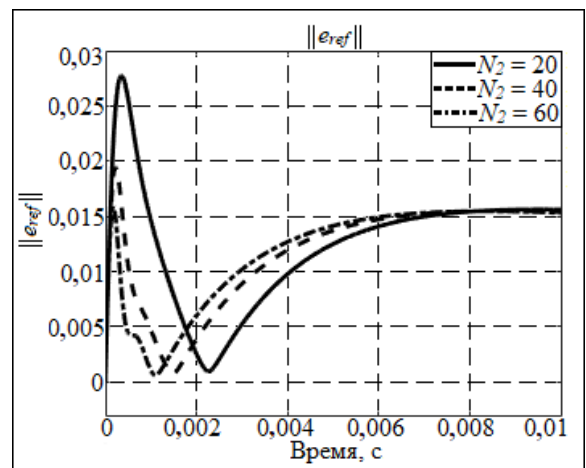
Откуда, в свою очередь, проведя аналогичные (2.15)–(2.20) рассуждения, имеем еще одно под-

тверждение теоретически доказанной в теореме 1 сходимости ошибки слежения e_{ref} в заданную неравенством (2.20) область. Рис. 3, б подтверждает возможность дальнейшего уменьшения ошибки слежения e_{ref} путем увеличения числа нейронов скрытого слоя N_2 .

Во втором эксперименте масса робота также была увеличена в два раза, но робот (2.1) начинал свое движение из начального состояния $x(0) = [0 \ 0,8 \ 0 \ 0]^T$, а поэтому неопределенность $\Lambda(z)$ была вызвана не только нестационарностью параметров робота, но и проявлением нелинейностей. На рис. 4 представлены норма ошибки $\Lambda(z) - \bar{B}v_{ad}$, а также сравнение норм ошибок e_{ref} и

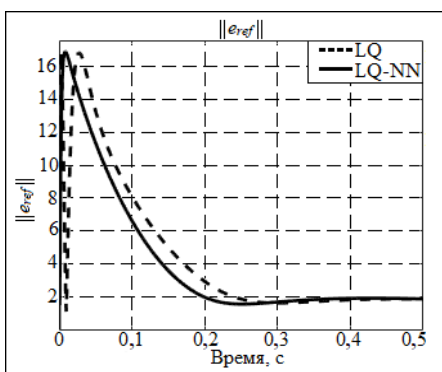


а

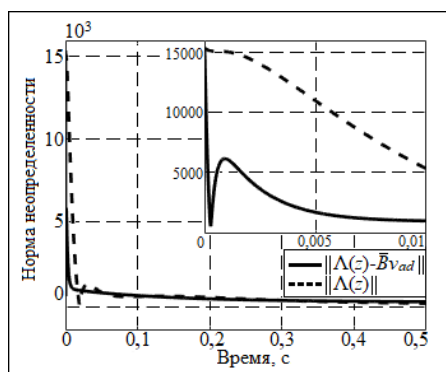


б

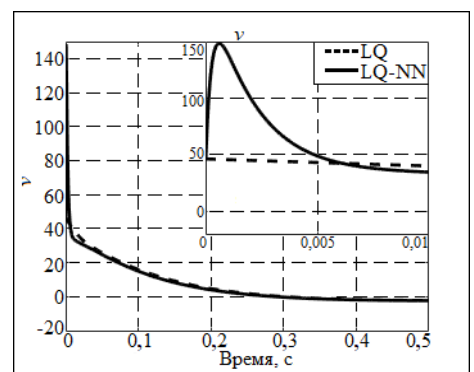
Рис. 3. Сравнение: а – неопределенностей $\|\Lambda(z)\|$ и $\|\Lambda(z) - \bar{B}v_{ad}\|$; б – норм ошибок слежения e_{ref} для различных значений N_2



а



б



в

Рис. 4. Сравнение: а – норм ошибок слежения e_{ref} для систем LQ-NN и LQ; б – неопределенностей $\|\Lambda(z)\|$ и $\|\Lambda(z) - \bar{B}v_{ad}\|$; в – управляющих воздействий v для систем LQ-NN и LQ

управляющих воздействий, полученных во втором эксперименте при использовании системы управления с нейросетью (LQ-NN) и без нее (LQ). Как следует из рис. 4, разработанная система оказывается способной компенсировать возмущение $\Lambda(z)$, вызванное не только параметрической неопределенностью, но и проявлениями нелинейностей робота. При этом, так как верхняя граница целевого множества (2.13), (2.20) задана траекториями замкнутой системы с LQ-регулятором (при $v_{ad} = 0$), то рис. 4, а подтверждает сходимость ошибки слежения e_{ref} к заданной области и для случая нелинейной неопределенности $\Lambda(z)$. Сравнивая затраты управляющего воздействия v на компенсацию линейной (см. рис. 2, б) и нелинейной неопределенности (см. рис. 4, б) можно сделать вывод о пропорциональном росте затрат управляющего воздействия v при усложнении функции $\Lambda(z)$.

ЗАКЛЮЧЕНИЕ

В работе предложена адаптивная нейросетевая система управления двухколесным балансирующим роботом, отличающаяся строгим аналитическим доказательством устойчивости замкнутого контура управления с оперативно обучаемым нейросетевым компенсатором.

Представленная процедура синтеза адаптивной нейросетевой системы управления может быть применена не только для управления двухколесным балансирующим роботом, но и для решения задач управления другими нелинейными объектами с дефицитом каналов управления (например, промышленными кранами [30, 31], манипуляторами [32], подводными аппаратами [33], самолетами с вертикальным взлетом [34, 35] и другими механическими системами [35, 36]).

ПРИЛОЖЕНИЕ

Доказательство теоремы 1.

Рассмотрим ограниченную квадратичную форму

$$V = e_{ref}^T P e_{ref} + \text{tr}(\tilde{W}^T \Gamma_W^{-1} \tilde{W}) + \text{tr}(\tilde{V}^T \Gamma_V^{-1} \tilde{V}),$$

$$\lambda_m \|\zeta\|^2 \leq V(\|\zeta\|) \leq \lambda_M \|\zeta\|^2, \quad (A.1)$$

$$\lambda_m = \min\{\lambda_{\min}(P), \lambda_{\min}(\Gamma_W^{-1}), \lambda_{\min}(\Gamma_V^{-1})\}, \quad \lambda_M = \max\{\lambda_{\max}(P), \lambda_{\max}(\Gamma_W^{-1}), \lambda_{\max}(\Gamma_V^{-1})\}.$$

Вычислим производную квадратичной формы (A.1) с учетом выражений (4.4) и (4.5):

$$\begin{aligned} \dot{V} = & -e_{ref}^T Q e_{ref} + 2e_{ref}^T P \tilde{B} \left[\tilde{W}^T (\sigma(\hat{V}^T \bar{z}) - \sigma'(\hat{V}^T \bar{z}) \hat{V}^T \bar{z}) + \tilde{W}^T \sigma'(\hat{V}^T \bar{z}) \tilde{V}^T \bar{z} + \varepsilon(z) - d \right] + 2e_{ref}^T P [\Lambda(z) - \bar{B} v_{ad}^*] - \\ & - 2\text{tr}(\tilde{W}^T \Gamma_W^{-1} \Gamma_W \left[(\sigma(\hat{V}^T \bar{z}) - \sigma'(\hat{V}^T \bar{z}) \hat{V}^T \bar{z}) e_{ref}^T P \tilde{B} - \sigma_W \hat{W} \right] - 2\text{tr}(\tilde{V}^T \Gamma_V^{-1} \Gamma_V \left[\bar{z} e_{ref}^T P \tilde{B} \tilde{W}^T \sigma'(\hat{V}^T \bar{z}) - \sigma_V \hat{V} \right]) = \\ & = -e_{ref}^T Q e_{ref} + 2e_{ref}^T P \tilde{B} [\varepsilon(z) - d] + 2e_{ref}^T P [\Lambda(z) - \bar{B} v_{ad}^*] + 2\text{tr}(\tilde{W}^T \sigma_W \hat{W}) + 2\text{tr}(\tilde{V}^T \sigma_V \hat{V}). \end{aligned} \quad (A.2)$$

С учетом выражений (2.12) и (3.6), верхняя оценка производной (A.2) имеет вид:

$$\begin{aligned} \dot{V} \leq & -\lambda_{\min}(Q) \|e_{ref}\|^2 + 2\|e_{ref}\| \lambda_{\max}(P) \|\tilde{B}\| (\alpha_1 \|\tilde{Z}\| + \alpha_2) + 2\|e_{ref}\| \lambda_{\max}(P) \varepsilon_\Lambda + 2\sigma_W \|\tilde{W}\| \|\hat{W}\| + 2\sigma_V \|\tilde{V}\| \|\hat{V}\| = \\ & = -\lambda_{\min}(Q) \|e_{ref}\|^2 + 2\|e_{ref}\| \lambda_{\max}(P) \|\tilde{B}\| (\alpha_1 \|\tilde{Z}\| + \alpha_2 + \alpha_3) + 2\sigma_W \|\tilde{W}\| \|\hat{W}\| + 2\sigma_V \|\tilde{V}\| \|\hat{V}\|, \end{aligned} \quad (A.3)$$

$$\text{где } \alpha_3 = \frac{\varepsilon_\Lambda}{\|\tilde{B}\|}.$$

Для слагаемых в выражении (A.3), в соответствии с неравенством (2.18), будут справедливы оценки сверху:

$$\begin{aligned} & -\lambda_{\min}(Q) \|e_{ref}\|^2 + 2\|e_{ref}\| \lambda_{\max}(P) \|\tilde{B}\| (\alpha_1 \|\tilde{Z}\| + \alpha_2 + \alpha_3) \leq \\ & \leq \frac{1}{2} \left[-\lambda_{\min}(Q) \|e_{ref}\|^2 - \left(\sqrt{\lambda_{\min}(Q)} \|e_{ref}\| - \frac{2\lambda_{\max}(P) \|\tilde{B}\|}{\sqrt{\lambda_{\min}(Q)}} (\alpha_1 \|\tilde{Z}\| + \alpha_2 + \alpha_3) \right)^2 \right] + \\ & + \frac{4\lambda_{\max}^2(P) \|\tilde{B}\|^2 (\alpha_1 \|\tilde{Z}\| + \alpha_2 + \alpha_3)^2}{\lambda_{\min}(Q)} \leq \frac{-\lambda_{\min}(Q) \|e_{ref}\|^2}{2} + \frac{2\lambda_{\max}^2(P) \|\tilde{B}\|^2}{\lambda_{\min}(Q)} (\alpha_1 \|\tilde{Z}\| + \alpha_2 + \alpha_3)^2, \end{aligned} \quad (A.4)$$

$$\|\tilde{W}\| \|\hat{W}\| = -\|\tilde{W}\|^2 + \|\tilde{W}\| \|\hat{W}\| \leq -\frac{1}{2} \|\tilde{W}\|^2 + \frac{1}{2} \|\hat{W}\|^2, \quad \|\tilde{V}\| \|\hat{V}\| = -\|\tilde{V}\|^2 + \|\tilde{V}\| \|\hat{V}\| \leq -\frac{1}{2} \|\tilde{V}\|^2 + \frac{1}{2} \|\hat{V}\|^2.$$



С учетом выражения (А.4), верхняя граница (А.3) приобретает вид:

$$\dot{V} \leq \frac{-\lambda_{\min}(Q)\|e_{ref}\|^2}{2} - \sigma_W \|\tilde{W}\|^2 - \sigma_V \|\tilde{V}\|^2 + \frac{2\lambda_{\max}^2(P)\|\bar{B}\|^2}{\lambda_{\min}(Q)} (\alpha_1 \|\tilde{Z}\| + \alpha_2 + \alpha_3)^2 + \sigma_W W_M^2 + \sigma_V V_M^2. \quad (A.5)$$

Так как верно

$$\sigma_W \|\tilde{W}\|^2 + \sigma_V \|\tilde{V}\|^2 \geq \underbrace{\min\{\sigma_W, \sigma_V\}}_{\sigma_Z} \|\tilde{Z}\|^2,$$

то выражение (А.5) приобретает вид:

$$\begin{aligned} \dot{V} &\leq \frac{-\lambda_{\min}(Q)\|e_{ref}\|^2}{2} - \sigma_Z \|\tilde{Z}\|^2 + \frac{2\lambda_{\max}^2(P)\|\bar{B}\|^2}{\lambda_{\min}(Q)} \left[\alpha_1^2 \|\tilde{Z}\|^2 + 2\alpha_1(\alpha_2 + \alpha_3)\|\tilde{Z}\| + (\alpha_2 + \alpha_3)^2 \right] + \sigma_W W_M^2 + \sigma_V V_M^2 \leq \\ &\leq \frac{-\lambda_{\min}(Q)\|e_{ref}\|^2}{2} - \left(\sigma_Z - \frac{2\lambda_{\max}^2(P)\|\bar{B}\|^2}{\lambda_{\min}(Q)} \alpha_1^2 \right) \|\tilde{Z}\|^2 + \frac{2\lambda_{\max}^2(P)\|\bar{B}\|^2}{\lambda_{\min}(Q)} (2\alpha_1(\alpha_2 + \alpha_3)\|\tilde{Z}\| + (\alpha_2 + \alpha_3)^2) + \sigma_W W_M^2 + \sigma_V V_M^2. \end{aligned} \quad (A.6)$$

Аналогично выражению (А.4), дополнив в формуле (А.6) слагаемые, содержащие $\|\tilde{Z}\|$, до полного квадрата, имеем:

$$\begin{aligned} \dot{V} &\leq \frac{-\lambda_{\min}(Q)\|e_{ref}\|^2}{2} - \underbrace{\frac{1}{2} \left(\sigma_Z - \frac{2\lambda_{\max}^2(P)\|\bar{B}\|^2}{\lambda_{\min}(Q)} \alpha_1^2 \right)}_{\gamma_1} \|\tilde{Z}\|^2 + \\ &+ \underbrace{\frac{2\lambda_{\max}^4(P)\|\bar{B}\|^4}{\lambda_{\min}^2(Q)} \gamma_1^{-1} (2\alpha_1(\alpha_2 + \alpha_3))^2 + \frac{2\lambda_{\max}^2(P)\|\bar{B}\|^2}{\lambda_{\min}(Q)} (\alpha_2 + \alpha_3)^2 + \sigma_W W_M^2 + \sigma_V V_M^2}_{\gamma_2} = \\ &= -\frac{1}{2} \left(\lambda_{\min}(Q)\|e_{ref}\|^2 + \gamma_1 \|\tilde{W}\|^2 + \gamma_1 \|\tilde{V}\|^2 \right) + \gamma_2 \leq -\kappa V + \gamma_2, \quad \kappa = \frac{1}{2\lambda_M} \min\{\lambda_{\min}(Q), \gamma_1\}. \end{aligned} \quad (A.7)$$

Здесь γ_1 и γ_2 – специально введенные обозначения для компактной записи (А.7). Выражения для них определены в рамках формулы (А.7).

Применив лемму о сравнении [21], получим решение неравенства (А.7):

$$\|\zeta(t)\| \leq \sqrt{\frac{\lambda_M}{\lambda_m}} e^{-\kappa t} \|\zeta(0)\| + \frac{\gamma_2}{\kappa \lambda_m}. \quad (A.8)$$

Откуда следует равномерная предельная ограниченность [3, 10, 21] ошибки ζ .

Для доказательства асимптотической сходимости ошибки слежения e_{ref} в заданную область из неравенства (А.8), с учетом $t \rightarrow \infty$ и определения величины γ_2 , получим предельную оценку ошибки слежения:

$$\begin{aligned} \|e_{ref}\| &\leq \sqrt{\frac{2\lambda_{\max}(P)}{\lambda_{\min}(Q)\lambda_{\min}(P)}} \gamma_2 \leq 2\varepsilon_\Lambda \frac{\lambda_{\max}(P)}{\lambda_{\min}(Q)} \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} + \\ &+ \underbrace{\sqrt{\frac{2\lambda_{\max}(P)}{\lambda_{\min}(Q)\lambda_{\min}(P)}} \left[\frac{\sqrt{2\lambda_{\max}^2(P)\|\bar{B}\|^2}}{\lambda_{\min}(Q)} \gamma_1^{-\frac{1}{2}} (2\alpha_1(\alpha_2 + \alpha_3)) + \frac{\sqrt{2}\lambda_{\max}(P)\|\bar{B}\|}{\lambda_{\min}^{\frac{1}{2}}(Q)} \alpha_2 + \sqrt{\sigma_W} W_M + \sqrt{\sigma_V} V_M \right]}_{\alpha_4}. \end{aligned} \quad (A.9)$$

Подставив полученное выражение в неравенство (2.20), проверим выполнение неравенства

$$2\varepsilon_\Lambda \frac{\lambda_{\max}(P)}{\lambda_{\min}(Q)} \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} \leq 2\varepsilon_\Lambda \frac{\lambda_{\max}(P)}{\lambda_{\min}(Q)} \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} + \alpha_4 < 2\|\Lambda(z)\| \frac{\lambda_{\max}(P)}{\lambda_{\min}(Q)} \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}. \quad (A.10)$$

Неравенство (А.10) будет справедливо в том и только том случае если величина α_4 достаточно мала:

$$0 \leq \alpha_4 < 2\left[\|\Lambda(z)\| - \varepsilon_\Lambda\right] \frac{\lambda_{\max}(P)}{\lambda_{\min}(Q)} \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}}.$$

По определению (А.9) значение коэффициента α_4 зависит от значений коэффициентов σ_V, σ_W и α_1, α_2 . В свою очередь, коэффициенты α_1, α_2 , в соответствии с формулами (3.1) и (3.6), обратно пропорционально зависят от числа нейронов скрытого слоя N_2 . Поэтому уменьшить значение коэффициента α_4 и тем самым обеспечить выполнение неравенства (А.10), а значит, и асимптотическую сходимость ошибки слежения e_{ref} в заданную область (2.13), возможно, увеличивая число нейронов скрытого слоя N_2 и уменьшая значения коэффициентов σ_W, σ_V , что и требовалось доказать.

ЛИТЕРАТУРА

1. *Spong, M.W., Hutchinson, S., Vidyasagar, M.* Robot Modeling and Control. – N.-Y.: John Wiley & Sons, 2020. – 419 p.
2. *Zhang, D., Wei, B.* A Review on Model Reference Adaptive Control of Robotic Manipulators // *Annual Reviews in Control.* – 2017. – Vol. 43. – P. 188–198.
3. *Nguyen, N.T.* Model-Reference Adaptive Control. – Cham: Springer, 2018. – 448 p.
4. *Wise, K.A., Lavretsky, E., Hovakimyan, N.* Adaptive Control of Flight: Theory, Applications, and Open Problems // *American Control Conference.* – IEEE, 2006. – P. 5966–5971.
5. *Korzonek, M., Tarchala, G., Orlowska-Kowalska T.* A Review on MRAS-Type Speed Estimators for Reliable and Efficient Induction Motor Drives // *ISA Transactions.* – 2019. – Vol. 93. – P. 1–13.
6. *Cartes, D., Wu, L.* Experimental Evaluation of Adaptive Three-Tank Level Control // *ISA Transactions.* – 2005. – Vol. 44, no. 2. – P. 283–293.
7. *Lim, Y., Ulsoy, A.G., Venugopal, R.* Process Control for Sheet-Metal Stamping. – New York: Springer, 2013. – 146 p.
8. *Еременко Ю.И., Полеценко Д.А., Глущенко А.И., Солодов С.В.* Исследование эффективности применения нейросетевого оптимизатора параметров ПИД-регулятора при решении задач управления нагревательными объектами в металлургии // *Известия Высших Учебных Заведений. Черная Металлургия.* – 2015. – Т. 57, № 7. – С. 61–65. [*Eremenko, Yu.I., Polechenko, D.A., Glushchenko, A.I., Solodov, S.V.* The Appliance Efficiency Estimation of PID-Regulator Parameters of Neural Optimizer for Solving of Control Problem of Metallurgical Heating Plants // *Izvestiya vuzov. Chernaya metallurgiya.* – Vol. 57, no. 7. – P. 61–65. (In Russian)]
9. *Zhou, K., Doyle, J. C.* Essentials of Robust Control. – Upper Saddle River, NJ: Prentice hall, 1998. – 430 p.
10. *Narendra, K.S., Annaswamy, A.M.* Stable Adaptive Systems. – North Chelmsford, MA: Courier Corporation, 2012. – 509 p.
11. *Lavretsky, E.* Combined/Composite Model Reference Adaptive Control // *IEEE Transactions on Automatic Control.* – 2009. – Vol. 54, no. 11. – P. 2692–2697.
12. *Chowdhary, G., Yucelen, T., Muhlegg, M., Johnson, E.N.* Concurrent Learning Adaptive Control of Linear Systems with Exponentially Convergent Bounds // *International Journal of Adaptive Control and Signal Processing.* – 2013. – Vol. 27, no. 4. – P. 280–301.
13. *Cao, C., Hovakimyan, N.* Design and Analysis of a Novel L1 Adaptive Control Architecture with Guaranteed Transient Performance // *IEEE Transactions on Automatic Control.* – 2008. – Vol. 53, no. 2. – P. 586–591.
14. *Aranovskiy, S., Bobtsov, A., Ortega, R., Pyrkin, A.* Performance Enhancement of Parameter Estimators via Dynamic Regressor Extension and Mixing // *IEEE Transactions on Automatic Control.* – 2016. – Vol. 62, no. 7. – P. 3546–3550.
15. *Gerasimov, D.N., Ortega, R., Nikiforov, V.O.* Relaxing the High-Frequency Gain Sign Assumption in Direct Model Reference Adaptive Control // *European Journal of Control.* – 2018. – Vol. 43. – P. 12–19.
16. *Kanellakopoulos, I., Kokotovic, P.V., Morse, A.S.* Systematic Design of Adaptive Controllers for Feedback Linearizable Systems // *American Control Conference.* – IEEE, 1991. – P. 649–654.
17. *Nikiforov, V.O., Voronov, K.V.* Adaptive Backstepping with a High-Order Tuner // *Automatica.* – 2001. – Vol. 37, no. 12. – P. 1953–1960.
18. *Quindlen, J.F., Chowdhary, G., How, J.P.* Hybrid Model Reference Adaptive Control for Unmatched Uncertainties // *American Control Conference (ACC).* – IEEE, 2015. – P. 1125–1130.
19. *Yayla, M., Kutay, A.T.* Adaptive Control Algorithm for Linear Systems with Matched and Unmatched Uncertainties // *55th Conference on Decision and Control (CDC).* – IEEE, 2016. – P. 2975–2980.
20. *Joshi, G., Chowdhary, G.* Hybrid Direct-Indirect Adaptive Control of Nonlinear System with Unmatched Uncertainty // *6th International Conference on Control, Decision and Information Technologies (CoDIT).* – IEEE, 2019. – P. 127–132.
21. *Khalil, H.K.* Nonlinear Systems, 3rd Edition. – Upper Saddle River: Prentice Hall, 2002. – 766 p.
22. *Olfati-Saber, R., Megretski, A.* Controller Design for a Class of Underactuated Nonlinear Systems // *Proceedings of the 37th IEEE Conference on Decision and Control.* – IEEE, 1998. – Vol. 4. – P. 4182–4187.
23. *Olfati-Saber, R.* Normal Forms for Underactuated Mechanical Systems with Symmetry // *IEEE Transactions on Automatic Control.* – 2002. – Vol. 47, no. 2. – P. 305–308.
24. *Hauser, J., Sastry, S., Kokotovic, P.* Nonlinear Control via Approximate Input-Output Linearization: The Ball and Beam Example // *IEEE Transactions on Automatic Control.* – 1992. – Vol. 37, no. 3. – P. 392–398.
25. *Spong, M.W.* Partial Feedback Linearization of Underactuated Mechanical Systems // *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'94).* – IEEE, 1994. – Vol. 1. – P. 314–321.
26. *Funahashi, K.I.* On the Approximate Realization of Continuous Mappings by Neural Networks // *Neural Networks.* – 1989. – Vol. 2, no. 3. – P. 183–192.
27. *Yamamoto, Y.* NXTway-GS Model-Based Design-Control of Self-Balancing Two-Wheeled Robot Built with Lego Mindstorms nxt // *Cybernet Systems Co., Ltd.* – 2008. – 65 p.
28. *Lewis, F.L., Yesildirek, A., Liu, K.* Multilayer Neural-Net Robot Controller with Guaranteed Tracking Performance // *IEEE Transactions on Neural Networks.* – 1996. – Vol. 7, no. 2. – P. 388–399.
29. *Hovakimyan, N., Nardi, F., Calise, A., Kim, N.* Adaptive Output Feedback Control of Uncertain Nonlinear Systems Using Single-Hidden-Layer Neural Networks // *IEEE Transactions on Neural Networks.* – 2002. – Vol. 13, no. 6. – P. 1420–1431.
30. *Zhao, Y., Gao, H.* Fuzzy-Model-Based Control of an Overhead Crane with Input Delay and Actuator Saturation // *IEEE Transactions on Fuzzy Systems.* – 2011. – Vol. 20, no. 1. – P. 181–186.
31. *He, W., Ge, S.* Cooperative Control of a Nonuniform Gantry Crane with Constrained Tension // *Automatica.* – 2016. – Vol. 66. – P. 146–154.



32. De Luca, A., Iannitti, S., Mattone, R., Oriolo, G. Control Problems in Underactuated Manipulators // IEEE International Conference on Advanced Intelligent Mechatronics. – 2001. – Vol. 2. – P. 855–861.
33. Wang, J., Wang, C., Wei, Y., Zhang, C. Command Filter Based Adaptive Neural Trajectory Tracking Control of an Underactuated Underwater Vehicle in Three-Dimensional Space // Ocean Engineering. – 2019. – Vol. 180. – P. 175–186.
34. Hauser, J., Sastry, S., Meyer, G. Nonlinear Control Design for Slightly Non-minimum Phase Systems: Application to V/STOL Aircraft // Automatica. – 1992. – Vol. 28, no. 4. – P. 665–679.
35. Liu, Y., Yu, H. A Survey of Underactuated Mechanical Systems // IET Control Theory & Applications. – 2013. – Vol. 7, no. 7. – P. 921–935.
36. Spong, M. Underactuated Mechanical Systems // Control Problems in Robotics and Automation. – 1998. – P. 135–150.

Статья представлена к публикации членом редколлегии
С.А. Красновой.

Поступила в редакцию 15.03.2021,
после доработки 17.08.2021.
Принята к публикации 24.08.2021.

Глушенко Антон Игоревич – д-р техн. наук, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва,
✉ strondutt@mail.ru

Петров Владислав Анатольевич – канд. техн. наук, Старооскольский технологический институт им. А.А. Угарова (филиал) НИТУ «МИСиС», ✉ petrov.va@misis.ru

Ласточкин Константин Андреевич – инженер, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва,
✉ lastconst@yandex.ru

ADAPTIVE NEURAL-NETWORK-BASED CONTROL OF NONLINEAR UNDERACTUATED PLANTS: AN EXAMPLE OF A TWO-WHEELED BALANCING ROBOT

A.I. Glushchenko¹, V.A. Petrov², and K.A. Lastochkin¹

¹Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

²Stary Oskol Technological Institute, National University of Science and Technology MISIS, Stary Oskol, Russia

✉ strondutt@mail.ru, ✉ petrov.va@misis.ru, ✉ lastconst@yandex.ru

Abstract. This paper proposes a new method to control nonlinear underactuated plants for eliminating unmatched parametric uncertainties. The method is based on a model reference adaptive control. The controller consists of a basic LQ one and an adaptive compensator reducing the uncertainty norm under certain assumptions. The compensator involves a multilayer neural network due to its universal approximation properties. The network is trained online. The equations to tune the compensator's neural network parameters are derived using Lyapunov's second method and the backpropagation algorithm. The asymptotic convergence of the tracking error (the difference between the plant's and reference model's outputs) to a given domain is proved. The theoretical results are validated by numerical experiments with the developed control system for the mathematical model of a balancing LEGO EV3 robot in MATLAB.

Keywords: model reference adaptive control, balancing robot, suppression of unmatched parametric uncertainties, neural networks, online training, stability.

Funding. This work was partially supported by the Russian Foundation for Basic Research, project no. 18-47-310003-r_a.

УПРАВЛЕНИЕ РЕАКТОРОМ С НЕПРЕРЫВНЫМ ПЕРЕМЕШИВАНИЕМ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ И С УЧЕТОМ ОГРАНИЧЕНИЙ НА ФАЗОВЫЕ ПЕРЕМЕННЫЕ И УПРАВЛЕНИЯ¹

С.И. Гулюкина, В.А. Уткин

Аннотация. Рассматривается задача синтеза управления, обеспечивающего поддержание температуры в рубашке резервуарного реактора с непрерывным перемешиванием. Стандартная математическая модель, описывающая работу реактора, расширена с помощью введения динамики исполнительного устройства. Проблема учета ограничений на компоненты вектора состояния и управление решается путем нелинейной замены переменных исходной модели объекта управления с использованием линейных функций с насыщением. В преобразованной системе синтез обратной связи «автоматически» учитывает заданные ограничения. Применение при синтезе обратной связи блочного подхода позволило осуществить линеаризацию по обратной связи в результате последовательного решения подзадач синтеза первого порядка. В условиях неполной информации о векторе состояния и при воздействии внешних возмущений синтезирован наблюдатель состояний и возмущений, позволяющий получить оценки неизвестных сигналов с заданной точностью, что обусловлено наличием внешних возмущений. Предложенные алгоритмы показали высокую эффективность при моделировании в системе Matlab-Simulink.

Ключевые слова: резервуарный реактор с непрерывным перемешиванием (РРНП), задача слежения, блочный подход, наблюдатель состояний и возмущений, ограничения на фазовые переменные и управление.

ВВЕДЕНИЕ

Резервуарный реактор с непрерывным перемешиванием (РРНП, англ. *continuous stirred tank reactor*, CSTR) широко используется в химической промышленности. Динамика реактора с непрерывным перемешиванием обычно описывается двумя нелинейными дифференциальными уравнениями первого порядка [1,2], которые часто используются в качестве эталонной модели для применения и тестирования новых алгоритмов управления.

Совершенствование управления реактором с непрерывным перемешиванием является в настоящее время актуальной проблемой, которая привлекает внимание многих специалистов по теории и практике управления [3–8].

Большое количество публикаций на эту тему основывается на использовании техники скользящих режимов [9–16], поскольку последняя позволяет обеспечить робастные свойства замкнутых систем и инвариантность к внешним возмущениям, действующим в каналах управления. Отметим, что часто в работах с использованием скользящих режимов в качестве управления принимаются переменные объекта управления, которые физически не могут быть разрывными функциями, например, скорость потока охлаждающей жидкости в рубашке РРНП. Это существенно снижает практическую значимость использования техники скользящих режимов при автоматизации различных технологических процессов.

В целях получения информации о компонентах вектора состояния и возмущениях широко используются наблюдатели состояния на основе теории скользящих режимов и в рамках систем с глубокими обратными связями [17–21]. Отметим, что в присутствии возмущений решение задачи получе-

¹Работа выполнена при поддержке РФФИ (20-01-00363 А).

ния оценок как собственно компонент вектора, так и возмущений, по-видимому, возможно только в рамках указанных подходов.

Недостаточно изученной в теории управления представляется проблема учета физических ограничений на компоненты вектора состояния и управления. Например, в работах [22–24] учитываются ограничения только на управления.

В настоящей работе предложено комплексное решение задач управления РРНП, развивающее предложенный в работе [25] метод решения задач управления механическими системами с учетом ограничений. В качестве методологической основы применяется блочный подход в управлении [26], позволяющий декомпозировать задачи высокой размерности на независимо решаемые подзадачи меньшей размерности при синтезе обратной связи и наблюдателей состояния. Трактуя в качестве фиктивных управлений компоненты вектора состояний для каждой из этих подзадач, прежде всего, с этой точки зрения выполняются условия согласования возмущений (принадлежность возмущений пространству управлений). Кроме того, использование при синтезе локальных обратных связей функций с насыщением обеспечивает ограниченность координат вектора состояния и управлений.

Работа имеет следующую структуру. В § 1 кратко описывается принцип работы РРНП, формулируется математическая модель и дается постановка задачи. В § 2 синтезируется наблюдатель вектора состояния и возмущений с разрывными и непрерывными корректирующими воздействиями с большими коэффициентами. В § 3 разработаны алгоритмы синтеза обратной связи с использованием синтеза локальных обратных связей в виде непрерывных функций с насыщением и разрывного управления напряжением якоря двигателя постоянного тока (ДПТ). В § 4 работоспособность предложенных алгоритмов подтверждается с помощью моделирования в среде Matlab-Simulink.

1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОБЪЕКТА УПРАВЛЕНИЯ. ПОСТАНОВКА ЗАДАЧИ

Резервуарный реактор с непрерывным перемешиванием представляет собой ключевой элемент оборудования, необходимый для завершения химических реакций во многих отраслях химической и биохимической промышленности. В РРНП происходит сложная химическая реакция, напри-

мер, превращение опасных химических отходов (реагента) в приемлемое химическое вещество (продукт). Принципиальная схема реактора приведена на рис. 1.

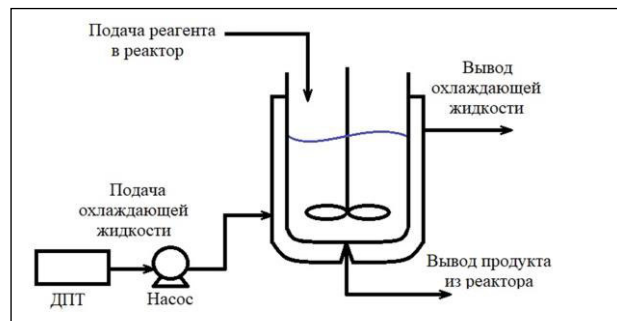


Рис. 1. Реактор РРНП (CSTR)

Предполагается, что в резервуаре происходит необратимая экзотермическая реакция первого порядка $A \rightarrow B + \Delta H$, где A – реагент; B – продукт; ΔH – тепловой эффект химической реакции (энтальпия реакции).

Предполагается, что объем резервуарного реактора с непрерывным перемешиванием равен V . На вход реактора со скоростью потока q подается реагент, имеющий концентрацию C_{Af} , температуру T_f и плотность ρ .

Температура в реакторе поддерживается на определенном уровне с помощью задания температуры жидкости в рубашке T_c путем управления скоростью потока охлаждающей жидкости. Выходящий из реактора продукт характеризуется температурой T , концентрацией C_A и, исходя из условия постоянства объема вещества в реакторе, скоростью потока q .

В предположениях о неизменности объема вещества в реакторе, идеальном перемешивании и постоянстве плотности вещества в реакторе в соответствии с законом сохранения массы и законом сохранения энергии может быть сформулирована динамическая модель РРНП [16]:

$$\begin{aligned} \dot{C}_A(t) &= \frac{q}{V}(C_{Af}(t) - C_A(t)) - k_0 C_A(t) e^{-E/RT(t)}, \\ \dot{T}(t) &= \frac{q}{V}(T_f(t) - T(t)) + \frac{(-\Delta H)}{\rho C_p} k_0 C_A(t) e^{-E/RT(t)} + \\ &\quad + \frac{UA}{\rho V C_p}(T_c - T(t)). \end{aligned} \quad (1)$$

Описание параметров модели (1) приведено в табл. 1.

Таблица 1

**Параметры модели РРНП, их значения
и единицы измерения**

Обозначение	Значение	Единицы измерения
C_A	Концентрация продукта В	Кмоль/м ³
T	Температура в реакторе и температура продукта	К
q	Скорость потока реагента	м ³ /мин
V	Объем реактора	м ³
C_{af}	Концентрация реагента	Кмоль/м ³
k_0	Постоянная скорости реакции первого порядка	мин ⁻¹
E	Энергия активации	Дж/моль
R	Универсальная газовая постоянная	Дж/(моль·К)
T_f	Температура реагента	К
ΔH	Энтальпия реакции	Кал/Кмоль
ρ	Плотность реагента	г/м ³
T_c	Температура охлаждающей жидкости	К
C_p	Удельная теплоемкость реагента	кал/(К·г)
U	Коэффициент теплопередачи	Вт/(м ² ·К)
A	Площадь поверхности теплоотдачи	м ²

Цель управления РРНП заключается в регулировке температуры охлаждающей жидкости T_c путем изменения скорости потока охлаждающей жидкости таким образом, чтобы температура в реакторе T соответствовала желаемым значениям.

Предполагается, что подача охлаждающей жидкости в реактор обеспечивается насосом с ДПТ. Система уравнений, описывающая работу ДПТ [27], имеет вид:

$$\begin{aligned}\dot{x}_3(t) &= a_{21}(gx_4(t) - m_L(t)), \\ \dot{x}_4(t) &= a_{32}(u_2(t) - gx_3(t) - a_{31}x_4(t)),\end{aligned}\quad (2)$$

где $x_3(t)$ – частота вращения вала двигателя; $x_4(t)$ – ток якоря; $u_2(t)$ – напряжение якоря; $g = \text{const}$ – магнитный поток; $m_L(t)$ – момент нагрузки; $a_{ij} = \text{const} > 0$ – конструктивные параметры двигателя.

Ставится задача слежения относительно выходной переменной $T(t)$ за заданной температурой охлаждающей жидкости в рубашке $T_d(t)$, являющейся функцией времени:

$$e_2(t) = T(t) - T_d(t) \rightarrow 0.$$

В предположении о «насосной» нагрузке $m_L = mx_3^2(t)$, $m = \text{const}$, и в предположении, что

температура в рубашке пропорциональна скорости потока охлаждающей жидкости $T_c(t) - T(t) = x_3(T_{c0} - T(t))$, где $T_{c0} = \text{const} > 0$ – температура охлаждающей жидкости на входе рубашки, запишем модель объекта управления (1), (2) в новых переменных:

$$\begin{aligned}\dot{x}_1(t) &= -ax_1(t) - f_1(x_2)x_1(t) + \xi_1(t), \\ \dot{e}_2(t) &= -ae_2(t) + bf_1(x_2)x_1(t) + \\ &+ \beta x_3(T_{c0} - T_d(t) - e_2) + \xi_2(t), \\ \dot{x}_3(t) &= a_{21}(gx_4(t) - mx_3^2(t)), \\ \dot{x}_4(t) &= a_{32}(u_2(t) - gx_3(t) - a_{31}x_4(t)),\end{aligned}\quad (3)$$

где $x_1(t) = C_A(t)$, $x_2(t) = T(t)$, $e_2(t) = x_2(t) - T_d(t)$, $\xi_1(t) = aC_{Af}(t)$, $\xi_2(t) = a(T_f - T_d) - \dot{T}_d$, $f_1(x_2) = k_0 e^{-\gamma/x_2(t)}$, $a = \frac{q}{V}$, $\gamma = \frac{E}{R}$, $b = \frac{\Delta H}{C_p \rho}$, $\beta = \frac{UA}{\rho VC_p}$.

Относительно модели процесса управления (3) предполагается, что для измерения доступны выходная переменная $y(t) = e_2(t)$ и ток якоря x_4 .

Возмущение $\xi_1(t)$, зависящее от концентрации входного реагента $C_{Af}(t)$, в режиме реального времени измерить проблематично, в то время как температуру поступающего реагента $T_f(t)$ измерить сравнительно просто. Будем далее полагать, что возмущение $\xi_1(t)$ недоступно для измерения, а значение сигнала $\xi_2(t)$ измеряется.

С учетом особенностей технологического процесса на переменные системы накладываются ограничения вида

$$\begin{aligned}x_i &\in [0, X_i], i = 2, 3; |x_4| \leq X_4; \\ |u_2| &\leq U, X_i, U = \text{const} > 0.\end{aligned}\quad (4)$$

Проблема учета ограничений на компоненты вектора состояния и управления в настоящее время недостаточно изучена в теории управления. В данной работе предложено учесть эти ограничения в математической модели объекта управления с помощью введения замены переменных в виде линейной функции с насыщением (см. § 3).

Цель управления заключается в поддержании температуры и концентрации выходного продукта в соответствии с заданными значениями. При этом в качестве управления в контуре поддержания температуры выходного продукта рассматривается напряжение u_2 на якоре ДПТ. В контуре поддержания концентрации выходного продукта в качестве управления рассматривается скорость подачи реагента в реактор q . Далее основное внимание уделяется синтезу контура управления



температурой в реакторе, а величина $a = \frac{q}{V}$ полагается известной. В следующем разделе решается задача получения оценок переменных вектора состояний и возмущений в системе (3).

2. СИНТЕЗ НАБЛЮДАТЕЛЯ СОСТОЯНИЙ И ВОЗМУЩЕНИЙ

Рассмотрим задачу получения оценок вектора состояния и возмущений в системе (3) по измеряемым выходным переменным e_2, x_4 в предположении, что сигналы q, ξ_2 измеряются.

Наблюдаемость системы (3) относительно выходных переменных e_2, x_4 устанавливается исходя из следующих соображений (а синтез наблюдателя осуществляется именно в указанной ниже последовательности):

- Модель ДПТ (два последних уравнения системы (3)) не зависит от переменных модели реактора (первые два уравнения системы (3)) и ее наблюдаемость относительно переменной x_4 очевидна в силу того, что она имеет блочную форму наблюдаемости [28]. В результате синтеза наблюдателя с использованием модели ДПТ имеем оценку переменной x_3 .

- Применительно к модели реактора с учетом того, что далее переменные e_2, x_3, ξ_2 доступны для измерения, ее структура совпадает с блочной формой наблюдаемости с учетом возмущений [21] и, как следствие, также является наблюдаемой.

Сформулируем результаты теории систем с глубокими обратными связями, которые используются в дальнейшем изложении.

Лемма. Пусть дана система первого порядка

$$\dot{\varepsilon}(t) = u + \xi(t), \quad (5)$$

где $\varepsilon(t), u(t), \xi(t) \in R$ – переменные состояния, управление и возмущение соответственно, $|\xi(t)| \leq E = \text{const}$, $|\dot{\xi}(t)| \leq \bar{E} = \text{const}$, $\forall t \geq 0$.

Тогда существует управление $u(t) = -\alpha \varepsilon(t)$, $\alpha = \text{const} > 0$, такое, что для любых заданных $\Delta_0, \bar{\Delta}_0, \alpha, \bar{\alpha} = \text{const} > 0$ за конечное время $t_0 > 0$ выполняются соотношения:

- 1) $|\varepsilon(t)| \leq \Delta_0$,
- 2) $|\dot{\varepsilon}(t)| < \bar{\Delta}_0$,
- 3) $|\alpha \varepsilon - \xi| < \bar{\Delta}_0$.

Доказательство.

1) Вопрос о сходимости переменной состояния системы (5) ε в заданную окрестность нуля $|\varepsilon| \leq \Delta_0$

решается путем выбора кандидата функции Ляпунова в виде $V = 0,5\varepsilon^2$. Определим коэффициент обратной связи в виде $\alpha_0 = \frac{E}{\Delta_0}$.

Тогда условие $\dot{V} = \varepsilon \dot{\varepsilon} = \varepsilon(-\alpha_0 \varepsilon + \xi) \leq |\varepsilon|(-\alpha_0 |\varepsilon| + E) < 0 \Rightarrow -\alpha_0 |\varepsilon| + E < 0$ выполняется вне области $|\varepsilon| \leq \frac{E}{\alpha_0} = \Delta_0$. При этом для случая $|\varepsilon(0)| \leq \Delta_0$ переменная не выходит за пределы заданной окрестности нуля $|\varepsilon(t)| \leq \Delta_0, \forall t > 0$, а для случая $|\varepsilon(0)| > \Delta_0$ может неограниченно приближаться извне к этой окрестности. Исходя из этих утверждений, следует, что выбор параметра, удовлетворяющего неравенству $\alpha > \alpha_0$, гарантирует сходимость переменной состояния в заданную окрестность $|\varepsilon(t)| \leq \Delta_0$ за конечное время. Действительно, в силу $\Delta = \frac{E}{\alpha} < \Delta_0 = \frac{E}{\alpha_0}$ переменная будет

оставаться в новой окрестности нуля $|\varepsilon(t)| \leq \Delta, \forall t > 0$, если $|\varepsilon(0)| \leq \Delta$, и неограниченно приближаться к новой окрестности нуля, если $|\varepsilon(0)| > \Delta$, а значит, за конечное время окажется в окрестности $\Delta_0 \leq \frac{E}{\alpha_0}$. Из оценки решения системы (5) $|\varepsilon(t)| \leq |\varepsilon(0)e^{-\alpha t}| + \left| e^{-\alpha t} \int_0^t e^{\alpha \tau} \xi(\tau) d\tau \right| \leq$

$|\varepsilon(0)|e^{-\alpha t} + \frac{E}{\alpha}(1 - e^{-\alpha t})$ получаем оценку времени сходимости в область $|\varepsilon| \leq \Delta_0, \forall t \geq t_0$, где t_0 определяется так:

$$(|\varepsilon(0)| - \Delta)e^{-\alpha t} + \Delta = \Delta_0 \Rightarrow t_0 = \frac{1}{\alpha} \ln \left(\frac{|\varepsilon(0)| - \Delta}{\Delta_0 - \Delta} \right),$$

$|\varepsilon(0)| > \Delta_0$.

2) Аналогичный результат справедлив и для производной переменной состояния применительно к системе $\dot{\varepsilon} = -\alpha \varepsilon + \dot{\xi}$, получаемой в результате дифференцирования обеих частей уравнения (5). Пусть задана окрестность $|\dot{\varepsilon}| \leq \bar{\Delta}_0 = \frac{\bar{E}}{\bar{\alpha}_0}$. Выберем параметр из неравенства $\alpha > \bar{\alpha}_0$ и обозначим $\bar{\Delta} = \frac{\bar{E}}{\alpha} < \bar{\Delta}_0 = \frac{\bar{E}}{\bar{\alpha}_0}$. Тогда за конечное время выполняется соотношение $|\dot{\varepsilon}| \leq \bar{\Delta}_0, t \geq \bar{t}_0$, где $\bar{t}_0 = \frac{1}{\bar{\Delta}_0} \ln \left(\frac{|\dot{\varepsilon}(0)| - \bar{\Delta}}{\bar{\Delta}_0 - \bar{\Delta}} \right), |\dot{\varepsilon}(0)| > \bar{\Delta}_0$.

3) Наконец, в силу равенства (5) с учетом пункта 2) справедлива оценка вида $|\alpha \varepsilon(t) - \xi(t)| \leq \bar{\Delta}_0, \forall t \geq \bar{t}_0$. Следовательно, можно получить оценку возмущения с заданной точностью: $\alpha \varepsilon(t) = \xi(t) + \delta(t), \delta(t) \leq \bar{\Delta}_0, \forall t \geq \bar{t}_0$. Отметим, что для случая постоянных возмущений си-

стема $\ddot{\varepsilon} = -\alpha \dot{\varepsilon} + \dot{\xi}$, $\dot{\xi} = 0$, имеет решение вида $\ddot{\varepsilon} = \dot{\varepsilon}(0)e^{-\alpha t}$ и, следовательно, имеем асимптотически сходящуюся оценку возмущения $\alpha \varepsilon \rightarrow \dot{\xi}$, $t \rightarrow \infty$.

Выбор значения параметра α из условия $\alpha \geq \max\{E/\Delta_0, \bar{E}/\bar{\Delta}_0\}$ гарантирует выполнение условий леммы, сходимость переменной $\varepsilon(t)$ и ее производной $\dot{\varepsilon}(t)$ в заданные области 1) $|\varepsilon(t)| \leq \Delta_0$, 2) $|\dot{\varepsilon}(t)| \leq \bar{\Delta}_0$, 3) $|\alpha \varepsilon(t) - \dot{\xi}(t)| \leq \bar{\Delta}_0$, $\forall t \geq \max\{t_0, \bar{t}_0\}$.

Лемма доказана. ♦

Для решения задачи оценивания вектора состояний системы (3) синтезируем наблюдатель вида

$$\begin{aligned}\dot{z}_1 &= -(a + f_1(x_2))z_1 + v_1, \\ \dot{z}_2 &= -az_2 + \beta z_3(T_{c0} - T_d - e_2) + \xi_2 + v_2, \\ \dot{z}_3 &= a_{21}gx_4 + v_3, \\ \dot{z}_4 &= a_{32}(u_2 - a_{31}z_4) + v_4,\end{aligned}\quad (6)$$

где v_i – корректирующие воздействия наблюдателя, определяемые далее.

Запишем систему в невязках $\varepsilon_i = x_i - z_i$, $i = 1, 3, 4$, $\varepsilon_2 = e_2 - z_2$ с учетом формул (3) и (6):

$$\begin{aligned}\dot{\varepsilon}_1 &= -(a + f_1(x_2))\varepsilon_1 + \xi_1(t) - v_1, \\ \dot{\varepsilon}_2 &= -a\varepsilon_2 + \beta(T_{c0} - T_d - e_2)\varepsilon_3 + bf_1(x_2)x_1(t) - v_2, \\ \dot{\varepsilon}_3 &= -a_{21}mx_3^2 - v_3, \\ \dot{\varepsilon}_4 &= -a_{32}a_{31}\varepsilon_4 - a_{32}gx_3 - v_4.\end{aligned}\quad (7)$$

Приведем пошаговую процедуру построения наблюдателя состояния и возмущений (7) в рамках каскадного синтеза наблюдателей [28].

1. Выбором корректирующего воздействия в последней подсистеме системы (7) в виде разрывной функции $v_4 = l_4 \text{sign}(\varepsilon_4)$ (где $l_4 = \text{const} > |a_{32}gx_3|$) обеспечивается скользящий режим по прямой $\varepsilon_4 = 0$. При этом среднее (эквивалентное) значение разрывного сигнала равно $v_{4eq} = -a_{32}gx_3$. Отметим, что на практике эквивалентное значение разрывного управления может быть получено с выхода фильтра первого порядка $\mu \dot{t} = -\tau + v_4$, $\mu > 0$, $v_{4eq} \approx \tau$ [28].

2. Используя полученное эквивалентное значение (с учетом $x_3 = -v_{4eq} / (a_{32}g)$), построим корректирующее воздействие в третьей подсистеме системы (7) в виде $v_3 = -a_{21}m \times (v_{4eq} / (a_{32}g))^2 + l_3[-v_{4eq} / (a_{32}g) - z_3]$. В итоге третья система (7) примет вид $\dot{\varepsilon}_3 = -l_3\varepsilon_3$ и выбором коэффициента $l_3 > 0$ обеспечивается асимпто-

тическое стремление переменной в нуль: $\varepsilon_3 \rightarrow 0 \Rightarrow z_3 \rightarrow x_3$.

3. В предположении $\varepsilon_3 \rightarrow 0 \Rightarrow z_3 \rightarrow x_3$ второе уравнение системы (7) принимает вид

$$\dot{\varepsilon}_2 = -a\varepsilon_2 + bf_1(x_2)x_1 - v_2.$$

Выбор корректирующего воздействия в виде $v_2 = l_2 \text{sign}(\varepsilon_2)$, $l_2 = \text{const} > bf_1(x_2)x_1$, гарантирует возникновение скользящего режима по плоскости $\varepsilon_2 = 0$. При этом среднее значение разрывного сигнала равно $v_{2eq} = bf_1(x_2)x_1$.

4. На последнем шаге положим $v_1 = (-f_1(x_2) + l_1)\left(\frac{v_{2eq}}{bf_1(x_2)} - z_1\right) = (-f_1(x_2) + l_1)\varepsilon_1$.

Тогда первое уравнение системы (7) принимает вид $\dot{\varepsilon}_1 = -(a + l_1)\varepsilon_1 + \xi_1$ и в предположении $|\xi_1(t)| \leq E_1$, $|\dot{\xi}_1(t)| \leq \bar{E}$, согласно лемме справедливы

$$\text{соотношения } |\varepsilon_1| \leq \frac{E_1}{a + l_1} = \Delta_1 \text{ и } |\dot{\varepsilon}_1| \leq \frac{\bar{E}_1}{a + l_1} = \bar{\Delta}_1.$$

Таким образом, выбором коэффициента $l_1 > 0$ обеспечивается за конечное время заданная точность стабилизации переменной $|\varepsilon_1| \leq \Delta_1$ и, как следствие, получение оценки $x_1(t) = z_1(t) + \delta_1(t)$, $|\delta_1(t)| \leq \Delta_1$, а также оценки возмущения с заданной точностью $\xi_1(t) = (a + l_1)\varepsilon_1(t) + \bar{\delta}_1(t)$, $|\bar{\delta}_1(t)| \leq \bar{\Delta}_1$. Отметим, что с ростом коэффициента усиления $l_1 \rightarrow \infty$ точность полученных оценок неограниченно возрастает: $\delta_1, \bar{\delta}_1 \rightarrow 0$.

Таким образом, решена задача получения оценок компонент вектора состояния системы (3) и возмущения $\xi_1(t)$.

В представленной выше процедуре синтеза наблюдателя вектора состояния и возмущений использовались как разрывные, так и непрерывные корректирующие воздействия. Как видно, синтез на основе скользящих режимов проще с вычислительной точки зрения, но следует помнить, что средние (эквивалентные) значения корректирующих воздействий получаются на выходе фильтров первого порядка [28], что приводит к динамическому расширению пространства состояний исходной модели. Техника линеаризации уравнений в невязках (7) с помощью непрерывных корректирующих воздействий, продемонстрированная выше на четвертом шаге процедуры синтеза наблюдателя, может быть использована по аналогии на всех других шагах. При этом принцип каскадного синтеза [28] позволяет декомпозиро-



вать процедуру синтеза наблюдателя на последовательно решаемые подзадачи размерности один с наперед заданной точностью получаемых оценок без расширения пространства состояний замкнутой системы.

Далее разрабатывается процедура синтеза управления в системе (7) при наличии полной информации (с учетом полученных в этом разделе оценок вектора состояния и возмущения).

3. СИНТЕЗ ОБРАТНОЙ СВЯЗИ

Представим решение задачи в общем виде в рамках блочного подхода в условиях полной информации о компонентах вектора состояния и возмущений, получаемой с помощью наблюдателя (6). Приведем схему пошаговой процедуры синтеза обратной связи при наличии полной информации о векторе состояния и возмущениях на основе блочного подхода [21] и с учетом ограничений на компоненты вектора состояния и управление (4).

Шаг 1. Перепишем второе уравнение системы (3) в виде

$$\dot{e}_2 = -ae_2(t) + \beta d(t)x_3 + \bar{\xi}_2(t), \quad (8)$$

где $\bar{\xi}_2 = bf_1(x_2)x_1 + \xi_2$, $d(t) = T_{c0} - T_d(t) - e_2(t) < 0$, $|d| \leq D = \text{const}$.

В системе (8) в качестве управляющего воздействия рассматривается частота вращения вала двигателя x_3 . Учитывая, что при насосной нагрузке частота вращения вала двигателя положительна, введем замену переменной

$$e_3 = \beta d(t)x_3 + M_3 \text{sat}^+(s_3), \quad (9)$$

где $s_3 = k_2 e_2 + \bar{\xi}_2$. Решение задачи стабилизации переменной $e_3 = \beta d(t)x_3 + M_3 \text{sat}^+(s_3) = 0$ позволяет учесть ограничение (4) на переменную $x_3 \in [0, X_3]$ выбором $0 < M_3 < \beta DX_3$.

Определение. Пусть $M = \text{const} > 0$, $b = \text{const}$. Тогда $M\text{sat}(s) = \min(M, |s|) \text{sign}(s)$, $M\text{sat}^+(s) = M\text{sat}(s)0,5[1 + \text{sign}(s)]$.

После подстановки суммы (9) в формулу (8) второе уравнение системы преобразуется к виду

$$\dot{e}_2 = -ae_2 + e_3 - M_3 \text{sat}^+(s_3) + \bar{\xi}_2. \quad (10)$$

При попадании переменной s_3 в линейную зону $0 < s_3 < M_3$ уравнение (10) описывается выражением

$$\dot{e}_2 = -(a + k_2)e_2 + e_3,$$

где выбор параметра $k_2 > 0$ определяет сходимость переменной данной подсистемы в заданную

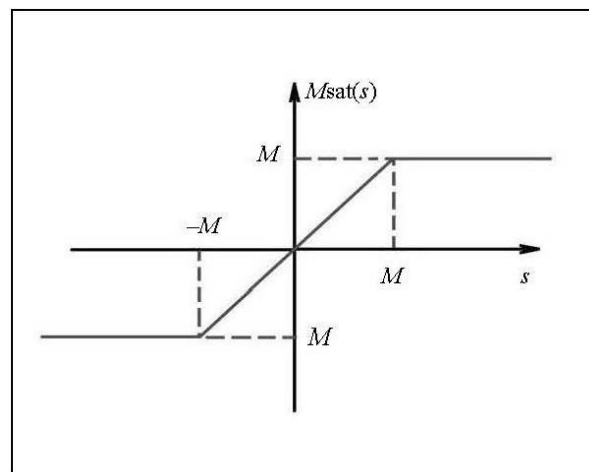


Рис. 2. К определению $M\text{sat}(s)$ -функции

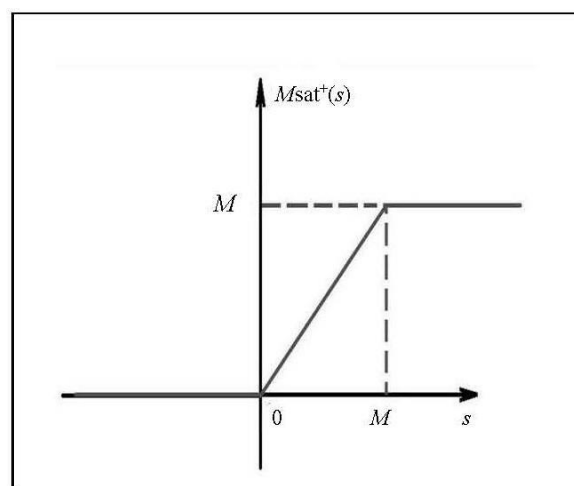


Рис. 3. К определению $M\text{sat}^+(s)$ -функции

окрестность нуля, определяемую согласно Лемме выражением $|e_2| \leq \Delta_3 / (a + k_2)$ в предположении $|e_3| \leq \Delta_3 = \text{const}$.

Условия попадания в линейную зону определяются выбором амплитуды M_3 исходя из соотношений $s_3 \dot{s}_3 < 0$ в нелинейной зоне $s_3 \notin (0, M_3)$ (см. рис. 3). Запишем уравнение относительно переменной s_3 :

$$\dot{s}_3 = -(as_3 - \bar{\xi}_2) + \dot{\bar{\xi}_2} + k_2[-M_3 \text{sat}^+(s_3) + e_3 + \bar{\xi}_2].$$

Следующий анализ позволяет определить параметры M_3 и k_2 функции $M_3 \text{sat}^+(s_3)$.

• При $s_3 \in [M_3, \infty]$ и, следовательно, $-M_3 \text{sat}^+(s_3) = -M_3$ должно выполняться неравенство $\dot{s}_3 = -a(s_3 - \bar{\xi}_2) + \dot{\bar{\xi}_2} + k_2[-M_3 + e_3 + \bar{\xi}_2] < 0$,

откуда следует неравенство на выбор амплитуды $M_3 > \frac{1}{k_2} [a\bar{\xi}_2 + \dot{\xi}_2] + e_3 + \bar{\xi}_2$.

• При $s_3 \in (-\infty, 0)$ и, следовательно, $M_3 \text{sat}^+(s_3) = 0$ должно выполняться неравенство $\dot{s}_3 = -a(s_3 - \bar{\xi}_2) + \dot{\xi}_2 + k_2[e_3 + \bar{\xi}_2] > 0$, откуда в предположении $e_3 + \bar{\xi}_2 > 0$ и $|e_3| \leq \Delta_3 = \text{const}$ следует ограничение на выбор коэффициента $k_2 > \frac{|a\bar{\xi}_2 + \dot{\xi}_2|}{e_3 + \bar{\xi}_2}$. Отметим, что требование $\bar{\xi}_2 + e_3 = bf_1(x_2)x_1 + a(T_f(t) - T_d) - \dot{T}_d + e_3 > 0$ в силу $bf_1(x_2)x_1 + a(T_f(t) - T_d) > 0$ накладывает ограничение на заданную точность стабилизации переменной $|e_3| \leq \Delta_3$ (решаемую на следующем шаге) и на скорость изменения задания окрестности $\Delta_3 < bf_1(x_2)x_1 + a(T_f(t) - T_d(t))$, $|\dot{T}_d(t)| < bf_1(x_2)x_1 + a(T_f(t) - T_d(t)) - \Delta_3$. С точки зрения физики, последнее неравенство ограничивает скорость изменения задания окрестности до уровня, когда задача слежения имеет решение.

Шаг 2. На втором шаге обеспечим попадание переменной e_3 из уравнения (10) в окрестность нуля $|e_3| \leq \Delta_3$. Согласно равенству (9), уравнение, описывающее поведение переменной e_3 , имеет вид:

$$\dot{e}_3 = \beta d(t)a_{21}gx_4 + \xi_3. \quad (11)$$

Рассматривая переменную x_4 в системе (11) в качестве фиктивного управления, положим его равным

$$e_4 = \beta d(t)a_{21}gx_4 + M_4 \text{sat}(s_4), \quad (12)$$

где $M_4 \text{sat}(s_4) = \min(M_4, |s_4|) \text{sign}(s_4)$ (см. рис. 2), $s_4 = k_3 e_3 + \xi_3$. Решение задачи стабилизации переменной $e_4 \rightarrow 0$ решает задачу учета ограничения $|x_4| \leq X_4$ с помощью выбора амплитуды из соотношения $M_4 < \beta Da_{21}gX_4$.

Уравнение (11), «замкнутое» локальной обратной связью (12), имеет вид

$$\dot{e}_3 = e_4 - M_4 \text{sat}(s_4) + \xi_3$$

и при функционировании в линейной зоне ($|s_4| < M_4 \Rightarrow M_4 \text{sat}(s_4) = s_4$) преобразуется к виду

$$\dot{e}_3 = -k_3 e_3 + e_4.$$

Выбор амплитуды M_4 , обеспечивающей попадание переменной s_4 в линейную зону, определяется на основе второго метода Ляпунова.

Запишем производную функции s_4 в виде

$$\dot{s}_4 = k_3 [-M_4 \text{sat}(s_4) + \varphi_4(\cdot)] + \dot{\xi}_3,$$

где $\dot{\xi}_3 = 2\beta da_{21}mx_3\dot{x}_3 + \frac{d^2}{dt^2} [M_3 \text{sat}^+(s_3)]$, $\varphi_4 = e_4 + \beta da_{21}mx_3^2 + \frac{d}{dt} [M_3 \text{sat}^+(s_3)]$.

Выберем кандидата на функцию Ляпунова $V = 0,5s_4^2$. Тогда требование $\dot{V} = s_4\dot{s}_4 < 0$ вне линейной зоны $|s_4| \geq M_4$ дает производную функции Ляпунова вида $\dot{V} = s_4 \{k_3 [-M_4 \text{sign}(s_4) + \varphi_4(\cdot)] + \dot{\xi}_3\} < 0$, откуда следует выражение для выбора амплитуды $M_4 > \Phi_4 + \bar{E}_3 / k_3$, где $|\dot{\xi}_3| \leq \bar{E}_3 = \text{const}$ и $|\varphi_4| \leq \Phi_4 = \text{const}$.

Шаг 3. На последнем шаге решается задача стабилизации переменной (12), описываемой уравнением вида

$$\dot{e}_4 = \beta da_{21}ga_{32}u_2 + \xi_4, \quad (13)$$

где $\xi_4 = \beta \dot{d}(t)a_{21}gx_4 - \beta d(t)a_{21}ga_{32}(gx_3 + a_{31}x_4) + \frac{d}{dt} [M_4 \text{sat}(s_4)]$.

Выберем управление в виде разрывной функции

$$u_2 = M_2 \text{sign}(e_4). \quad (14)$$

В случае выполнения условия существования скользящего режима $M_2 > \left| \frac{\xi_4}{\beta da_{21}ga_{32}} \right|$ в системе (13) за конечное время возникает скользящий режим по плоскости $e_4 = 0$.

Для наглядности выпишем уравнения движения системы (5) в скользящем режиме:

$$\begin{aligned} \dot{x}_1 &= -ax_1 - f_1(T_d)x_1 + \xi_1(t), \\ \dot{e}_2 &= -(a + k_2)e_2 + e_3, \quad \dot{e}_3 = -k_3e_3, \quad e_4 = 0. \end{aligned} \quad (15)$$

В системе (15) осуществляется последовательное асимптотическое стремление к нулю переменных $e_4 = 0 \Rightarrow e_3 \rightarrow 0 \Rightarrow e_2 \rightarrow 0$, в то время как первая подсистема представляет собой уравнение нулевой динамики. Первое уравнение системы (15), описывающее нулевую динамику, устойчиво в силу $f_1(T_d) > 0$.

Таким образом, выбор управления разрывным (см. формулу (14)) за конечное время обеспечивает скользящее движение по плоскости $e_4 = 0$, ко-



торое описывается устойчивой системой линейных дифференциальных уравнений (15).

Отметим, что если для стабилизации подсистемы (13) использовать непрерывную обратную связь

$$\beta da_{21}ga_{32}u_2 + \xi_4 = -k_4e_4 \Rightarrow u_2 = -\frac{\xi_4 + k_4e_4}{\beta da_{21}ga_{32}},$$

последнее уравнение системы (15) примет вид $\dot{e}_4 = -k_4e_4$ и устойчивость замкнутой системы также будет обеспечена. Дополнительными требованиями к реализации такого непрерывного управления будут необходимость получения информации о переменной ξ_4 и использование устройства широтно-импульсной модуляции для управления инвертором напряжения ДПТ. Отметим, что для реализации разрывного управления (14) требовалась лишь оценка сверху переменной $|\xi_4| \leq E_4 = \text{const}$.

Первая подсистема (15), рассматриваемая как уравнение нулевой динамики, согласно лемме имеет ограниченное решение $|x_1| \leq \frac{E_1}{a + f_1(T_d)}$ при $|\xi_1(t)| = aC_{Af} \leq E_1 = \text{const}$. В частности, в предположении $\xi_1 = E_1, x_2 = T_d, E_1, T_d = \text{const}$, согласно лемме имеем (с учетом $a = \frac{q}{V}$) $x_1 \rightarrow \frac{qC_{Af}}{q + Vf_1(T_d)}$.

Полагая параметр q в качестве управления в контуре поддержания концентрации выходного продукта, имеем предельные соотношения $q \rightarrow \infty \Rightarrow C_A \rightarrow C_{Af}$ и $q \rightarrow 0 \Rightarrow C_A \rightarrow 0$. Значит, задавая разумные границы изменения концентрации выходного продукта $C_A \in [C_{A1}, C_{A2}]$, можно поддерживать концентрацию выходного продукта в данном диапазоне с помощью изменения скорости его подачи в реактор.

Видно, что с ростом скорости поступления реагента в реактор концентрация выходного продукта уменьшается, а с уменьшением – увеличивается.

Рассмотрим три набора параметров, в которых значения $C_{Af} = 0,9$ и $q = 0,9$ неизменны, а желаемая температура варьируется: 1) $T_d = 350$, 2) $T_d = 380$, 3) $T_d = 400$. Согласно приведенному выше соотношению $x_1 \rightarrow \frac{qC_{Af}}{q + Vf_1(T_d)}$, для каждого из наборов параметров получим: 1) $x_1 \rightarrow 0,3251$; 2) $x_1 \rightarrow 0,3214$; 3) $x_1 \rightarrow 0,3192$.

Таким образом, с увеличением температуры в реакторе концентрация продукта уменьшается, что согласуется с результатами моделирования из § 4 (см. рис. 10).

4. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Эффективность предложенного подхода проверена посредством численного моделирования системы «реактор с непрерывным перемешиванием – двигатель постоянного тока» в системе Matlab-Simulink. При моделировании системы (3), наблюдателя (6) и управления (14) были выбраны параметры из табл. 2.

Таблица 2

Параметры модели

Группа параметров	Значения параметров
Параметры РРНП (CSTR)	$q = 0,9, V = 1, \beta = 0,003, \gamma = 80,$ $b = 5, k_0 = 2, T_{c0} = 300,$ $C_{Af} = 0,9 + 0,005 \sin(0,03\pi t),$ $T_f = 395 + 0,01 \sin(0,05\pi t).$
Параметры ДПТ	$a_{21} = 0,8, g = 0,7, m = 0,0001,$ $a_{31} = 12,5, a_{32} = 2$
Начальные условия и задание	$x_1(0) = 0,3, x_2(0) = 400,$ $x_3(0) = 100, x_4(0) = 20$
Сценарий моделирования	$T_d(0) = 350$ при $t \in [0, t_1],$ $T_d(t_1) = 380$ при $t \in (t_1, \infty], t_1 = 75$
Параметры наблюдателя	$z_i(0) = 0, i = \overline{1,4},$ $l_1 = 100, l_2 = 3, l_3 = 100, l_4 = 500$
Параметры контроллера	$M_2 = 400, M_3 = 50, M_4 = 170,$ $k_2 = 0,5, k_3 = 0,1$
Физические ограничения на фазовые переменные	$C_A \in [0,1], T \in [0,400],$ $x_3 \in [0, 100], x_4 < 150, u_2 = \pm 400$

Предполагается, что в объекте для измерения доступны температура в рубашке реактора x_2 , ток якоря x_4 , скорость поступающего реагента q и его температура T_f . Для получения полной информации о фазовых переменных объекта РРНП–ДПТ и о действующих на него внешних возмущениях был построен наблюдатель состояния (6), позволяющий получить оценки неизвестных сигналов с заданной точностью.

На рис. 4–7 представлены ошибки наблюдения $\varepsilon_i = x_i - z_i, i = \overline{1,4}$.

На рис. 8 представлен график внешнего возмущения $\xi_1 = 0,9(0,9 + 0,005\sin(0,03\pi t))$, восстановленного с помощью наблюдателя (6). Прерывистая линия соответствует реальным значениям

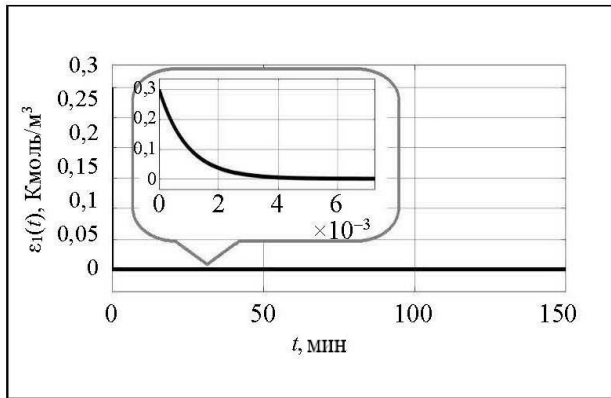


Рис. 4. Ошибка наблюдения $\varepsilon_1(t)$

возмущения, сплошная – восстановленным. На рис. 9 показан график изменения температуры в рубашке реактора.

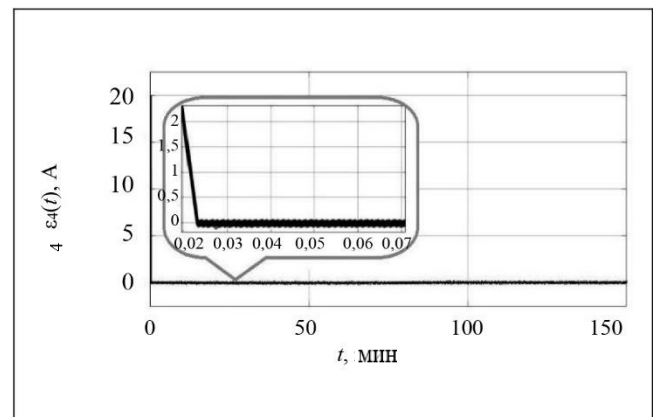


Рис. 7. Ошибка наблюдения $\varepsilon_4(t)$

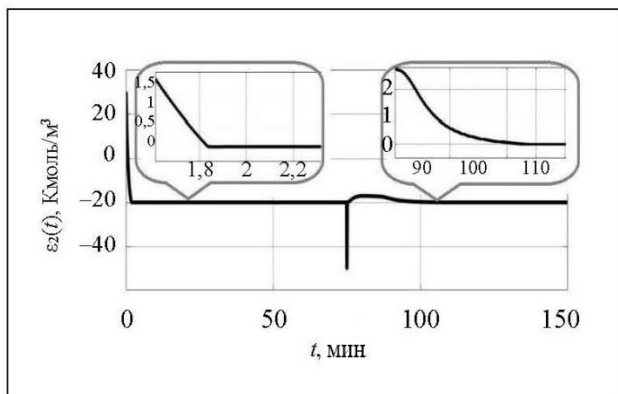


Рис. 5. Ошибка наблюдения $\varepsilon_2(t)$

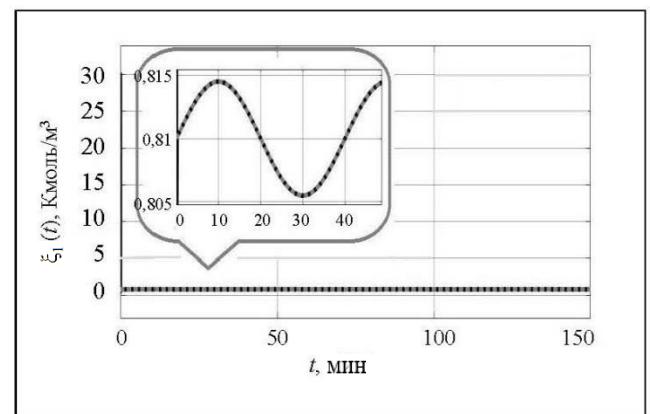


Рис. 8. Внешнее возмущение $\xi_1(t)$

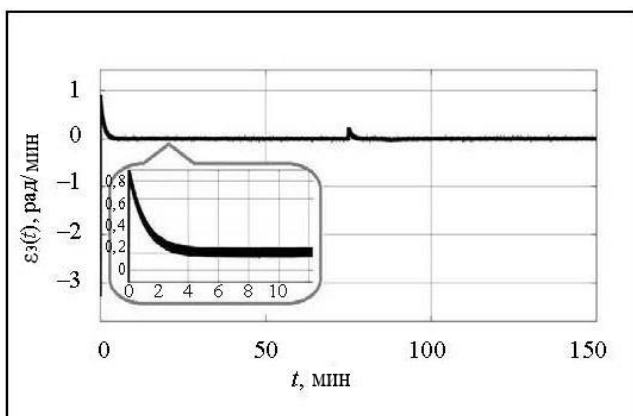


Рис. 6. Ошибка наблюдения $\varepsilon_3(t)$

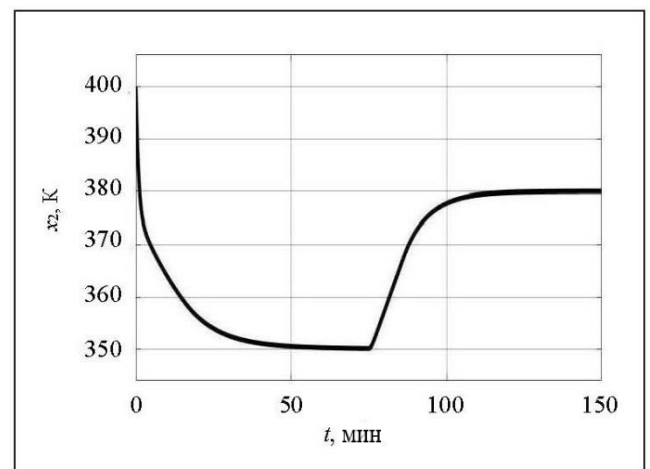


Рис. 9. Температура в рубашке $x_2(t)$

На рис. 10–13 представлены графики зависимости от времени концентрации продукта $x_1(t)$, угловой частоты вращения вала двигателя $x_3(t)$, тока якоря $x_4(t)$ и напряжения якоря $u_2(t)$ соответственно.

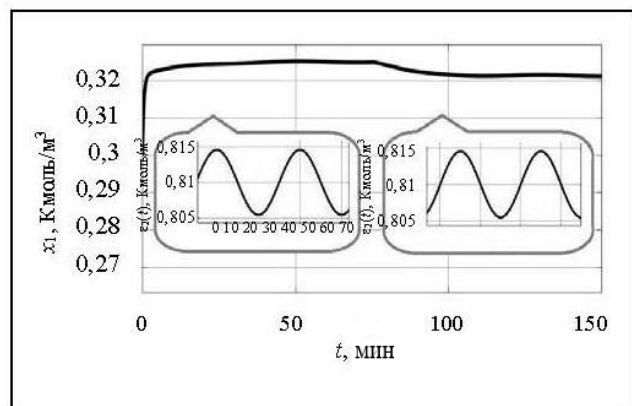


Рис. 10. Концентрация продукта $x_1(t)$

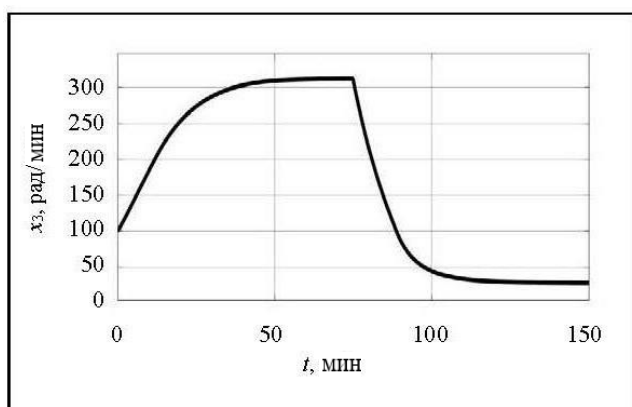


Рис. 11. Угловая частота вращения вала двигателя $x_3(t)$

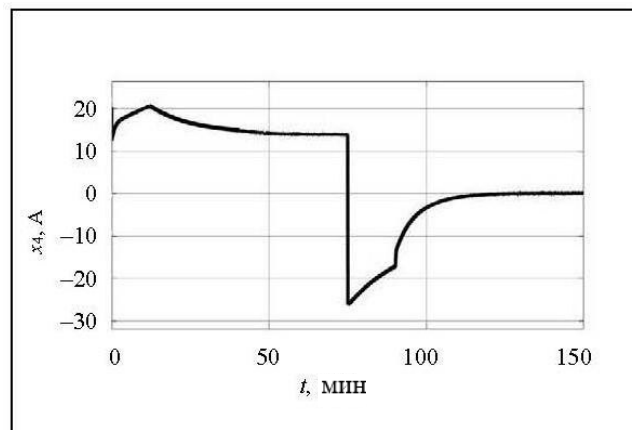


Рис. 12. Ток якоря $x_4(t)$

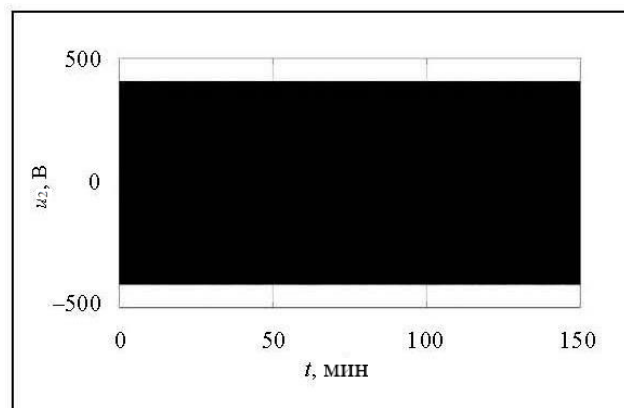


Рис. 13. Напряжение якоря $u_2(t)$

ЗАКЛЮЧЕНИЕ

Математическая модель резервуарного реактора с непрерывным перемешиванием была расширена с помощью введения динамики исполнительного устройства – двигателя постоянного тока, что вполне обоснованно позволило применять теорию скользящих режимов путем переключения ключей инвертора напряжения.

Для получения информации о неизмеряемых компонентах вектора состояния и возмущениях разработан наблюдатель со смешанными корректирующими воздействиями (разрывными и непрерывными).

В предложенной схеме синтеза обратной связи учитывается тот факт, что как в реальном скользящем режиме, так и при использовании глубоких обратных связей оценивание реальных сигналов с помощью наблюдателя осуществляется лишь с заданной точностью. Ключевой особенностью данной работы является синтез обратной связи с использованием методологии блочного подхода, позволяющий учесть ограничения на фазовые переменные и управление применительно к реактору с непрерывным перемешиванием.

Работоспособность предложенного алгоритма подтверждена как аналитически, так и с помощью моделирования в системе Matlab-Simulink.

ЛИТЕРАТУРА

1. Saravanathamizhan, R., Paranthaman, R., and Balasubramanian, N. Tanks in Series Model for Continuous Stirred Tank Electrochemical Reactor // *Industrial & Engineering Chemistry Research*. – 2018. – No. 47(9). – P. 2976–2984.
2. Flores-Tlacuahuac, A., and Grossmann, I.E. Simultaneous Cyclic Scheduling and Control of a Multiproduct CSTR // *Industrial & Engineering Chemistry Research*. – 2006. – No. 45(20). – P. 6698–6712.

3. Li, S., Sun, H., Yang, J., and Yu, X. Continuous Finite-Time Output Regulation for Disturbed Systems under Mismatching Condition // IEEE Trans. Autom. Control. – 2015. – No. 60(1). – P. 277–282.
4. Lagos, B., and Cipriano, A. Performance Evaluation of a Distributed MPC Strategy Applied to the Continuous Stirred Tank Reactor // IEEE Latin America Transactions – 2015. – No. 13(6). – P. 1921–1926.
5. Delbari, M., Salahshoor, K., Moshiri, B. Adaptive Generalized Predictive Control and Model Reference Adaptive Control for CSTR Reactor // 2010 Intelligent Control and Information Processing (ICICIP). – Dalian, China, 2010. – P. 165–169.
6. Al-Jehani, N.A., Nounou, H.N. and Nounou, M.N. Fuzzy Control of a CSTR Process // Mechatronics and its Applications, 2012. – P. 1–6.
7. Antonelli, R., and Astolfi, A. Continuous Stirred Tank Reactors: Easy to Stabilise // Automatica. – 2003. – No. 39(10). – P. 1817–1827.
8. Graichen, K., Hagenmeyer, V., and Zeitz, M. Design of Adaptive Feedforward Control under Input Constraints for a Benchmark CSTR Based on a BVP solver // Computers and Chemical Engineering. – 2009. – No. 33(2). – P. 473–483.
9. Chiu, C.S. Derivative and Integral Terminal Sliding Mode Control for a Class of MIMO Nonlinear Systems // Automatica. – 2012. – No. 48(2). – P. 316–326.
10. Zhao, D., Zhu, Q., and Dubbeldam, J. Terminal Sliding Mode Control for Continuous Stirred Tank Reactor // Chemical Engineering Research and Design. – 2015. – No. 94. – P. 266–274.
11. Ma, H., Wu, J., and Xiong, Z. A Novel Exponential Reaching Law of Discrete-Time Sliding-Mode Control // IEEE Transactions on Industrial Electronics. – 2017. – No. 64(5). – P. 3840–3850.
12. Ma, H., Wu, J., and Xiong, Z. Discrete-Time Sliding-Mode Control with Improved Quasi-sliding-mode Domain // IEEE Transactions on Industrial Electronics. – 2016. – No. 63(10). – P. 6292–6304.
13. Yan, X.-G., Spurgeon, S.K. and Edwards, C. State and Parameter Estimation for Nonlinear Delay Systems Using Sliding Mode Techniques // IEEE Transactions on Automatic Control. – 2013. – No. 58(4). – P. 1023–1029.
14. Boudjellal, M., and Illoul, R. High-Order Sliding Mode and High-Gain Observers for State Estimation and Fault Reconstruction for a Nonlinear CSTR // 6th International Conference on Systems and Control (ICSC). – Batna, Algeria, 2017. – P. 231–236.
15. Ignaciuk, P. Nonlinear Inventory Control with Discrete Sliding Modes in Systems with Uncertain Delay // IEEE Transactions on Industrial Informatics. – 2014. – No. 10(1). – P. 559–568.
16. Ma, L., Zhao, D., and Spurgeon, S.K. Disturbance Observer Based Discrete Time Sliding Mode Control for a Continuous Stirred Tank Reactor // 15th Workshop on Variable Structure Systems, 2018. – P. 372–377.
17. Chen, W.H., Yang, J., Guo, L., and Li, S. Disturbance-Observer-Based Control and Related Methods: An Overview // IEEE Transactions on Industrial Electronics. – 2016. – No. 63(2). – P. 1083–1095.
18. Rios, H., Efimov, D., Moreno, J.A., and Perruquetti, W. Time-Varying Parameter Identification Algorithms: Finite and Fixed-Time Convergence // IEEE Transactions on Automatic Control. – 2017. – No. 62(7). – P. 3671–3678.
19. Уткин В.А., Уткин А.В. Задача слежения в линейных системах с параметрическими неопределенностями при неустойчивой нулевой динамике // Автоматика и телемеханика. – 2014. – № 9. – С. 62–81. [Utkin, V.A., Utkin, A.V. Problem of Tracking in Linear Systems with Parametric Uncertainties under Unstable Zero Dynamics // Autom. Remote Control. – 2014. – Vol. 75, no. 9. – P. 1577–1592].
20. Krasnova, S.A., and Utkin, A.V. Sigma Function in Observer Design for States and Perturbations // Automation and Remote Control. – 2016. – No. 77(9). – P. 1676–1688.
21. Краснова С.А., Уткин В.А., Уткин А.В. Блочный подход к анализу и синтезу инвариантных нелинейных систем слежения // Автоматика и телемеханика. – 2017. – № 12. – С. 26–53. [Krasnova, S.A., Utkin, V.A., Utkin, A.V. Block Approach to Analysis and Design of the Invariant Nonlinear Tracking Systems // Autom. Remote Control. – 2017. – Vol. 78, no. 12. – P. 2120–2140].
22. Han, J.S., Bahn, W., and Kim, T.I. Decoupled Disturbance Compensation under Control Saturation with Discrete-Time Variable Structure Control Method in Industrial Servo Systems // The 16th International Conference on Control, Automation and Systems. – Gyeongju, Korea, 2016. – P. 1453–1457.
23. Wu, W. Nonlinear Bounded Control of a Nonisothermal CSTR // Industrial and Engineering Chemistry Research. – 2000. – No. 39(10). – P. 3789–3798.
24. Уткин А.В., Уткин В.А. Синтез систем стабилизации при односторонних ограничениях на управляющие воздействия // Проблемы управления. – 2020. – № 3. – С. 3–4. [Utkin, A.V., Utkin, V.A. The Synthesis of Stabilization Systems under One-Sided Restrictions on Control Actions // Control Sciences. – 2020. – No. 3. – P. 3–14. (In Russian)].
25. Антипов А.С., Краснова С.А. Блочный синтез системы слежения для двухроторной электромеханической системы при ограничениях на переменные состояния // Прикладная математика и механика. – 2021. – № 85(1). – С. 3–20. [Antipov, A.S., Krasnova, S.A. Block-Based Synthesis of a Tracking System for a Twin-Rotor Electromechanical System with Constraints on State Variables // Mech. Solids. 2021. – Vol. 56, no. 7. – P. 43–56].
26. Дракунов С.В., Изосимов Д.Б., Лукьянов А.Г. и др. Принцип блочного управления. I // Автоматика и телемеханика. – 1990. – No. 5. – С. 3–13. [Drakunov, S.V., Izosimov, D.B., Luk'yanov, A.G., et al. The Block Control Principle. I // Automation and Remote Control. – 1990. – Vol. 51, no. 5. – P. 601–609].
27. Чиликин М.Г., Ключев В.И., Сандлер А.С. Теория автоматизированного электропривода. – М.: Энергия, 1979. [Chilikin, M.G., Klyuchev, V.I., Sandler, A.S. Teoriya avtomatizirovannogo ehlektroprivoda. – M.: Energiya, 1979. (In Russian)].
28. Краснова С.А., Уткин В.А. Каскадный синтез наблюдателей состояния динамических систем. – М.: Наука, 2006. – 272 с. [Krasnova, S.A., Utkin, V.A. Cascade Synthesis State Observer of Dynamical System. – M.: Nauka, 2006. – 272 p. (In Russian)].

Статья представлена к публикации членом редколлегии
В.Н. Афанасьевым.

Поступила в редакцию 24.05.2021,
после доработки 28.07.2021.
Принята к публикации 24.08.2021.

Гулюкина Светлана Игоревна – мл. науч. сотрудник
✉ gulyukina.s.i@mail.ru,

Уткин Виктор Анатольевич – д-р техн. наук
✉ vicutkin@ipu.ru,

Институт проблем управления им. В.А. Трапезникова РАН,
г. Москва.



A BLOCK APPROACH TO CSTR CONTROL UNDER UNCERTAINTY, STATE-SPACE AND CONTROL CONSTRAINTS

S.I. Gulyukina¹ and V.A. Utkin²

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

¹✉ gulyukina.s.i@mail.ru, ²✉ vicutkin@ipu.ru

Abstract. This paper designs a control law to maintain the temperature in the jacket of a continuous stirred tank reactor (CSTR). The standard mathematical model describing the reactor operation is extended by introducing the actuator's dynamics. The state-space and control constraints are considered by a nonlinear change of the variables of the plant's initial model using linear sat functions. In the transformed system, these constraints are considered by feedback control law design. The block approach allows linearizing the feedback control law by sequentially solving the first-order design subproblems. Under incomplete information on the state vector and the effect of exogenous disturbances, an observer of the state vector and disturbances is constructed to estimate the unknown signals with a given accuracy. The effectiveness of the proposed approach is illustrated by simulating the CSTR–DC motor system in MATLAB.

Keywords: CSTR, tracking problem, block approach, observer of state vector and disturbances, state-space and control constraints.

Funding. This work was supported by the Russian Foundation for Basic Research, project no. 20-01-00363 A.

ОЦЕНКА ВЛИЯНИЯ ФАКТОРОВ ОКРУЖАЮЩЕЙ СРЕДЫ НА ПОКАЗАТЕЛИ СМЕРТНОСТИ НАСЕЛЕНИЯ СТАРШИХ ВОЗРАСТНЫХ ГРУПП НА ПРИМЕРЕ Г. КРАСНОЯРСКА¹

О.В. Тасейко, Д.А. Черных

Аннотация. Выполнен анализ и прогноз негативного воздействия климатических параметров и качества атмосферного воздуха на смертность населения г. Красноярска. Проанализирована многолетняя динамика климатических показателей для г. Красноярска. Статистика по заболеваниям, чувствительным к климатическим изменениям, рассматривалась за период с 2011 по 2014 г. Методами многомерного статистического анализа установлена взаимосвязь между климатическими параметрами, уровнем загрязнения атмосферного воздуха и количеством смертей. Выявлены факторы, вносящие наибольший вклад в показатели смертности по основным группам климатозависимых заболеваний; проанализированы первоначальные причины смерти, связанные с негативным влиянием факторов окружающей среды; определены возрастные группы населения, которые подвергаются наибольшему воздействию климатических параметров и загрязнения атмосферного воздуха. Показана применимость нелинейной регрессии Пуассона для прогнозирования смертности, довольно близкой к фактической.

Ключевые слова: качество атмосферного воздуха, климатические параметры, многомерная регрессионная модель, смертность населения, социально-природно-техногенная система.

ВВЕДЕНИЕ

Совместное воздействие антропогенной деятельности и природных факторов на здоровье населения вызывает рост рисков устойчивого развития территориальных образований, при этом проявляются тенденции урбанизации природной среды, что создает новые типы технологических и природных угроз.

В силу складывающихся экономических и социальных условий в РФ в ближайшем будущем основным объектом урбанизации будут являться территории Сибири и Арктики. Принимая во внимание уникальность природных систем и глобальную значимость этих регионов для устойчивого развития страны, исследование природных и технологических угроз и рисков развития с использованием информационных технологий нового поко-

ления представляется особо важной научной проблемой. Устойчивое развитие территорий связано с количественной оценкой комплексной безопасности и с управлением регионом с применением риск-ориентированного подхода [1–4]. Территория промышленного региона представляет собой единую сложную социально-природно-техногенную (СПТ) систему, элементы которой оказывают друг на друга взаимное влияние и характеризуются наличием связанных видов рисков [5]. Функционирование социосферы как элемента СПТ системы сопряжено с индивидуальными стратегическими рисками потери жизни и здоровья. К наиболее значимым факторам, формирующим индивидуальные стратегические риски, относят загрязнение окружающей среды (прежде всего, атмосферного воздуха) и климат, имеющий тенденцию к глобальному изменению. Конкретные метеорологические явления, оказывающие влияние на здоровье населения, являются кратковременными, тем не менее их повторяемость определяется климатическими особенностями территории. Поэтому в данной постановке используется как термин «метеорологические условия», так и термин «климатические параметры».

¹Исследование выполнено при финансовой поддержке РФФИ, Правительства Красноярского края и Краевого фонда науки в рамках научного проекта № 19-413-240013 «Методология оценки риска от влияния факторов окружающей среды на здоровье и смертность населения промышленных агломераций».



Наличие зависимости между уровнями смертности и климатическими факторами, такими как температурные волны, установлено в ряде исследований, выполненных для таких городов, как Архангельск, Якутск, Астрахань, Красноярск, Москва и др. [6–11]. Для г. Красноярска оценка относительного риска смертности от негативного воздействия волн холода и волн жары выполнена с помощью регрессионной модели Пуассона [7].

Периоды сильной жары в г. Архангельске (порог температуры 21°C) связаны с увеличением смертности как от сердечно-сосудистых заболеваний, так и от всех естественных причин среди лиц старше 65 лет, а также от всех внешних причин среди лиц старше 30 лет (за период с 1999 по 2008 г.). С 1999 по 2008 г. зафиксировано 110 дополнительных смертей из-за волн жары и 179 дополнительных смертей из-за волн холода [6].

Температурные волны были впервые выявлены и исследованы в 1881–1884 гг. и описывались как область высоких или низких температур атмосферного воздуха, которая в течение длительного времени не остается на одном месте [12]. Температурные волны связаны не только с серьезными экологическими и медицинскими проблемами, но могут повлечь за собой также значительный экономический ущерб. Высокие температуры могут вызвать термический стресс и ухудшить качество воздуха, что, в свою очередь, приводит к неблагоприятным последствиям для здоровья, особенно для более уязвимых групп населения, в том числе лиц, страдающих сердечно-сосудистыми или респираторными заболеваниями, и пожилых людей [13].

Экстремально низкие температурные показатели зимой и высокие летом характерны для климата территорий Центральной Сибири, Якутии, а также других регионов и связаны с резко континентальным типом погоды. При таких уникальных климатических характеристиках для крупномасштабных промышленных образований существует необходимость в анализе синергетического воздействия факторов окружающей среды на смертность населения. Разработка методологии оценки рисков многофакторного воздействия на уровень смертности населения с применением регрессионных статистических моделей, а также разработка мероприятий по защите населения от негативного воздействия факторов окружающей среды позволят в некоторой степени решить данную проблему. При этом методология должна учитывать структурные особенности населения и региона.

Методы оценки рисков для здоровья населения разрабатываются уже в течение трех десятилетий.

Негативное влияние высоких уровней загрязнения воздуха на заболеваемость и смертность населения доказано многими исследованиями. Дети и подростки наиболее подвержены влиянию химического загрязнения атмосферного воздуха [14–16]. Дыхательная система ребенка является основной мишенью для загрязнителей воздуха. Они вызывают широкий спектр острых и хронических эффектов либо как один фактор риска, либо, чаще всего, в сочетании с другими внешними агентами и/или характеристиками восприимчивости ребенка. Важную роль в период воздействия ингаляционных загрязняющих веществ играет возраст человека [17–19]. Младенцы более восприимчивы к поражению легких токсинами одного и того же вида, чем взрослые, даже в дозах ниже оптимального уровня для взрослого человека [20–21].

Анализ смертности населения г. Красноярска за 2000–2018 гг. (по данным Территориального органа Федеральной службы государственной статистики по Красноярскому краю) показал, что общая смертность снижалась в период до 2009 г. (на 17 %) и увеличивалась до 2018 г. (на 9 %) при росте численности населения более чем на 20 % (рис. 1).

Смертность от цереброваскулярных болезней за период с 2006 по 2015 г. снизилась на 9,9 % (рис. 2). Доля умерших от ишемической болезни сердца, напротив, возросла с 18 % в 2000 г. до 30 % в 2018 г.

Для выявления взаимосвязи между смертностью населения и факторами окружающей среды рекомендована пуассоновская регрессионная модель [22]. В зарубежных эпидемиологических исследованиях применение модели Пуассона прослеживается со второй половины XX в. С помощью данной модели была изучена взаимосвязь между смертностью и факторами окружающей среды: загрязнением атмосферного воздуха [23], средней температурой [24], температурными волнами [25–27]. Существуют исследования по выявлению зависимости между заболеваемостью онкологией и загрязненным воздухом, а также вредными привычками [28–29]. Работы, в которых рассматривается применение данной модели для г. Москвы, включали три влияющих фактора – среднесуточную температуру воздуха, концентрацию озона и концентрацию взвешенных частиц PM_{10} . Было выявлено увеличение смертности от всех причин, за исключением внешних (происшествия, условия и обстоятельства в качестве причины травмы, отравления и другого неблагоприятного воздействия) при повышении количества дней с аномально низкими и высокими температурами [30].

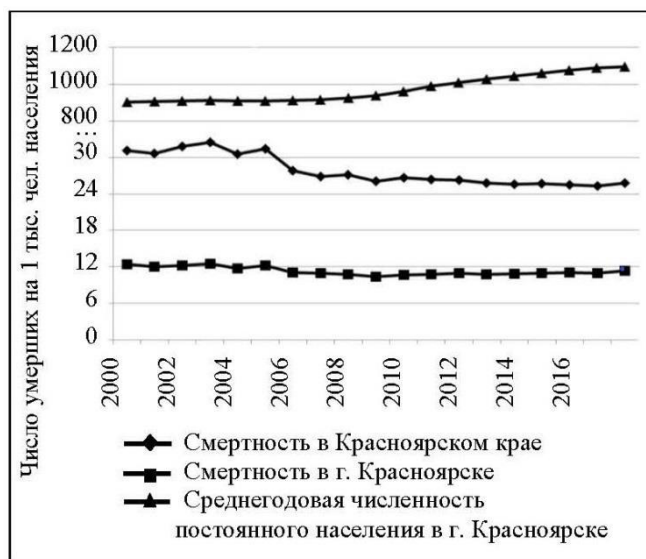


Рис. 1. Динамика смертности и среднегодовой численности населения г. Красноярска и Красноярского края



Рис. 2. Динамика смертности населения г. Красноярска от ишемической болезни сердца и цереброваскулярных болезней

Целью настоящего исследования является оценка относительных рисков социосферы СПТ системы г. Красноярска. В качестве влияющих факторов рассматриваются характеристики климатических параметров и уровни загрязнения атмосферного воздуха, в качестве прогнозируемого показателя используется риск преждевременной смертности населения г. Красноярска от совместного влияния климатических и экологических факторов. Основная задача исследования включает в себя оценку наличия взаимосвязи между климатическими параметрами, загрязненным атмосферным воздухом и показателями смертности населения с помощью методов многомерного статистического

анализа. Выполнялся также анализ значимости вклада каждого из исследуемых факторов в уровни смертности по основным группам климатозависимых заболеваний.

1. МАТЕРИАЛЫ И МЕТОДЫ

Было проведено исследование показателей смертности населения г. Красноярска от **первоначальных причин**, характеризующихся наибольшей чувствительностью к климатическим факторам, согласно Международной классификации болезней (МКБ-10) [22] в двух возрастных группах (60–74 лет и старше 75 лет), разделенных по полу за период с 2011 по 2014 г.:

- ишемическая болезнь сердца (I20–I25),
- цереброваскулярные болезни (I60–I69),
- болезни органов дыхания (J00–J22, J30, J40–J44, J45).

Определение и оценка климатических характеристик осуществлялись с использованием данных государственных наблюдательных сетей метеостанций [31, 32]. С целью стандартизации временных рядов было проведено нормирование делением на стандартное отклонение [33]. Заполнение пропусков по значениям данных факторов выполнялось методом скользящего среднего.

Для построения нелинейной модели с пуассоновской регрессией, описывающей зависимость смертности от воздействия факторов окружающей среды, использовались данные о ежедневной смертности населения г. Красноярска за период с 2011 по 2014 г. (предоставленные Территориальным органом Федеральной службы государственной статистики по Красноярскому краю) [22]. Модель имеет вид:

$$\log(\mu_t) = \beta_0 + \beta_1 X_{1,t-L} + \dots + \beta_k X_{k,t-L} + \beta_{k+1} Z,$$

$$\mu_t = e^{\beta_0} e^{\beta_1 X_{1,t-L}} \dots e^{\beta_k X_{k,t-L}} e^{\beta_{k+1} Z},$$

где μ_t – прогнозируемый параметр (пуассоновская переменная); $X_{1,t}, \dots, X_{k,t}$ – влияющие переменные; Z – краткосрочный временной тренд по дням недели; L – отставание негативного эффекта во времени (лаг); $\beta_1, \dots, \beta_{k+1}$ – параметры модели; β_0 – свободный член.

В эпидемиологических исследованиях определяют связь между концентрацией загрязнителей воздуха за один день и последствиями для здоровья в некоторые лаг-дни. Лаг подразумевает под собой отставание эффекта во времени. То есть при воздействии негативного явления его эффект мо-



жет проявиться не в тот же день, а лишь через некоторое время [22]. В настоящем исследовании использовался лаг от 1 до 15 дней.

Для интерпретации коэффициента регрессии применялся показатель относительного риска (англ. *relative risk*, *RR*). Относительный риск представляет собой меру связи между независимой переменной (например, концентрация загрязняющих веществ в воздухе) и риском определенного результата (например, количество людей с респираторным повреждением). Относительный риск показывает, во сколько раз увеличивается смертность населения в период негативного явления относительно фоновой смертности (т. е. смертности в отсутствие негативного воздействия факторов окружающей среды). Для регрессии Пуассона относительный риск определяется так [13, 34]:

$$RR = \exp(\beta_i),$$

где β_i – коэффициент регрессии.

В табл. 1 представлена характеристика исследуемых факторов окружающей среды в динамике с 2000 по 2018 г. Данные о ежедневных концентрациях загрязняющих веществ в атмосфере г. Красноярска были предоставлены ФГБУ «Среднесибирское УГМС».

Выбор загрязняющих веществ основывался на доказанности фактов их влияния на показатели заболеваемости и смертности населения (табл. 2). Одним из наиболее опасных загрязнителей атмосферы является озон (O_3). Это вещество первого класса опасности, являющееся одним из главных компонентов фотохимического смога. Повышенный уровень приземного озона отмечается лишь при солнечной жаркой погоде. Ингаляционное воздействие вещества на организм сопровождается раздражением органов дыхания, снижением легочных функций, развитием астмы и аллергии, а также значительным снижением иммунитета к инфекциям [22, 35–37]. В связи с тем, что в городах Красноярского края на постах государственной наблюдательной сети измерения озона не выполняются, в настоящем исследовании использовались концентрации диоксида азота и формальдегида как предшественников озона.

В процессе обработки данных и проведения численного эксперимента выполнялся анализ на наличие коллинеарности или мультиколлинеарности факторов, а также проводилась корректировка краткосрочных временных трендов (по дням недели) путем добавления дополнительного коэффициента в модель.

Таблица 1

Динамика факторов окружающей среды

Год	Волны жары				Холодовые волны				Температура, °С	Относительная влажность, %	Экстремальные перепады температуры в течение суток			Средняя концентрация загрязняющих веществ		
	Число волн, единицы	Продолжительность волн, дни	Амплитуда волн, °С		Число волн, единицы	Продолжительность волн, дни	Амплитуда волн, °С				Число событий, единицы	Диапазон мощности, °С		Взвешенные вещества, мг/м ³	NO ₂ – диоксид азота, мг/м ³	F – формальдегид, мг/м ³
			min	max			min	max				min	max			
2000	6	37	7,2	20,2	5	52	−38	−14	0,8	73	49	10	16,6	0,254	0,024	–
2003	4	29	6,4	20,6	3	30	−14	−10	2,5	71	73	10	31	0,265	0,042	0,008
2006	2	24	13,6	20,4	4	49	−24	−10	0,8	71	15	10,1	20	0,198	0,051	0,0043
2009	3	19	8,3	21	9	82	−24	−7,7	0,2	73	15	10,1	22,2	0,173	0,062	0,0094
2012	4	31	8,3	21,1	4	78	−21	−9	0,6	71	14	10,1	30,3	0,228	0,051	0,0185
2015	7	42	7,5	20,8	1	14	−11	−11	3,8	66	16	10	13,7	0,1380	0,0360	0,0120
2018	4	40	7,7	20,7	8	60	−20	−11	1,3	70	26	10,1	19,4	0,1063	0,0379	0,0168

Таблица 2

Эффект воздействия загрязняющих веществ на организм [38]

Вещество	Класс опасности	Критический орган	Критический эффект
Взвешенные вещества	3	Органы дыхания	– увеличение общей смертности, – смертность от заболеваний сердечно-сосудистой системы, органов дыхания, – частота симптомов со стороны верхних и нижних отделов дыхательных путей, – обращаемость по поводу респираторных заболеваний, – частота обострения бронхиальной астмы
Озон (O ₃) [22, 35]	1	Органы дыхания	– увеличение общей смертности, – раздражение органов дыхания, – снижение легочных функций, – развитие астмы и аллергии, – значительное снижение иммунитета к инфекциям
Диоксид азота (NO ₂)	2	Органы дыхания, кровь и кроветворные органы	– увеличение частоты возникновения и продолжительности заболеваний верхних и нижних отделов дыхательных путей, – увеличение количества болезней нижних дыхательных путей у детей
Формальдегид	2	Органы дыхания, органы зрения, иммунная система	– увеличение частоты возникновения заболеваний верхних и нижних отделов дыхательных путей, – воспалительные процессы в легких, – заболевания иммунной системы, включая аллергические реакции, – заболевания органов зрения

2. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Уровень воздействия конкретного фактора на уровень смертности в исследуемых моделях определяется с помощью коэффициента регрессии, который указывает, на сколько единиц изменится результат при изменении исследуемого фактора на одну единицу [39].

У женщин в возрастной группе от 60 до 74 лет на смертность от болезней органов дыхания в наибольшей степени влияют температурные волны на девятый день воздействия (лаг равен 9). Влияние поллютантов представлено воздействием формальдегида:

$$\ln(\mu_t) = -7,4 + 0,9F_t - 4,9Temp_{t-2} - 3,3Temp_{t-12} + 0,9Diff_{t-1} + 0,9Diff_{t-12} + 1,3Hum_{t-11} + 2,8W_{t-9}, \quad (1)$$

где F – концентрация формальдегида в атмосферном воздухе, мг/м³; $Temp$ – температура, °C; $Diff$ – перепад температуры, °C; Hum – относительная влажность, %; W – температурные волны, ед.

Небольшое значение свободного члена β_0 (по сравнению с суммарным значением весов учтенных в модели факторов) свидетельствует о несущественном влиянии неучтенных факторов в данной модели. Согласно выражению (1), одни и те же влияющие факторы могут оказывать негативный эффект при различных лагах. Влияние среднесуточной температуры проявляется как на второй день, так и на 12-й день после его ингаляционного воздействия, перепад температуры в течение суток – на первый и на 12-й дни.

На рис. 3 приведено распределение фактической смертности и смертности, рассчитанной с использованием пуассоновской модели. Видно, что рассчитанный показатель смертности хорошо описывает фактическую (с коэффициентом корреляции 0,8) на примере смертности от болезней органов дыхания у женщин в возрастной группе от 60 до 74 лет.

В результате обработки имеющегося статистического массива данных за период с 2011 по 2014 г. для возрастных категорий «от 60 до 74 лет» и «старше 75 лет» было получено 12 моделей прогноза смертности для трех основных непосредственных причин, относящихся к категории климатозависимых. Относительный риск смертности от воздействия факторов окружающей среды, лаги и показатели корреляции для каждой из моделей приведены в табл. 3.

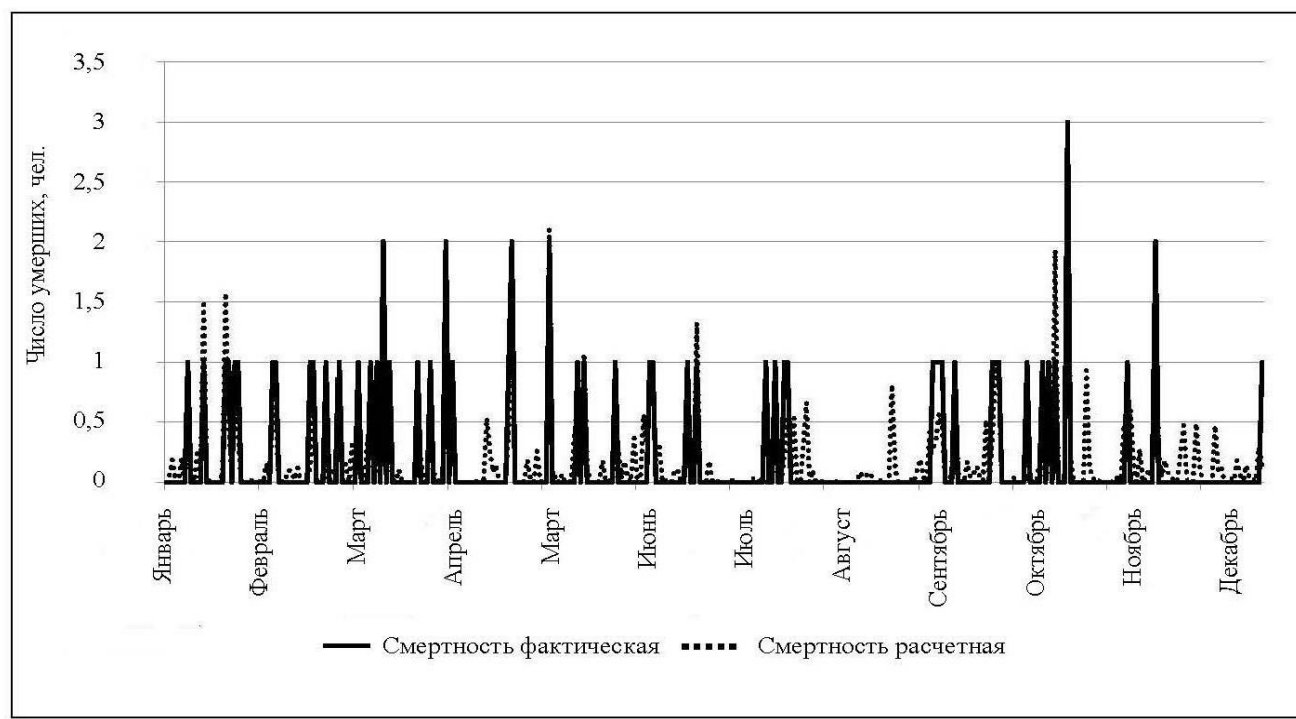


Рис. 3. Фактическая и расчетная смертности от болезней органов дыхания у женщин в возрастной группе от 60 до 74 лет за 2014 г.

Таблица 3

Относительный риск смертности населения

Влияющие факторы	Пол	Ишемическая болезнь сердца	Церебро- васкуляр- ные болезни	Болезни органов дыхания	Ишемическая болезнь сердца	Церебро- васкулярные болезни	Болезни органов дыхания
		Возрастная группа					
		От 60 до 74 лет			Старше 75 лет		
		Значение коэффициента корреляции между фактической и расчетной смертностью					
	Ж	0,69	0,66	0,8	0,63	0,53	0,66
	М	0,6	0,63	0,6	0,55	0,64	0,66
Относительный риск смертности (relative risk)							
Диоксид азота	Ж	1,2 / L = 5	1,3 / L = 15	—	—	—	1,3
	М	1,1 / L = 5	1,2 / L = 13	1,4	—	—	1,4 / L = 14
Формальдегид	Ж	1,2 / L = 9	—	2,5	—	1,2	1,4 / L = 11
	М	—	1,5 / L = 4	—	0,7 / L = 14	0,7 / L = 2	1,6 / L = 3
Взвешенные ве- щества	Ж	1,2 / L = 5	1,3 / L = 8	—	—	0,8	—
	М	1,2 / L = 15	1,3 / L = 5	—	1,2 / L = 3	—	1,4 / L = 8
Температура	Ж	0,4 / L = 11	—	0,04 / L = 12	0,7 / L = 4	—	3,7 / L = 5
	М	0,6 / L = 15	—	—	—	0,4 / L = 12	3,3 / L = 10
Относительная влажность	Ж	1,3 / L = 6	—	3,6 / L = 11	0,8 / L = 11	—	1,6 / L = 2
	М	1,1	—	—	1,2 / L = 5	1,5 / L = 4	—
Резкие перепады температуры	Ж	—	—	2,5 / L = 12	—	—	—
	М	—	—	1,4 / L = 9	—	0,8 / L = 14	1,4 / L = 1
Температурные волны	Ж	1,8 / L = 2	—	16,4 / L = 9	1,5 / L = 13	—	0,5 / L = 2
	М	—	0,4 / L = 10	—	—	—	—

3. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Анализ смертности населения от негативного воздействия факторов окружающей среды с использованием пуассоновской регрессионной модели позволил получить такие результаты:

- Анализ распределения лагов показывает, что большая часть негативных эффектов у мужчин от ишемической болезни сердца и цереброваскулярных болезней проявляется до шести дней после воздействия совокупности факторов, от болезней органов дыхания – после седьмого дня, в то время как у женщин негативные эффекты для здоровья накапливаются более равномерно в течение двух недель (рис. 4).

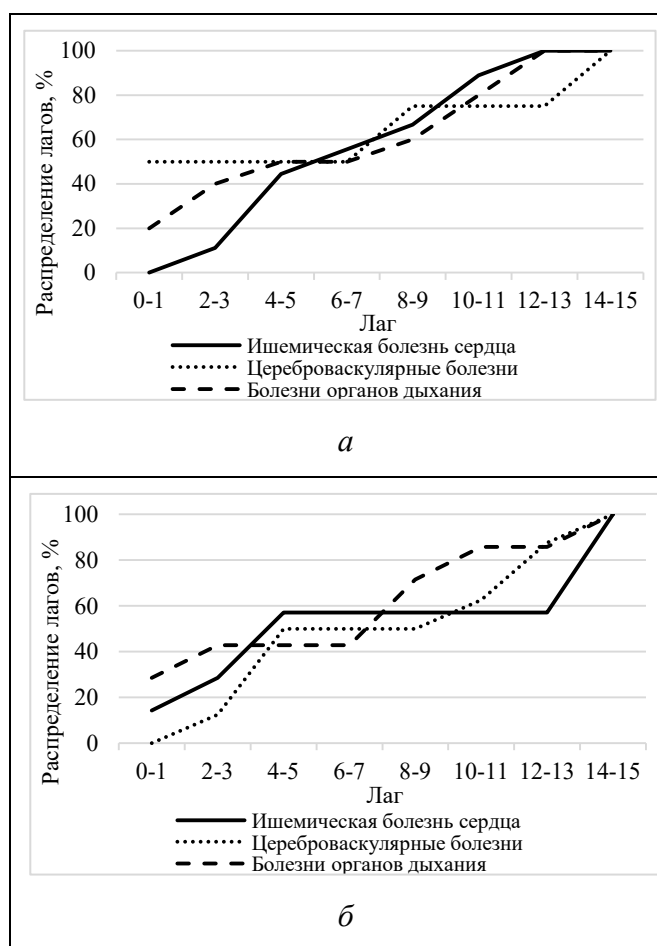


Рис. 4. Распределение отставания отрицательного эффекта во времени: а – среди женщин, б – среди мужчин

- Анализ полученных моделей выявил наибольший вклад воздействия климатических факторов в смертность населения всех возрастных и половых групп, что и показано в большинстве полученных моделей в соотношении 58,1 % (влияние климатических факторов) к 41,9 % (влияние

загрязненного атмосферного воздуха). Также показано, что женщины в наибольшей степени подвергаются негативному воздействию факторов окружающей среды.

- Изучение относительного вклада в уровни смертности поллютантов показало преобладающее влияние формальдегида. Из климатических факторов наибольший вред оказывают температурные волны.

- Оценка тесноты связи спрогнозированной и фактической смертностей показала, что наибольшие коэффициенты корреляции приходятся на смертность от болезней органов дыхания.

- Показано, что среди возрастных групп наиболее чувствительными к воздействию комплекса анализируемых факторов оказались пожилые люди в возрасте от 60 до 74 лет.

ЗАКЛЮЧЕНИЕ

Применение нелинейной регрессии Пуассона позволило доказать применимость данного метода для прогнозирования смертности населения от факторов окружающей среды довольно близкой к фактической. Это подтверждает возможность применения данного подхода для исследований по оценке совместного воздействия климатических параметров и качества атмосферного воздуха на смертность населения. С целью развития представленного подхода к оценке рисков предполагается исключение незначащих факторов и более детальное исследование характера влияния значащих факторов по группам в зависимости от направления выявленных связей.

ЛИТЕРАТУРА

- Безопасность России. Правовые, социально-экономические и научно-технические аспекты. Региональные проблемы безопасности. Красноярский край / под ред. В.Ф. Шабанова. – М.: МГФ «Знание», 2001. – 576 с. [Bezopasnost' Rossii. Pravovye, sotsial'no-ehkonomicheskie i nauchno-tekhnicheskie aspekty. Regional'nye problemy bezopasnosti. Krasnoyarskii krai / pod red. V.F. Shabanova. – M.: MGF «ZnaniE», 2001. – 576 p. (In Russian)]
- Винокуров Ю.И., Лепихин А.М., Москвичев В.В. и др. Геоинформационные технологии и математические модели для мониторинга и управления экологическими и социально-экономическими системами / под ред. Ю.И. Шокина. – Б.: Пять плюс, 2011. – 250 с. [Vinokurov, Yu.I., Lepikhin, A.M., Moskvichev, V.V., et al. Geoinformatsionnye tekhnologii i matematicheskie modeli dlya monitoringa i upravleniya ehkologicheskimi i sotsial'no-ehkonomicheskimi sistemami / pod red. Yu.I. Shokina. – B.: Pyat' plus, 2011. – 250 p. (In Russian)]



3. Левкевич В.Е., Лепихин А.М., Москвичев В.В. и др. Безопасность и риски устойчивого развития территорий. – К.: Сибирский федеральный ун-т, 2014. – 224 с. [Levkevich, V.E., Lepikhin, A.M., Moskvichev, V.V., et al. Bezopasnost' i riski ustoichivogo razvitiya territorii. – K.: Sibirskii federal'nyi un-t, 2014. – 224 p. (In Russian)]
4. Акимов В.А., Быков А.А., Шевченко А.В., Стрелко С.В. Развитие научно-методических основ государственного управления с использованием показателей стратегических рисков, в том числе и рисков чрезвычайных ситуаций (на примере Ульяновской области) // Стратегия гражданской защиты: проблемы и исследования. – 2013. – Т. 3, № 2. – С. 736–780. [Akimov, V.A., Bykov, A.A., Shevchenko, A.V., Strelko, S.V. Razvitie nauchno-metodicheskikh osnov gosudarstvennogo upravleniya s ispol'zovaniem pokazatelei strategicheskikh riskov, v tom chisle i riskov chrezvychainykh situatsii (na primere Ul'yanovskoi oblasti) // Strategiya grazhdanskoj zashchity: problemy i issledovaniya. – 2013. – Vol. 3, no. 2. – P. 736–780. (In Russian)]
5. Москвичев В.В., Бычков И.В., Потанов В.П. и др. Информационная система территориального управления рисками развития и безопасностью // Вестник Российской академии наук. – 2017. – Т. 87, № 8. – С. 696–705. [Moskvichev, V.V., Bychkov, I.V., Potanov, V.P. et al. Informatsionnaya sistema territorial'nogo upravleniya riskami razvitiya i bezopasnost'yu // Vestnik Rossijskoi akademii nauk. – 2017. – Vol. 87, no. 8. – P. 696–705. (In Russian)]
6. URL: <https://www.who.int/publications/i/item/protecting-health-from-climate-change-a-seven-country-initiative>.
7. Ревич Б.А., Шапошников Д.А. Особенности воздействия волн холода и жары на смертность в городах с резко-континентальным климатом // Сибирское медицинское обозрение. – 2017. – № 2. – С. 84–90. [Revich, B.A., Shaposhnikov, D.A. Osobennosti vozdeistviya voln kholoda i zhary na smertnost' v gorodakh s rezko-kontinental'nym klimatom // Sibirskoe meditsinskoe obozrenie. – 2017. – No. 2. – P. 84–90. (In Russian)]
8. Ревич Б.А., Шапошников Д.А. Волны холода в южных городах европейской части России и преждевременная смертность населения // Проблемы прогнозирования. – 2016. – № 2. – С. 125–131. [Revich, B.A., Shaposhnikov, D.A. Volny kholoda v yuzhnykh gorodakh evropeiskoi chasti Rossii i prezhdevremennaya smertnost' naseleniya // Problemy prognozirovaniya. – 2016. – No. 2. – P. 125–131. (In Russian)]
9. Ревич Б.А., Харьков Т.Л., Подольная М.А. Динамика смертности и ожидаемой продолжительности жизни населения арктического/приарктического региона России в 199–2014 годах // Экология человека. – 2017. – № 9. – С. 48–58. [Revich, B.A., Khar'kova, T.L., Podol'naya, M.A. Mortality Dynamics and Life Expectancy of Population of Arctic/Subarctic Region of the Russian Federation in 1999–2014 // Ekologiya Cheloveka (Human Ecology). – 2017. – No. 9. – P. 48–58. (In Russian)]
10. Шапошников Д.А., Ревич Б.А. О некоторых подходах к вычислению рисков температурных волн для здоровья // Анализ риска здоровью. – 2018. – № 1. – С. 22–31. [Shaposhnikov, D.A., Revich, B.A. On Some Approaches to Calculation of Health Risks Caused by Temperature Waves // Health Risk Analysis. – 2018. – No. 1. – P. 22–31. (In Russian)]
11. Кваша Е.А., Ревич Б.А., Харьков Т.Л. Сходство и различия смертности населения в 4-х мегаполисах России // Бюллетень национального научно-исследовательского института общественного здоровья имени Н.А. Семашко. – 2017. – № 4. – С. 69–75. [Kvasha, E.A., Revich, B.A., Khar'kova, T.L. Skhodstvo i razlichiya smertnosti naseleniya v 4-kh megapolisakh Rossii // Byulleten' natsional'nogo nauchno-issledovatel'skogo instituta obshchestvennogo zdorov'ya imeni N.A. Semashko. – 2017. – No. 4. – P. 69–75. (In Russian)]
12. URL: http://dic.academic.ru/dic.nsf/brokgauz_efron/110731/Холода.
13. Tadano, Y.S., Ugaya, C.M.L., and Franco, A.T., Methodology to Assess Air Pollution Impact on Human Health Using the Generalized Linear Model with Poisson Regression, in *Air Pollution - Monitoring, Modelling and Health*, Khare, M., Ed., Rijeka, Croatia: IntechOpen. – 2012. – P. 281–304. – DOI: 10.5772/33385.
14. Колесникова Л.И., Долгих В.В., Астахова Т.А. и др. Оценка нарушений здоровья и микроанамалий развития детей // Бюллетень СО РАМН. – 2008. – № 1(129). – С. 26–29. [Kolesnikova, L.I., Dolgikh, V.V., Astakhova, T.A., et al. Evaluation of Health Impairment and Malformations of Development of Children // Byulleten' SO RAMN. – 2008. – No. 1(129). – P. 26–29. (In Russian)]
15. Пинегин Б.В., Баймуканова Г.П., Печуркина Н.С. Экологический иммунодефицит: иммуногенетические аспекты его развития и коррекции // Вестник РАМН. – 1994. – № 4. – С. 20–28. [Pinegin, B.V., Baimukanova, G.P., Pechurkina, N.S. Ehkologicheskii immunodefitsit: immunogeneticheskie aspekty ego razvitiya i korrektsii // Vestnik RAMN. – 1994. – No. 4. – P. 20–28. (In Russian)]
16. Савилов Е.Д., Ильина О.В. Инфекционная патология в условиях техногенного загрязнения окружающей среды. – Н.: Наука, 2010. – 248 с. [Savilov, E.D., Il'ina, O.V. Infektsionnaya patologiya v usloviyakh tekhnogennoho zagryazneniya okruzhayushchei sredy. – N.: Nauka, 2010. – 248 p. (In Russian)]
17. Dezateux, C., Lum, S., Hoo, A., et al. Low Birth Weight for Gestation and Airway Function in Infancy: Exploring the Fetal Origins Hypothesis // Thorax. – 2004. – Vol. 59. – P. 60–66.
18. Hoo, A.F., Dezateux, C., Henschen, M., et al. Development of Airway Function in Infancy after Preterm Delivery // Journal of Pediatrics. – 2002. – Vol. 141. – P. 652–658.
19. Dezateux, C., Stocks, J., Dundas, I., Fletcher, M.E. Impaired Airway Function and Wheezing in Infancy: The Influence of Maternal Smoking and a Genetic Predisposition to Asthma // American Journal of Respiratory and Critical Care Medicine. – 1999. – Vol. 159. – P. 403–410.
20. Fanucchi, M.V. Pulmonary Developmental Responses to Toxicants // Comprehensive Toxicology: Toxicology of the Respiratory System. Oxford, Pergamon Press. – 1997. – P. 203–220.
21. Smiley-Jewell, S.M., Liu, F.J., Weir, A.J., Plopper, C.G. Acute Injury to Differentiating Clara Cells in Neonatal Rabbits Results in Age-Related Failure of Bronchiolar Epithelial // Toxicologic Pathology. – 2000. – Vol. 28. – P. 267–276.
22. МР 2.1.10.0057-12.2.1.10. Состояние здоровья населения в связи с состоянием окружающей среды и условиями проживания населения. Оценка риска и ущерба от климатических изменений, влияющих на повышение уровня заболеваемости и смертности в группах населения повышенного риска. утв. 17.01.2012. – 36 с. [MR 2.1.10.0057-12.2.1.10. Sostoyanie zdorov'ya naseleniya v svyazi s sostoyaniem okruzhayushchei sredy i usloviyami prozhivaniya naseleniya. Otsenka riska i ushcherba ot klimaticheskikh izmenenii, vliyayushchikh na povyshenie urovnya zabolevaemosti i smertnosti v gruppakh naseleniya povyshennogo riska. utv. 17.01.2012. – 36 p. (In Russian)]

23. Tang, G., Zhao, P., Wang, Y., Gao, W. Mortality and Air Pollution in Beijing: The Long-Term Relationship // *Atmospheric Environment*. – 2017. – No. 150. – P. 238–243.
24. Leone, M., D'Ippoliti, D., De Sario, M., Analitis, A., et al. A Time Series Study on the Effects of Heat on Mortality and Evaluation of Heterogeneity into European and Eastern-Southern Mediterranean Cities: Results of EU CIRCE Project // *Environmental Health*. – 2013. – No. 12:55. – P. 1–12.
25. D'Ippoliti, D., Michelozzi, P., Marino, C., de' Donato, F., et al. The Impact of Heat Waves on Mortality in 9 European Cities: Results from the EuroHEAT Project // *Environmental Health*. – 2010. – No. 9:37. – P. 1–9.
26. Almeida, S.P., Casimiro, E., Calheiros, J.M. Effects of Apparent Temperature on Daily Mortality in Lisbon and Oporto, Portugal // *Environmental Health*. – 2010. – No. 9:12. – P. 1–7.
27. Heo, S., Lee, E., Kwon, B.Y., et al. Long-Term Changes in the Heat-Mortality Relationship According to Heterogeneous Regional Climate: A Time-Series Study in South Korea // *BMJ Open*. – 2016. – No. 6. – P. 1–10.
28. Eitan, O., Yuval, Barchana, M., et al. Spatial Analysis of Air Pollution and Cancer Incidence Rates in Haifa Bay, Israel // *Science of the Total Environment*. – 2010. – No. 408. – P. 4429–4439.
29. Frome, E.L. The Analysis of Rates Using Poisson Regression Models // *Biometrics*. – 1983. – No. 39. – P. 665–674.
30. Ревич Б.А., Шапошников Д.А., Семутникова Е.Г. Климатические условия и качество атмосферного воздуха как факторы риска смертности населения Москвы // *Медицина труда и промышленная экология*. – 2008. – № 7. – С. 29–35. [Revich, B.A., Shaposhnikov, D.A., Semutnikova, E.G. Klimaticheskie usloviya i kachestvo atmosfornogo vozdukha kak faktory riska smertnosti naseleniya Moskvy // *Meditsina truda i promyshlennaya ehkologiya*. – 2008. – No. 7. – P. 29–35. (In Russian)]
31. URL: <http://meteo.ru/>
32. URL: https://rp5.ru/Погода_в_мире
33. Басовский Л.Е. История и методология экономической науки: учебное пособие для студентов высших учебных заведений, обучающихся по направлению 38.04.01 «Экономика» (магистратура) и экономическим специальностям. – М.: ИНФРА-М, 2017. – 212 с. [Basovskii, L.E. Istoriya i metodologiya ehkonomicheskoi nauki : uchebnoe posobie dlya studentov vysshikh uchebnykh zavedenii, obuchayushchikhsya po napravleniyu 38.04.01 «Ehkonomika» (magistratura) i ehkonomicheskim spetsial'nostyam. – М.: INFRA-M, 2017. – 212 p. (In Russian)]
34. Guangyong, Z. A Modified Poisson Regression Approach to Prospective Studies with Binary Data // *American Journal of Epidemiology*. – 2004. – Vol. 159, no. 7. – P. 702–706. – DOI: 10.1093/aje/kwh090.
35. Милыев В.А., Котельников С.Н. Ядовитый озон. Новая экологическая угроза для России // *Экология и жизнь*. – 2008. – № 2(75). – С. 52–56. [Milyaev, V.A., Kotel'nikov, S.N. Yadovityi Poisonous Ozone: Another Environmental Threat To Russia // *Ehkologiya i zhizn'*. – 2008. – No. 2(75). – P. 52–56. (In Russian)]
36. Zhang, J.(J.), Wei, Y., Fang, Zh. Ozone Pollution: A Major Health Hazard Worldwide // *Frontiers in Immunology*. – 2019. – 10:2518.
37. Balajee, K.L., Babu, S., Suliankatchi, R.A., Meena, S. Characteristics of the Ozone Pollution and Its Health Effects in India // *International Journal of Medicine and Public Health*. – 2017. – Vol. 7(1). – P. 56–60.
38. URL: <https://cfpub.epa.gov/ncea/iris/search/index.cfm?keyword=ozone>.
39. Наищенко Н.И. Лекции по эконометрике: учебное пособие для студентов высших учебных заведений, обучающихся по специальности «Прикладная информатика (в экономике)». – У.: УлГТУ, 2008. – 139 с. [Nashchenko, N.I. Lektsii po ehkonometrike : uchebnoe posobie dlya studentov vysshikh uchebnykh zavedenii, obuchayushchikhsya po spetsial'nosti «Prikladnaya informatika (v ehkonomike)». – U.: ULGTU, 2008. – 139 p. (In Russian)]

Статья представлена к публикации членом редколлегии
А.И. Михальским.

Поступила в редакцию 17.06.2021,
после доработки 19.07.2021.
Принята к публикации 24.08.2021.

Тасейко Ольга Викторовна – канд. физ.-мат. наук, СибГУ имени академика М.Ф. Решетнева; Красноярский филиал ФИЦ ИВТ, г. Красноярск, ✉ taseiko@gmail.com,

Черных Дарья Александровна – мл. науч. сотрудник, Красноярский филиал ФИЦ ИВТ; СибГУ имени академика М.Ф. Решетнева, ✉ dachernykh93@gmail.com.



ESTIMATING THE INFLUENCE OF ENVIRONMENTAL FACTORS ON MORTALITY IN ELDER AGE GROUPS: AN EXAMPLE OF KRASNOYARSK

O.V. Taseiko^{1,2} and D.A. Chernykh^{1,2}

¹Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russia,

²Siberian Branch, Russian Academy of Sciences, Krasnoyarsk, Russia

✉ taseiko@gmail.com, ✉ dachernykh93@gmail.com

Abstract. This paper analyzes and forecasts the influence of climatic parameters and air quality on the mortality of the Krasnoyarsk industrial agglomeration. The long-term climatic data for the city of Krasnoyarsk are studied. The diseases sensitive to climate change are considered for the period from 2011 to 2014. The relationship between climatic parameters, air pollution, and the number of deaths is established using multivariate statistical analysis. The factors with the greatest contribution to mortality for climate-dependent diseases are identified. The mortality causes associated with negative environmental factors are examined. The age groups most affected by climatic parameters and air pollution are determined. As shown below, the nonlinear Poisson regression model predicts population mortality quite close to the factual data.

Keywords: air quality, climatic parameters, generalized linear model with Poisson regression, population mortality, social-natural-technogenic (S-N-T) system.

Funding. The reported study was funded by Russian Foundation for Basic Research, Government of Krasnoyarsk Territory, Krasnoyarsk Regional Fund of Science, project number 19-413-240013 “Risk assessment methodology caused by environmental factors on population health and mortality in industrial agglomerations”.

НЕБЛОКИРУЕМЫЕ ОТКАЗОУСТОЙЧИВЫЕ ДУАЛЬНЫЕ ФОТОННЫЕ КОММУТАТОРЫ ШИРОКОЙ МАСШТАБИРУЕМОСТИ

В.С. Подлазов

Аннотация. Продолжена работа по построению принципиально нового класса системных сетей – дуальных фотонных сетей, которые являются неблокируемыми сетями со статической самомаршрутизацией, обладают широкой масштабируемостью при максимальном достижимом на них быстродействии и сложностью, меньшей сложности полного коммутатора, а также допускают балансировку соотношений масштабируемость/быстродействие и сложность/быстродействие. Наконец, такие сети выполнены в расширенном схемном базисе, состоящем из дуальных фотонных коммутаторов и отдельных фотонных мультиплексоров и демultipлексоров. Предложен метод построения отказоустойчивой дуальной сети с указанными свойствами на основе сетей с топологией квазиполного графа и квазиполного орграфа и метода инвариантного расширения с внутренним распараллеливанием. Предложен метод расширения построенной ранее двухкаскадной дуальной сети в четырехкаскадную и восьмикаскадную дуальные сети с широким масштабированием при сохранении периода исходной сети и уменьшении ее показательной сложности.

Ключевые слова: фотонный коммутатор, дуальный коммутатор, фотонные мультиплексоры и демultipлексоры, многокаскадный коммутатор, бесконфликтная самомаршрутизация, неблокируемый коммутатор, статическая самомаршрутизация, квазиполный орграф, квазиполный граф, инвариантное расширение сетей, коммутационные свойства, прямые каналы, масштабируемость и быстродействие.

ВВЕДЕНИЕ

Данная работа развивает методику построения принципиально нового класса системных сетей [1–6] – дуальных фотонных сетей, которые являются неблокируемыми сетями со статической самомаршрутизацией [1–3, 5] и могут иметь заданную степень канальной отказоустойчивости [6].

В работе предложена методика построения неблокируемых самомаршрутизируемых фотонных сетей широкой масштабируемости. Это так называемые дуальные сети, в основе которых лежит неблокируемый дуальный коммутатор $p \times p$ с периодом разрядов в p длительностей сигнала (тактов) [1–3]. Дуальный коммутатор совмещает шинный способ разрешения конфликтов сигналов (разведение по разным тактам в одном канале) и коммутаторный способ (разведение по разным каналам). Дуальный коммутатор является неблокируемым коммутатором на любом входном трафике, если

информационные разряды передаются с периодом в p тактов. Дуальный коммутатор был разработан коллегами автора [1, 2], а затем применен и именован в совместных работах [3–5]. Фактически его использование оказалось необходимым условием построения неблокируемых сетей широкой масштабируемости и приемлемой сложности.

В дуальном фотонном коммутаторе применяется способ передачи сигнальной и управляющей информации параллельно на разных частотах для каждого разряда данных. Этот способ позволяет снять проблему синхронизации сигналов разных каналов.

Фотонная специфичность рассматриваемых сетей состоит в использовании внутриразрядной виртуализации каналов посредством обратных связей через линии задержки длительностью в один такт и, кроме того, в использовании управляющих сигналов на разных частотах для выполнения маршрутизации отдельных разрядов. Разведение информационных сигналов по разным тактам со-



проводится разведением по тем же тактам и сопровождающей управляющей информацией. По ней осуществляется маршрутизация разрядов посредством перемещения их между разными каналами без изменения установившихся номеров тактов.

По всей статье термин «дуальный коммутатор» или «дуальная сеть, построенная с использованием дуальных коммутаторов», подразумевает использование в них разрядов с периодом в p тактов. Эти разряды обеспечивают неблокируемость сети в ее первом каскаде, а в остальных каскадах остаются по «инерции», уже без применения шинного способа разрешения конфликтов.

Масштабируемость дуальных сетей осуществляется с использованием сетей с топологией квазиполных графа или орграфа [4], которые выполняются в расширенном схемном базисе, состоящем из дуальных фотонных коммутаторов и отдельных фотонных мультиплексоров и демultipлексоров. В работах [1–3, 5] широкая масштабируемость достигается посредством применения метода инвариантного расширения сетей с большим числом дополнительных демultipлексоров и мультиплексоров.

В статьях [5, 6] был разработан и впервые применен новый метод расширения дуальных сетей посредством их внутреннего распараллеливания без использования дополнительных устройств. В частности, в работе [5] была построена двухкаскадная неблокируемая сеть, которая состоит в каждом каскаде из сетей с топологией квазиполного орграфа, имеющих $N = p^2$ каналов, тогда как двухкаскадная неблокируемая сеть имеет N^2 каналов. С другой стороны, в работе [6] была построена двухкаскадная неблокируемая сеть с $(\sigma - 1)$ -канальной отказоустойчивостью, которая состоит в каждом каскаде из сетей с топологией квазиполного графа, имеющих $N = p(p - 1)/\sigma + 1$ каналов, а двухкаскадная отказоустойчивая неблокируемая сеть имеет N^2 каналов.

В настоящей работе строятся четырех- и восьмикаскадные отказоустойчивые сети посредством развития и применения обобщенного метода внутреннего распараллеливания. При этом достигается такая же степень масштабирования сетей, как и в инвариантном методе с использованием внешних демultipлексоров и мультиплексоров, но без их применения, и строятся сети с существенно меньшей сложностью.

В § 1 дается обзор наличия свойств неблокируемости и канальной отказоустойчивости у ряда современных системных сетей. В § 2 вслед за статьями [6] вводятся понятия прямой и обратной p -перестановки, которые существенно используются при доказательстве неблокируемости четырех- и восьмикаскадных сетей, и повторяются доказа-

тельства неблокируемости двухкаскадных сетей. В § 3 данной части строятся четырехкаскадные неблокируемые самомаршрутизируемые коммутаторы с одноканальной и двухканальной отказоустойчивостью и оцениваются их характеристики. При этом метод внутреннего распараллеливания из работ [5, 6] обобщается для четырехкаскадных коммутаторов.

В § 4 сравниваются характеристики четырехкаскадных неблокируемых самомаршрутизируемых коммутаторов на базе коммутаторов с топологией дуальных квазиполных графов и орграфов. Наконец, в § 5 строятся восьмикаскадные неблокируемые самомаршрутизируемые коммутаторы на базе коммутаторов с топологией дуальных квазиполных графов и орграфов. При этом метод из § 3 обобщается для восьмикаскадных коммутаторов. В § 6 обсуждаются свойства построенных сетей в сравнении с другими неблокируемыми сетями и недостатки построенных сетей, а также намечаются способы преодоления этих недостатков.

В заключении дается разбор сути обобщенного метода внутреннего распараллеливания, который представляет собой ядро методики построения дуальных неблокируемых сетей широкой масштабируемости и невысокой удельной сложности. Здесь выделяются три основных компонента предложенной методики – неблокируемый дуальный коммутатор, коммутатор с топологией квазиполного графа или орграфа на базе дуального коммутатора и метод внутреннего распараллеливания.

1. НЕБЛОКИРУЕМОСТЬ И ОТКАЗОУСТОЙЧИВОСТЬ В СИСТЕМНЫХ СЕТЯХ

Задача построения неблокируемых отказоустойчивых системных сетей суперкомпьютеров до настоящего времени не имеет полного решения.

Системная сеть является неблокируемой, если в ней для любой перестановки пакетов можно проложить бесконфликтные пути от источников к приемникам. Системная сеть является самомаршрутизируемой, если бесконфликтные пути можно проложить локально по узлам сети без их взаимодействия только на основе маршрутной информации в пакетах. Наконец, самомаршрутизация является статической, если любой источник может самостоятельно наметить бесконфликтные пути к своему приемнику без взаимодействия с другими источниками.

Существование неблокируемых сетей было доказано еще Клозом [7, 8]. Пока еще не построены процедуры самомаршрутизации неблокируемых сетей Клоза, которые, однако, могут быть мерилем качества других неблокируемых сетей.

Неблокируемой является сеть в виде двумерного обобщенного гиперкуба с топологией квазиполного орграфа, например, в коммутаторах YARK и ROSETTA, использованных в ряде сетей разной структуры: перестраиваемой сети Клоза [9], трехмерного тора [10], иерархии полных и квазиполных орграфов [11–13]. К сожалению, квазиполный орграф имеет малое число каналов $N = p^2$, где p – это степень внутренних коммутаторов, и большую коммутационную сложность $S \geq N^2$, которая больше сложности полного орграфа и заведомо больше сложности неблокируемой сети Клоза.

В настоящее время в литературе широко представлены системные сети со структурой толстого дерева (в частности, перестраиваемые сети Клоза), со структурой обобщенного гиперкуба, со структурой многомерного тора и сети с иерархией полных и квазиполных орграфов.

Сети со структурой толстого дерева являются перестраиваемыми сетями [9, 14, 15], в которых бесконфликтная передача осуществляется только по заранее составленным расписаниям для конкретных перестановок пакетов. Для произвольных перестановок эти сети оказываются блокируемыми, в них произвольная перестановка полностью осуществляется за несколько скачков между узлами сети. Максимальное число таких скачков задает диаметр сети. В перестраиваемых сетях Клоза диаметр равен числу каскадов сети.

Сети со структурой обобщенного гиперкуба [16–19] не являются даже перестраиваемыми сетями [20, 21]. Их можно сделать перестраиваемыми посредством увеличения числа каналов в некоторых измерениях. Обобщенные кубы имеют диаметр, равный числу измерений или на единицу меньше в расширенном гиперкубе [17, 18]. Обобщенные гиперкубы с удвоенным числом каналов в каждом измерении являются перестраиваемыми сетями сразу для двух перестановок. Отметим, что попытка использования обобщенного гиперкуба в качестве неблокируемой сети для фотонного компьютера [22] представляется очень сомнительной затеей.

Сети со структурой многомерного тора для произвольных перестановок вообще не имеют возможности передавать пакеты по прямым каналам [11, 23–25]. В них произвольные перестановки осуществляются только за несколько скачков между узлами сети. Многомерные торы являются самыми простыми, но и самыми медленными сетями вследствие их больших диаметров. Так, в сетях, рассмотренных в работах [11, 23–25], диаметр измеряется десятками скачков.

Наоборот, сети со структурой иерархии полных или квазиполных орграфов [10, 12, 26] имеют са-

мый маленький диаметр в три скачка. Сетей с малыми диаметрами в последнее время появилось довольно много [27–32]. Все они имеют серьезные проблемы с выравниванием загрузки сети при отказах каналов.

Канальной отказоустойчивостью считается возможность сохранения полнодоступности сети при отказах каналов с сохранением ее исходных характеристик (неблокируемость сети, задержки передачи или диаметра сети).

В чистом виде канальной отказоустойчивостью обладают, по-видимому, только сети с топологией квазиполных графов. Эти сети изоморфны такому математическому объекту, как неполная уравновешенная симметричная блок-схема [33]. Эти сети имеют элементную базу из коммутаторов $p \times p$, демультиплексоров $1 \times p$ и мультиплексоров $p \times 1$ и являются неблокируемыми сетями со статической самомаршрутизацией. Они имеют прямые каналы между $N = p(p-1)/\sigma + 1$ абонентами сети и имеют σ разных каналов между любыми двумя абонентами [4].

В других сетях восстановление полнодоступности сети при отказах каналов сопровождается так или иначе увеличением задержек передачи по сети. Так, в перестраиваемой сети Клоза при отказах каналов нагрузка на оставшиеся каналы увеличивается, что, в свою очередь, увеличивает число конфликтов и задержки передачи части пакетов.

Сеть TOFY со структурой трехмерного тора [25] использует еще три измерения для создания резервных каналов. В ней при отказе части колец сети их целостность восстанавливается с увеличением диаметра сети на единицу.

Обобщенные гиперкубы с удвоенным числом каналов в каждом измерении являются 1-отказоустойчивыми сетями с неизменным диаметром [19].

В сетях с иерархией полных или квазиполных орграфов [10, 12, 26] при отказе части каналов полнодоступность сети восстанавливается путем использования обходных путей с длительностью в пять скачков, т. е. с увеличением задержки передачи в 5/3 раза.

2. ДУАЛЬНЫЙ КВАЗИПОЛНЫЙ ГРАФ, А ПЕРЕСТАНОВКИ И ДУАЛЬНЫЙ ДВУХКАСКАДНЫЙ КОММУТАТОР

Дуальный коммутатор KPN_1 с топологией квазиполного графа $KPG(N_1, p, \sigma)$ состоит из $N_1 = p(p-1)/\sigma + 1$ дуальных коммутаторов $p \times p$, обозначаемых как ДКр, N_1 входных демультиплексоров $1 \times p$ и N_1 выходных мультиплексоров $p \times 1$ [6]. Схема их межсоединений строится комбинатор-

Таблица 2

Характеристики дуальных коммутаторов КПМ₁ на основе квазиполного орграфа

p	N_1	T_1	S_1	L_1
2	4	2	$48 = N_1^{2,79}$	$16 = N_1^2$
4	16	4	$640 = N_1^{2,33}$	$128 = N_1^{1,75}$
6	36	6	$3\,024 = N_1^{2,24}$	$432 = N_1^{1,69}$
8	64	8	$9\,216 = N_1^{2,19}$	$1\,024 = N_1^{1,67}$

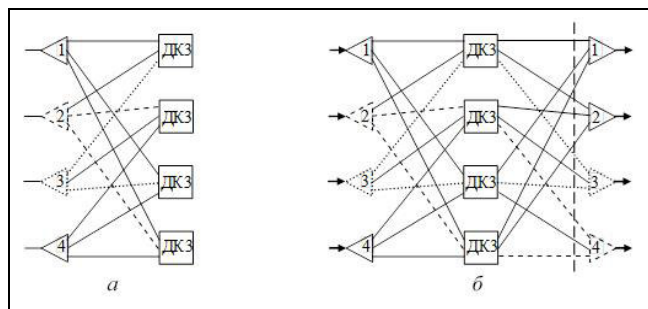


Рис. 1. Дуальный с периодом разрядов в три такта квазиполный коммутатор КП4 в виде графа КПП(4, 3, 2): а – в исходном виде с дуплексными каналами, б – в применяемом виде с симплексными каналами; пунктирами и точками изображены разные пути между выбранными входами и выходами

Таблица 1

Характеристики дуальных коммутаторов КПМ₁ с одноканальной отказоустойчивостью

p	N_1	T_1	S_1	L_1
2	2	2	$24 = N_1^{4,58}$	$8 = N_1^3$
4	7	4	$280 = N_1^{2,9}$	$56 = N_1^{2,07}$
6	15	6	$1\,260 = N_1^{2,64}$	$180 = N_1^{1,92}$
8	27	8	$3\,888 = N_1^{2,51}$	$432 = N_1^{1,84}$

Любой дуальный коммутатор КПМ₁ имеет тот же период разрядов T_1 , что и входящий в него дуальный коммутатор ДКр. Для коммутатора КПМ₁ рассчитываются такие характеристики, как коммутационная сложность S_1 , выраженная в числе точек коммутации, и канальная сложность, выраженная в числе каналов. Они выражаются в показательном виде, через число каналов, и именуются в статье показательными сложностями¹ (см. табл. 1).

Для дуального коммутатора КПМ₁ с топологией квазиполного орграфа [5] характеристики задаются табл. 2, которые лучше чем в табл. 1, но без обеспечения канальной отказоустойчивости.

Введем понятие p -разбиения пакетов, передаваемых через заданное сечение сети на входах в мультиплексоры. Все пакеты разделяются на группы переменного состава, содержащие не более p

пакетов в каждой. При обычной перестановке пакетов на входе и выходе коммутатора имеет место 1-разбиение. Назовем p -перестановкой вариант передачи, при котором на входе сети имеет место 1-разбиение, а на заданном сечении – p -разбиение.

Для дуального коммутатора КПМ₁ с топологией дуального квазиполного графа КПП(N_1, p, σ) указанное сечение проводится через входы мультиплексоров и называется выходным сечением. На рис. 1 выходное сечение обозначено вертикальным пунктиром. По свойству дуального коммутатора ДКр на выходном сечении при любом трафике имеет место p -разбиение. Имеет место

Лемма 1 [6]. Дуальный коммутатор КПМ₁ с периодом разрядов p тактов является неблокируемым коммутатором со статической самомаршрутизацией на любой обычной перестановке и обладает при этом $(\sigma - 1)$ -канальной отказоустойчивостью.

В работах [5, 6] была разработана методика построения двухкаскадных неблокируемых коммутаторов с N_1^2 каналами. Здесь она излагается на примере дуального коммутатора КП2 с топологией дуального квазиполного графа КПП(2, 2, 2), представленного на рис. 2.

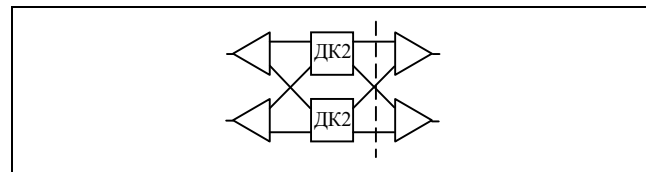


Рис. 2. Дуальный неблокируемый коммутатор КП2 с одноканальной отказоустойчивостью

На его основе строится четырехканальная двухкаскадная сеть C_24 , которая содержит по два коммутатора КП2 в каждом каскаде, соединенных обменными связями (рис. 3). Эта сеть является блокируемой на мультиплексорах первого каскада, выделенных заливкой, и на них же утрачивает канальную отказоустойчивость.

Сеть C_24 может быть превращена в неблокируемый коммутатор K_24 с одноканальной отказоустойчивостью посредством метода внутреннего

¹ Термин автора.

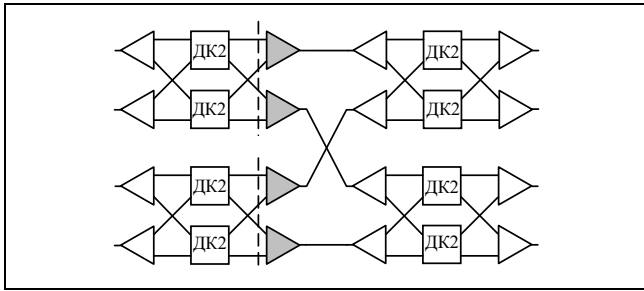


Рис. 3. Дуальная двухкаскадная блокируемая сеть C_24 с обменными связями

распараллеливания [5, 6]. В нем во втором каскаде образуются две копии второго каскада сети C_24 . В первом каскаде мультиплексоры на разрезах удаляются, а их входы подсоединяются к входам копий второго каскада: нечетные – к первой копии, а четные – ко второй копии. Эти соединения сохраняют порядок подсоединения тех каналов, которые в сети C_24 шли ко второму каскаду. Сами вырезанные мультиплексоры перемещаются для объединения выходов копий второго каскада, образуя выходные мультиплексоры коммутатора K_24 (рис. 4), который оказывается неблокируемым коммутатором со статической самомаршрутизацией и одноканальной отказоустойчивостью.

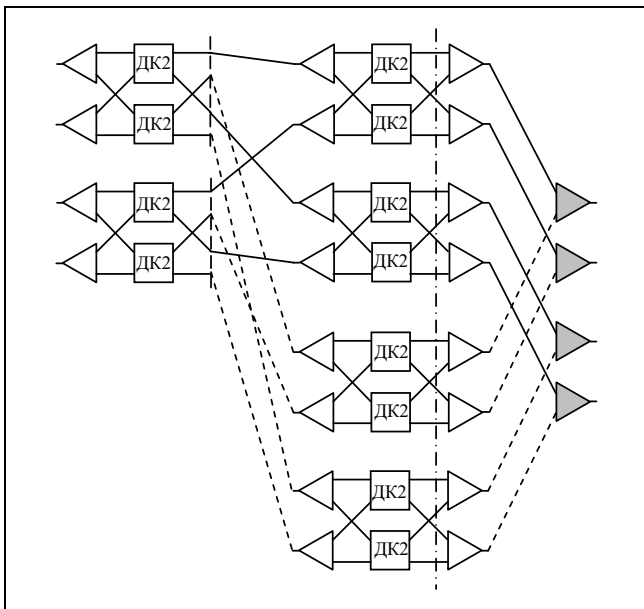


Рис. 4. Дуальный неблокируемый самомаршрутизируемый коммутатор K_24 с одноканальной отказоустойчивостью

В общем случае ($p \geq 2$) дуальный коммутатор KPN_1 имеет топологию дуального квазиполного графа $KPN(N_1, p, \sigma)$ с периодом разрядов в p тактов. На его основе строится двухкаскадная блокируемая сеть C_2N_2 ($N_2 = N_1^2$), в каждом каскаде ко-

торой содержится N_1 коммутаторов KPN_1 с обменными связями между каскадами. Для применения метода внутреннего распараллеливания образуется p копий второго каскада сети C_2N_2 , а мультиплексоры первого каскада используются для объединения одноименных выходов копий второго каскада.

На разрезах первого каскада имеется pN_2 входов в мультиплексоры. Они перенумеровываются сверху вниз как I ($1 \leq I \leq pN_2$), и входы с номерами $i = I(\bmod p) + 1$ подсоединяются к одноименным входам i -й копии второго каскада, сохраняя при этом порядок размещения по коммутаторам K_1N_1 , как в сети C_2N_2 .

Лемма 2. Дуальный коммутатор K_2N_2 имеет p -перестановку на заданном сечении. Он является неблокируемым коммутатором со статической маршрутизацией на любой обычной перестановке и обладает $(\sigma - 1)$ -канальной отказоустойчивостью.

Коммутационная сложность коммутатора K_2N_2 задается по построению по рекуррентным формулам как $S_2 = N_1S_1 + pN_1S_1$, а канальная – как $L_2 = N_1L_1 + pN_1L_1$. Численные значения характеристик коммутаторов K_2N_2 для $\sigma = 2$ представлены в табл. 3. Отметим, что имеет место уменьшение значений показательных сложностей по сравнению с табл. 1.

Таблица 3

Характеристики дуальных коммутаторов K_2N_2 с одноканальной отказоустойчивостью

p	N_1	$N_2 = N_1^2$	$T_2 = p$	S_2	L_2
2	2	4	2	$N_2^{3,58}$	$N_2^{2,9}$
4	7	49	4	$N_2^{2,37}$	$N_2^{1,97}$
6	15	225	6	$N_2^{2,18}$	$N_2^{1,84}$
8	27	729	8	$N_2^{2,09}$	$N_2^{1,77}$

При построении дуального коммутатора K_2N_2 на базе коммутатора с топологией квазиполного орграфа [5] его характеристики задаются табл. 4, которые существенно лучше, чем в табл. 2, но без обеспечения канальной отказоустойчивости.

Таблица 4

Характеристики коммутаторов K_2N_2 на основе квазиполного орграфа

p	N_1	$N_2 = N_1^2$	$T_2 = p$	S_2	L_2
2	4	16	2	$N_2^{2,29}$	$N_2^{1,95}$
4	16	256	4	$N_2^{1,96}$	$N_2^{1,68}$
6	36	1 296	6	$N_2^{1,89}$	$N_2^{1,63}$
8	64	4 096	8	$N_2^{1,86}$	$N_2^{1,6}$

Отметим также, что дуальный коммутатор K_2N_2 имеет два каскада выходных мультиплексоров, содержащих pN_2 и N_2 мультиплексоров соответственно. Для целей следующего раздела в коммутаторе K_2N_2 проводится новый штрихпунктирный разрез (см. рис. 4), проходящий через входы первого каскада мультиплексоров.

3. ЧЕТЫРЕХКАСКАДНЫЙ ОТКАЗОУСТОЙЧИВЫЙ НЕБЛОКИРУЕМЫЙ САМОМАРШРУТИЗИРУЕМЫЙ КОММУТАТОР С ДВУМЕРНЫМ ВНУТРЕННИМ РАСПАРАЛЛЕЛИВАНИЕМ

В работах [1–3, 5] дальнейшее увеличение числа каналов неблокируемого коммутатора осуществлялось с помощью метода инвариантного расширения сетей с внешним распараллеливанием, при котором не происходит изменения периода сигналов. Однако он оказался малоприменим для отказоустойчивых неблокируемых коммутаторов [6].

В данном разделе увеличение числа каналов без изменения периода сигналов осуществляется с помощью обобщенного метода внутреннего распараллеливания сети посредством построения четырехкаскадных коммутаторов K_4N_4 из двухкаскадных коммутаторов K_2N_2 с числом каналов $N_4 = N_2^2$ и периодом разрядов $T_4 = T_2 = p$.

Построение сети осуществляется на примере использования коммутаторов K_24 (см. рис. 4). Сначала создается двухкаскадная сеть C_416 . Каждый ее каскад состоит из четырех копий коммутатора K_24 , а каскады соединены обменными связями (рис. 5).

В реальности сеть C_416 состоит из четырех каскадов КПП(2, 2, 2), что задает нижний индекс в ее обозначении. Сеть C_416 является блокируемой вследствие возможности возникновения конфликтов сигналов в двух каскадах выходных мультиплексоров М2, обозначенных заливкой. Имеется два слоя этих мультиплексоров общим числом $W_4 = 48$. Кроме того, в них нарушается и канальная отказоустойчивость. Через входы в первый каскад проводится штрихпунктирный разрез (см. рис. 5), для которого в предыдущем разделе сформулировано понятие p -перестановки.

Затем создается сеть $\underline{C}_416$, которая содержит первый каскад сети C_416 и две копии второго каскада сети C_416 . В сети $\underline{C}_416$ создается одна параллельная схема первого измерения (рис. 6). Для этого сначала вырезается внешний слой со светлой заливкой общим числом $W_2^* = 16$ мультиплексоров М2. Они остаются пока без подсоединений. Затем

вырезаются мультиплексоры М2 внутреннего слоя с темной заливкой и их нечетные входы разводятся по входам двух копий второго каскада сети $\underline{C}_416$. При этом $W_{4,1} = 16$ вырезанных мультиплексоров объединяют выходы этих двух копий.

Оставшиеся $W_{4,2} = 16$ мультиплексоров М2 с темной заливкой используются для создания второй параллельной схемы первого измерения аналогичным образом (рис. 7). Для этого их четные входы разводятся по входам еще двух копий второго каскада сети $\underline{C}_416$.

Забегая вперед, отметим, что на рис. 6 и 7 обозначены и новые соединения мультиплексоров первого и второго слоев. Они задают объединение схем первого измерения в схему второго измерения.

В результате построены две схемы первого измерения, каждая из которых состоит из двух коммутаторов K_24 , включенных параллельно (рис. 8 без соединений второй схемы первого измерения). Две схемы первого измерения образуют двумерную схему. Выходы двумерной схемы объединяют $W_2^* = 16$ мультиплексоров со светлой заливкой, образуя выходы коммутатора K_416 . На рис. 8 последние соединения представлены точечными линиями и полностью только в одном экземпляре из-за недостатка площади рисунка (они есть на рис. 6 и 7).

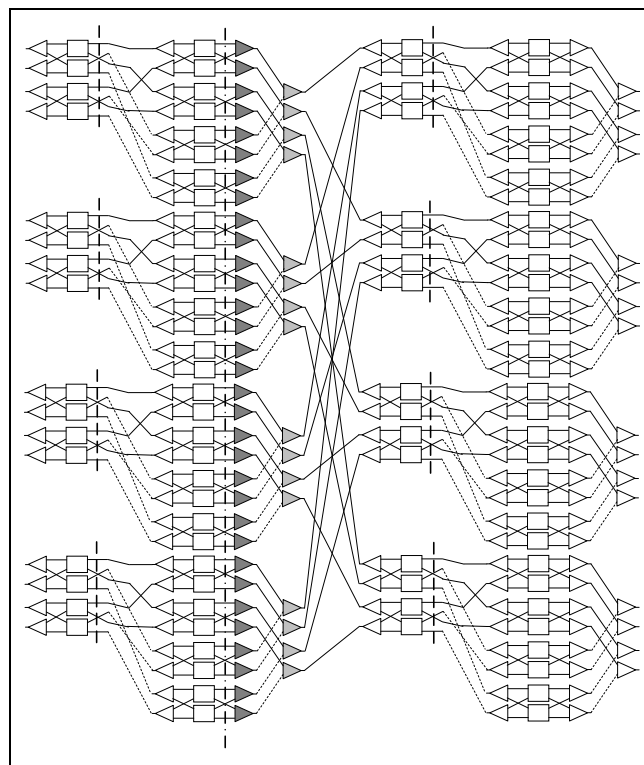


Рис. 5. Блокируемая дуальная сеть C_416 без канальной отказоустойчивости

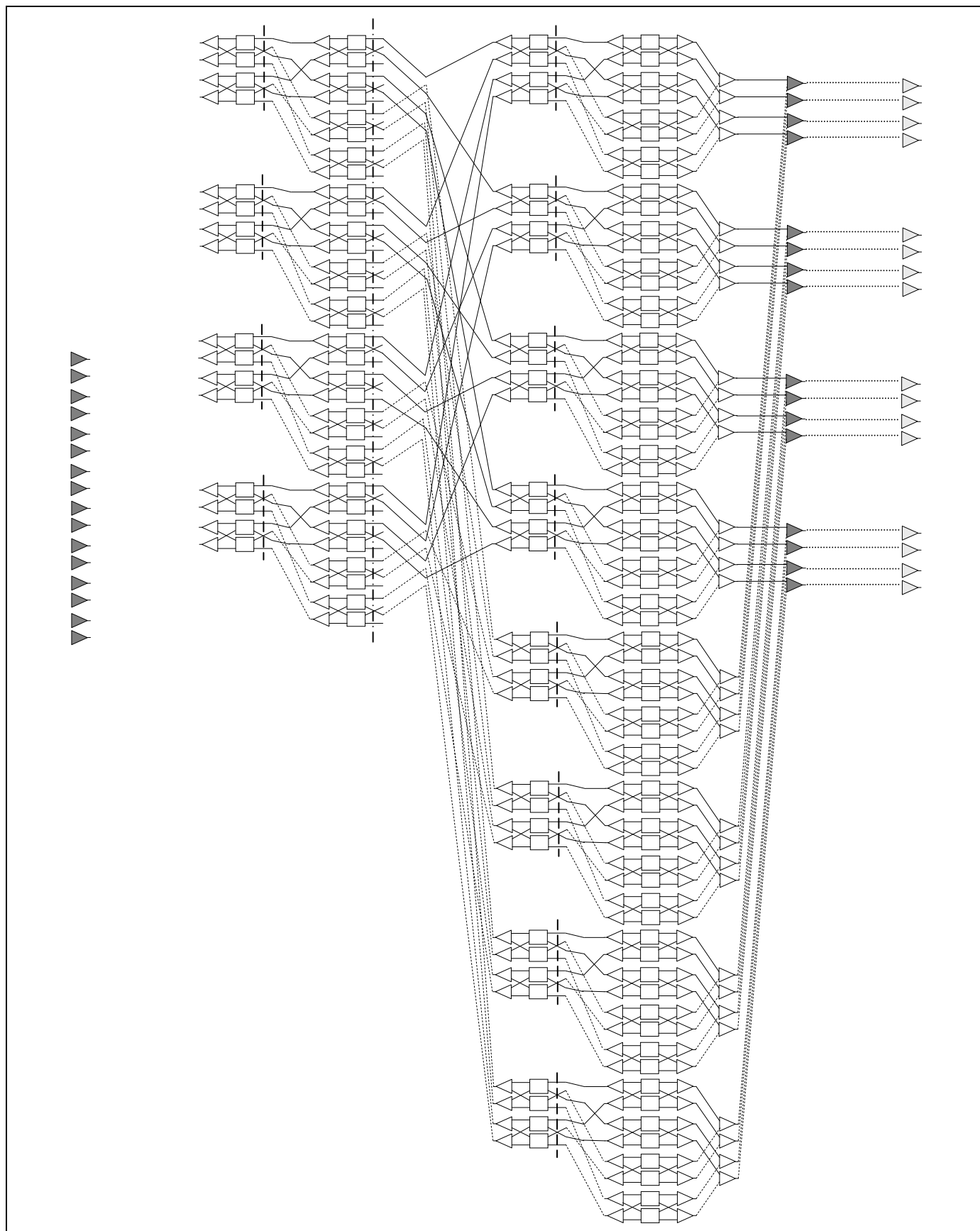


Рис. 6. Построение первой схемы первого измерения. Неиспользованные в этой схеме мультиплексоры сети С416 приведены слева

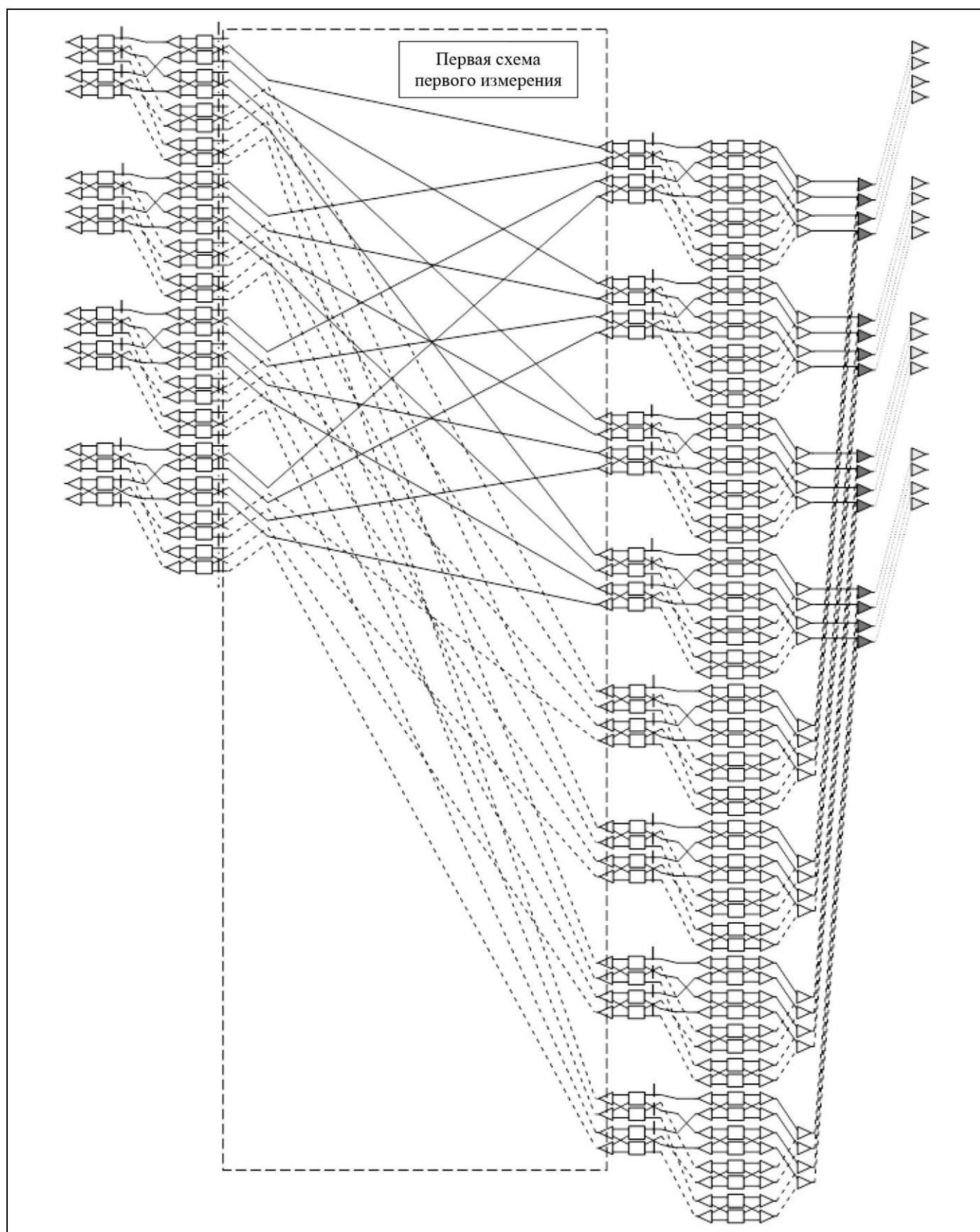


Рис. 7. Построение второй схемы первого измерения посредством использования мультиплексоров, которые не использовались при построении первой схемы первого измерения

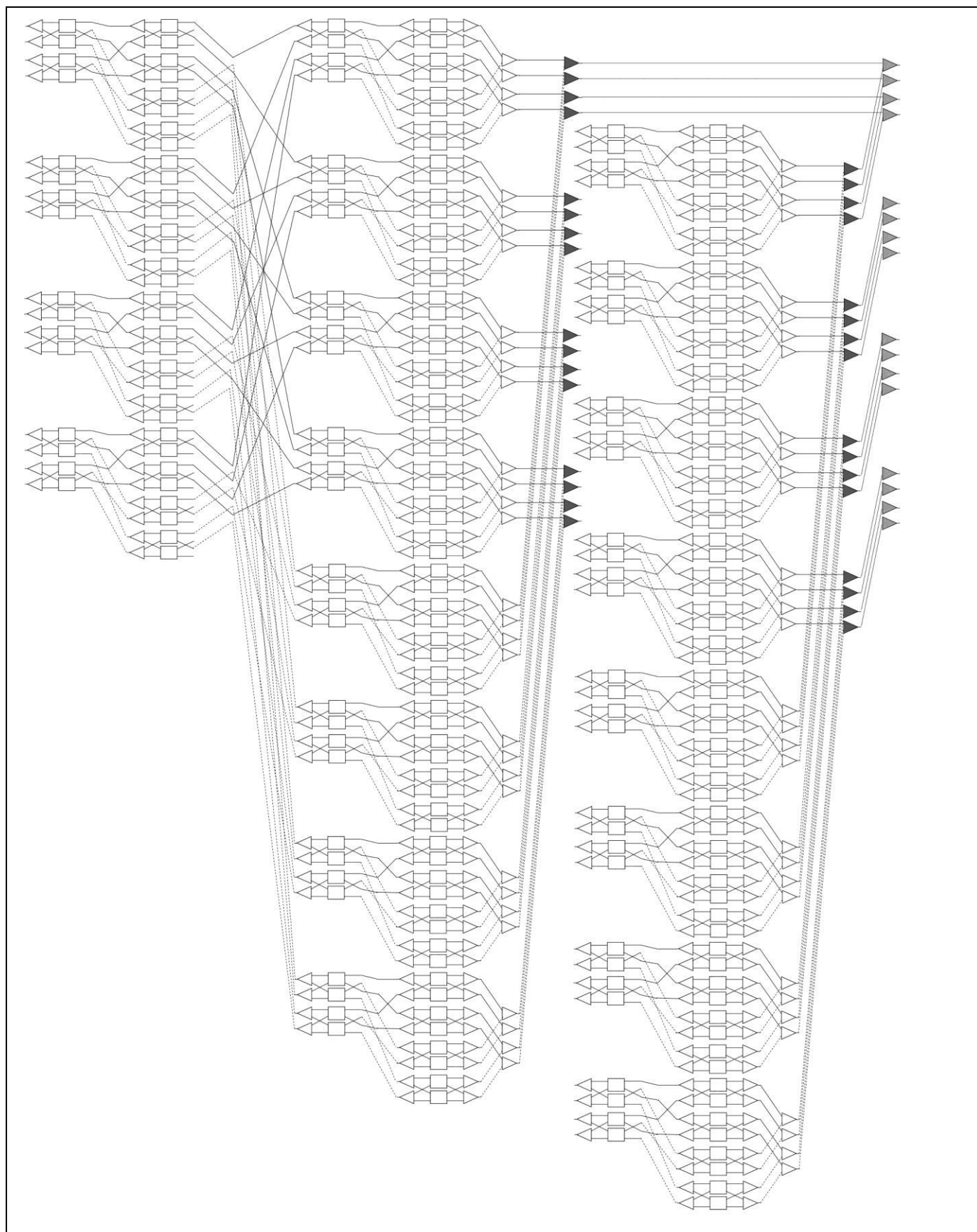


Рис. 8. Неблокируемый дуальный коммутатор K416 с одноканальной отказоустойчивостью



Построенная таким образом 16-канальная сеть состоит из 16 копий коммутатора K_24 , включенных параллельно. На их входы поступают разреженные альтернативные (не пересекающиеся ни по входам, ни по выходам) прямые p -перестановки, которые по лемме 2 реализуются бесконфликтно по единому статическому расписанию. Пути между источниками и приемниками в коммутаторах K_24 проходят по двум путям через разные схемы первого измерения. Поэтому в коммутаторе K_24 имеет место одноканальная отказоустойчивость, так как $p = \sigma = 2$.

В общем случае ($p > 2$ и $\sigma \geq 2$) сначала строится сеть C_4N_4 , которая состоит из двух каскадов по N_2 коммутаторов K_2N_2 в каждом, соединенных обменными связями. Она имеет $N_4 = N_2^2$ каналов и допускает возможность конфликта сигналов на выходных мультиплексорах M_p . Поэтому она является блокируемой сетью и не обладает канальной отказоустойчивостью.

Первый каскад сети C_4N_4 имеет два слоя выходных мультиплексоров M_p общим числом $W_4 = N_2V_2 = N_4(p+1)$. Первый внутренний слой мультиплексоров M_p содержит $W_{4,1} = pN_4$ мультиплексоров M_p , которые в совокупности имеют p^2N_4 входов.

Затем создается сеть $\underline{C}_4N_4$, которая содержит первый каскад сети C_4N_4 и p^2 копий второго каскада сети C_4N_4 . Сеть $\underline{C}_4N_4$ создается с соблюдением следующей структуры: копии второго каскада сети C_4N_4 образуют p схем первого измерения, а все вместе они образуют схему второго измерения.

Сеть $\underline{C}_4N_4$ содержит p^2N_2 коммутаторов K_2N_2 , имеющих в совокупности p^2N_4 входов. Далее вырезаются оба слоя мультиплексоров M_p первого каскада сети C_4N_4 , а входы первого слоя мультиплексоров M_p подсоединяются к входам коммутаторов K_2N_2 . Это возможно, так как число этих входов совпадает.

Копии второго каскада сети C_4N_4 в сети $\underline{C}_4N_4$ числом p^2 разобьем на p групп и обозначим через G ($1 \leq G \leq p$) номер группы и через I ($1 \leq I \leq p$) – номер копии. Фактически G задает номер схемы первого измерения, а I – номер копии в схеме первого измерения.

Обозначим через J ($1 \leq J \leq N_2$) номер коммутатора K_2N_2 в схеме первого измерения. Обозначим через K ($1 \leq K \leq N_2$) номер входа каждого такого коммутатора K_2N_2 . Таким образом, вход любого коммутатора K_2N_2 задается составным индексом G, I, J, K .

Обозначим через i ($0 \leq i \leq p-1$) номер входа по mod p каждого мультиплексора M_p в первом их

слое. В каждом коммутаторе первого каскада сети C_4N_4 содержится pN_2 этих мультиплексоров M_p . Разобьем их на N_2 групп и обозначим через g ($1 \leq g \leq N_2$) номер группы, а через j ($1 \leq j \leq p$) номер мультиплексора M_p в группе. Обозначим через k ($1 \leq k \leq N_2$) номер коммутатора K_2N_2 в первом каскаде сети C_4N_4 . Таким образом, вход любого мультиплексора M_p в первом слое задается составным индексом i, g, j, k .

Произвольный вход мультиплексора M_p с индексом i, g, j, k переключается на вход коммутатора K_2N_2 с индексом G, I, J, K , в котором $G = i + 1$, $I = j$, $J = g$ и $K = k$. В результате образуется p^2N_2 параллельно включенных коммутаторов K_2N_2 , на входы которых поступают разреженные непересекающиеся прямые p -перестановки.

Вырезанные мультиплексоры M_p первого слоя объединяют выходы схем первого измерения. Вырезанные мультиплексоры M_p второго слоя объединяют выходы p схем первого измерения, образуя выходы схемы второго уровня, которые также являются выходами и коммутатора K_4N_4 .

Для коммутаторов K_4N_4 справедлива

Лемма 3. *Дуальный коммутатор K_4N_4 является неблокируемым коммутатором со статической маршрутизацией на любой обычной перестановке при любом p . Он обладает $(\sigma - 1)$ -канальной отказоустойчивостью.*

Доказательство. Первое утверждение опирается на использование коммутатора K_2N_2 и справедливость для него леммы 2. Второе утверждение опирается на неблокируемость коммутатора K_2N_2 и тот факт, что p -перестановка на сечении состоит из разреженных 1-перестановок, разведенных по разным каналам и тактам.

Согласно назначению $G = i + 1$, разные входы одного мультиплексора первого слоя подсоединяются к разным одномерным схемам (разным копиям второго каскада сети C_4N_4), а входы разных мультиплексоров – к входам разных коммутаторов K_2N_2 второго каскада сети C_4N_4 . Иначе говоря, имеется σ разных путей коммутаторов K_2N_2 по p разным путям в коммутаторе K_4N_4 , откуда следует его $(\sigma - 1)$ -канальная отказоустойчивость, так как $p \geq \sigma$. ♦

В результате неблокируемый самомаршрутизируемый коммутатор K_4N_4 с $N_4 = N_2^2$ каналами и $(\sigma - 1)$ -канальной отказоустойчивостью имеет характеристики, приведенные в табл. 5 и табл. 6. Они рассчитываются по рекуррентным формулам $S_4 = N_2S_2 + p^2N_2S_2$ и $L_4 = N_2L_2 + p^2N_2L_2$. Коммутатор K_4N_4 имеет четыре слоя выходных мультиплексоров M_p общим числом $V_4 = N_4(p^4 - 1)/(p - 1)$.

Отметим дальнейшее снижение показательных коммутационной и канальной сложностей комму-

татора K_4N_4 (см. табл. 4) по сравнению с коммутатором K_2N_2 (см. табл. 2).

Если отказаться от требования канальной отказоустойчивости, то на основе коммутатора с топологией квазиполного орграфа можно построить неблокируемый самомаршрутизируемый коммутатор K_4N_4 с характеристиками, заданными в табл. 7. Видно, что по сравнению с отказоустойчивыми вариантами он имеет большее число каналов и меньшую сложность. Кроме того, весь набор коммутаторов имеет коммутационную и канальную сложности меньше, чем у двухкаскадного коммутатора на базе коммутатора с топологией полного

графа (ср. с табл. 3) и меньше, чем у коммутатора с топологией полного графа (*switchboard*).

4. ВОСЬМИКАСКАДНЫЕ НЕБЛОКИРУЕМЫЕ КОММУТАТОРЫ С ЧЕТЫРЕХМЕРНЫМ ВНУТРЕННИМ РАСПАРАЛЛЕЛИВАНИЕМ НА ОСНОВЕ ГРАФА И ОРГРАФА

Метод расширения двухкаскадных коммутаторов K_2N_2 в четырехкаскадные коммутаторы K_4N_4 можно обобщить и на построение восьмикаскадных коммутаторов K_8N_8 из четырехкаскадных коммутаторов K_4N_4 .

Таблица 5

Характеристики дуальных коммутаторов K_4N_4 с одноканальной отказоустойчивостью

p	N_1	$N_4 = N_1^4$	$T_4 = p$	S_4	L_4
2	2	16	2	$2\,720 = N_4^{2,85}$	$1\,120 = N_4^{2,53}$
3	4	256	3	$238\,080 = N_4^{2,23}$	$69\,120 = N_4^{2,01}$
4	7	2\,401	4	$8\,000\,132 = N_4^{2,04}$	$1\,795\,948 = N_4^{1,85}$
5	11	14\,641	5	$1,35E+08 = N_4^{1,95}$	$24\,743\,290 = N_4^{1,77}$
6	15	65\,536	6	$1,41E+09 = N_4^{1,9}$	$2,18E+08 = N_4^{1,73}$
7	21	194\,481	7	$8,64E+09 = N_4^{1,88}$	$1,16E+09 = N_4^{1,71}$
8	27	531\,441	8	$4,45E+10 = N_4^{1,86}$	$5,25E+09 = N_4^{1,7}$

Таблица 6

Характеристики дуальных коммутаторов K_4N_4 с двухканальной отказоустойчивостью

p	N_1	$N_4 = N_1^4$	$T_4 = p$	S_4	L_4
3	3	81	3	$75\,330 = N_4^{2,56}$	$21\,870 = N_4^{2,27}$
4	5	625	4	$2\,082\,500 = N_4^{2,26}$	$467\,500 = N_4^{2,03}$
5	7	2\,401	5	$22\,161\,230 = N_4^{2,17}$	$4\,057\,690 = N_4^{1,95}$
6	11	14\,641	6	$3,15E+08 = N_4^{2,04}$	$48\,754\,530 = N_4^{1,85}$
7	15	50\,625	7	$2,25E+09 = N_4^{1,99}$	$3,01E+08 = N_4^{1,8}$
8	19	130\,321	8	$1,09E+10 = N_4^{1,96}$	$1,29E+09 = N_4^{1,78}$

Таблица 7

Характеристики дуальных коммутаторов K_4N_4 на основе орграфа

p	N_1	$N_2 = N_1^2 = p^4$	$N_4 = N_2^4 = p^8$	$T_4 = p$	S_4	L_4
2	4	16	256	2	$43\,520 = N_4^{1,93}$	$17\,920 = N_4^{1,77}$
3	9	81	6\,561	3	$6\,101\,730 = N_4^{1,78}$	$1\,771\,470 = N_4^{1,64}$
4	16	256	65\,536	4	$2,18E+08 = N_4^{1,73}$	$49\,020\,928 = N_4^{1,60}$
5	25	625	390\,625	5	$3,61E+09 = N_4^{1,71}$	$6,6E+08 = N_4^{1,58}$
6	36	1\,296	1\,679\,616	6	$3,62E+10 = N_4^{1,70}$	$5,59E+09 = N_4^{1,57}$
7	49	2\,401	5\,764\,801	7	$2,56E+11 = N_4^{1,69}$	$3,43E+10 = N_4^{1,56}$
8	64	4\,096	16\,777\,216	8	$1,4E+12 = N_4^{1,68}$	$1,66E+11 = N_4^{1,55}$



ют схему первого измерения, а p схем первого измерения образуют схему второго измерения.

Аналогично, схема третьего измерения состоит из p схем второго измерения, а схема четвертого измерения состоит из p схем третьего измерения.

Сеть $\underline{C}_8N_8$ содержит p^4N_4 коммутаторов K_4N_4 , имеющих в совокупности p^4N_8 входов. Далее вырезаются все четыре слоя мультиплексоров Mr первого каскада сети $\underline{C}_8N_8$, и входы первого слоя мультиплексоров Mr подсоединяются к входам коммутаторов K_4N_4 . Это возможно, так как число этих входов совпадает. Копии второго каскада сети $\underline{C}_8N_8$ в сети $\underline{C}_8N_8$ числом p^4 разобьем на p^3 групп и обозначим через G ($1 \leq G \leq p^3$) номер группы и через I ($1 \leq I \leq p$) – номер в группе. Фактически G задает номер схемы первого измерения, а I – номер копии K_4N_4 в схеме первого измерения.

Обозначим через J ($1 \leq J \leq N_4$) номер коммутатора K_4N_4 в схеме первого измерения. Обозначим через K ($1 \leq K \leq N_4$) номер входа каждого такого коммутатора K_4N_4 . Таким образом, вход любого коммутатора K_4N_4 задается составным индексом G, I, J, K .

Обозначим через i ($0 \leq i \leq p-1$) номер входа по $\bmod p$ каждого мультиплексора Mr в первом их слое. В каждом коммутаторе первого каскада сети $\underline{C}_8N_8$ содержится p^3N_4 этих мультиплексоров Mr . Разобьем их на N_4 групп и обозначим через g ($1 \leq g \leq N_4$) номер группы, а через j ($1 \leq j \leq p$) – номер мультиплексора Mr в группе. Обозначим через k ($1 \leq k \leq N_4$) номер коммутатора K_4N_4 в первом каскаде сети $\underline{C}_8N_8$. Таким образом, вход любого мультиплексора Mr в первом слое задается составным индексом i, g, j, k .

Для этого вырезаются все слои мультиплексоров Mr первого каскада сети $\underline{C}_8N_8$. Первый внутренний слой содержит $W_{8,1} = N_4p^3$ мультиплексоров Mr , входы которых соединяются с входами коммутаторов K_4N_4 в сети $\underline{C}_8N_8$.

Вход мультиплексора Mr первого слоя с индексом i, g, j, k переключается на вход коммутатора K_4N_4 с индексом G, I, J, K , в котором $G = i + 1$, $I = j$, $J = g$ и $K = k$. В результате образуется p^4N_4 параллельно включенных коммутаторов K_4N_4 , на входы которых поступают разреженные непересекающиеся прямые p -перестановки.

Вырезанные мультиплексоры Mr первого слоя объединяют выходы p^3 схем первого измерения. Вырезанные мультиплексоры Mr второго слоя объединяют выходы схем первого измерения с одинаковым индексом G , образуя выходы схем

второго уровня с этим индексом. Вырезанные мультиплексоры Mr второго слоя объединяют выходы схем второго измерения с одинаковыми индексами. Вырезанные мультиплексоры Mr третьего слоя объединяют выходы схем второго измерения с одинаковыми индексами G , образуя выходы схем третьего уровня с тем же индексом. Вырезанные мультиплексоры Mr четвертого слоя объединяют выходы схем третьего измерения, образуя выходы коммутатора K_8N_8 .

Для коммутаторов K_8N_8 справедлива

Лемма 4. *Дуальный коммутатор K_8N_8 является неблокируемым коммутатором со статической маршрутизацией на любой обычной перестановке при любом p . Он обладает $(\sigma-1)$ -канальной отказоустойчивостью.*

Доказательство. Первое утверждение опирается на использование коммутатора K_4N_4 и справедливость для него леммы 3. Второе утверждение опирается на неблокируемость коммутатора K_4N_4 и тот факт, что p -перестановка состоит на сечении из разреженных 1-перестановок, разведенных по разным каналам и тактам.

Канальная отказоустойчивость следует из того, что пути между источниками и приемниками в коммутаторе K_4N_4 проходят через разные схемы каждого измерения и того, что $p \geq \sigma$.

В результате образуется неблокируемый самомаршрутизируемый коммутатор K_8N_8 с $N_8 = N_4^2$ каналами, обладающий $(\sigma-1)$ -канальной отказоустойчивостью. ♦

Коммутатор K_8N_8 имеет восемь слоев выходных мультиплексоров Mr общим числом $V_8 = N_8(p^8 - 1)/(p - 1)$.

Характеристики наиболее быстродействующих вариантов коммутаторов K_8N_8 приводятся в табл. 8, табл. 9 и табл. 10. В них даются характеристики коммутаторов для $\sigma = 2$ и $\sigma = 3$, которые рассчитываются по рекуррентным формулам $S_8 = N_4S_4 + p^4N_4S_4$ и $L_8 = N_4L_4 + p^4N_4L_4$. Отметим, что коммутационную и канальную сложность коммутатора K_8N_8 на основе орграфа можно сделать существенно меньше, чем у сосредоточенного коммутатора с топологией полного графа.

5. АНАЛИЗ РЕЗУЛЬТАТОВ И ВОЗМОЖНОСТИ ПРАКТИЧЕСКОГО РАЗВИТИЯ ПОСТРОЕННЫХ СЕТЕЙ

В работе предложена методика построения нового вида неблокируемых самомаршрутизируемых фотонных сетей широкой масштабируемости. Это так называемые дуальные сети, в основе которых

лежит неблокируемый дуальный коммутатор $p \times p$ с периодом разрядов в p длительностей сигнала (тактов).

Дуальный коммутатор используется как составная часть неблокируемого самомаршрутизируемого коммутатора $N_1 \times N_1$ $K_1 N_1$ с топологией квазиполного графа или орграфа. В первом случае число каналов $N_1 = p(p-1)/\sigma + 1$ и существует возможность иметь $(\sigma-1)$ -канальную отказоустойчивость. Во втором случае число каналов $N_1 = p^2$ и имеется возможность увеличить число каналов. Коммутатор с топологией квазиполного (ор)графа состоит из N_1 дуальных коммутаторов $p \times p$ вместе с N_1 демультиплексорами Dp $1 \times p$ и мультиплексорами Mp $p \times 1$ без линий задержки. Коммутационная сложность коммутатора $K_1 N_1$ оценивается как $S_1 = N_1(S_0 + 2p)$. Период сигналов T_1 коммутатора $K_1 N_1$ равен периоду сигналов у дуального коммутатора, т. е. $T_1 = p$.

Из коммутаторов $K_1 N_1$ составляются два каскада для построения блокируемой сети $N_2 \times N_2$ $C_2 N_2$ с $N_2 = N_1^2$ каналами. Каждый каскад состоит из N_1 коммутаторов $N_1 \times N_1$, а каналы между каскадами

прокладываются по схеме с обменными связями. Сеть $C_2 N_2$ преобразуется в неблокируемый самомаршрутизируемый двухкаскадный коммутатор $K_2 N_2$ посредством ее одномерного внутреннего распараллеливания.

Если коммутатор $N_1 \times N_1$ был построен на основе квазиполного графа, то коммутатор $K_2 N_2$ обладает $(\sigma-1)$ -канальной отказоустойчивостью, поскольку $p \geq \sigma$. Коммутационная сложность коммутатора $K_2 N_2$ оценивается по рекуррентным формулам как $S_2 = N_1 S_1 + p N_1 S_1$, а канальная сложность – как $L_2 = N_1 L_1 + p N_1 L_1$. По построению период сигналов T_2 коммутатора $K_2 N_2$ равен периоду сигналов коммутатора $K_1 N_1$, т. е. $T_2 = T_1 = p$.

Если коммутатор $N_1 \times N_1$ был построен на основе квазиполного графа, то четырехкаскадный коммутатор $K_4 N_4$ обладает $(\sigma-1)$ -канальной отказоустойчивостью, поскольку $p \geq \sigma$. Коммутационная сложность коммутатора $K_4 N_4$ оценивается по рекуррентным формулам как $S_4 = N_2 S_2 + p^2 N_2 S_2$, а канальная сложность – как $L_4 = N_2 L_2 + p^2 N_2 L_2$. По построению период сигналов T_4 коммутатора $K_4 N_4$ равен периоду сигналов коммутатора $K_2 N_2$, т. е.

$$T_4 = T_2 = p.$$

Таблица 8

Характеристики дуальных коммутаторов $K_8 N_8$ с одноканальной отказоустойчивостью

p	N_1	$N_8 = N_1^8$	$T_8 = p$	S_8	L_8
2	2	256	2	$739\,840 = N_8^{2,44}$	$304\,640 = N_8^{2,28}$
3	4	65\,536	3	$4,998E + 09 = N_8^{2,01}$	$1,451E + 09 = N_8^{1,9}$
4	7	5\,764\,801	4	$4,937E + 12 = N_8^{1,88}$	$1,108E + 12 = N_8^{1,78}$

Таблица 9

Характеристики дуальных коммутаторов $K_8 N_8$ с двухканальной отказоустойчивостью

p	N_1	$N_8 = N_1^8$	$T_8 = p$	S_8	L_8
3	3	6\,561	3	$500\,341\,860 = N_8^{2,28}$	$145\,260\,540 = N_8^{2,14}$
4	5	390\,625	4	$3,345E + 11 = N_8^{2,06}$	$7,509E + 10 = N_8^{1,94}$
5	7	5\,764\,801	5	$3,331E + 13 = N_8^2$	$6,099E + 12 = N_8^{1,89}$

Таблица 10

Характеристики дуальных коммутаторов $K_8 N_8$ на основе орграфа

p	N_1	$N_8 = N_1^8$	$T_8 = p$	S_8	L_8
2	4	65\,536	2	$189\,399\,040 = N_8^{1,72}$	$77\,987\,840 = N_8^{1,64}$
3	9	5\,764\,801	3	$3,283E + 12 = N_8^{1,64}$	$9,531E + 11 = N_8^{1,57}$
4	16	4,29E+09	4	$3,678E + 15 = N_8^{1,62}$	$8,256E + 14 = N_8^{1,55}$

Аналогично из коммутаторов $K_4 N_4$ составляются два каскада для построения блокируемой сети $N_8 \times N_8$ $C_8 N_8$ с $N_8 = N_4^2 = N_1^8$ каналами.

Каждый каскад состоит из N_4 коммутаторов $K_4 N_4$, а каналы между каскадами прокладываются по схеме с обменными связями.

Сеть $C_8 N_8$ преобразуется в неблокируемый самомаршрутизируемый двухкаскадный коммутатор $K_8 N_8$ посредством ее четырехмерного внутреннего распараллеливания. Если коммутатор $N_1 \times N_1$ был построен на основе квазиполного графа, то и восьмикаскадный коммутатор $K_8 N_8$ обладает $(\sigma-1)$ -канальной отказоустойчивостью. Коммутационная сложность коммутатора $K_8 N_8$ оценивается по рекуррентным формулам



как $S_8 = N_4 S_4 + p^4 N_4 S_4$, а канальная сложность – как $L_8 = N_4 L_4 + p^4 N_4 L_4$. По построению период сигналов T_8 коммутатора $K_8 N_8$ равен периоду сигналов коммутатора $K_4 N_4$, т. е. $T_8 = T_4 = p$.

Характеристики коммутаторов $K_2 N_2$, $K_4 N_4$ и $K_8 N_8$ обладают несколькими степенями свободы. Прежде всего, при увеличении основания p число каналов увеличивается, а быстродействие уменьшается. Кроме того, показательная сложность уменьшается при увеличении основания p и имеется возможность разменивать быстродействие на сложность. К тому же, увеличение числа каналов посредством увеличения числа каскадов также сопровождается уменьшением показательной сложности.

Предложенная методика позволяет строить неблокируемые самомаршрутизируемые сети с самоподобной структурой. Коммутатор $K_2 N_2$ состоит из дуальных коммутаторов $K_1 N_1$ с топологией дуального графа или дуального орграфа и использует одномерное внутреннее распараллеливание. Коммутатор $K_4 N_4$ состоит из коммутаторов $K_2 N_2$ и использует двумерное внутреннее распараллеливание. Наконец, коммутатор $K_8 N_8$ состоит из коммутаторов $K_4 N_4$ и использует четырехмерное внутреннее распараллеливание. Все эти коммутаторы наследуют базовые свойства коммутатора $K_1 N_1$, такие как неблокируемость при статической самомаршрутизации и канальную отказоустойчи-

вость (при необходимости), но при существенно меньшей сложности.

Широкая масштабируемость неблокируемых коммутаторов также может быть достигнута многократным применением метода инвариантного расширения к коммутатору $K_1 N_1$ с топологией орграфа, в котором используется обычный коммутатор $p \times p$. Такие расширенные коммутаторы имеют период разрядов в один такт, но обладают повышенной сложностью. В табл. 11 сравнивается коммутационная сложность дуальных коммутаторов $K_4 N_4$ и $K_8 N_8$ и расширенных коммутаторов $K_1 N_1$. Видно, что дуальные коммутаторы имеют на несколько порядков меньшую коммутационную сложность.

Заметим, что дуальный коммутатор разрешает конфликты шинным способом только в первом каскаде в коммутаторе $K_1 N_1$. Все остальные конфликты во всех каскадах просто предотвращаются благодаря использованию внутреннего распараллеливания, и в них дуальные коммутаторы используются как обычные коммутаторы $p \times p$. Поэтому, может быть, имеет смысл использовать дуальный коммутатор в его первоначальном виде [1–3] – в виде пары «мультиплексор – демультиплексор», коммутационная сложность которой в p раз меньше. Это даст уменьшение коммутационной сложности дуальных коммутаторов $K_4 N_4$ и $K_8 N_8$ еще в несколько раз ($1,5 \div 4,5$).

Таблица 11

Сравнительная сложность дуальных коммутаторов и расширенных орграфов

Коммутационные сложности неблокируемых четырехкаскадных коммутаторов (S_4 , д) и расширенных коммутаторов с топологией квазиполных орграфов (S_{PO})				
p	N_4	Дуальный коммутатор $K_4 N_4$ $S_{4, д}$	Расширенный коммутатор $K_1 N_1$ S_{PO}	Отношение $S_{PO}/S_{4, д}$
2	256	$46\,080 = N_4^{1,94}$	$261\,120 = N_4^{2,25}$	5,67
3	6 561	$6\,298\,560 = N_4^{1,78}$	$129\,120\,480 = N_4^{2,12}$	20,5
4	65 536	$22\,282\,400 = N_4^{1,74}$	$11\,453\,071\,360 = N_4^{2,09}$	514
Коммутационные сложности неблокируемых восьмикаскадных коммутаторов (S_8 , д) и расширенных коммутаторов с топологией квазиполных орграфов (S_{PO})				
p	N_8	Дуальный коммутатор $K_8 N_8$ $S_{8, д}$	Расширенный коммутатор $K_1 N_1$ S_{PO}	Отношение $S_{PO}/S_{8, д}$
2	65 536	$18\,939\,9040 = N_8^{1,72}$	$17\,179\,607\,040 = N_8^{2,12}$	85,7
3	43 046 721	$3,283E + 12 = N_8^{1,64}$	$5,55822E + 15 = N_8^{2,06}$	1 640
4	4,29E + 09	$3,678E + 15 = N_8^{1,62}$	$4,91906E + 19 = N_8^{2,04}$	13 107

Отметим, что сложность отказоустойчивых коммутаторов $K_2 N_2$ и $K_4 N_4$ при малых p больше сложности полного графа, а при больших p – меньше. При этом сложность коммутаторов $K_8 N_8$ оказывается существенно меньше сложности полного графа при любых p . Специально хотелось бы выделить характеристики коммутатора $K_8 N_8$ при $p = 2$ и $\sigma = 1$. При числе каналов $N_8 = 65\,536$ и вдвое меньшем быстродействии его коммутационная сложность сопоставима со сложностью пятикаскадной небло-

кируемой сети Клоза на базе 64-канального маршрутизатора YARC [9] с числом каналов $N = 32\,768$, при условии, что она построена как неблокируемая сеть [7, 8]. Оценка сложности этой неблокируемой сети Клоза дает величину $S = N^{1,73}$. Однако эта сеть не имеет параллельных процедур ни статической, ни динамической самомаршрутизации.

Остальные коммутаторы K_8N_8 с $p > 2$ и $\sigma = 1$ имеют еще меньшую коммутационную сложность и еще большую масштабируемость, но при меньшем быстродействии.

Пониженное в p раз быстродействие коммутаторов K_2N_2 , K_4N_4 и K_8N_8 может быть скомпенсировано разными протокольными способами. Можно использовать процессоры с p независимыми портами, делить пакеты на p частей и передавать их параллельно. Широкая масштабируемость построенных коммутаторов легко обеспечивает такой режим работы, но посредством уменьшения в p раз числа абонентов и увеличения сложности сети. Также можно использовать параллельно-последовательный способ передачи пакетов по p линиям как в протоколе PCI-express уже без уменьшения числа абонентов.

Одним из недостатков коммутаторов K_1N_1 , K_2N_2 , K_4N_4 и K_8N_8 является их оптимизация под бесконфликтную реализацию произвольных перестановок. А каким будет их поведение на произвольном трафике? Для его определения можно придать мультиплексорам в выходных каскадах свойство «одноканальности». Мультиплексор с таким свойством из нескольких входных пакетов пропускает на выход только один и блокирует прохождение остальных пакетов. Передача блокированных пакетов, не подтвержденных приемниками, повторно осуществляется источниками.

Существенным недостатком предложенной методики является необходимость параллельной передачи сигнальной и управляющей информации, что существенно увеличивает необходимую полосу пропускания. Для фотонных коммутаторов этот недостаток не является роковым, так как оптический кабель может нести одновременно сотни разных частот. Однако в общем случае этот недостаток можно ликвидировать, если обеспечить поразрядную синхронизацию сигналов разных каналов. Это можно сделать на основе метода [34, 35] с локацией взаимного положения источников и приемников и соответствующих задержек передач от источников. В этом случае управляющую информацию для дуальных коммутаторов и демультиплексоров можно передавать, как обычно, в виде наборов разрядов в заголовке пакетов.

Приятным бонусом поразрядной синхронизации является возможность построения в канале на входе каждого приемника АЛУ только сетевыми средствами. Такие АЛУ были разработаны для метода вычислений в общем канале [36]. Для их реализации требуется передача по каналу значений разрядов чисел в парафазном виде – по двум линиям с активными сигналами для значений 0 и 1 в каждом. В сетевом АЛУ производится операция над числом, поступающим по каналу, и числом, находящимся у приемника, с образованием результата в канале после АЛУ. В канале возможно выполнение операций сложения, умножения, любых поразрядных логических операций, в том числе нахождения максимума (минимума).

ЗАКЛЮЧЕНИЕ

В работе предложена методика построения неблокируемых отказоустойчивых фотонных сетей широкой масштабируемости, рассмотренных в статье [5], но со значительной меньшей сложностью. Эта методика базируется на трех основных компонентах:

- использование p -канального дуального коммутатора с периодом разрядов в p тактов, который оказывается неблокируемым на любом входном трафике – именно это является необходимым условием неблокируемости более сложных сетей;
- использование коммутатора с топологией квазиполного графа или орграфа с дуальным коммутатором внутри, в результате чего сохраняется неблокируемость, обеспечивается канальная отказоустойчивость и более высокая масштабируемость при каскадировании по сравнению с чистым дуальным коммутатором;
- применение метода внутреннего распараллеливания для сохранения неблокируемости посредством предотвращения конфликтов и сохранения отказоустойчивости, которое обеспечивает широкую масштабируемость при каскадировании неблокируемых сетей.

В работе [5] масштабирование осуществляется путем каскадного применения метода инвариантного расширения посредством использования дополнительных внешних мультиплексоров и демультиплексоров. В настоящей работе масштабирование осуществляется путем каскадирования неблокируемых сетей меньшего размера и применения обобщенного метода внутреннего распараллеливания на каждом шаге каскадирования.

Каскадирование неблокируемой сети с N каналами осуществляется посредством построения



блокируемой сети с N^2 каналами. Эта сеть состоит из двух каскадов с обменными связями с N исходными неблокируемыми сетями в каждом. Блокировки в этой двухкаскадной сети возникают на выходных мультиплексорах первого каскада. Эти блокировки предотвращаются посредством разведения конфликтующих каналов по нескольким копиям второго каскада и перемещения мультиплексоров на выходы той частью второго каскада, которая выполняет маршрутизацию пакетов. В этой части сети конфликты не возникают, так как она состоит из копий неблокируемых подсетей, которые выполняют маршрутизации разреженных перестановок. Объединение разреженных перестановок в полную перестановку на сети с N^2 каналами осуществляется перемещенными каскадами мультиплексоров бесконфликтно.

При первом каскадировании [5] внутреннее распараллеливание осуществляется с использованием p копий второго каскада и использованием однослойного каскада выходных мультиплексоров. Второе каскадирование при построении неблокируемой сети с N^4 каналами выполняется с использованием p^2 копий второго каскада и использованием двухслойного каскада выходных мультиплексоров. Третье каскадирование при построении неблокируемой сети с N^8 каналами выполняется с использованием p^4 копий второго каскада и использованием четырехслойного каскада выходных мультиплексоров. Таким образом, были построены неблокируемые двух-, четырех- и восьмикаскадные сети с каскадами, состоящими из неблокируемых дуальных сетей с топологией квазиполного графа или орграфа.

Внутреннее распараллеливание при каждом каскадировании сохраняет период разрядов и сопровождается уменьшением удельной сложности создаваемой неблокируемой сети. В частности, были построены неблокируемые сети, имеющие удельную сложность не больше, чем у теоретической неблокируемой сети Клоза.

Рассмотренная методика может служить основой для построения практических проектов неблокируемых коммутаторов широкой масштабируемости со статической самомаршрутизацией и канальной отказоустойчивостью.

ЛИТЕРАТУРА

1. Барabanова Е.А., Вытовтов К.А., Подлазов В.С. Многокаскадные коммутаторы для оптических и электронных суперкомпьютерных систем // Материалы 8-го Национального Суперкомпьютерного Форума (НСКФ-2019). – Переславль-Залесский, 2019. – http://2019.nscf.ru/TesisAll/02_Apparatura/037_BarabanovaEA.pdf. [Barabanova, E.A., Vytovtov, K.A., Podlazov, V.S. Mnogokaskadnye kommutatory

- dlya opticheskikh i elektronnykh superkomp'yuternykh sistem // Materialy 8-go Nacional'nogo Superkomp'yuternogo Forum (NSKF-2019). – Pereslavl'-Zalesskiy, 2019. – http://2019.nscf.ru/TesisAll/02_Apparatura/037_BarabanovaEA.pdf. (In Russian)]
2. Барabanова Е.А., Вытовтов К.А., Вишнеvский В.М., Подлазов В.С. Новый принцип построения оптических устройств обработки информации для информационно-измерительных систем // Датчики и системы. – 2019. – № 9. – С. 3–9. [Barabanova, E.A., Vytovtov, K.A., Vishnevskiy, M., Podlazov, V.S. The New Principle for the Construction of Optical Information Processing Devices for Information-Measuring Systems // Sensors and Systems. – 2019. – No. 9. – P. 3–9 (In Russian)]
3. Barabanova, E., Vytovtov, K., Vishnevskiy, V., Podlazov, V. Model of Optical Non-blocking Information Processing System for Next-generation Telecommunication Networks // Proceedings of the 22nd International Conference on Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN-2019). – Moscow, 2019. – P. 188–198. – DOI: 10.1007/978-3-030-36625-4_16.
4. Каравай М.Ф., Подлазов В.С. Метод инвариантного расширения системных сетей многопроцессорных вычислительных систем. Идеальная системная сеть // Автоматика и телемеханика. – 2010. – № 12. – С. 166–176. [Karavay, M.F., Podlazov, V.S. An Invariant Extension Method for System Area Networks of Multicore Computational Systems. An Ideal System Network // Automation and Remote Control. – 2010. – Vol. 71, no. 12. – P. 2644–2654.]
5. Барabanова Е.А., Вытовтов К.А., Подлазов В.С. Двухкаскадные дуальные фотонные коммутаторы в расширенном схемном базисе // Проблемы управления. – 2021. – № 1. – С. 69–80. [Barabanova, E.A., Vytovtov, K.A., Podlazov, V.S. Two-Stage Dual Photonic Switches in an Extended Circuit Basis // Control Sciences. – 2021. – No. 1. – P. 69–80.]
6. Барabanова Е.А., Вытовтов К.А., Подлазов В.С. Непрокируемые отказоустойчивые двухкаскадные дуальные фотонные коммутаторы // Проблемы управления. – 2021. – № 4. – С. 82–92. [Barabanova, E.A., Vytovtov, K.A., Podlazov, V.S. Non-blocking Fault-Tolerant Two-Stage Dual Photon Switches. – Control Sciences. – 2021. – No. 4. – P. 67–76.]
7. Clos, C. A Study of Non-locking Switching Networks // Bell System Tech. J. – 1953. – Vol. 32. – P. 406–424.
8. Бенеш В.Э. Математические основы теории телефонных сообщений. – М.: Связь. – 1968. – С. 83–150. [Benesh, V.E. Matematicheskie osnovy teorii telefonnykh soobshhenij. – M.: Svjaz'. – 1968. – P. 83–150. (In Russian)]
9. Scott, S., Abts, D., Kim, J., and Dally, W. The Black Widow High-radix Clos Network // Proc. 33rd Intern. Symp. Comp. Arch. (ISCA'2006). – Boston, 2006. – URL: https://www.researchgate.net/publication/4244660_The_Black_Widow_High-Radix_Clos_Network.
10. De Sensi, D., Di Girolamo, S., H. McMahon, K., Roweth, D. An In-Depth Analysis of the Slingshot Interconnect // arXiv: 2008.08886v1. – 2020. – URL: https://www.researchgate.net/publication/343786515_An_In-Depth_Analysis_of_the_Slingshot_Interconnect.
11. Alverson, R., Roweth, D., and Kaplan, L. The Gemini System Interconnect // 18th IEEE Symposium on High Performance Interconnects. – Santa Clara, CA, 2009. – P. 83–87.
12. Alverson, R., Roweth, D., Kaplan, L., and Roweth, D. Cray XC® Series Network. – URL: <http://www.cray.com/Assets/PDF/products/xcr/CrayXC30Networking.pdf>.
13. Kim, J., Dally, W. J., Scott, S. and Abts, D. Technology-Driven, Highly-Scalable Dragonfly Topology // Proceedings of the 35th Annual International Symposium on Computer Architecture (ISCA'2008). – Beijing, 2008. – P. 77–88. – URL:

- <http://users.ece.gatech.edu/~sudha/academic/class/Networks/Lectures/4%20-%20Topologies/papers/dragonfly.pdf>.
14. Mellanox OFED for Linux User Manual. Rev 2.3-1.0.1. – Mellanox Technologies, Ltd.: 2014. – URL: https://dldnets.asus.com/pub/ASUS/mb/accessory/PEM-FDR/Manual/Mellanox_OFED_Linux_User_Manual_v2_3-1_0_1.pdf.
 15. Pipenger, N. On Rearrangeable and Non-blocking Switching Networks // J. Comput. Syst. Sci. – 1978. – Vol. 17. – P. 307–311.
 16. Bhuyan, L.N. and Agrawal, D.P. Generalized Hypercube and Hyperbus Structures for a Computer Network // IEEE Trans. on Computers. – 1984. – Vol. C-33, no 4. – P. 323–333.
 17. Tzeng, N., Wei, S. Enhanced Hypercubes // IEEE Trans. Computers. – 1991. – Vol. 40, no. 3. – P. 284–294.
 18. Efe, K. A Variation on the Hypercube with Lower Diameter // IEEE Trans. Computers. – 1991. – Vol. 40, no. 11. – P. 1312–1316.
 19. Kim, J., Dally, W.J. Flattened Butterfly Topology for On-Chip Networks // IEEE Computer Architecture Letters. – 2007. – Vol. 6, no. 2. – P. 37–40.
 20. Gu, Q.P. and Tamaki, H. Routing a Permutation in Hypercube by Two Sets of Edge-Disjoint Paths // J. of Parallel and Distributed Comput. – 1997. – Vol. 44, no. 2. – P. 147–152.
 21. Lubiwi, A. Counterexample to a Conjecture of Szymanski on Hypercube Routing // Inform. Proc. Let. 1990. – Vol. 35(2). – P. 57–61.
 22. Stepanenko, S. Structure and Implementation Principles of a Photonic Computer // EPJ Web of Conferences. – 2019. – Vol. 224. – DOI: <https://doi.org/10.1051/epjconf/201922404002>.
 23. Жабин И.А., Макагон Д.В., Поляков Д.А. и др. Первое поколение высокоскоростной коммуникационной сети «Ангара» // Наукоемкие технологии. – 2014. – № 1. – С. 21–27. [Zhabin, I.A., Makagon, D.V., Polyakov, D.A., et al. Pervoe pokolenie vysokoskorostnoi kommunikatsionnoi seti «Angara» // Naukoemkie tekhnologii. – 2014. – No. 1. – P. 21–27. (In Russian)]
 24. Stegailov, V., Agarkov, A., Biryukov, V., et al. Early Performance Evaluation of the Hybrid Cluster with Torus Interconnect Aimed at Molecular Dynamics Simulations // International Conference on Parallel Processing and Applied Mathematics. – Cham: Springer, 2017. – P. 327–336.
 25. Ajima, Y., Inoue, T., Hiramoto, S., Shimiz, T. Tofu: Interconnect for the K Computer // Fujitsu Scientific & Technical Journal. – Vol. 48, no. 3. – July 2012. – P. 280–285. – URL: https://www.researchgate.net/publication/265227674_Tofu_Interconnect_for_the_K_computer.
 26. Arimili, B., Arimilli, A., Chung, V., et al. The PERCS High-Performance Interconnect // 18th IEEE Symposium on High Performance Interconnects. – New York, 2009. – P. 75–82.
 27. Kathareios, G., Minkenberg, C., Prisacari, B., et al. Cost-Effective Diameter-Two Topologies: Analysis and Evaluation // SC'15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC'15). – 2015. – P. 1–11.
 28. Besta, M. and Hoefler, T. Slim fly: A Cost Effective Low-Diameter Network Topology // SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. – 2014. – P. 348–359.
 29. Flajslik, M., Borch, E., and Parker, M.A. Megaflly: A topology for exascale systems // Yokota, R., Weiland, M., Keyes, D., and Trinitis, C. eds., High Performance Computing. Cham. Springer International Publishing. – 2018. – P. 289–310.
 30. Ahn, J. H., Binkert, N., Davis, A., et al. Hyperx: Topology, Routing, and Packaging of Efficient Large-Scale Networks // Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. – 2009. – P. 1–11.
 31. Domke, J., Matsuoka, S., Ivanov, I.R., et al. Hyperx Topology: First At-scale Implementation and Comparison to the Fat-Tree // Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC'19). New York, USA, Association for Computing Machinery. – 2019.
 32. Singla, A., Hong, C.-Y., Popa, L., and Godfrey, P.B. Jellyfish: Networking Data Centers Randomly // Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12). San Jose. CA. USENIX. – 2012. – P. 225–238.
 33. Холл М. Комбинаторика. Главы 10-12. – М. Мир. – 1970. – 424 с. [Hall, M. Combinatorial Theory. – Blaisdell Publishing Company. Waltham. – 1967.]
 34. Стецюра Г.Г. Компьютерная сеть с быстрой распределенной перестройкой своей структуры и обработкой данных в процессе их передачи // Проблемы управления. – 2017. – № 1. – С. 47–56. http://pu.mtas.ru/archive/Stetsyura_117.pdf [Stetsyura, G.G. Computer Network with the Fast Distributed Reorganization of Its Structure and Data Processing During Their Transmission // Control Sciences. – 2017. – No. 1. – P. 47–56.] – URL: (In Russian)]
 35. Стецюра Г.Г. Компьютерные кластеры с быстрым аппаратным выполнением синхронизации сообщений и распределенных вычислений сетевыми средствами // Проблемы управления. – 2020. – № 4. – С. 61–69. [Stetsyura, G.G. The Computer Clusters with Fast Synchronization of Messages and with Fast Distributed Computing by the Network Hardware // Control Sciences. – 2020. – No. 4. – P. 61–69. (In Russian)]
 36. Прангшвили И.В., Подлазов В.С., Стецюра Г.Г. Локальные микропроцессорные вычислительные сети. Глава шестая. – М.: Наука, 1984. – 175 с. [Prangishvili, I.V., Podlazov, V.S., Stetsyura, G.G. Lokal'nye mikroprocessornye vychislitel'nye seti. Glava shestaya. – M.: Nauka, 1984. – 175 p. (In Russian)]
- Статья представлена к публикации членом редколлегии В.М. Вишневым.
- Поступила в редакцию 25.03.2021,
после доработки 12.08.2021.
Принята к публикации 24.08.2021
- Подлазов Виктор Сергеевич – д-р техн. наук, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва,
✉ podlazov@ipu.ru,



NON-BLOCKING FAULT-TOLERANT DUAL PHOTON SWITCHES WITH HIGH SCALABILITY

V.S. Podlazov

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

✉ podlazov@ipu.ru

Abstract. This paper continues the construction of a fundamentally new class of system area networks (dual photon networks) with the following features: non-blocking property and static self-routing, high scalability with the maximum achievable speed and a small complexity compared to a full switch, and balancing the scalability-speed and complexity-speed ratios. These networks are implemented in an extended circuit basis consisting of dual photon switches and separate photon multiplexers and demultiplexers. We propose a method for constructing a fault-tolerant dual network with the indicated properties based on networks with the quasi-complete graph and quasi-complete digraph topologies and the invariant extension method with internal parallelization. Also, we propose a method for extending the two-stage dual network designed previously into four-stage and eight-stage dual networks with high scalability while maintaining the original network period and reducing its exponential complexity.

Keywords: photon switch, dual switch, photon multiplexers and demultiplexers, multistage switch, conflict-free self-routing, non-blocking switch, static self-routing, quasi-complete digraph, quasi-complete graph, invariant extension of networks, switching properties, direct channels, scalability and speed.

СЦЕНАРНЫЕ МЕТОДЫ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ РЕАЛИЗАЦИИ ЖИЗНЕННОГО ЦИКЛА ПРОГРАММНО-ЦЕЛЕВОГО УПРАВЛЕНИЯ (АНАЛИЗ КОНЦЕПЦИИ)

И.В. Чернов

Аннотация. Исследованы методологические и прикладные проблемы повышения эффективности процессов программно-целевого планирования и управления социально-экономическими системами на базе сценарного подхода. Рассмотрены основные особенности методологии и механизмов программно-целевого управления. Решение задач управления и контроля реализации целевых программ предложено проводить с применением сценарного подхода и моделей жизненного цикла. Для повышения эффективности программно-целевого планирования и управления разработана сценарная W-образная модель жизненного цикла, реализующая сквозную методологию упреждающего управления реализацией долгосрочных социально-экономических программ в условиях неопределенности и риска. Методологическое ядро предложенной модели составляют механизмы сценарного анализа, моделирования, прогнозирования, планирования и группового управления. Наибольшую эффективность предлагаемый подход обеспечивает в условиях возрастающей неопределенности благодаря его ориентации на предвидение будущих условий и альтернатив развития социально-экономических систем.

Ключевые слова: планирование, управление, эффективность, целевая программа, сценарный подход, жизненный цикл, неопределенность.

ВВЕДЕНИЕ

Усиление во втором десятилетии XXI в. негативных процессов в мировом развитии, углубление вызванного пандемией коронавируса COVID-19 глобального экономического кризиса, разрушение международных хозяйственно-экономических связей, усиление санкционной направленности антироссийской политики стран Запада, а также ухудшение связанных с внутренними проблемами бюджетных показателей и спад производства приводят к существенному росту различного рода рисков и неопределенности, что ужесточает требования к качеству и эффективности управления социально-экономическими системами (СЭС), и, соответственно, существенно сужает область применения традиционных подходов и методов планирования и управления их устойчивым развитием.

В сложившихся нестабильных условиях возрастает роль и значение базирующихся на программно-целевом подходе методов сценарного анализа и моделирования, результаты которого обеспечивают возможность снижения неопределенности и

учета спектра внешних и внутренних угроз достижению поставленных целей с помощью оценки наиболее вероятных и целесообразных направлений развития динамических процессов в исследуемых сегментах, характеризующихся сложной пространственной и административно-управленческой организацией социально-экономических систем, их устойчивости и других желательных и нежелательных свойств по информации об их структурных особенностях.

1. СОВЕРШЕНСТВОВАНИЕ МЕХАНИЗМОВ ПРОГРАММНО-ЦЕЛЕВОГО ПЛАНИРОВАНИЯ И УПРАВЛЕНИЯ

Программно-целевое планирование и управление представляет собой специальный метод государственного регулирования, предназначенный для применения в первую очередь на критически важных направлениях развития СЭС страны и позволяющий сочетать интересы общества, государства и экономических субъектов различных форм соб-



ственности. Рассматриваемая методология обеспечивает возможность решения крупномасштабных и сложных комплексных проблем развития СЭС путем структуризации и взаимоувязки составляющих элементов, а также организации взаимозаинтересованного партнерства государства, бизнеса, науки, образования и гражданского общества [1].

В значительной степени программно-целевое планирование опирается на методологию анализа иерархий [2, 3], что обеспечивает возможность достижения главной цели путем ее декомпозиции и создания таким образом системы частных подцелей, достижение которых осуществляется на основе результатов реализации комплекса взаимосвязанных программных мероприятий выделенным множеством исполнителей.

Сложность решения проблем повышения эффективности программно-целевого планирования и управления определяется различием методов и подходов к решению управленческих задач на разных стадиях рассматриваемого процесса. Кроме того, реализация любого крупномасштабного долгосрочного проекта связана со значительным числом как видоизменяющихся, так и вновь возникающих рисков, а также внешних и внутренних угроз различной природы, которые необходимо учитывать в процессе программно-целевого планирования и управления. Данные риски и неопределенности имеют различный характер и существенно влияют на специфику решаемых задач на разных стадиях и этапах программно-целевого управления.

В процесс достижения поставленных целей крупномасштабных проектов и программ вовлекается большое количество экономических субъектов различных форм собственности и отраслевой принадлежности, как принимающих непосредственное участие в мероприятиях государственных или федеральных целевых программ, так и прямо или косвенно работающих по сути на достижение рассматриваемой главной цели на инициативной основе, ориентируясь на проявляемый государством интерес к решению конкретных задач социально-экономического развития и рассчитывая на экономическую востребованность результатов своей работы, в том числе с опорой на рыночные механизмы хозяйствования [4].

В данной ситуации при использовании традиционных подходов к управлению возникает угроза «падения» социально-экономического эффекта от полученных результатов реализации конкретных проектов и программ в силу возникновения возможных ошибок в процессе анализа, прогнозиро-

вания и целеполагания, определения реальных потребностей в ресурсах, отсутствия должной координации и рассогласования деятельности соответствующих управляющих подструктур и исполнителей программных мероприятий.

Одним из путей повышения эффективности программно-целевого планирования и управления является разработка и практическое внедрение методологии, базирующейся на процессном и сценарном подходах [5], обеспечивающей возможность снижения неопределенности и учета широкого спектра внешних и внутренних угроз достижению поставленных целей на разных этапах разработки и реализации целевых программ и проектов, а также обеспечение эффективной координации и контроля деятельности большой группы экономических субъектов по реализации программных мероприятий в рамках решения стратегических задач развития, характеризующихся сложной пространственной и административно-управленческой организацией социально-экономических систем.

2. АНАЛИЗ ПРОЦЕССОВ УПРАВЛЕНИЯ РЕАЛИЗАЦИЕЙ ЦЕЛЕВЫХ ПРОГРАММ НА ОСНОВЕ МОДЕЛЕЙ ЖИЗНЕННОГО ЦИКЛА

Любая целевая программа проходит несколько стадий своего жизненного цикла: выявление требующих решения проблем, формирование системы целей и критериев эффективности, планирование и разработка механизмов достижения поставленных целей в условиях неопределенности (труднопредсказуемых изменений во внутренней и внешней среде), управление реализацией программных мероприятий (в том числе по отклонениям), оценка полученных результатов и на последнем этапе – завершение программы или проекта.

Термин «жизненный цикл» (ЖЦ) в настоящее время является общеупотребительным и широко используется в естественных, технических, гуманитарных и др. науках. При этом можно отметить структурное и в основных чертах содержательное сходство определений ЖЦ в различных предметных областях [6]. Системотехническое определение модели полного ЖЦ приведено в работе [7]: «Модель полного жизненного цикла отдельного объекта представляет собой описание последовательности всех фаз, этапов его существования от замысла и появления («рождения») до исчезновения («отмирания»)». Следовательно, для модели полного ЖЦ характерна определенная структурная

инвариантность – состав фаз ЖЦ существенно не зависит от того, какой конкретный объект описывается, что делает ее достаточно универсальной и широко применимой, поскольку в общем случае практически любой объект управления (в данном случае – программа, проект и т. д.) проходит все основные фазы ЖЦ: разработку, реализацию и развитие, завершение. При этом наиболее общими параметрами ЖЦ являются моменты начала и окончания цикла, общая длительность, длительность каждой фазы (стадии, этапа, задачи и т. д.), порядок чередования стадий и этапов, вид и форма внутрифазовых изменений, совокупность характеризующих состояние объекта показателей (по фазам, стадиям, этапам) и т. д. Одновременно с этим каждому этапу ЖЦ соответствуют свои специфические управленческие задачи и особенности.

По своей сути модель ЖЦ представляет собой базирующийся на системном подходе прикладной инструмент организационного управления, позволяющий:

- обеспечить соответствие поставленным целям и адекватность понимания процессов управления путем представления его конечных результатов в виде кумулятивного итога решения более простых для понимания и оценки промежуточных задач;
- декомпозировать процесс управления на относительно самостоятельные и одновременно с этим логически взаимосвязанные временные блоки;
- структурировать процесс управления;
- поэтапно определять содержание управленческих задач, технологию их решения, а также частные критерии эффективности и ресурсно-временные ограничения с учетом поставленных целей и динамики развития объекта управления;
- как следствие – совершенствовать методологическую базу и принципы организации системы управления.

В настоящее время для решения широкого класса задач используется значительное количество разнообразных моделей ЖЦ, среди которых наиболее широко применяемыми являются каскадная, инкрементная, спиральная, V-образная и др. модели. Каждая из них ориентирована на определенные предметные области и сферы применения и, соответственно, имеет свои достоинства и недостатки. Одновременно с этим, пожалуй, только последняя из перечисленных выше моделей (V-образная) в наибольшей степени соответствует за-

дачам и особенностям программно-целевого управления [6].

Специфика программно-целевого управления заключается в том, что на длительном жизненном цикле реализации программ (3–5 и более лет) требуется решать сложный комплекс взаимосвязанных и при этом различных по своей природе проблем в условиях неопределенности, риска и непрерывных изменений во внешней и внутренней средах (целеполагание, анализ рисков, прогнозирование, разработка программных мероприятий, финансовое и ресурсное обеспечение, оперативное управление реализацией мероприятий, мониторинг и оценка результативности реализации программы, а также эффективности использования выделенных ресурсов и т. д.). Причем неэффективные решения, принятые на любом этапе, влекут за собой целый «букет» проблем на последующих (вплоть до необходимости возврата на начальные этапы целеполагания и планирования с целью внесения существенных корректировок, которые, в свою очередь, приведут к необходимости внесения изменений по всему управленческому циклу).

Отметим также, что в настоящее время большинство систем организационного управления различного уровня и целевого назначения во многом ориентированы на внутренние процессы развития объекта управления, в силу чего наиболее успешно обеспечивается решение в основном оперативных и частично тактических задач. По этой же причине труднопредсказуемые и во многих случаях существенные изменения во внутренней среде и практически все, что происходит вне системы (во внешней среде), в основном является только входными данными для нее.

Одновременно с этим «негативной» и требующей немедленной реакции входной информацией для подобных систем управления в основном являются лишь результаты непосредственного внешнего или внутреннего влияния на объект управления, когда в большинстве практических случаев ущерб (в самом широком понимании данного термина) уже фактически нанесен. Вследствие этого, если система управления ограничивается лишь противодействием последствиям неблагоприятных ситуаций, она принципиально неспособна обеспечить эффективное решение стратегических и среднесрочных задач управления социально-экономическим развитием государства и общества. Таким образом, указанные обстоятельства представляют собой основной источник снижения эф-



фективности процессов управления развитием социально-экономических систем.

В этой ситуации возрастает роль и значение сценарного подхода в программно-целевом управлении, применение которого позволяет проводить опережающую оценку качества принимаемых решений на различных этапах жизненного цикла крупных проектов и программ, а также снижать неопределенность путем учета широкого спектра внешних и внутренних угроз достижению поставленных целей. Неопределенность, под которой понимают ситуацию, когда частично или полностью отсутствует информация о структуре и возможных состояниях исследуемой СЭС и (или) ее внешней среды, является одним из ключевых методологических понятий сценарного подхода. Построение сценариев преследует две цели в отношении неопределенности: максимально возможное в рамках данного подхода ее снижение и описание не устраненной части неопределенности с помощью ряда сценарных вариантов. Тем самым закладывается основа для последующего уменьшения уровня неопределенности развития ситуации применительно к процессам поиска наилучших решений возникших проблем.

Технологии сценарного анализа, прогнозирования и планирования, а также группового управления позволяют комплексно оценивать эффективность и согласованность множества управленческих решений при выборе и реализации программ развития СЭС, а также учитывать и в значительной степени координировать цели и направления деятельности принимающих участие в рамках реализации программных мероприятий хозяйствующих субъектов, которые самостоятельно на средне- и краткосрочном временном горизонте планируют и осуществляют свои тактические действия в рамках определенных общих долгосрочных целевых установок, а также правил и обязательств, регламентируемых соответствующими целевыми программами, проектами и т. п.

Для повышения эффективности программно-целевого планирования и управления предлагается модифицировать V-образную модель ЖЦ, добавив связующую ветвь, отражающую методологическую базу решения рассматриваемых проблем обеспечения устойчивого поступательного социально-экономического развития государства и общества в условиях неопределенности, требующих реакции на труднопредсказуемые изменения во внутренней среде, а также влияние негативных явлений и процессов, происходящих во внешней среде (см. рис.).

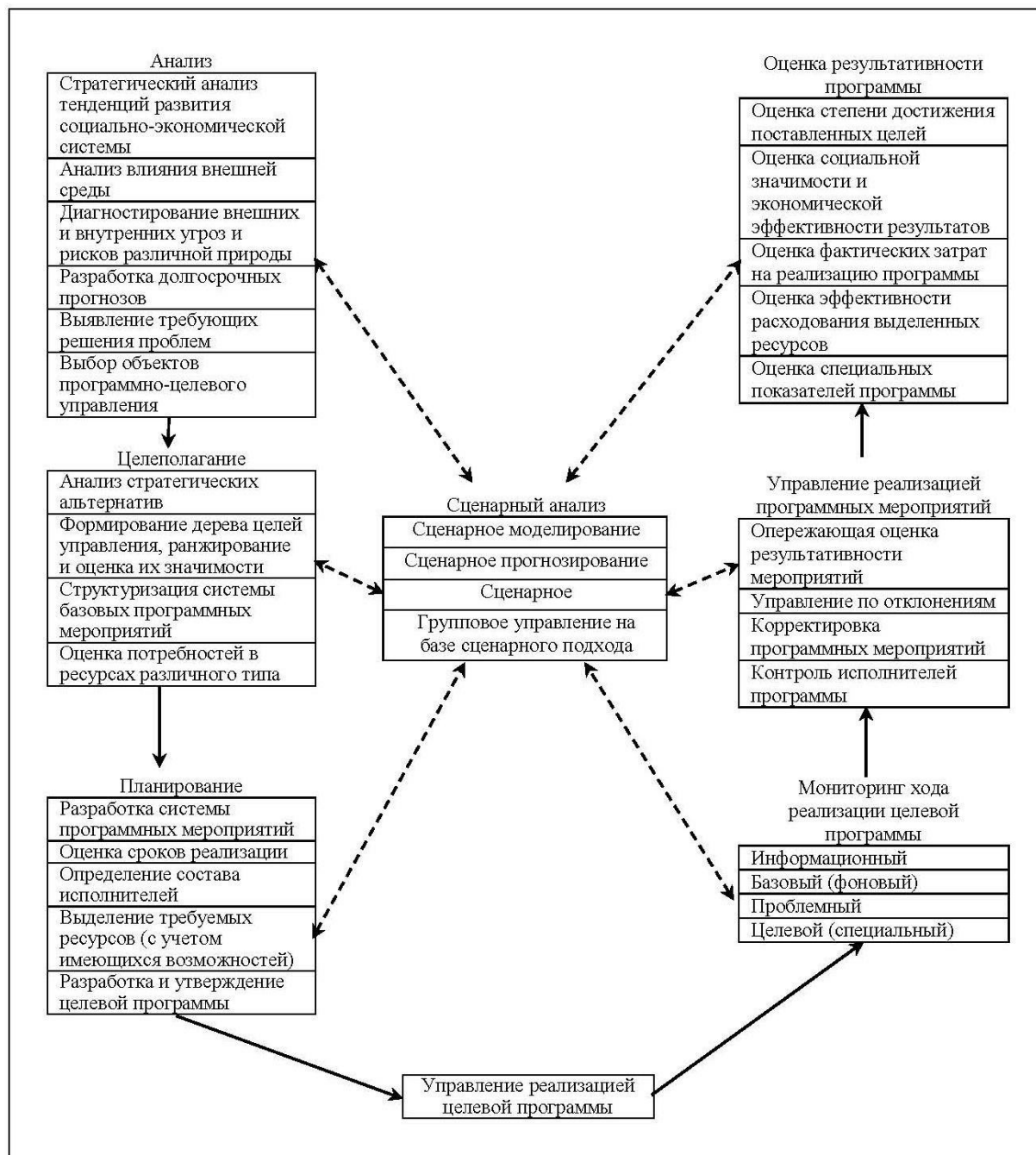
Если в классической V-образной модели ЖЦ особое значение придается действиям и процеду-

рам, направленным на верификацию и аттестацию результатов планирования и управления проектами, то W-образная модель наряду с традиционными аналитическими технологиями предоставляет возможность получения упреждающей или прогнозной оценки как конечных, так и наиболее значимых с точки зрения поставленных целей промежуточных результатов, что позволяет перейти от упомянутого выше пассивного учета уже свершившихся событий к упреждающему управлению реализацией целевых программ.

Рассматриваемая сквозная методология сценарного планирования и группового управления сложными системами в условиях неопределенности обеспечивает повышение эффективности управления группами объектов сложных СЭС (в качестве которых выступают экономические субъекты) путем формирования и анализа широкого спектра сценариев развития исследуемых социально-экономических систем и выделенных множеств составляющих их сегментов (в том числе объектов управления), учитывающих влияние внешней среды и сопутствующие риски различной природы.

Наибольшую эффективность предлагаемый подход обеспечивает в современных условиях возрастающей неопределенности, которая, по сути, и определяет потребность в инструментах и механизмах сценарного анализа, моделирования, прогнозирования, планирования и группового управления, поскольку именно они позволяют в значительной мере предвидеть будущие условия и соответствующие альтернативы развития СЭС на различных (преимущественно длительных) временных горизонтах, диагностировать внешние и внутренние угрозы и риски, исследовать возможные пути развития ситуации во внутренней и внешней среде.

Фактически рассматриваемая методология принципиально позволяет свести объект анализа в процессе подготовки управленческих решений к ограниченному числу альтернатив (в том числе наиболее вероятных), отражающих наиболее существенные с точки зрения поставленных целей риски и угрозы на любом из этапов жизненного цикла целевой программы. Более того, данная методология позволяет разрабатывать оптимистические и пессимистические сценарии и строить прогнозы развития заданных сегментов СЭС на различных уровнях управления, корректировать стратегическое видение ситуации и выделять желаемые направления ее развития, а также оценивать уровень скоординированности принимаемых решений на различных стадиях жизненного цикла целевых программ и крупных проектов.



Принципиальная схема W-образной модели жизненного цикла процессов программно-целевого планирования и управления

ЗАКЛЮЧЕНИЕ

В настоящее время накоплен значительный опыт разработки формализованной методологии анализа и синтеза моделей и методов, реализующих сценарный подход, а также решения прикладных задач в области анализа, прогнозирования, планирования и управления развитием различных сегментов социально-экономических систем. При этом разрабатываемые решения в абсолютном большинстве ориентированы на специфику кон-

кретных задач и фактически ограничены их рамками. Одновременно с этим практически отсутствуют работы, посвященные комплексному применению сценарного подхода при решении задач повышения эффективности программно-целевого управления разработкой и реализацией крупномасштабных территориально-распределенных проектов на протяжении их жизненного цикла, а также качества и скоординированности управленческих решений в условиях неопределенности и рисков различной природы.



Дальнейшее развитие фундаментальных и прикладных исследований в рассматриваемой предметной области обеспечит возможность решения широкого круга задач повышения эффективности процессов программно-целевого планирования и управления, а также результативности реализации приоритетных национальных проектов и социально-экономических программ.

ЛИТЕРАТУРА

1. Кульба В.В., Шульц В.Л., Шелков А.Б., Чернов И.В. Методы управления эффективностью реализации социально-экономических целевых программ // Тренды и управление. – 2013. – № 4. – С. 4–28. [Kul'ba, V.V., Shul'ts, V.L., Shelkov, A.B., Chernov, I.V. Metody upravleniya effektivnost'yu realizatsii sotsial'no-ekonomicheskikh tselevykh programm // Trendy i upravlenie. – 2013. – No. 4. – S. 4–28. (In Russian)]
2. Саати Т. Принятие решений. Метод анализа иерархий. – М.: Радио и связь, 1989. – 316 с. [Saaty, T.L. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. – N.-Y.: McGraw-Hill International Book Company, 1980. – 287 p.]
3. Саати Т., Кернс К. Аналитическое планирование. Организация систем – М.: Радио и связь, 1991. – 224 с. [Saaty, T.L., Kearns, K.P. Analytical Planning: The Organization of Systems. – Oxford: Pergamon, 1985. – 216 p.]
4. Управление и контроль реализации социально-экономических целевых программ. / Под ред. д.т.н., проф. Кульбы В.В. и д.т.н. Ковалевского С.С. – М.: Книжный дом

«ЛИБРОКОМ», 2009. – 400 с. [Upravlenie i kontrol' realizatsii sotsial'no-ekonomicheskikh tselevykh programm. / Pod red. d.t.n., prof. Kul'by V.V. i d.t.n. Kovalevskogo S.S. – M.: Knizhnyi dom «LIBROKOM», 2009. – 400 s. (In Russian)]

5. Модели и методы анализа и синтеза сценариев развития социально – экономических систем: в 2-х кн. / Под ред. В.Л. Шульца, В.В. Кульбы. – М.: Наука, 2012. – Кн. 1 – 304 с., кн. 2 – 358 с. [Modeli i metody analiza i sinteza stsensariiev razvitiya sotsial'no – ekonomicheskikh sistem: v 2-kh kn. / Pod red. V.L. Shul'tsa, V.V. Kul'by. – M.: Nauka, 2012. – Kn. 1 – 304 s., kn. 2 – 358 s. (In Russian)]
6. Берг Д.Б., Ульянова Е.А., Добряк П.В. Модели жизненного цикла: учеб. пособие. – Екатеринбург: Изд-во Урал. ун-та, 2014. – 74 с. [Berg, D.B., Ul'yanova, E.A., Dobryak, P.V. Modeli zhiznennogo tsikla: ucheb. posobie. – Ekaterinburg: Izd-vo Ural. un-ta, 2014. – 74 s. (In Russian)]
7. Бурков В.Н., Ириков В.А. Модели и методы управления организационными системами. – М.: Наука, 1994. – 269 с. [Burkov, V.N., Irikov, V.A. Modeli i metody upravleniya organizatsionnymi sistemami. – M.: Nauka, 1994. – 269 s. (In Russian)]

Статья представлена к публикации членом редколлегии В.Н. Бурковым.

Поступила в редакцию 16.08.2021,

Принята к публикации 31.08.2021.

Чернов Игорь Викторович – канд. техн. наук, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, ichernov@gmail.com.

SCENARIO METHODS TO IMPROVE THE EFFICIENCY OF IMPLEMENTING THE LIFE CYCLE OF PROGRAM-TARGET MANAGEMENT: A CONCEPTUAL ANALYSIS

I.V. Chernov

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

ichernov@gmail.com

Abstract. This paper is devoted to methodological and applied problems of improving the efficiency of program-target planning and management in socio-economic systems based on a scenario approach. The main features of the methodology and mechanisms of program-target management are considered. The idea proposed below is to manage and control the implementation of target programs using a scenario approach and life cycle models. A scenario W-shaped life cycle model is developed to improve the efficiency of program-target planning and management. This model involves the end-to-end methodology of proactive management to implement long-term socio-economic programs under uncertainty and risk. The model's methodological core consists of the following mechanisms: scenario analysis, simulation, forecasting, planning, and group management. The new approach is most effective under increasing uncertainty due to its focus on anticipating future conditions and alternatives for the development of socio-economic systems.

Keywords: planning, management, efficiency, target program, scenario approach, life cycle, uncertainty.