

ISSN 1684-8853 (print); ISSN 2541-8610 (online)

# ИНФОРМАЦИОННО- УПРАВЛЯЮЩИЕ СИСТЕМЫ

НАУЧНЫЙ ЖУРНАЛ

4(113)/2021

4(113)/2021

PEER REVIEWED JOURNAL

# INFORMATSIONNO- UPRAVLIAIUSHCHIE SISTEMY (INFORMATION AND CONTROL SYSTEMS)

**Founder**

«Information and Control Systems», Ltd.

**Publisher**Saint Petersburg State University  
of Aerospace Instrumentation**Editor-in-Chief**M. Sergeev  
Dr. Sc., Professor, Saint Petersburg, Russia**Deputy Editor-in-Chief**E. Krouk  
Dr. Sc., Professor, Moscow, Russia**Executive secretary**

O. Muravtsova

**Editorial Board**

S. Andreev  
Dr. Sc., Tampere, Finland

V. Anisimov  
Dr. Sc., Professor, Saint Petersburg, Russia

B. Bezruchko  
Dr. Sc., Professor, Saratov, Russia

N. Blaunstein  
Dr. Sc., Professor, Beer-Sheva, Israel

M. Buzdalov,  
PhD, Researcher, Saint Petersburg, Russia

C. Christodoulou  
PhD, Professor, Albuquerque, New Mexico, USA

A. Dudin  
Dr. Sc., Professor, Minsk, Belarus

I. Dumer  
PhD., Professor, Riverside, USA

M. Favorskaya  
Dr. Sc., Professor, Krasnoyarsk, Russia

L. Fortuna  
PhD, Professor, Catania, Italy

A. Fradkov  
Dr. Sc., Professor, Saint Petersburg, Russia

A. Hramov  
Dr. Sc., Professor, Innopolis, Russia

L. Jain  
PhD, Professor, Canberra, Australia

V. Khimenko  
Dr. Sc., Professor, Saint Petersburg, Russia

G. Matvienko  
Dr. Sc., Professor, Tomsk, Russia

A. Myllari  
PhD, Professor, Grenada, West Indies

Y. Podoplyokin  
Dr. Sc., Professor, Saint Petersburg, Russia

K. Samouylov  
Dr. Sc., Professor, Moscow, Russia

J. Seberry  
PhD, Professor, Wollongong, Australia

A. Shalyto  
Dr. Sc., Professor, Saint Petersburg, Russia

A. Shepeta  
Dr. Sc., Professor, Saint Petersburg, Russia

Yu. Shokin  
RAS Academician, Dr. Sc., Novosibirsk, Russia

A. Smirnov  
Dr. Sc., Professor, Saint Petersburg, Russia

T. Sutikno  
PhD, Associate Professor, Yogyakarta, Indonesia

Z. Yuldashev  
Dr. Sc., Professor, Saint Petersburg, Russia

R. Yusupov  
RAS Corr. Member, Dr. Sc., Professor, Saint Petersburg, Russia

A. Zeifman  
Dr. Sc., Professor, Vologda, Russia

**Editor:** A. Larionova**Proofreader:** T. Zvertanovskaia**Design:** M. Chernenko, Yu. Umnitsyna**Layout and composition:** Yu. Umnitsyna**Contact information**

The Editorial and Publishing Center, SUAI  
67, B. Morskaia, 190000, Saint Petersburg, Russia  
Website: <http://i-us.ru/en>, e-mail: [i-us.spb@gmail.com](mailto:i-us.spb@gmail.com)  
Tel.: +7 - 812 494 70 02

**THEORETICAL AND APPLIED MATHEMATICS**

**Balotin N. A., Sergeev A. M., SinitSYna O. I.** Finite field and group algorithms for orthogonal sequence search 2

**INFORMATION PROCESSING AND CONTROL**

**Fomin V. V., Aleksandrov I. V.** Indexing algorithm based on storing additional distances in metric space for multi-vantage-point tree 18

**INFORMATION AND CONTROL SYSTEMS**

**Melekhin V. B., Khachumov M. V.** Planning polyphase behavior of autonomous intelligent mobile systems in uncertain environments 28

**Afanasev A. A., Biktimirov S. N.** CubeSat formation architecture for small space debris surveillance and orbit determination 37

**SYSTEM AND PROCESS MODELING**

**Smirnov A. V., Levashova T. V., Petrov M. V.** Scenario model of intelligent decision support based on user's digital life models 47

**HARDWARE AND SOFTWARE RESOURCES**

**Parasich A. V., Parasich V. A., Parasich I. V.** Training set formation in machine learning tasks. Survey 61

**INFORMATION CODING AND TRANSMISSION**

**Trofimov A. N., Taubin F. A.** Evaluation of the union bound for the decoding error probability using characteristic functions 71

**INFORMATION ABOUT THE AUTHORS**

86

Submitted for publication 01.07.21. Passed for printing 23.08.21. Format 60×84<sub>1/8</sub>.  
Phototype SchoolBookC. Digital printing.

Layout original is made at the Editorial and Publishing Center, SUAI.  
67, B. Morskaia, 190000, Saint Petersburg, Russia  
Printed from slides at the Editorial and Publishing Center, SUAI.  
67, B. Morskaia, 190000, Saint Petersburg, Russia

The journal is indexed in Scopus.

Free distribution.

4(113)/2021

ИНФОРМАЦИОННО-  
УПРАВЛЯЮЩИЕ  
СИСТЕМЫ

РЕЦЕНЗИРУЕМОЕ ИЗДАНИЕ

## Учредитель

ООО «Информационно-управляющие системы»

## Издатель

Санкт-Петербургский государственный университет  
аэрокосмического приборостроения

## Главный редактор

М. Б. Сергеев,

д-р техн. наук, проф., Санкт-Петербург, РФ

## Зам. главного редактора

Е. А. Крук,

д-р техн. наук, проф., Москва, РФ

## Ответственный секретарь

О. В. Муравцова

## Редакционная коллегия:

С. Д. Андреев,

д-р техн. наук, Тампере, Финляндия

В. Г. Анисимов,

д-р техн. наук, проф., Санкт-Петербург, РФ

Б. П. Безручко,

д-р физ.-мат. наук, проф., Саратов, РФ

Н. Блаунштейн,

д-р физ.-мат. наук, проф., Беэр-Шева, Израиль

М. В. Буздалов,

канд. техн. наук, научный сотрудник, Санкт-Петербург, РФ

Л. С. Джайн,

д-р наук, проф., Канберра, Австралия

А. Н. Дудин,

д-р физ.-мат. наук, проф., Минск, Беларусь

И. И. Думер,

д-р наук, проф., Риверсайд, США

А. И. Зейфман,

д-р физ.-мат. наук, проф., Вологда, РФ

К. Кристодолу,

д-р наук, проф., Альбукерке, Нью-Мексико, США

Г. Г. Матвиенко,

д-р физ.-мат. наук, проф., Томск, РФ

А. А. Мюллери,

д-р наук, профессор, Гренада, Вест-Индия

Ю. Ф. Подоплёкин,

д-р техн. наук, проф., Санкт-Петербург, РФ

К. Е. Самуйлов,

д-р техн. наук, проф., Москва, РФ

Д. Себерри,

д-р наук, проф., Волонгонг, Австралия

А. В. Смирнов,

д-р техн. наук, проф., Санкт-Петербург, РФ

Т. Сутикну,

д-р наук, доцент, Джокьякарта, Индонезия

М. Н. Фаворская,

д-р техн. наук, проф., Красноярск, РФ

Л. Фортуна,

д-р наук, проф., Катания, Италия

А. Л. Фрадков,

д-р техн. наук, проф., Санкт-Петербург, РФ

В. И. Хименко,

д-р техн. наук, проф., Санкт-Петербург, РФ

А. Е. Храмов,

д-р физ.-мат. наук, Иннополис, РФ

А. А. Шальто,

д-р техн. наук, проф., Санкт-Петербург, РФ

А. П. Шепета,

д-р техн. наук, проф., Санкт-Петербург, РФ

Ю. И. Шокин,

акад. РАН, д-р физ.-мат. наук, проф., Новосибирск, РФ

З. М. Юлдашев,

д-р техн. наук, проф., Санкт-Петербург, РФ

Р. М. Юсупов,

чл.-корр. РАН, д-р техн. наук, проф., Санкт-Петербург, РФ

## Редактор: А. Г. Ларионова

Корректор: Т. В. Звертановская

Дизайн: М. Л. Черненко, Ю. В. Умницына

Компьютерная верстка: Ю. В. Умницына

Адрес редакции: 190000, Санкт-Петербург,

Б. Морская ул., д. 67, ГУАП, РИЦ

Тел.: (812) 494-70-02, эл. адрес: ius.spb@gmail.com,

сайт: http://i-us.ru

Журнал зарегистрирован в Министерстве РФ по делам печати, телерадиовещания и средств массовых коммуникаций.  
Свидетельство о регистрации ПИ № 77-12412 от 19 апреля 2002 г.  
Перерегистрирован в Роскомнадзоре.  
Свидетельство о регистрации ПИ № ФС77-49181 от 30 марта 2012 г.

© Коллектив авторов, 2021

## ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ МАТЕМАТИКА

**Балонин Н. А., Сергеев А. М., Синицына О. И.** Алгоритмы конечных полей и групп поиска ортогональных последовательностей 2

## ОБРАБОТКА ИНФОРМАЦИИ И УПРАВЛЕНИЕ

**Fomin V. V., Aleksandrov I. V.** Indexing algorithm based on storing additional distances in metric space for multi-vantage-point tree 18

## ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

**Melekhin V. B., Khachumov M. V.** Planning polyphase behavior of autonomous intelligent mobile systems in uncertain environments 28**Afanasev A. A., Biktimirov S. N.** CubeSat formation architecture for small space debris surveillance and orbit determination 37

## МОДЕЛИРОВАНИЕ СИСТЕМ И ПРОЦЕССОВ

**Смирнов А. В., Левашова Т. В., Петров М. В.** Базовый сценарий интеллектуальной поддержки принятия решений на основе моделей жизни пользователей в цифровой среде 47

## ПРОГРАММНЫЕ И АППАРАТНЫЕ СРЕДСТВА

**Парасич А. В., Парасич В. А., Парасич И. В.** Формирование обучающей выборки в задачах машинного обучения. Обзор 61

## КОДИРОВАНИЕ И ПЕРЕДАЧА ИНФОРМАЦИИ

**Трофимов А. Н., Таубин Ф. А.** Вычисление аддитивной границы вероятности ошибки декодирования с использованием характеристических функций 71

## СВЕДЕНИЯ ОБ АВТОРАХ

86

Журнал входит в БД SCOPUS и в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук.

Сдано в набор 01.07.21. Подписано в печать 23.08.21. Формат 60×84/8.  
Гарнитура SchoolBookC. Печать цифровая.  
Усл. печ. л. 10,4. Уч.-изд. л. 14,1. Тираж 1000 экз (1-й завод 50 экз.). Заказ № 370.

Оригинал-макет изготовлен в редакционно-издательском центре ГУАП.  
190000, Санкт-Петербург, Б. Морская ул., 67.

Отпечатано с готовых диапозитивов в редакционно-издательском центре ГУАП.  
190000, Санкт-Петербург, Б. Морская ул., 67.

УДК 519.614

doi:10.31799/1684-8853-2021-4-2-17

## Алгоритмы конечных полей и групп поиска ортогональных последовательностей

Н. А. Балонин<sup>а</sup>, доктор техн. наук, профессор, [orcid.org/0000-0001-7338-4920](https://orcid.org/0000-0001-7338-4920), [korbendfs@mail.ru](mailto:korbendfs@mail.ru)

А. М. Сергеев<sup>а</sup>, канд. техн. наук, доцент, [orcid.org/0000-0002-4788-9869](https://orcid.org/0000-0002-4788-9869)

О. И. Сеницына<sup>а</sup>, аспирант, [orcid.org/0000-0002-2819-4682](https://orcid.org/0000-0002-2819-4682)

<sup>а</sup>Санкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ

**Введение:** матрицы Адамара, состоящие из элементов 1 и  $-1$ , представляют собой идеальный объект для наглядного приложения конечномерной математики, оперирующей конечным числом адресов элементов  $-1$ . Системы нотаций методов абстрактной алгебры, в отличие от устоявшейся матричной алгебры, интенсивно менялись и, к тому же, не были повсеместно распространены, вызывая потребность пересмотреть и систематизировать накопленный опыт. **Цель:** описать набор алгоритмов конечных полей и групп в единых обозначениях для облегчения восприятия обширного материала, способствующего нахождению ортогональных и субортогональных последовательностей. **Результаты:** предложены формулы расчетов малоизвестных алгоритмов (и их версий) Скарпи, Зингера, Секереша, Гетхальса – Зейделя, Нобору Ито, а также полиномиальные уравнения, используемые для доказательства теорем существования конечномерных решений. Устранен существенный недостаток информации как в отечественной литературе (большинство затрагиваемых вопросов освещается у нас впервые), так и зарубежной систематизацией обширных знаний. **Практическая значимость:** ортогональные последовательности и методы их эффективного нахождения теорией конечных полей и групп имеют непосредственное практическое значение для задач помехоустойчивого кодирования, сжатия и маскирования видеоинформации.

**Ключевые слова** – матрицы Адамара, ортогональные матрицы, конечномерная математика, поля Галуа, конечные группы, алгоритм Скарпи, подход Зингера, метод Секереша, алгоритм Гетхальса – Зейделя, подход Нобору Ито.

Для цитирования: Балонин Н. А., Сергеев А. М., Сеницына О. И. Алгоритмы конечных полей и групп поиска ортогональных последовательностей. *Информационно-управляющие системы*, 2021, № 4, с. 2–17. doi:10.31799/1684-8853-2021-4-2-17

For citation: Balonin N. A., Sergeev A. M., Sinitsyna O. I. Finite field and group algorithms for orthogonal sequence search. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 2–17 (In Russian). doi:10.31799/1684-8853-2021-4-2-17

### Введение

Теория ортогональных матриц Адамара с двумя значениями элементов (уровней) 1 и  $-1$  закономерно относится к сфере привлечения теории конечных полей и групп, органично учитывающих фиксированные размеры матриц и их блоков. В циклических структурах положение отрицательных элементов этих матриц связано с характеристиками конечных полей — символами Лежандра и Якоби. Тем не менее, несмотря на эти знания, теория малоуровневых матриц далека от завершения.

Такие матрицы, как правило, имеют не один, а несколько характерных для них орнаментов (узоров) — взаимных расположений элементов с положительными и отрицательными знаками на «портретах» матриц. В зависимости от вида орнамента должен меняться алгоритм их нахождения. Однако многоклеточный, трехблочный и двухблочный орнаменты с одной или двумя каймами имеют существенно различающиеся между собой фиксированные размеры.

В научной литературе по данной тематике наблюдается дефицит информации относительно

вариативного употребления конечномерной математики. Если размер блока может убывать и не только на единицу, возникает неопределенность: какой вид индикаторов положения отрицательного элемента в орнаменте матрицы применять. Использовать символы Лежандра повсеместно становится невозможно. Это обстоятельство редко или вовсе не учитывается в существующих научных работах. Более того, оно является причиной сегодняшних научных исследований.

Ортогональные матрицы нашли множество применений в современной технике, поэтому вопрос, как их найти, включая высокие порядки и вариации орнаментов, важен. Вместе с тем соответствующую классификацию того, что происходит при изменении размеров блоков, никто в общей совокупности возникающих при этом задач не проводил. Еще один факт заключается в том, что сама по себе наука о конечных объектах не является чем-то вполне законченным. За столетие сам ее язык меняется, меняются обозначения, причем настолько, что прежде полученное знание снова утрачивается ввиду малочисленности статей и малопонятного, не закрепившегося на практике специфического языка описания алгоритмов.

Языки науки естественно повторяют путь древнегреческого, латинского и прочих языков, со временем отмирая и выходя из употребления. Работая со статьей, нам пришлось переписать и перевести на современный более понятный язык забытые ныне алгоритмы или алгоритмы, которые не известны в нашей стране и описаны в редких статьях за рубежом. Таковы, например, оригинальные алгоритмы Н. Ито, описанные весьма кратко, а также алгоритмы, сложившиеся под влиянием продуктивной группы математиков, близких к П. Эрдшу, таких как Д. Секереш, Дж. Себери, Д. Джокович и др.

Описание нами орнаментов матриц в их тесной связи с положениями точек Гаусса на квадратичных поверхностях закономерно подводит к этапу, когда обобщенных параметров орнаментов (координат целочисленных точек) становится мало. Нам нужен сам орнамент, а чтобы его найти, приходится выполнить большую и напряженную работу по сбору необходимой информации.

В такой совокупности настоящая статья будет нова и для российской, и для зарубежной аудитории. В ней ставится цель привести общие универсальные знания об орнаментах и симметриях, а также свойствах чисел и числовых последовательностей как порядках матриц, иллюстрируемые портретами ортогональных матриц. Часть приводимых алгоритмов при этом имеет значение для пограничных областей математики, далеко выходящих за пределы потребностей построения ортогональных базисов.

### Введение в конечные поля и группы

Абстрактную систему элементов (матриц, векторов, полиномов) вместе с четырьмя операциями называют полем. Убрав умножение и деление, получаем кольцо. Оставив одну операцию умножения, получим группу, что не мешает размышлять свойства этой составной операции, делая ее похожей на сложение [1–5].

Поля с конечным числом элементов — поля Галуа (Galois fields) обозначают как  $GF(q)$ , где  $q$  — простое число  $p$  или его степень. Числа  $0, 1, 2, \dots, p - 1$  с операциями, выполняемыми по модулю  $p$ , дают пример простого поля  $GF(p)$ . Таблицы сложения и умножения в поле отличаются от привычных лишь тем, что содержат остатки от деления на  $p$ , поэтому элементы поля принято называть вычетами.

Если вычет образован квадратом некоторого числа поля, он называется *квадратичным вычетом*, в противном случае — *невычетом*. Роль корней квадратных в конечномерной математике выполняют *инволюции* — элементы, квадрат которых равен единичному элементу.

**Пример 1.** Наименьшее число элементов, образующих поле, равно двум. Такое поле  $GF(2)$  содержит два опорных элемента:  $A = 0$  относительно операции сложения и  $B = 1$  относительно операции умножения. Их можно записывать конкретно как 0 и 1 или абстрактно как  $A$  и  $B$  и исследовать таблицы на предмет соблюдения, для операций с ними, всех правил арифметики (рис. 1).

**Пример 2.** Попытка построить поле  $GF(2^2) = GF(4)$  из  $A = 0, B = 1, C = 2, D = 3$ , дает расширенные таблицы (рис. 2). На них видно, что в центре таблицы умножения появился нулевой элемент  $A$ , и в этой строке сдвоились  $CB = CD = B, B$  — единица поля. Выходит, что на роль обратного к  $C$  элемента претендуют два кандидата:  $B$  и  $D$ .

Первая из таблиц, согласно классификации 1884 г., носит название четверной группы Клейна и обозначается как  $V_4$ . Чтобы отремонтировать вторую, в качестве элементов конечных полей  $GF(p^m)$  рассматриваются полиномы или векторы их коэффициентов размера  $m$ . Полиномы удобны тем, что, задав всего одно дополнительное полиномиальное уравнение, формирующее поле, с его помощью можно убрать последствия повышения степени при произведениях, сумма таких хлопот не вызывает.

Поле  $GF(p^2)$  связано по количеству востребованных коэффициентов с линейными функциями  $a + bx$  или векторами  $(a, b)$  с параметрами, определенными в поле  $GF(p)$ . Произведение такого сорта элементов неприятно лишь в связи с появлением полинома второй степени: от  $x^2$

+	A	B
A	A	B
B	B	A

×	A	B
A	A	A
B	A	B

■ **Рис. 1.** Таблицы сложения и умножения в  $GF(2)$   
 ■ **Fig. 1.** Addition and multiplication tables in  $GF(2)$

+	A	B	C	D
A	A	B	C	D
B	B	A	D	C
C	C	D	A	B
D	D	C	B	A

×	A	B	C	D
A	A	A	A	A
B	A	B	C	D
C	A	B	A	B
D	A	D	C	B

■ **Рис. 2.** Таблицы с дефектом умножения по месту  $A = 0$   
 ■ **Fig. 2.** Tables with multiplication defect by location  $A = 0$

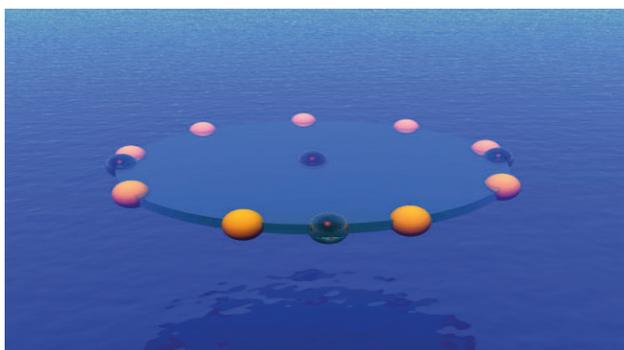
несложно избавиться, выразив через степени меньшего порядка. Образующий переносом всех членов этого уравнения влево полином не должен разлагаться на произведение полиномов первого порядка, поэтому его называют *неприводимым*. Этих сведений достаточно, чтобы конструировать конечные поля.

Механическая интерпретация *циклической группы* известна по устройству часов, напоминающих расположением отметок времени другую распространенную конструкцию — позиции на комплексной плоскости корней из единицы: решений  $g^n = 1$  (рис. 3).

Другой популярный источник полей и групп — матрицы — появился, потому что в качестве элементов можно брать члены прогрессии — показательной функции  $1, g, g^2, \dots, g^{n-1}$ , где  $g$  — примитивный элемент (генератор).

Заметим, что операция умножения матриц составная, для получения элементов произведения работает конвейер из цепочек умножений и сложений, применяемых к элементам сомножителей. В теории циклических групп, оперирующей набором элементов с единственной операцией, обозначаемой знаком умножения « $\times$ », разрешено и вовсе не строить цепочек, подменяя операцию умножения сложением порядковых номеров  $a, b$  элементов, так как  $g^a g^b = g^{a+b}$ , различие этих двух операций относительное. Очевидно, что порядок циклической группы может быть любым, четным или нечетным. Это выгодно отличает группу от конечного поля, размеры которого лимитированы возможностями полиномиальной арифметики. Кажется, что для построения циклических групп нет препятствий. Однако это не так. Недостаток слишком просто устроенных групп прямо вытекает из их достоинства — простоты, которая не позволяет моделировать с их помощью более сложно устроенные математические объекты.

Сложнее устроена мультипликативная группа  $GF(q)^*$ , построенная усечением количества



■ Рис. 3. Расположение корней из 1  
 ■ Fig. 3. Root location from 1

элементов конечного поля  $GF(q)$ , где  $q$  — простое число  $p$  или его степень, выбрасыванием ненужного группы 0. Соответственно, размер такой циклической группы не может быть любым, она обременена выброшенным сложением (а также вычитанием и делением). К мультипликативным группам примыкают по смыслу близкие к ним группы порядков  $n = p^t u$ , где множитель  $u$  взаимно прост с  $p$ , а  $p$  — простое число. Для абелевых групп (произведение коммутативно) принято выделять и использовать в доказательствах теорем циклические  $p$ -подгруппы Силова размера  $p^t$ .

### Тактика использования конечных полей и групп

В настоящее время теория групп кажется точной солидной наукой. Обилие книг. Некоторая торжественность и чопорность изложения. Однако есть одно смущающее обстоятельство. В теории принято одним и тем же символом обозначать очень разные вещи, если эти вещи считаются отражениями одного объекта. Например, элемент группы может быть номером элемента или матрицей, выбирай, что хочешь, обозначаем  $g$ .

Смысл конкретного наполнения элемента в теории групп и полей зависит от характера его употребления. Более шадят читателя разнесенные обозначения для индексов и матриц, но тогда его голову начинает сушить введенное в оборот большое количество символов, смысл которых не успевает закрепиться в сознании. Подчеркнем, что двусмысленность является рабочим методом решения задач, поскольку, следуя иному содержанию термина, мы иначе организуем вычисление, и это неожиданно может дать новое решение задачи, обладающее новыми свойствами. Отстригать такую приятную неожиданность — действовать против себя.

Например, в алгоритме Скарпи вычисления матриц Адамара [6, 7] фигурируют произведение и сумма индексов строк. Можно к двум прибавить три и получить пять. Но можно действовать иначе! Можно 2-й элемент поля сложить с 3-м элементом поля, получить элемент суммы, у него есть номер, и этот номер вовсе не обязан быть 5-м. В сложных полях это будет иной номер. Алгоритм Скарпи, разработанный его автором для простых полей [8], «неожиданно» даст верный результат, матрицу Адамара, если в более общем случае мы изменим ход вычислений. Что подсказало такой прием? Этот ход вычислений подсказала принципиальная двойственность обозначений.

Не только смысл элементов двойственен. Смысл функции тоже может переключаться в зависимости от характера ее употребления. Ведь

уже в алгоритме Скарпи мы имеем два итога вычислений, ход вычислений зависит от обрабатываемого материала.

Учитывая тезис о соответствии элементов группы символам или порядковым номерам  $0, 1, 2, \dots$  в таблице умножения, сводящейся к таблице сложения показателей степеней, при таком подходе приходится перешагивать через привычку аннулировать произвольный элемент умножением на единичный элемент, обозначаемый, в том числе, и  $0$ .

С обозначениями в аддитивной по своему характеру арифметике, занятой умножениями, возникает сложность восприятия ввиду замены вычитания делением или умножением на обратный элемент. Так что с  $ab^{-1}$  вы освоитесь заметно быстрее, если будете иметь в виду, что это вполне знакомое вам  $a - b$ . В старых примерах таким (разностью) и бывшее, а потом переписанное в иных терминах в новые времена.

Разумеется, это вызывает внутренний протест, но что делать, если это так действенно и действительно сильно сокращает формулы.

Все это говорит о том, что употребление теории требует у читателя, привычного к иной работе с обозначениями, известной доли терпения и понимания, зачем это все нужно. Ортогональные и экстремальные по детерминанту матрицы являются превосходным иллюстративным материалом для демонстрации прикладной стороны абстрактной арифметики. Впрочем, всерьез делить, где здесь иллюстрация, а где инструмент, нет смысла, поскольку понятие ортогонального базиса первичное в математике, и речь идет, скорее, о тесной взаимной связи ее разделов.

Операция умножения чисел *по модулю*, не равному простому числу, не дает гарантии наличия среди соответствующих цепочек элементов генератора подгрупп силовского размера, хотя кажется почти очевидным, что они там должны быть. Подгруппы прочих размеров можно встретить, выделить и использовать в алгоритмах поиска матриц. Можно иначе строить подмножества, но тогда мы те подгруппы потеряем.

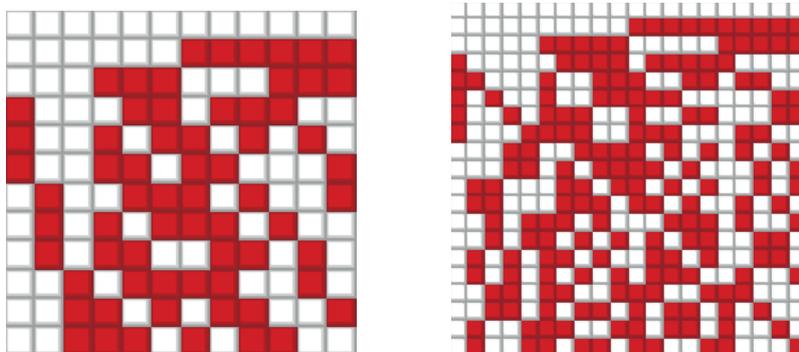
### Алгоритм Скарпи

*Определение.* Квадратная матрица  $\mathbf{H}$  порядка  $n$  с элементами  $\{1, -1\}$  и ортогональными столбцами  $\mathbf{H}^T \mathbf{H} = n\mathbf{I}$ , где  $\mathbf{I}$  — единичная матрица, называется матрицей Адамара [6, 7].

Ввиду наличия только двух значений элементов, порядок матриц Адамара может быть равен только  $1, 2$  или  $4t$ , где  $t$  — натуральное число. Известно, что на классе матриц с элементами, не превышающими по модулю  $1$ , матрицы Адамара имеют максимальный детерминант.

Вскоре после публикации Адамаром первых двух матриц  $\mathbf{H}$  порядков  $12$  и  $20$ , отличающихся от силовских итераций  $\begin{pmatrix} \mathbf{H} & \mathbf{H} \\ \mathbf{H} & -\mathbf{H} \end{pmatrix}$  с началом  $\mathbf{H} = 1$ , оригинальным размещением элементов (рис. 4), появился первый и более общий метод синтеза, предложенный алгебраистом итальянского происхождения Умберто Скарпи [8].

Мы излагаем этот метод с использованием (введенной позднее) нормальной формы матрицы Адамара в виде  $\mathbf{H} = \begin{pmatrix} -1 & \mathbf{e}^T \\ \mathbf{e} & \mathbf{M} \end{pmatrix}$ , где  $\mathbf{e}$  — вектор единичных элементов каймы, и *произведения Скарпи* — кронекерова



■ *Рис. 4.* Матрицы конструкции Адамара порядков  $12$  и  $20$   
 ■ *Fig. 4.* Construction of Hadamard matrices of orders  $12$  and  $20$

произведения с коррекцией знака каймы и нарастающим циклическим смещением строк основы, обозначаемого как « $\times$ ». Тогда он описывается одной формулой (а не цепочкой преобразований)

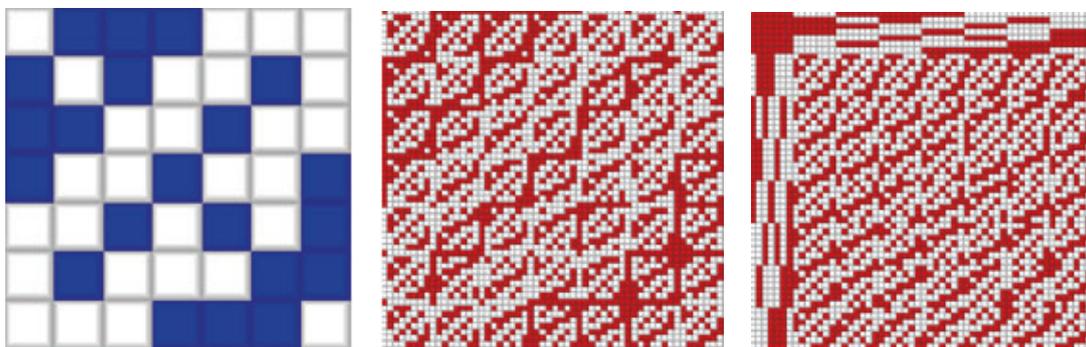
$$\mathbf{M} \times \mathbf{M} = \begin{pmatrix} \begin{pmatrix} -1 & m_{00}\mathbf{e}^T \\ m_{00}\mathbf{e} & \mathbf{M} \end{pmatrix} & \begin{pmatrix} -1 & m_{01}\mathbf{e}^T \\ m_{01}\mathbf{e} & \mathbf{M} \end{pmatrix} & \dots & \begin{pmatrix} -1 & m_{0(v-1)}\mathbf{e}^T \\ m_{0(v-1)}\mathbf{e} & \mathbf{M} \end{pmatrix} \\ \begin{pmatrix} -1 & m_{10}\mathbf{e}^T \\ m_{10}\mathbf{e} & \mathbf{M} \end{pmatrix} & \begin{pmatrix} -1 & m_{11}\mathbf{e}^T \\ m_{11}\mathbf{e} & \mathbf{T}\mathbf{M} \end{pmatrix} & \dots & \begin{pmatrix} -1 & m_{1(v-1)}\mathbf{e}^T \\ m_{1(v-1)}\mathbf{e} & \mathbf{T}^{v-1}\mathbf{M} \end{pmatrix} \\ \dots & \dots & \ddots & \dots \\ \begin{pmatrix} -1 & m_{(v-1)0}\mathbf{e}^T \\ m_{(v-1)0}\mathbf{e} & \mathbf{M} \end{pmatrix} & \begin{pmatrix} -1 & m_{(v-1)1}\mathbf{e}^T \\ m_{(v-1)1}\mathbf{e} & \mathbf{T}^{v-1}\mathbf{M} \end{pmatrix} & \dots & \begin{pmatrix} -1 & m_{(v-1)(v-1)}\mathbf{e}^T \\ m_{(v-1)(v-1)}\mathbf{e} & \mathbf{T}^{(v-1)(v-1)}\mathbf{M} \end{pmatrix} \end{pmatrix}.$$

Формула Скарпи работает корректно, если размер блока  $v = n - 1$  представляет собой простое число,  $\mathbf{T}$  — матрица циклического смещения строк  $\mathbf{M}$ .

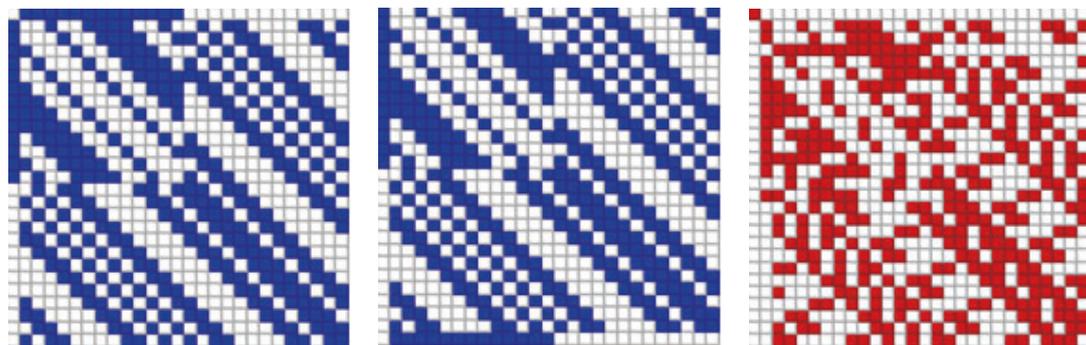
В кронекеровом произведении матриц

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \dots & a_{nn}\mathbf{B} \end{pmatrix},$$

которым предлагал пользоваться Адамар, знаки элементов первого сомножителя влияют на всю матрицу второго сомножителя. Для сравнения на рис. 5 рядом приведены основа (ее называют в прикладной



■ Рис. 5. Основа и матрицы Адамара конструкции Скарпи порядка 56  
 ■ Fig. 5. Basis and the Scarpy design Hadamard matrices of order 56



■ Рис. 6. Основа  $\mathbf{M}$ , смещенная основа и смещение в алгебре полей Галуа  $GF(27)$   
 ■ Fig. 6. Basis  $\mathbf{M}$ , basis offset and offset in algebra of Galois fields  $GF(27)$

литературе *core*, она выделена синим цветом)  $M$  порядка  $v = 7$  матрицы Адамара  $H$  порядка  $n = 8$ , произведение Скарпи и типичное для оригинала алгоритма размещение составной каймы, на которой отражаются знаки элементов основы, что дает две матрицы Адамара порядка  $vn = 56$ .

Заметим, что в алгоритме Скарпи циклическое смещение строки  $k$  основы  $M$  на величину  $i \times j$  определяют индексы номера строки  $i$  и столбца  $j$  блока с нею, отсчитываемые от 0 — для блока каймы смещение отсутствует.

Поправка к алгоритму формирования адреса смещаемой строки  $k + i \times j$  на случай  $v = p^k$  — степень простого числа, согласно концепции теории групп, не отражается на формуле расчета (рис. 6).

Как видно, при произведении в сложном поле меняется интерпретация номеров строк и столбцов номерами перемножаемых и потом складываемых элементов  $GF(p^k)$ . Итоги смещения строки матрицы  $M$  размера  $v = 27$  обычным и модифицированным алгоритмами отличаются — во втором случае ее адрес определяется номером результирующего элемента в конечном поле.

### Орнаментальные инварианты и разности Зингера

В алгоритме Скарпи [8] проявилась польза от деления матрицы на блок с каймой или блоки, используемые наравне с матрицами Адамара в операции, подобной кронекерову произведению. Моноблок  $M$  порядка  $v = n - 1$  матрицы с каймой можно характеризовать двумя параметрами: числом  $-1$  в каждой строке  $k$  и числом  $-1$  в каждой паре строк  $\lambda$ .

Эти инварианты отвечают уравнению реализуемости орнамента  $k(k - 1) = \lambda(v - 1)$ , сводимому к каноническому виду  $x^2 = 1$ ,  $k = (v - x)/2$ , учитывая, что для ортогональных матриц  $\lambda = k - n/4$ , что дает параметры  $v = 7$ ,  $k = 3$ ,  $\lambda = 1$  первой матрицы на рис. 5.

Следующий пример кососимметрической циклической матрицы  $M$  (во всех таких случаях для краткости опускают уточнение — с точностью до диагонали) со строкой  $a = [1, -1, 1, -1, -1, 1, 1, 1, -1, 1]$  длины  $v = 11$  кочует из книги в книгу, он принадлежит Зингеру [9], заметившему соответствие инвариантов  $k = (v - x)/2 = 5$  и  $\lambda = k - n/4 = 2$  инвариантам не связанной с матрицами Адамара системы разностей чисел  $D$ , образующей некоторое множество  $G$ .

**Пример 3.** Возьмем из работы [10] дифференциальный набор  $k = 5$  чисел  $D = \{1, 3, 4, 5, 9\}$  по mod 11. Он порождает (кроме 0) множество  $G = \{1, 2, \dots, 10\}$ :  $1 - 3 = -2 = 9$ ;  $1 - 4 = -3 = 8$ ;  $1 - 5 = -4 = 7$ ;  $1 - 9 = -8 = 3$ ;  $3 - 1 = 2$ ;  $3 - 4 = -1 = 10$ ;  $3 - 5 = -2 = 9$ ;  $3 - 9 = -6 = 5$ ;  $4 - 1 = 3$ ;  $4 - 3 = 1$ ;

$4 - 5 = -1 = 10$ ;  $4 - 9 = -5 = 6$ ;  $5 - 1 = 4$ ;  $5 - 3 = 2$ ;  $5 - 4 = 1$ ;  $9 - 1 = 8$ ;  $9 - 3 = 6 = 5$ ;  $9 - 4 = 5$ ;  $9 - 5 = 4$ . Число совпадений разностей  $\lambda = 2$ .

Учитывая кососимметрию  $M$ , из нее можно построить матрицу Адамара добавлением каймы из 1 ее первого столбца и  $-1$  (над блоком) ее первой строки. Заметим, что параметры набора  $D$  в точности совпадают с адресами  $-1$  последовательности  $a$ .

Тем самым Зингер ввел в обиход дифференциальные наборы чисел, дав старт новому направлению исследования матриц Адамара при помощи групп, которыми такого сорта разности можно описать (напомним, что умножение в группе весьма условно), не опираясь на ресурсы поля. В своем решении он не избавился от перебора, поскольку метода формирования дифференциального набора он не предложил. Очевидно, что его подход игнорирует некоторую дополнительную информацию, которую несет в себе удачный для синтеза матриц Адамара выбор группы.

Потребность в полях или группах может исчезнуть совсем при симметрировании и перестановке порядка сомножителей с разницей их порядков, не большей 4. Соответствующая модификация алгоритма Скарпи [8] рассмотрена в работе [11]. Хороший вопрос, можно ли естественный для конечномерной математики ход вычислений, предопределенный изначально двусмысленностью записей абстрактной арифметики, считать *обобщением*, неведомым такому специалисту по теории чисел, как Скарпи? Скорее всего, Скарпи знал о таком продолжении. Тем более спустя тридцать лет об этом знал Пэли [12]. Пэли заметил, что комбинаторные алгоритмы, закрывая новые и новые порядки, кратные 4, фрагментарно пересекаются и никогда не закрывают область в целом.

Что свидетельствует о неограниченных способностях квадратичной задачи к усложнению. Похоже, решения в принципе нельзя охватить единой формулой или комбинаторным алгоритмом, хотя такие матрицы, безусловно, существуют вне исследованной зоны и, подчеркнем, поддаются изучению методами, связанными с максимумом детерминанта адамаровых матриц [13–15]. Теперь нам становится интересен пример, в котором блок или блоки ортогональной матрицы синтезируются на основе хотя бы мультипликативных групп  $GF(q)^*$ .

### Алгоритм Секереша

В отличие от Скарпи и Зингера, Дьердь Секереш занялся алгоритмами поиска матриц Адамара спустя более половины столетия [16, 17]. Он руководствовался инвариантами канонического

уравнения  $x^2 + y^2 = 2$ ,  $k_1 = (v - x)/2$ ,  $k_2 = (v - y)/2$ ,  $\lambda = k_1 + k_2 - n/4 = (v - 2(x + y) + 1)/2$ ,  $n = 2v + 2$ , где  $v$  — размер блока бициклической

основы (бицикла)  $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & -\mathbf{A}^T \end{pmatrix}$  матрицы Адамара

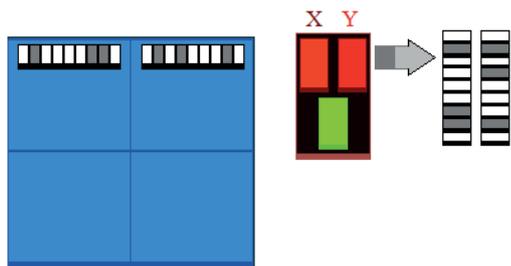
с парной каймой.

В таком случае размер порождающего мультипликативную группу  $GF(q)^*$  поля  $q = v$  соотносим с размерами клетки или всей матрицы за вычетом 1. Базовая идея первого алгоритма состоит в том, чтобы разделить элементы четной по размерам циклической группы на четыре части, сформировав из них два множества  $X$  и  $Y$ , пересекающиеся наполовину (рис. 7).

Синтез множеств происходит в два этапа. Сначала синтезируются общие части  $X$  и  $Y$  в виде последовательности  $g^{4k}$  длиной  $(q - 1)/4$ , где  $g$  — примитивный элемент группы. Умножив последовательность на  $g$  и на  $g^3$ , получим  $g^{4k+1}$  и  $g^{4k+3}$ , образующие два расходящихся между собой и дистанцирующихся от начала набора элементов для  $X$  и  $Y$ . Верхние части образуемой вилки можно поменять местами, но в таком случае нижнюю часть (ручку вилки) надо заменить оставшейся  $1/4$  элементов группы.

Теперь обратим внимание, что каждый элемент блока  $\mathbf{A}$  или  $\mathbf{B}$  описывается разностью его индексов, что касается первых строк, то это попросту индекс элемента без смещения. Разность индексов  $(i - j)$  элемента блока — целое число, не имеющее прямого отношения к элементам мультипликативной группы. Но здесь начинает работать магия теоретико-группового подхода, не запрещающая считать его номером элемента группы. Если выбранный элемент группы принадлежит набору  $X$  (или  $Y$ ), то блок  $\mathbf{A}$  (или  $\mathbf{B}$ ) имеет вместе с заданными координатами значение  $-1$ . Идея Секереша в обращении со степенями изящна, она имеет ясный и понятный смысл в совокупности с простотой реализации.

Для циклических блоков достаточно посчитать элементы первых строк блоков и получить из них смещением кососимметричный бицикл

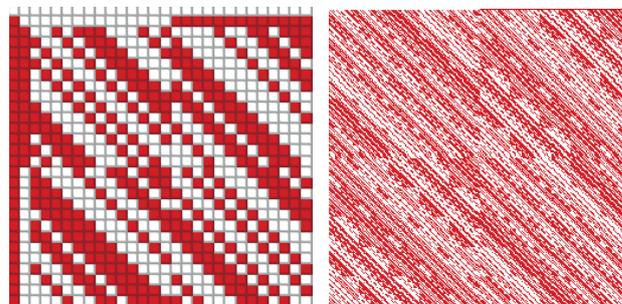


■ Рис. 7. Схема алгоритма Секереша  
 ■ Fig. 7. Seceresh algorithm diagram

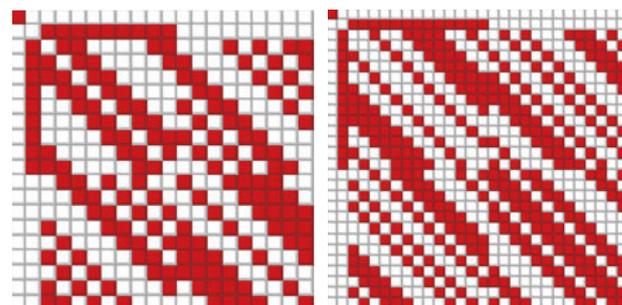
Адамара с парной каймой. Если группа сложная  $GF(q^m)^*$ , то и орнамент блоков  $\mathbf{A}$  и  $\mathbf{B}$  сложен, но матрица остается строго кососимметричной (рис. 8). К сожалению, именно поэтому она существует не всегда. Строгая структуризация приводит к появлению «дыр» в порядках матриц: алгоритм позволяет находить матрицы порядков  $n = 4(4t - 1)$ .

Если  $q = n - 1$  — простое число или степень простого числа, то в  $GF(q)^*$  генерируется подгруппа  $X = g^k$  размера  $v$ . Элементы  $-1$  последовательностей  $a$  и  $b$  отвечают пересечениям с  $X$  и  $e = X$ , где  $e$  — единственный элемент группы. Порядок одной из приведенных на рис. 9 матриц равен 28, но он отвечает сложной группе (узор матрицы не усложняется). Как видно, эти решения отличаются типом симметрии синтезируемых матриц. Перестановкой блоков их можно сделать как симметричными, так и кососимметричными (см. рис. 9).

Очевидно, Секереш развил идею Зингера построением разностных семейств в ином, отличном от него, направлении. Это уже полноценные алгоритмы синтеза матриц Адамара, использующие четный размер мультипликативных групп для деления элементов на непересекающиеся подмножества. Поскольку размер группы совпадает здесь с удвоенным размером блока, очевидно, что



■ Рис. 8. Матрицы Секереша порядков 28 и 256  
 ■ Fig. 8. Seceresh matrices of orders 28 and 256



■ Рис. 9. Матрицы Секереша порядков 20 и 28  
 ■ Fig. 9. Seceresh matrices of orders 20 and 28

основой построения является циклическая подгруппа, что объясняет относительно простой вид матрицы Адамара порядка 28 (27 — степень простого числа).

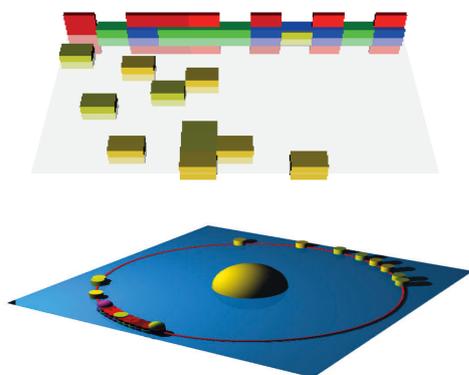
### Алгоритмы Гетхальса — Зейделя

Переход математики от римского представления чисел к десятичной позиционной системе счисления дал преимущество — алгоритмы расчета стали короче. Усложнение конструкции числа облегчает вычисления. То же самое произошло при расширении поля вещественных чисел — формулы для вычисления корней полиномов стали проще.

Излагаемая далее идея [18] противостоит идее Секереша: давайте не сокращать, а, наоборот, увеличивать размер поля  $GF(q^2)$ , где  $q = n - 1$ ,  $n$  — порядок матрицы Адамара. На рис. 10 слева приведена форма представления конечной модели комплексных чисел, отображаемых кубиками песочно-желтого цвета, стартовый ряд состоит из начальных  $q$  элементов. При этом форму представления можно менять (представление модели справа).

Прогрессия  $g^k$  масштабируется умножением на  $g^v$ . Песочно-желтые элементы  $X = \{g^v, g^{v+1}, g^{v+2}, \dots, g^{2v-1}\}$  операция сложения  $Y = X + X^q$  проецирует в более ограниченную область — конечную модель вещественных чисел, отображаемых на рис. 10 элементами красного цвета.

Что касается генерации квадратичных вычетов, то красные элементы на рис. 10 нам нужно рассортировать (элементы со значениями 1 и -1) по факту совпадения их с квадратичными вычетами и невычетами. Для этого используем элемент  $\omega = g^n$ , чьи четные степени будем интерпретировать как вычеты (зеленый цвет на рис. 10), а нечетные — как невычеты (синий цвет).



■ **Рис. 10.** Формы представления конечной модели комплексных чисел  
 ■ **Fig. 10.** Representation forms of the complex numbers finite model

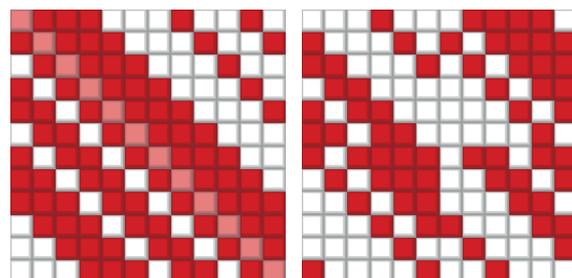
В силу расширения пространства группы квадраты элементов не реверсируются в ограниченную область, например, вычислениями по модулю  $q$  (размер поля вырос до  $q^2$ ) и образуют семейство четных членов данной прогрессии.

В итоге получаем алгоритм, который завершается вычислением первой половины строки матрицы, стартовый элемент — нулевой. Вторая половина, ввиду симметрии, восстанавливается выписыванием элементов в реверсном порядке, игнорируя стартовый, при этом у негациклических матриц *знак каждого второго элемента инвертируется*. У конференц-матриц вся хвостовая часть еще раз инвертируется. Разделением четных и нечетных строк и столбцов матрица приводится к кососимметрическому виду с двумя блоками **A** и **B**, с исправлением диагонали в 1 (рис. 11).

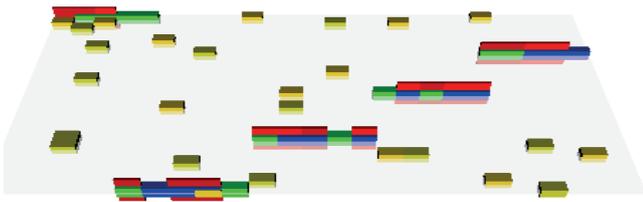
Заметим, что во всех таких случаях мы всегда применяем один и тот же прием — вычисляем прогрессию. У негациклического блока строки получают сдвиг с размещением слева вытесняемых элементов справа, но с инверсией знака. Отчетливо выраженных орнаментальных инвариантов такая матрица не имеет, с чем связана, возможно, продуктивность этого подхода.

Для четных порядков матриц, кратных 2, но не делящихся на 4, метод ведет к так называемым конференц-матрицам, когда компенсировать ноль на диагонали не получается по соображениям совместности. После инвертирования знаков четных строк и столбцов блоков **A** и **B** они становятся циклическими симметричными и кососимметричными блоками. Дополнительно циклическое смещение второго блока на половину его размера решает проблему обеспечения блочной симметрии, характерной для этих бициклов.

В результате получается много ортогональных матриц двухблочной конструкции, что позволяет отнести этот метод едва ли не к основным и постулировать, что если поля нет, то соответствующим образом устроенная матрица Адамара все равно существует, теряя, разве что, вид симметрии.



■ **Рис. 11.** Негациклическая матрица и негациклический бицикл  
 ■ **Fig. 11.** Negacyclic matrix and negacyclic bicycle



■ **Рис. 12.** Представление модели при работе в сложных полях  
 ■ **Fig. 12.** Model representation when working in complex fields

При работе предлагаемого алгоритма в сложных полях меняется логика размещения расчетных элементов (рис. 12), но не тип матрицы, как это было с алгоритмами Секереша.

Заметим, что нам здесь пришлось воспользоваться двумя операциями поля, поскольку проецирование осуществляется сложением.

### Алгоритмы Нобору Ито

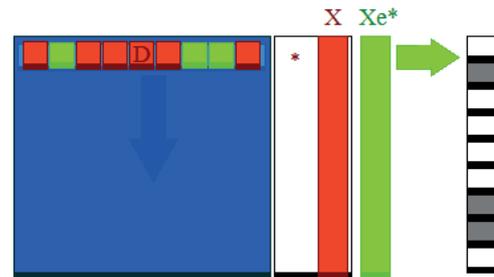
Ито начал работать в рассматриваемой тематике в 1993 г. [19, 20] с примера тетраэдральной группы  $G = \{X, Y\}$ , построенной с помощью трех генераторов  $a, b$  и  $c$  (бывает и такое). Первую ее половину составляет  $X = \{e, a, b, ab, c, ac, bc, abc, c^2, ac^2, bc^2, abc^2\}$ , вторую  $Y = Xe^*$ . Здесь  $e$  — единичный элемент,  $e^* = a^2 = b^2$  — инволюция. Это близко идее Секереша в алгоритме, использующем для селекции знаков пару множеств (рис. 13).

Цель Ито состоит в том, чтобы ясно показать, что ортогональная матрица Адамара отвечает симметриям, описываемым группой тетраэдра (рис. 14). Чтобы убедиться, что это именно группа, необходимо сконструировать умножением по модулю 11 все  $n = 12$  ее элементов, опираясь на следующее матричное представление  $a, b$  и  $c$ :

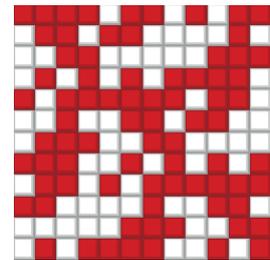
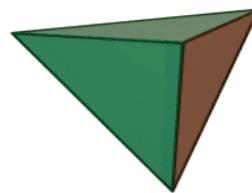
$$A = \begin{pmatrix} 0 & 10 \\ 1 & 0 \end{pmatrix}; B = \begin{pmatrix} 1 & 3 \\ 3 & 10 \end{pmatrix}; C = \begin{pmatrix} 4 & 7 \\ 8 & 6 \end{pmatrix},$$

и проверить свойства  $c^3 = e, b^{-1}ab = ae^*, c^{-1}ac = b, c^{-1}bc = ab$ , обеспечивающие основное качество группы: нет такого сочетания элементов, которое не сводилось бы к элементам группы  $G$ . В качестве  $e$  берется единичная матрица  $I, e^* = a^4 = b^2$  — инволюция здесь и далее  $(n - 2)I$ .

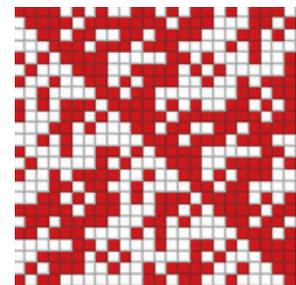
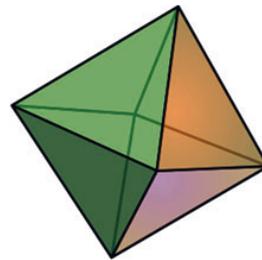
$D = \{e^*, ae^*, be^*, abe^*, c, ace^*, bce^*, abc, c^2e^*, ac^2, bc^2e^*, abc^2e^*\}$  задает первую строку кососимметричной матрицы Адамара тем, что элементы, входящие в первую половину группы, генерируют 1 и  $-1$ , если во вторую. Заметим, что если не учитывать элемент  $e^*$ , она повторяет порядок следования элементов группы.



■ **Рис. 13.** Схема алгоритма Нобору Ито  
 ■ **Fig. 13.** Noboru Ito algorithm diagram



■ **Рис. 14.** Тетраэдр и матрица Адамара с его симметриями  
 ■ **Fig. 14.** Tetrahedron and Hadamard matrix with its symmetries



■ **Рис. 15.** Октаэдр и матрица Адамара с его симметриями  
 ■ **Fig. 15.** Octahedron and Hadamard matrix with its symmetries

Остальные строки определяются последовательными произведениями  $D$  на элементы  $g$  первой половины группы. Перед синтезом элементы  $Dg$  тоже переставляются в соответствии с порядком их следования в группе. Если отказаться от перестановки, элементы  $-1$  строки кососимметричной матрицы определяются вхождением  $Y = Xe^*$  в  $Dg$ , поскольку  $Y$  содержит нужный порядок.

Следующий пример разнообразит ассоциативный ряд фигура-матрица группой октаэдра (рис. 15) с  $D = \{e^*, ae^*, a^2e^*, a^3e^*, b, abe^*, a^2b, a^3be^*, c, ac, a^2ce^*, a^3ce^*, bc, abce^*, a^2bc, a^3bce^*, c^{-1}e^*, ac^{-1}, a^2c^{-1}e^*, a^3c^{-1}, bc^{-1}e^*, abc^{-1}e^*, a^2bc^{-1}e^*, a^3bc^{-1}\}$ .

Остается сконструировать умножением по модулю 23 все  $n = 24$  элемента, опираясь на следующее матричное представление  $a, b, c$  и  $c^{-1}$ :

$$A = \begin{pmatrix} 7 & 19 \\ 4 & 11 \end{pmatrix}; B = \begin{pmatrix} 10 & 9 \\ 22 & 13 \end{pmatrix};$$

$$C = \begin{pmatrix} 18 & 11 \\ 19 & 4 \end{pmatrix}; C^{-1} = \begin{pmatrix} 4 & 12 \\ 4 & 18 \end{pmatrix},$$

и проверить свойства  $c^3 = e, b^{-1}ab = a^3e^*, c^{-1}a^2c = b, c^{-1}bc = a^2be^*$  и  $(ab)^{-1}c^2b = c^{-1}$ , обеспечивающие качество сочетания элементов быть группой.

Ито построил пример с группой икосаэдра (рис. 16).

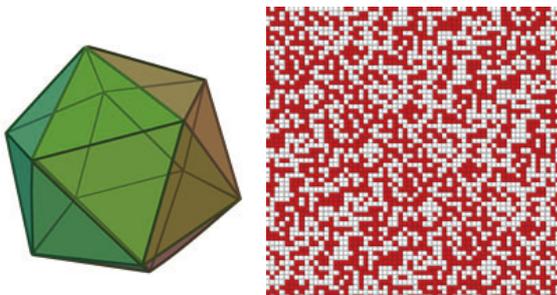
Остается сконструировать умножением по модулю 59 все  $n = 60$  элементов, опираясь на следующее матричное представление  $a, b, c$  и  $d$ :

$$A = \begin{pmatrix} 6 & 8 \\ 47 & 53 \end{pmatrix}; B = \begin{pmatrix} 17 & 51 \\ 51 & 42 \end{pmatrix};$$

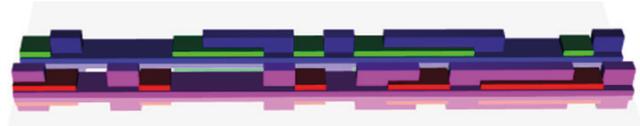
$$C = \begin{pmatrix} 51 & 9 \\ 33 & 7 \end{pmatrix}; D = \begin{pmatrix} 46 & 3 \\ 56 & 46 \end{pmatrix},$$

и проверить свойства  $c^3 = e, d^5 = e, (da)^3 = e^*, d^2a = cde^*$ , обеспечивающие качество сочетания элементов быть группой, инволюция  $e^* = a^4 = b^2, D = \{e^*, a, be^*, abe^*, c, ac, bc, abc, c^2e^*, ac^2, bc^2, abc^2, d, ad, bd, abd, cde^*, acde^*, bcd, abcd, c^2d, ac^2de^*, bc^2de^*, abc^2d, d^2e^*, ad^2e^*, bd^2e^*, abd^2, cd^2, acd^2e^*, bcd^2e^*, abcd^2e^*, c^2d^2, ac^2d^2e^*, bc^2d^2e^*, abc^2d^2e^*, d^3, ad^3, bd^3e^*, abd^3, cd^3, acd^3e^*, bcd^3, abcd^3e^*, c^2d^3e^*, ac^2d^3, bc^2d^3, abc^2d^3, d^4e^*, ad^4, bd^4e^*, abd^4, cd^4e^*, acd^4e^*, bcd^4, abcd^4e^*, c^2d^4e^*, ac^2d^4e^*, bc^2d^4, abc^2d^4\}$ .

Для нахождения обратной матрицы применяется обычный алгоритм инверсии матрицы, описанный еще Гауссом, но в конечном поле заданного модулем размера. Этот небольшой, но яркий набор примеров показывает богатство групп и возможных ассоциаций с ними. В нем отчетливо видны особенности идеи, к которой Ито позднее пришел,



■ Рис. 16. Икосаэдр и матрица Адамара с его симметриями  
 ■ Fig. 16. Icosahedron and Hadamard matrix with its symmetries



■ Рис. 17. Трансверсали D и Dg  
 ■ Fig. 17. Transversals D and Dg

подыскивая универсальную группу на роль такого описания: минимальная конструкция должна обходиться двумя генераторами, причем для ортогональных матриц характерна дихотомия в разделии элементов группы на равные по размерам подмножества, что может быть использовано как признак ортогональности.

Дициклическая группа  $G = \{e, a, a^2, \dots, a^{2v-1}, b, ab, a^2b, \dots, a^{2v-1}b\}$ , образованная двумя генераторами  $a$  и  $b$ , годится на роль такого посредника. Ее элементы имеют вид  $a^k b^m$ , и эта конструкция получена сращиванием пары циклических групп. Конечной ее делают соглашения:  $a^{2v} = b, a^{4v} = e, bab^{-1} = a^{-1}$ . Элемент  $e^* = b^2 = a^{2v}$  обладает свойством  $(e^*)^2 = e$ , это единственная центральная (перестановочная с любым элементом) инволюция группы (у вещественных чисел это корень квадратный).

Трансверсалью  $D$  группы  $G$  по отношению к  $\langle e^* \rangle$  (единичному элементу и инволюции) называется такая половина ее элементов, что  $\{D, De^*\}$  содержит все элементы группы, половинки не пересекаются. Если есть трансверсаль такая, что множества  $D$  и  $Dg$  наполовину пересекаются для любого элемента группы  $g$ , не совпадающего с  $\langle e^* \rangle$ , то такой набор называется *адамаровым* (рис. 17).

На рис. 17 зелеными и красными модулями представлено размещение элементов трансверсали на парной подложке группы согласно их порядковым номерам. Высокие детали изображения показывают размещение элементов  $Dg$ .

Алгоритм Ито вычисления бинегациклической матрицы Адамара состоит в формировании половины элементов циклической группы  $a^*$  двух наборов  $a^*, ba^*$  в связанных с ними последовательностях, формирующих бинегациклическую матрицу Адамара, изображенную на рис. 11. Значения 1 или -1 обозначают пересечение их элементов с элементами трансверсали. Все познается в сравнении: после громоздких построений, описанных ранее, дициклическая группа в модели Ито уже не кажется сложной.

### Теоретико-групповой подход

После 70-х годов XX века комбинаторная теория начала испытывать недостаток новых общих идей, способствующих ее развитию. Одна из та-

ких идей состояла в том, чтобы «увидеть» матрицу Адамара прямо в таблице, но не в таблице инцидентности графа, а, например, в «таблице умножения»  $G \times G$  группы, благодаря некоторому преобразованию. Так появились *коциклические* матрицы [21], которые разберем на основе часто приводимого примера.

Допустим, для синтеза матриц Адамара мы решили обойтись одной таблицей четверной группы Клейна  $V_4$ . Теперь рассмотрим выражение

$$\psi: G \times G \rightarrow \langle 1 \rangle$$

и поясним, что если  $G = V_4$ , операция « $\times$ » — это привычное сложение, а преобразование  $\psi$  переводит элементы таблицы в числа 1 и -1, содержащиеся в сете  $\langle 1 \rangle$ , по следующему правилу:

$$\psi(A) = \psi(B) = 1, \psi(C) = -1, \psi(D) = 1.$$

Легко проверить, что получающаяся таким образом из таблицы матрица  $\mathbf{H}$  будет удовлетворять уравнению  $\mathbf{H}^T \mathbf{H} = n\mathbf{I}$ , т. е. будет матрицей Адамара, представленной на рис. 18.

В случае матриц порядка 4 элементарными преобразованиями (перестановкой первых двух строк вниз) матрица  $\mathbf{H}$  сводится к циклической форме, в которой ее чаще всего изображают. В целом это направление скорее иллюстративное, чем практическое. Дальнейшее развитие этой идеи потребовало разбирательства в том, каким уравнениям должна удовлетворять вспомогательная функция  $\psi$ .

Уже с появлением матриц у операций сложения и умножения появилась двойная нагрузка. С одной стороны, обе эти операции активно используются для манипуляций с элементами матрицы, позволяя сконструировать операцию умножения матриц. С другой стороны, есть сложение и умножение матриц, обозначаемое теми же символами и с теми же (если исключить коммутативность умножения) правилами употребления.

+	A	B	C	D
A	1	1	-1	1
B	1	1	1	-1
C	-1	1	1	1
D	1	-1	1	1

■ *Рис. 18.* Коциклическая матрица Адамара  
 ■ *Fig. 18.* Cocsyclic Hadamard matrix

С тех пор нагрузка на смысловое содержание операций только растет.

Теперь, внимание, к этой операции умножения вида « $\times$ », реализованной внутренним сложением, можно добавить внешнее сложение  $g^2 + g^3$ , порождая операцию группового кольца. Эти два типа сложения сосуществуют и вместе образуют то, что мы привыкли организовывать иным путем. Из школьного курса известна реализация операции умножения из ранее определенного сложения, но здесь сложение появляется после того, как мы определились с тем, что будем понимать под умножением.

### Полиномиальные уравнения орнаментов

Разность пары элементов  $a - b$  в абстрактной алгебре интерпретируют как операцию, противоположную операции « $\times$ » группы. Соответственно, ее обозначают не так, как выше, а делением или умножением на обратный элемент  $ab^{-1}$  (*right quotient*). Столь замысловатое обозначение можно было бы отнести к причудам абстрактной теории и не обращать на нее внимание, не получи она в настоящее время широкое распространение в литературе по ортогональным матрицам [22, 23].

Парадокс в том, что теперь этой операции « $\times$ » недостаточно. К циклической группе элементов  $G = \{1, g, g^2, \dots\}$ , порожденных последовательными степенями примитивного элемента  $g$ , мы намерены добавить формальное сложение «+», выводящее на элементы группового кольца. Напомним, что «+» не позволяет пользоваться термином группа, у группы нет «суммы» ее элементов, кроме того, этот знак, как и у матриц, соотносим с надстраиваемым сверху групповым кольцом.

Переход на новый понятийный уровень позволяет элегантно записать условие Зингера для моноблока  $\mathbf{M}$  порядка  $v = n - 1$  матрицы

$$\mathbf{H} = \begin{pmatrix} -1 & \mathbf{e}^T \\ \mathbf{e} & \mathbf{M} \end{pmatrix} \text{ с каймой в виде одного уравнения}$$

$$D D^{(-1)} = k - \lambda + \lambda G \tag{1}$$

и впредь оперировать уравнениями такого вида, связывающими количество -1 в каждой строке  $k$  и количество значений -1 в каждой паре строк  $\lambda$  с деталями орнамента матрицы. Для ортогональных матриц  $\lambda = k - n/4$ , так что может фигурировать просто  $n/4$ .

**Пример 4.** В самом деле, обращаясь к рассмотренному в примере 1 дифференциальному набору  $D = \{1, 3, 4, 5, 9\}$ , без угрозы путаницы обозначений запишем его как элемент группового кольца  $D = g + g^3 + g^4 + g^5 + g^9$ . Произведение

с  $D^{(-1)} = g^{-1} + g^{-3} + g^{-4} + g^{-5} + g^{-9}$  порождает все описанные ранее разности показателей степеней по модулю  $v = 11$ . Одинаковые разности встречаются  $\lambda = 2$  раза, порождая  $\lambda G$  в (1) за вычетом не встречаемых комбинаций самого первого ее элемента  $g^0 = 1$  (вычитаем  $\lambda$ ). Кроме того, из-за  $k$  вычитаний у степеней  $g$  и  $g^{-1}$ ,  $g^3$  и  $g^{-3}$  и т. п. наберется-таки  $k = 5$  элементов  $g^0$ , которые также учтены полиномиальной формулой.

Помимо модели (1) с ее детальным описанием узора  $D$ , есть модель попроще в виде условия реализуемости узора (орнамента)  $k(k - 1) = \lambda(v - 1)$ , которое можно переписать в каноническом виде как

$$x^2 = 1, k = (v - x)/2. \quad (2)$$

Это условие действительно дает параметры в приведенном примере с  $v = 11, k = 5, \lambda = 2$ .

Правило, по которому детальная модель, например матрица, переходит к более простому ее описанию, например детерминанту матрицы, называется гомоморфизмом, причем алгоритм вычисления детерминанта фигурирует как описание этого соответствия. В отличие от изоморфизма, описывающего взаимно-однозначную связь моделей одной сложности, матрицу можно изменить так, чтобы детерминант ее не изменился. Обратное неверно, изменение детерминанта однозначно связано с тем, что изменилась матрица. То есть гомоморфизм — это путь от сложной модели к простой.

Для перехода от модели (1) к модели (2) построим нужный нам гомоморфизм, именуя его для простоты *характеристикой*  $D$  такой, что  $\underline{x} = \chi(D) = \chi(D^{(-1)})$  — разность между числом *нечетных* и *четных* степеней, так что  $\chi(G) = 0$ . Благодаря этому можно записать почти нужное нам уравнение вида  $\underline{x}^2 = n/4$ . Замена переменных  $n\underline{x} = 4\underline{x}$  порождает уравнение  $x^2 = 1$ , описывающее решения в форме корней из единицы.

При рассмотрении матриц Адамара  $\mathbf{H} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & -\mathbf{A}^T \end{pmatrix}$  порядка  $n = 2v$  полиномиальные уравнения, которые дублируют более простые квадратичные уравнения связи орнаментальных инвариантов, получили дополнительную работу. Условие реализуемости узора с параметрами  $v/2 = k_1 + k_2 - \lambda$  такой матрицы имеет малое различие с (1)

$$D_0 D_0^{(-1)} + D_1 D_1^{(-1)} = k_1 + k_2 - \lambda + \lambda G \quad (3)$$

для ортогональных матриц с  $\lambda = k_1 + k_2 - n/4$ , где  $v$  — размер блока основы, так что в правой части может фигурировать просто  $n/4$  в сумме с  $\lambda G$ .

**Пример 5.** Для матрицы Адамара порядка 16:  $v = 8, k_1 = 2, k_2 = 4, \lambda = 2, D_0 = \{1, 7\}, D_1 = \{2, 3, 5, 6\}$ , группа  $G$  размера  $v$  содержит все числа от 0 до  $v - 1 = 7$ . Плечи матрицы описываются тут двумя элементами группового кольца  $D_0 = g^1 + g^7, D_1 = g^2 + g^3 + g^5 + g^6$ , произведения  $D_0 D_0^{(-1)} = (g^1 + g^7)(g^{-1} + g^{-7}) = 2 + g^{-6} + g^6 = 2 + g^2 + g^6, D_1 D_1^{(-1)} = (g^2 + g^3 + g^5 + g^6)(g^{-2} + g^{-3} + g^{-5} + g^{-6}) = 4 + 2g^1 + g^2 + 2g^3 + 2g^4 + 2g^5 + g^6 + 2g^7$ . Их сумма равна элементу  $G = 1 + g + g^2 + g^3 + g^5 + g^6 + g^7$ , умноженному на  $\lambda = 2$ , за вычетом  $\lambda$  (при 1), тогда как первые члены 2 и 4 произведений дадут значение  $k_1 + k_2 = 6$ .

При сравнении полиномов мы сравниваем показатели степеней и коэффициенты полиномов отдельно, так что запись совокупности элементов группы 1,  $g, g^2, \dots$  в виде одного элемента группового кольца  $1 + g + g^2 + \dots$  открывает возможность одним уравнением описать цепочку равенств. Как и в матричной алгебре, мы сокращаем количество уравнений. Теперь поясним, зачем все это нужно.

Пойдем проторенным путем, учтя, что, помимо модели (3) с ее детальным описанием узора, есть модель проще в виде уравнения реализуемости орнамента, которое можно переписать в каноническом виде уравнения окружности

$$x^2 + y^2 = n, k_1 = (v - x)/2, k_2 = (v - y)/2, \quad (4)$$

причем для перехода от модели (1) к модели (2) годится построенный ранее нами гомоморфизм в форме характеристики, примененной к двум полиномиальным плечам  $\underline{x} = \chi(D_0) = \chi(D_0^{(-1)})$ ,  $\underline{y} = \chi(D_1) = \chi(D_1^{(-1)})$ , выражающей разности между числом *нечетных* и *четных* степеней в полиномах, так что  $\chi(G) = 0$ . На первом этапе получим уравнение окружности меньшего радиуса  $\underline{x}^2 + \underline{y}^2 = n/4$ , которое переходит в нужное нам удваиванием квадрата радиуса, так что  $x = 2\underline{x}, y = 2\underline{y}$ .

В частности, для указанного выше примера  $\underline{x} = \chi(D_0) = 2, \underline{y} = \chi(D_1) = 0$ , так что алгоритм работает:  $x^2 + y^2 = 16$ .

Причина, по которой с этим формализмом пришлось подробнее разбираться, состоит в том, что нам хотелось бы, чтобы блоки  $\mathbf{A}$  и  $\mathbf{B}$  были циклическими, ввиду чего уравнение (4) становится необходимым, но не достаточным условием существования бицикла. Иными словами, помимо этого уравнения должно возникнуть дополнительное, ограничивающее область определения.

Это сужение того, из чего делается выбор переменных, описывающих точки Гаусса (точки с целыми координатами  $x, y$ ) на окружности, лимитирующей количества элементов со значением  $-1$  в каждой строке двух блоков бицикла  $k_1 = (v - x)/2, k_2 = (v - y)/2$ , а оно не может быть

дробным, и доопределяет задачу так, как нам хотелось бы ее доопределить. Предварительно известно, что согласно так называемой рождественской теореме Ферма целое число вида  $n = 4t + 1$  всегда разложимо на сумму двух квадратов:  $5 = 1 + 2 \times 2$ ,  $9 = 0 + 3 \times 3$  и т. д., и целое число вида  $4t + 3$  неразложимо:  $3, 7, 11, \dots$

Соответственно, если порядок бицикла  $n$  имеет делители последнего вида (или их нечетные степени), то уравнение окружности с точками Гаусса на них составить невозможно. Сомнение в возможности существования бицикла возникает при наличии делителей в виде четных степеней простых чисел вида  $p = 3 \pmod{4}$ . Во всех остальных случаях решение определено существует. Вспомогательное уравнение  $x^2 + y^2 = v/2$  требует, кстати, чтобы каждое слагаемое в нем было либо 0, либо разностью двух квадратов. Например, для случаев  $v/2 = 9 = 0 + 9 = 0 + (25 - 16)$ ,  $v/2 = 18 = 9 + 9$  и т. п. Остается вспомнить, что абелева группа  $G$  размера  $p^{2t}u$  имеет силовскую  $p$ -подгруппу  $P$ , и нас интересует случай, когда эта подгруппа циклическа. Тогда цепочка проецирований, гомоморфизмов, описанных выше, удлиняется на одно звено от исходных полиномов  $D_0, D_1$  уравнения (3) к новым полиномам  $C_0, C_1$  задачи, в которой вследствие первого сжатия размера появляются отличные от единичных коэффициенты при слагаемых группового кольца. Например,  $C(\lambda G)$ , равные  $\lambda u P$ , удовлетворяющие уравнению

$$C(D_0)C(D_0)^{(-1)} + C(D_1)C(D_1)^{(-1)} = k_1 + k_2 - \lambda + \lambda u P \quad (5)$$

и далее привычным  $x = \chi(C_0)$ ,  $y = \chi(C_1)$  в первом уравнении из (4). В работе [23], где проложен этот путь, уравнение записывается в максимальной общей форме с комплексными переменными. Обычно чем менее мы сужаем решение задачи, тем проще добраться до сколь-нибудь вразумительного ответа. Прием погружения аргументов в комплексное пространство часто используется для упрощения вида уравнений, однако здесь это преимущество не ощущается.

Поэтому читатель при желании может заменять в ссылках на (4) в двух следующих леммах [23], имеющих в реферируемой литературе своих авторов, область определения аргумента с реальной на комплексную. Однако после такой замены (если она имеет иной смысл, помимо усложнения, не нужного задаче) при реальной правой части уравнение описывает гиперсферу, а не окружность, теряя иллюстрируемость рис. 3, призванного отражать корни из 1. Возможно, этот ритуал подчеркивает, что в работах на выяснение существования решения важен не столько рабочий вид гомоморфизма, сколько

отыскиваемый признак разрешимости задачи с квадратичным уравнением общего вида, заданным над множеством решений уравнения (5).

**Лемма 1.** Если правая часть уравнения окружности из (4) имеет делитель  $p^{2t}$ ,  $p = 3 \pmod{4}$  — простое число, то  $x$  и  $y$  делятся на  $p^t$ .

Такие пересчеты масштабов координат возникли у нас и ранее. Лемма доказывается от обратного. Выделим  $p^t$  из  $x$  и  $y$ , элементарные правила арифметики позволяют вынести появившийся ввиду квадрирования общий множитель  $p^{2t}$  за скобки, правая часть (4) тоже должна его содержать. Условие реверсивно, если она его содержит, обратный процесс перераспределит навязываемое  $p^t$  по составным частям.

**Лемма 2.** Пусть  $x$  и  $y$  делятся на  $p^t$ , тогда  $C_0$  и  $C_1$  можно описать аддитивно значениями  $p^t X + YP$ , где множители  $X$  и  $Y$ , как это свойственно элементам окружности, взяты на том же множестве всех возможных решений уравнения (5).

Доказательство приведено у Ма [24]. Этот несколько искусственный прием позволяет сформулировать следующую теорему.

**Теорема.** Пусть  $v = p^{2t}u$ ,  $p = 3 \pmod{4}$  — простое число,  $p$  и  $u$  взаимно просты, причем  $(v/2 \parallel p^{2e})$ , т. е.  $v/2$  содержит делитель  $p^{2e}$  (причем остаток уже не делится на  $p$ ), тогда уравнение окружности (4) разрешимо, если  $u \geq p^e$ .

Заметим, что по условиям формирования коэффициенты полиномов  $D_0$  и  $D_1$  не превосходят 1, их сведение к полиномам меньшей степени  $C_0 = C(D_0)$  и  $C_1 = C(D_1)$  дает весовые коэффициенты при степенях  $1, g, g^2, \dots$  примитивного элемента, не большие  $u$ .

Поскольку  $v/2$  содержит делитель  $p^{2e}$ , то, согласно первой лемме, элементы  $C_0$  и  $C_1$  содержат делитель с вдвое меньшим показателем  $p^e$ . Соответственно, согласно второй лемме, их можно перегруппировать к приспособленной для доказательства аддитивной форме  $C_0 = p^e X_0 + Y_0 P$ ,  $C_1 = p^e X_1 + Y_1 P$ . Эта форма удобна тем, что, умножив обе части ее на множитель  $1 - g$ , где  $g$  — примитивный элемент циклической подгруппы Силовой  $P$ , вторым слагаемым можно пренебречь:  $P(1 - g) = 0$ . Смещение составляющих  $1, g, g^2, \dots$  суммы умножением на примитивный элемент не меняет подгруппы. Отсюда имеем  $C_0(1 - g) = p^e X_0(1 - g)$ ,  $C_1(1 - g) = p^e X_1(1 - g)$ .

Поскольку коэффициенты полиномов  $C_0$  и  $C_1$  не превышают  $u$ , то при  $u < p^e$  правые части формул придется аннулировать выбором настраиваемых множителей  $X_0, X_1$ . Тогда  $C_0$  и  $C_1$  обретают признаки полиномов, описывающих подгруппы, так как выходит, что  $C_0 = C_0 g$ ,  $C_1 = C_1 g$ . Это, в свою очередь, невозможно, так как окажется, что характеристики  $\chi(C_0) = 0$ ,  $\chi(C_1) = 0$ , что противоречит условию, при котором точки решения берутся на окружности, а не в центре. Данный

вывод дает нам в руки желаемый формализм решения задач на существование бициклов. Для матриц Адамара размер  $n = 2v$  кратен 4, соответственно, для четного  $v = p^{2t}u$  задача неразрешима при  $u < 2p$ ,  $p \equiv 3 \pmod{4}$ .

**Пример 6.** Перейдем теперь к конкретному порядку. Пусть  $n = 2v = 36$ ,  $v = p^{2t}u = 18$ ,  $p = 3$ ,  $t = 1$ ,  $u = 2$ , абелева группа  $G$  имеет  $p$ -подгруппу Силова размера  $p^2 = 9$ . При  $u = 2 < 2p = 2 \times 3 = 6$  задача неразрешима, причем удвоение размера бицикла до порядка 72 с  $u = 4$ , как видно, тоже ничего не дает. Бицикл порядка 144 реализуем и найден путем компьютерных экспериментов [25].

## Заключение

Теоретико-групповой подход позволяет находить матрицы высоких порядков, опираясь на неявно выраженные симметрии, описываемые группами. Это обстоятельство позволяет строить вычислительные алгоритмы, отличные по эффективности от переборных алгоритмов.

Неразрешимость задачи на частном примере не отвергает существование матрицы Адамара.

## Литература

1. Davenport H. *An Introduction to the Theory of Numbers*. Harper & Brothers, New York, 1952.
2. Manfred Schroeder. *Number Theory in Science and Communication*. Springer-Verlag, Berlin-Heidelberg, 2009. 431 p.
3. Манин Ю. И., Панчишкин А. А. Введение в теорию чисел. *Итоги науки и техники. Сер. Современные проблемы математики. Фундаментальные направления*, 1990, т. 49, с. 5–341.
4. Ball Rouse W. W. *A Short Account of the History of Mathematics*. NY, Dover Publications, 2001. 439 p.
5. Балонин Н. А. *Новый курс теории управления движением*. СПб.: СПбГУ, 2000. 160 с.
6. Hadamard J. Résolution d'une Question Relative aux Déterminants. *Bulletin des Sciences Mathématiques*, 1893, vol. 17, pp. 240–246.
7. Seberry J., Yamada M. Hadamard matrices, sequences and block designs. In: *Contemporary Design Theory: A Collection of Surveys*. Eds. J. H. Dinitz and D. R. Stinson. J. Wiley, New York, 1992. Pp. 431–560.
8. Scarpis U. Sui determinanti di valore Massimo. *Rendiconti della R. Istituto Lombardo di Scienze e Lettere*, 1898, ser. 2, vol. 31, fascicolo 20, pp. 1441–1446.
9. Singer J. A Theorem in finite projective geometry and some applications to number theory. *Transactions of the American Mathematical Society*, 1938, vol. 43, pp. 377–385.
10. Balonin N. A., Seberry Jennifer. Two level Cretan matrices constructed via Singer difference sets. *Информационно-управляющие системы*, 2014, № 6, с. 2–5.
11. Балонин Н. А., Сергеев М. Б. Матрицы Мерсенна и Адамара, произведения. *Информационно-управляющие системы*, 2016, № 5, с. 2–14, doi.org/10.15217/issn1684-8853.2016.5.2.
12. Paley R. E. A. C. On orthogonal matrices. *Journal of Mathematics and Physics*, 1933, vol. 12, pp. 311–320.
13. Балонин Н. А., Сергеев М. Б. Нормы обобщенных матриц Адамара. *Вестник СПбГУ. Сер. 10. Прикладная математика. Информатика. Процессы управления*, 2014, вып. 2, с. 5–11.
14. Балонин Н. А., Сергеев М. Б., Себерри Дж., Синицына О. И. Окружности на решетках и матрицы Адамара. *Информационно-управляющие системы*, 2019, № 3, с. 2–9. doi.org/10.31799/1684-8853-2019-3-2-9
15. Балонин Н. А., Сергеев М. Б., Себерри Дж., Синицына О. И. Окружности на решетках и матрицы максимального детерминанта. *Информационно-управляющие системы*, 2020, № 6, с. 2–11. doi.org/10.31799/1684-8853-2020-6-2-11
16. Szekeres G. Tournaments and Hadamard matrices. *L'Enseignement Math*, 1969, vol. 15, pp. 269–278.
17. Szekeres G. Cyclotomy and complementary difference sets. *Acta Arithmetica*, 1971, vol. 18, pp. 349–353. doi:10.4064/aa-18-1-349-353
18. Goethals J.-M., Seidel J. J. Orthogonal matrices with zero diagonal. *Canad. J. Math.*, 1967, vol. 19, pp. 1001–1010. doi:10.4153/CJM-1967-091-8

Матричная алгебра не располагает простыми средствами (пока так видится) определять, на каких порядках циклический орнамент блоков матрицы входит в противоречие с условием ортогональности вида  $A^T A + B^T B = nI$ . Однако судить об этом позволяет теоретико-групповой подход.

## Благодарности

Авторы выражают искреннюю благодарность профессорам Дж. Себерри и Д. Джоковичу за ценные советы и вспомогательные примеры, облегчившие нашу работу.

Кроме того, мы выражаем признательность Т. В. Балониной за техническую подготовку текста этой статьи.

## Финансовая поддержка

Статья подготовлена при финансовой поддержке Министерства науки и высшего образования Российской Федерации, соглашение № FSRF-2020-0004.

19. Ito N. Note on Hadamard groups of quadratic residue type. *Hokkaido Mathematical Journal*, 1993, vol. 22, pp. 373–378.
20. Ito N. On Hadamard groups III. *Kyushu J. Math.*, 1997, no. 51, pp. 369–379.
21. Horadam K. J. Hadamard matrices and their applications: Progress 2007–2010. *Cryptography and Communications*, 2010, no. 2, iss. 2, pp. 129–154.
22. Arasu K. T., Xiang Q. On the existence of periodic complementary binary sequences. *Designs, Codes and Cryptography*, 1992, vol. 2, pp. 257–262. doi:10.1007/BF00141970
23. Egan Ronan. On equivalence of negaperiodic Golay pairs. *Designs, Codes and Cryptography*, 2017, vol. 85, pp. 523–532. doi:10.1007/s10623-016-0320-6
24. Ma S. L. *Polynomial addition sets*. University of Hong Kong, 1985. 120 p. doi:10.5353 / TH\_B3123054
25. Балонин Н. А., Джокович Д. Ж. Симметрия двуциклических матриц Адамара и периодические пары Голея. *Информационно-управляющие системы*, 2015, № 3, с. 2–16. doi.org/10.15217/issn1684-8853.2015.3.2

UDC 519.614

doi:10.31799/1684-8853-2021-4-2-17

### Finite field and group algorithms for orthogonal sequence search

N. A. Balonin<sup>a</sup>, Dr. Sc., Tech., Professor, orcid.org/0000-0001-7338-4920, korbendfs@mail.ru

A. M. Sergeev<sup>a</sup>, PhD, Tech., Associate Professor, orcid.org/0000-0002-4788-9869

O. I. Sinitsyna<sup>a</sup>, Post-Graduate Student, orcid.org/0000-0002-2819-4682

<sup>a</sup>Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaja St., 190000, Saint-Petersburg, Russian Federation

**Introduction:** Hadamard matrices consisting of elements 1 and  $-1$  are an ideal object for a visual application of finite dimensional mathematics operating with a finite number of addresses for  $-1$  elements. The notation systems of abstract algebra methods, in contrast to the conventional matrix algebra, have been changing intensively, without being widely spread, leading to the necessity to revise and systematize the accumulated experience. **Purpose:** To describe the algorithms of finite fields and groups in a uniform notation in order to facilitate the perception of the extensive knowledge necessary for finding orthogonal and suborthogonal sequences. **Results:** Formulas have been proposed for calculating relatively unknown algorithms (or their versions) developed by Scarpis, Singer, Szekeres, Goethal — Seidel, and Noboru Ito, as well as polynomial equations used to prove the theorems about the existence of finite-dimensional solutions. This replenished the significant lack of information both in the domestic literature (most of these issues are published here for the first time) and abroad. **Practical relevance:** Orthogonal sequences and methods for their effective finding via the theory of finite fields and groups are of direct practical importance for noise-immune coding, compression and masking of video data.

**Keywords** — Hadamard matrices, orthogonal matrices, finite dimensional mathematics, Galois fields, finite groups, Scarpis algorithm, Singer approach, Szekeres method, Goethal — Seidel algorithm, Noboru Ito approach.

**For citation:** Balonin N. A., Sergeev A. M., Sinitsyna O. I. Finite field and group algorithms for orthogonal sequence search. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 2–17 (In Russian). doi:10.31799/1684-8853-2021-4-2-17

### References

- Davenport H. *An Introduction to the Theory of Numbers*. Harper & Brothers, New York, 1952.
- Manfred Schroeder. *Number Theory in Science and Communication*. Springer-Verlag, Berlin-Heidelberg, 2009. 431 p.
- Manin Yu. I., Panchishkin A. A. Introduction to number theory, *Itogi Nauki i Tekhniki. Ser. Sovrem. Probl. Mat. Fund. Napr.*, 1990, vol. 49, pp. 5–341.
- Ball Rouse W. W. *A Short Account of the History of Mathematics*. NY, Dover Publications, 2001. 439 p.
- Balonin N. A. *Novyj kurs teorii upravleniya dvizheniem*. [New Course in Motion Control Theory]. Saint-Petersburg, SPbGU Publ., 2000. 160 p. (In Russian).
- Hadamard J. Résolution d'une Question Relative aux Déterminants. *Bulletin des Sciences Mathématiques*, 1893, vol. 17, pp. 240–246 (In French).
- Seberry J., Yamada M. *Hadamard matrices, sequences and block designs*. In: *Contemporary Design Theory: A Collection of Surveys*. Eds. J. H. Dinitz and D. R. Stinson. J. Wiley, New York, 1992. Pp. 431–560.
- Scarpis U. Sui determinanti di valore Massimo. *Rendiconti della R. Istituzione Lombardo di Scienze e Lettere*, 1898, ser. 2, vol. 31, fascicolo 20, pp. 1441–1446 (In Italian).
- Singer J. A theorem in finite projective geometry and some applications to number theory. *Transactions of the American Mathematical Society*, 1938, vol. 43, pp. 377–385.
- Balonin N. A., Seberry Jennifer. Two level Cretan matrices constructed via Singer difference sets. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2014, no. 6, pp. 2–5.
- Balonin N. A., Sergeev M. B. Mersenne and Hadamard Matrices, Products. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2016, no. 5, pp. 2–14 (In Russian). doi.org/10.15217/issn1684-8853.2016.5.2
- Paley R. E. A. C. On orthogonal matrices. *Journal of Mathematics and Physics*, 1933, vol. 12, pp. 311–320.
- Balonin N. A., Sergeev M. B. The generalized Hadamard matrix norms. *Vestnik S.-Petersburg Univ. Ser. 10. Prikl. Mat. Inform.*, 2014, iss. 2, pp. 5–11 (In Russian).
- Balonin N. A., Sergeev M. B., Seberry J., Sinitsyna O. I. Circles on lattices and Hadamard matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2019, no. 3, pp. 2–9 (In Russian). doi.org/10.31799/1684-8853-2019-3-2-9
- Balonin N. A., Sergeev M. B., Seberry J., Sinitsyna O. I. Circles on lattices and maximum determinant matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 6, pp. 2–11 (In Russian). doi.org/10.31799/1684-8853-2020-6-2-11
- Szekeres G. Tournaments and Hadamard matrices. *L'Enseignement Math.*, 1969, vol. 15, pp. 269–278.

17. Szekeres G. Cyclotomy and complementary difference sets. *Acta Arithmetica*, 1971, vol. 18, pp. 349–353. doi:10.4064/aa-18-1-349-353
18. Goethals J.-M., Seidel J. J. Orthogonal matrices with zero diagonal. *Canad. J. Math.*, 1967, vol. 19, pp. 1001–1010. doi:10.4153/CJM-1967-091-8
19. Ito N. Note on Hadamard groups of quadratic residue type. *Hokkaido Mathematical Journal*, 1993, vol. 22, pp. 373–378.
20. Ito N. On Hadamard groups III. *Kyushu J. Math.*, 1997, no. 51, pp. 369–379.
21. Horadam K. J. Hadamard matrices and their applications: Progress 2007–2010. *Cryptography and Communications*, 2010, no. 2, iss. 2, pp. 129–154.
22. Arasu K. T., Xiang Q. On the existence of periodic complementary binary sequences. *Designs, Codes and Cryptography*, 1992, vol. 2, pp. 257–262. doi:10.1007/BF00141970
23. Egan Ronan. On equivalence of negaperiodic Golay pairs. *Designs, Codes and Cryptography*, 2017, vol. 85, pp. 523–532. doi:10.1007/s10623-016-0320-6
24. Ma S. L. *Polynomial Addition Sets*. University of Hong Kong, 1985. 120 p. doi:10.5353 / TH\_B3123054
25. Balonin N. A., Djokovic D. Z. Symmetry of two-circulant Hadamard matrices and periodic Golay pairs. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 3, pp. 2–16. doi.org/10.15217/issn1684-8853.2015.3.2

### УВАЖАЕМЫЕ АВТОРЫ!

Научная электронная библиотека (НЭБ) продолжает работу по реализации проекта SCIENCE INDEX. После того как Вы регистрируетесь на сайте НЭБ (<http://elibrary.ru/defaultx.asp>), будет создана Ваша личная страничка, содержание которой составят не только Ваши персональные данные, но и перечень всех Ваших печатных трудов, имеющих в базе данных НЭБ, включая диссертации, патенты и тезисы к конференциям, а также сравнительные индексы цитирования: РИНЦ (Российский индекс научного цитирования), h (индекс Хирша) от Web of Science и h от Scopus. После создания базового варианта Вашей персональной страницы Вы получите код доступа, который позволит Вам редактировать информацию, помогая создавать максимально объективную картину Вашей научной активности и цитирования Ваших трудов.

## Indexing algorithm based on storing additional distances in metric space for multi-vantage-point tree

V. V. Fomin<sup>a</sup>, Dr. Sc., Tech., Professor, [orcid.org/0000-0001-7040-5386](https://orcid.org/0000-0001-7040-5386), [v\\_v\\_fomin@mail.ru](mailto:v_v_fomin@mail.ru)

I. V. Aleksandrov<sup>b</sup>, Post-Graduate Student, [orcid.org/0000-0002-6258-8191](https://orcid.org/0000-0002-6258-8191)

<sup>a</sup>Herzen State Pedagogical University of Russia, 48, Moika Emb., 191186, Saint-Petersburg, Russian Federation

<sup>b</sup>The Bonch-Bruевич Saint-Petersburg State University of Telecommunications, 22-1, Bolshhevikov Pr., 193232, Saint-Petersburg, Russian Federation

**Introduction:** The similarity search paradigm is used in various computational tasks, such as classification, data mining, pattern recognition, etc. Currently, the technology of tree-like metric access methods occupies a significant place among search algorithms. The classical problem of reducing the time of similarity search in metric space is relevant for modern systems when processing big complex data. Due to multidimensional nature of the search algorithm effectiveness problem, local research in this direction is in demand, constantly bringing useful results. **Purpose:** To reduce the computational complexity of tree search algorithms in problems involving metric proximity. **Results:** We developed a search algorithm for a multi-vantage-point tree, based on the priority node-processing queue. We mathematically formalized the problems of additional calculations and ways to solve them. To improve the performance of similarity search, we have proposed procedures for forming a priority queue of processing nodes and reducing the number of intersections of same level nodes. Structural changes in the multi-vantage-point tree and the use of minimum distances between vantage points and node subtrees provide better search efficiency. More accurate determination of the distance from the search object to the nodes and the fact that the search area intersects with a tree node allows you to reduce the amount of calculations. **Practical relevance:** The resulting search algorithms need less time to process information due to an insignificant increase in memory requirements. Reducing the information processing time expands the application boundaries of tree metric indexing methods in search problems involving large data sets.

**Keywords** – complex data indexing, tree data structures, metric access methods, priority queue of nodes, multi-vantage-point tree.

**For citation:** Fomin V. V., Aleksandrov I. V. Indexing algorithm based on storing additional distances in metric space for multi-vantage-point tree. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 18–27. doi:10.31799/1684-8853-2021-4-18-27

### Introduction

When developing an applied analytical system [1] based on metric methods of artificial intelligence (methods of decision functions), the authors faced the classic problem of computational complexity. The task was to increase the speed of data processing by modifying the indexing algorithms. The emphasis is on modification of tree search algorithms as part of the task of overlapping tree nodes in order to reduce number of operations for calculating distances between objects. They use metric methods (K nearest neighbor's method, method of standards, method of correction increments, and others) to solve similarity search problems in practice. Distinctive features of these methods are: 1) any object from data set  $D$  is a point in space; 2) during the search, distances between two objects are repeatedly calculated; 3) distance processing takes place in metric space.

Let  $X$  — an arbitrary set. The function  $\text{dist}$  is called a metric on the  $X$ , if for all  $x, y, z \in X$  the following conditions are satisfied: a)  $\text{dist}(x, y) \geq 0$  (non negativity); b)  $\text{dist}(x, y) = 0$ , when  $x = y$  (identity to itself); c)  $\text{dist}(x, y) = \text{dist}(y, x)$  (symmetry); d)  $\text{dist}(x, y) \leq \text{dist}(x, z) + \text{dist}(z, y)$  (triangle inequality).

The presence of the triangle inequality (fourth property — subparagraph  $d$ ) is inherent in its metric spaces. Therefore, the authors in the article did not consider datasets in which the triangle inequality does not performed. At the same time, we note that there is a study on the modification of the VP-tree for non-metric spaces using the Bregman distance, which does not satisfy the triangle inequality principle [2].

To find the nearest object among the set of data  $D$  in metric space, data indexing technology metric access methods (MAM) is used [3, 4]. MAM uses a special data structure to speed up the search for the nearest object in database  $D$  by reducing number of calculated distances.

We will highlight the following features of MAM technology:

1) the presence of a metric (distance function) for any two objects [5];

2) during the search, MAM algorithms use the principle of triangle inequality [6], which makes it possible to exclude the calculation of additional distances;

3) the computational complexity of the search algorithm depends significantly on the number of calculated distances between two objects [3].

There are many types of MAM tree algorithms, including GHT [7], D-index [8], BP [9], MM-tree [10], Onion-Tree [11] and others. One of the first MAM algorithms successfully applied in practice was the M-Tree [12–14]. Among the disadvantages of M-Tree, we highlight the problem of overlapping tree nodes, which increases the number of distance calculations. An alternative approach to using vantage points to the M-Tree is the OMNI algorithm [15]. OMNI uses strategically positioned vantage points to increase efficiency. Another classical algorithm Geometric Near Access Tree (GNAT) [16] suggests taking into account not only the minimum and maximum distances to objects, but also the distances between pairs of vantage points during the search.

A characteristic feature of the M-Tree, OMNI, and GNAT algorithms reviewed is the problem of overlapping nodes of the same level with each other. We can highlight the following M-Tree modifications to solve this problem:

— Slim-tree [17], which reduces the number of intersections of nodes of the same level with each other by using the slim-down algorithm. The disadvantage of the algorithm used is the need to re-insert objects;

— PM-tree [18], which combines a compact partition and a set of reference points. In the PM-tree, one can form the areas of nodes by the intersection of multiple hyperspheres defined by a global set of vantage points.

An alternative to the M-Tree family of algorithms is the VP-tree [19–21] and its modification with several reference points in each node — the multi-vantage-point (MVP) tree [22, 23]. The features of the structures are the balance tree based on the median data partition and the absence of intersection of nodes of the same level. The development of VP-tree ideas is the algorithm Fixed-Queries Tree (FQT) [24], which suggests using a single vantage point for all nodes of the same level to reduce calculations. The absence of the problem of intersection of nodes of the same level has a significant impact on the search efficiency. Thus, research [25] has shown in practice that searching based on disjoint nodes in an MVP-tree is more efficient than in an M-tree.

Currently, there are theoretical studies on the use of VP / MVP-trees as a basic component to improve the efficiency of various search methods [19, 21, 23]. Experimental results in [26, 27] confirmed the competitiveness of VP / MVP-trees compared to other metric data indexing methods. Analysis of the results obtained in practice for VP / MVP-trees proves their efficiency in metric spaces of small, medium [26] and large [20] dimensions.

Now, there is a demand in the development of search improvements for VP / MVP-trees, which is

confirmed by various modern publications [20, 28]. Therefore, the research of reducing the number of calculated distances between two objects (search complexity) for the MVP-tree when solving problems of searching the nearest objects is relevant.

One of the ways to improve the search efficiency in the MVP-tree is to use a priority queue for processing nodes. As part of the adaptation of the priority queue to the MVP-tree, the authors have developed an algorithm for determining the intersection of the search area with the subtrees of a node: determination of intersection based on hierarchic processing of nodes (DIHPN).

### DIHPN algorithm

Feature considered MVP-tree is logarithmic nearest neighbor search complexity depending on the number of objects in the original data set  $D$ . Without loss of generality, consider and analyze a search algorithm for a classical MVP-tree with the following characteristics:

- 1)  $U = 2$  — branching of one vantage point;
- 2)  $Y = 2$  — number of vantage points in one node of the tree;
- 3)  $V = \{v1, v2\}$  — set of vantage points in each node of the tree, where  $v1$  — the primary point,  $v2$  — secondary;
- 4)  $B = U^Y = 4$  — number of children for each node of the tree;
- 5)  $A = \{A1, A2, A3, A4\}$  — set of subtrees in each node of the tree;
- 6)  $R = \{r1, r2, r3\}$  — set of median distances (radii) that defines the boundaries of the four regions of the node:  $A1 \in (v1, r1) \ \&\& \ A1 \in (v2, r2)$ ,  $A2 \in (v1, r1) \ \&\& \ A2 \notin (v2, r2)$ ,  $A3 \notin (v1, r1) \ \&\& \ A3 \in (v2, r3)$ ,  $A4 \notin (v1, r1) \ \&\& \ A4 \notin (v2, r3)$ , where  $(v, r)$  — hypersphere centered at point  $v$  and radius  $r$ ;
- 7)  $K = 1$  — branching of each leaf;
- 8)  $P = 0$  — number of stored distances in the leaves from each leaf point to the set of vantage points.

The research [29] have identified the following features of the classical algorithm for finding the nearest object  $n$  for the query point  $q$  in the MVP-tree:

1. The search for the nearest object is recursive. During each iteration, it directly processes the data of only one node of the tree.
2. The node processing queue is formed based on the depth-first search principle. When a node is processed, the search area is constrained by the current node and all of its parent nodes. Each node implicitly takes into account its own set of hierarchical search rules.
3. Due to the architecture of building an MVP-tree, the descendants of  $A1, A2, A3$ , and  $A4$  with-

in the same parent node do not intersect with each other.

4. Taking into account the set of hierarchical search rules and the architecture of building an MVP-tree allows you to set narrower boundaries for the search area.

5. The set of points  $D$  of the node  $Node$  of the tree  $T_{MVP}$  is defined by the intersection of two hyperspheres with centers at  $\{v1, v2\}$  and the corresponding two radii from  $R = \{r1, r2, r3\}$  of the parent node Parent.

6. Within the framework of parent node Parent, the fourth area  $A4$  has only minimal borders.

7. The MVP-tree is a balanced tree.

Research [30] suggests using the priority queue  $C$  for processing nodes to improve efficiency of the classical search for the nearest object. Sorting nodes by increasing the distance from them to the query point  $q$  allows you to set the priority of processing areas. During research, when using the priority queue  $C$  in the search for the closest object for an MVP-tree, we identified the problem of additional computations. The emergence of additional calculations is due to the need to recreate the set of hierarchical search rules for each node. The nodes extracted from the priority queue for processing can belong to different parts and levels of the MVP-tree. Note that not taking into account the set of hierarchical search rules can lead to an increase in the number of overlapping nodes. Let us introduce the following notation and characteristics:

- 1)  $D$  — input set of points for indexing;
- 2)  $N^D_{points}$  — number of points in a  $D$ ;
- 3)  $H$  — number of levels in the tree;
- 4)  $W$  — number of nodes at all levels in the tree;
- 5)  $L_k$  —  $k$ -th level in the tree.

Based on the notation introduced, we estimate influence of the number of overlapping nodes in the MVP-tree on the search efficiency. In previous author's studies [31], the following dependencies and estimates are given:

1) maximum estimate of the number of intersecting nodes  $NIL_k$  from the number  $k$  of the level  $L_k$  of the tree, when the set of hierarchical search rules is not taken into account:

$$NIL(k) = B^{2k}, \quad (1)$$

where  $B = U^Y = 4$  — number of children for each node in the tree;

2) based on formula (1), dependence of the total number of intersecting nodes NIT for the entire MVP-tree was calculated and presented as follows:

$$NIT \approx B^H \sim N^D_{points}. \quad (2)$$

Studies have derived the dependence of the efficiency of finding the nearest object to the point  $q$

with the search radius  $r_q$  for tree structures MAM [3]:

$$O(NN(q, r_q)) \approx \frac{1}{r_0} \sum_{k=0}^{H-1} \left(1 + fat(T_{MVP})(N^k_{nodes} - 1)\right) \times \left(N^D_{points}\right)^{\frac{k}{H}} \times \left(r^k_{aver} + r_q\right)^\Omega, \quad (3)$$

where  $r_0$  — radius required to cover the entire set of points at the root of the tree  $T_{MVP}$ ;  $\Omega \in R$  — internal dimension of the space;  $fat(T_{MVP})$  — absolute fat factor of the tree  $T_{MVP}$ , which characterizes the degree of intersection of nodes with each other;  $N^k_{nodes}$  — average number of nodes at the  $k$ -th level of the tree;  $r^k_{aver}$  — average radius of the node coverage at the  $k$ -th level, which is calculated by the formula:

$$r^k_{aver} = \sqrt[\Omega]{\left(N^D_{points}\right)^{\frac{-k}{H}}}. \quad (4)$$

The formula for calculating the parameter absolute fat factor [3] is as follows:

$$fat(T_{MVP}) = \frac{I_C - H \times N^D_{points}}{N^D_{points}} \times \frac{1}{(W - H)}, \quad (5)$$

where  $I_C$  displays the total number of accesses to the tree nodes required to respond to a point request for each of the  $N^D_{points}$  objects in the tree  $T_{MVP}$ :

$$I_C \sim NIT. \quad (6)$$

Taking into account the architectural features of the MVP-tree and formula (4), we rewrite formula (3) for the MVP-tree as the following dependency:

$$O(NN(q, r_q)) \sim fat(T_{MVP}). \quad (7)$$

Based on the fact that the range for the absolute factor  $fat(T_{MVP})$  is  $[0..1]$  and analysis of formula (5), we get the dependence:

$$fat(T_{MVP}) \sim I_C \sim NIT. \quad (8)$$

From formulas (1)–(8) it follows:

1) if the set of hierarchical search rules is not taken into account, there is a problem of overlapping nodes with each other;

2) the number of intersecting nodes within the entire tree depends on the number of indexed  $N^D_{points}$  in  $D$ ;

3) search efficiency is inversely proportional to the number of intersecting nodes.

As part of solving the problem of additional computations, we propose to use a modification of

the classical algorithm of the process of determining the fact of intersection of the search area with the current node Node. The following logical statement determines the fact that a node Node of type Type intersects with the search area ( $q, ln$ ) based on the classical method:

$$overlap(Node, (q, ln)) = \left( \begin{array}{l} \left( (Type = A1) \&\& ((ln + r1) \geq l1) \&\& ((ln + r2) \geq l2) \right) \parallel \\ \left( (Type = A2) \&\& ((ln + r1) \geq l1) \&\& ((ln + l2) > r2) \right) \parallel \\ \left( (Type = A3) \&\& ((ln + r3) \geq l2) \&\& ((ln + l1) > r1) \right) \parallel \\ \left( (Type = A4) \&\& ((ln + l1) > r1) \&\& ((ln + l2) > r2) \right) \parallel \end{array} \right), \quad (9)$$

where  $l1$  — distance from point  $q$  to point Node.v1;  $l2$  — distance from point  $q$  to point Node.v2.

Additional checks for intersections of the search area with parent nodes for Node at all higher levels of the tree is the center of the proposed modification — the algorithm DIHPN (Fig. 1). The algorithm DIHPN explicitly recreates the hierarchy of search rules, which allows you to set narrower boundaries of nodes and avoid overlapping nodes of the same level with each other. We present a generalized dependence of the number of overlapping nodes on the number of the current level  $k$  and the number of processed higher levels  $t$ :

$$NIL(k, t) = (B^{k+1-t})^2. \quad (10)$$

Based on formula (10), we concluded that an increase in the number of processed higher-level parents (parameter  $t$ ) leads to a decrease in the number of intersecting nodes in the entire tree. In the limiting case of the formula (10), when the algorithm requests all the higher parents for the current node, the number of intersecting nodes is equal to one (the root). Because no two nodes of the same level intersect each other due to the architectural.

The disadvantage of the DIHPN algorithm is the increase in the number of calculated distances between two objects. Determining the distances from the query point to the vantage points of the parent nodes is the reason for the increase in the computational complexity of the proposed algorithm. We can solve the problem by storing the previously calculated results in additional memory  $L$  in order to avoid repeated calculations.

As part of improving the efficiency of the DIHPN algorithm, we need to solve several problems:

1) the problem of effectively determining the distances from the query point  $q$  to the subtrees of node  $A1, A2, A3, A4$ . Note that based on the distance to the query point  $q$ , the priority of the node in queue  $C$  is calculated. To improve the efficiency of distance determination, the authors developed the algorithm: determination distance from request point to node area (DDRPNA);

2) the problem of increasing the efficiency of determining the fact of intersection of nodes with the search area [formula (1)]. Determining the fact of intersection is the basis of the procedure for adding a node to the priority queue  $C$ . We propose to modify the classic rules of intersection of MVP-tree regions [formula (9)] based on the use of minimum distances: modification of the rules for determining the intersection of nodes based on minimum distances (MRDINMD).

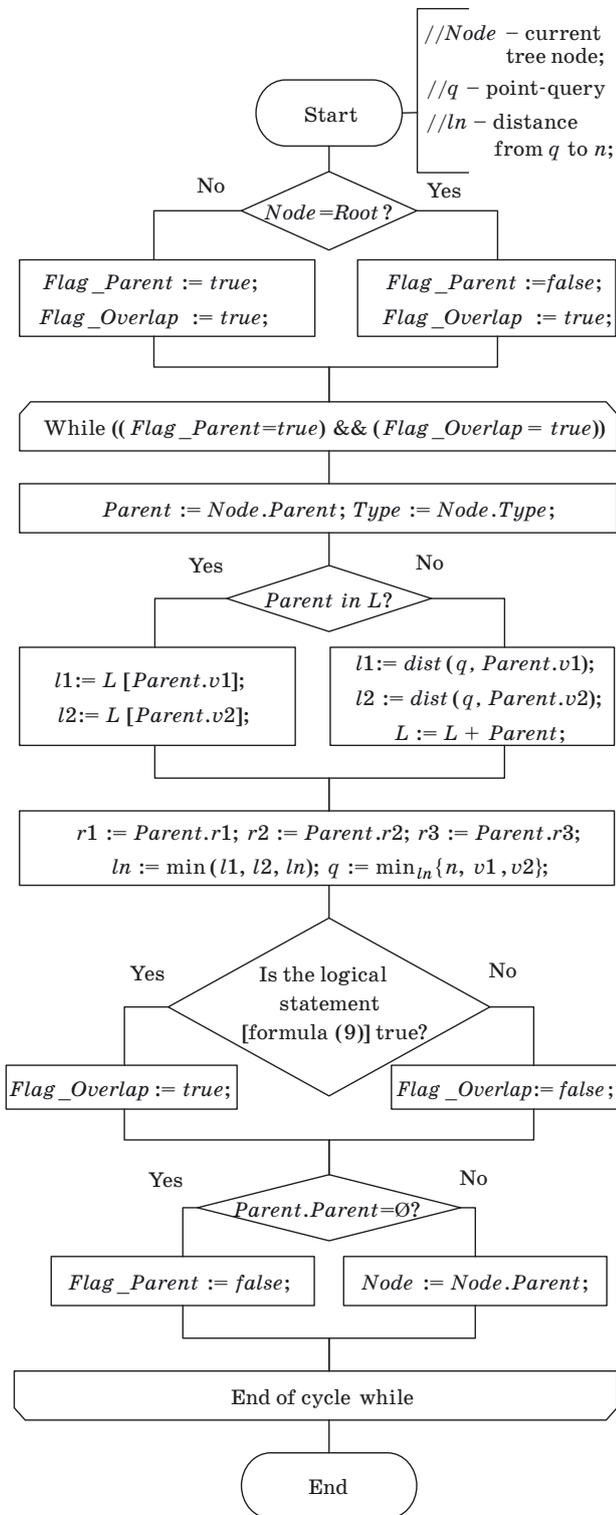
## DDRPNA algorithm

To assign priorities that are more precise to nodes in queue  $C$  based on the distance from the query point  $q$  to the subtrees of the MVP-tree node, we suggest to use the author's DDRPNA algorithm. The problem appears by using only two reference points  $\{v1, v2\}$  for four regions. Thus, the distances can only be determined to areas  $A3$  and  $A1/A2$ . The choice of a specific second area depends on which region  $A1$  or  $A2$  the vantage point  $v1$  belongs. Enter the following notation:

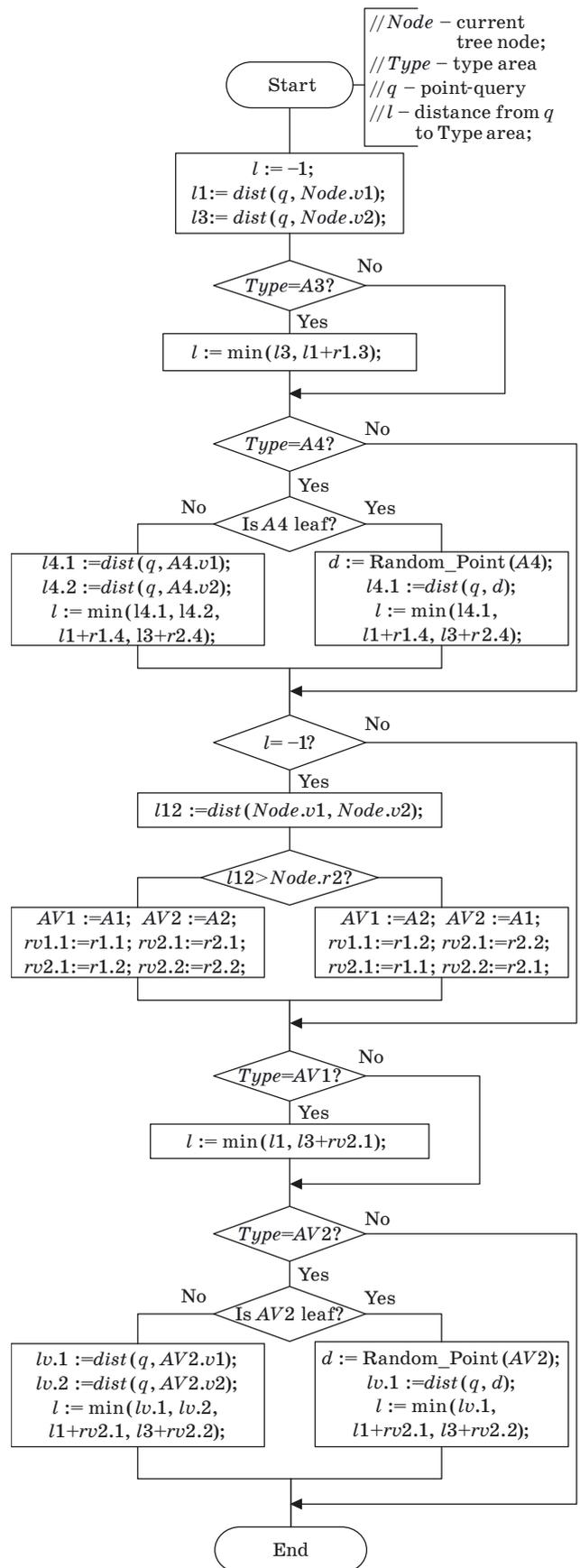
- 1)  $AV1$  — a region of  $A12 = A1 + A2$ , which belongs to the vantage point  $v1$ ;
- 2)  $AV2$  — the area of  $A12$  that does not contain the reference point  $v1$ , i. e.  $AV2 = A12 \cap AV1$ .

To determine the distance from the query point  $q$  to the subtree Area of the node Node, the authors propose to use the following algorithm:

- 1) if the subtree Area is an area Node.AV1 and an internal node of the tree:
  - determine the distance  $l1.1$  from the query point  $q$  to the primary vantage point of the node Node.v1;
  - set the desired distance  $l1$  as the distance  $l1.1$ ;



■ Fig. 1. Flow chart of DIHPN algorithm



■ Fig. 2. Flow chart of DDRPNA algorithm

- 2) if the subtree Area is an area Node.AV2 and an internal node of the tree:
  - determine the distance  $l2.1$  from the query point  $q$  to the primary vantage point of the node Node.v1;
  - calculate the distance  $l2.2$  from the query point  $q$  to the secondary vantage point of the node Area.v2;
  - select the minimum value among  $(l2.1, l2.2)$  as the desired distance  $l2$ ;
- 3) if the subtree Area is an area Node.A3 and an internal node of the tree:
  - determine the distance  $l3.1$  from the query point  $q$  to the primary vantage point of the node Node.v2;
  - set the desired distance  $l3$  as the distance  $l3.1$ ;
- 4) if the subtree Area is an area Node.A4 and an internal node of the tree:
  - determine the distance  $l4.1$  from the query point  $q$  to the primary vantage point of the node Area.v1;
  - calculate the distance  $l4.2$  from the query point  $q$  to the secondary vantage point of the node Area.v2;
  - select the minimum value among  $(l4.1, l4.2)$  as the desired distance  $l4$ ;
- 5) if the subtree Area is a leaf of the tree:
  - select a random point  $d$  from the Area;
  - set the value of the desired distance as the distance from the query point  $q$  to the point  $d$ .

To increase the efficiency of the algorithm, we propose to additionally calculate and use the minimum distances between the vantage points  $\{v1, v2\}$  and the subtrees of the node. Denote by  $ri, j$  the minimum distance between the  $i$ -th vantage point and the  $j$ -th area  $\{AV1, AV2, A3, A4\}$  of the current tree node. Then, for the regions AV1, AV2, A3 and A4, it is possible to determine more precise distances to the query point  $q$ :

- 1) for AV1, find the minimum among  $(l1, l3 + r2.1)$ ;
- 2) for AV2, find the minimum among  $(l2, l1 + r1.2, l3 + r2.2)$ ;
- 3) for A3, find the minimum among  $(l3, l1 + r1.3)$ ;
- 4) for A4, find the minimum among  $(l4, l1 + r1.4, l3 + r2.4)$ .

We have implemented the proposed ideas in the DDRPNA algorithm (Fig. 2). The architectural changes associated with the DDRPNA algorithm in the MVP-tree ( $Y*B = 8$  minimum distances at each node) increase the amount of memory for storing the indexing tree.

### Modification of the node intersection rules MRDINMD

Another problem of the DIHPN algorithm is the problem of efficiently determining the intersection of the search area with the subtrees of the current node. The less intersections of the search area with the nodes of the MVP-tree, the higher the search efficiency. This problem can be solved by using the previously introduced minimum distances  $ri, j$ .

Studies [32] proposed a method to improve the efficiency of determining the intersection of the search area and nodes for a PM-tree. The features of the proposed method are:

- 1) using both minimum and maximum distances from vantage points to areas of the current node;
- 2) set of vantage points is global and common to all nodes of the tree.

The method proposed in [32] for determining the intersection of the search area and tree nodes by the authors is adapted for the MVP-tree.

The features of the adaptation are:

- 1) the use of only a set of minimum distances from the vantage points to the areas of the tree node;
- 2) using local vantage points for each node of the tree instead of a set of global control points.

The modification of the classical MVP-tree area intersection rules [formula (9)] based on the minimum distances MRDINMD is represented as

$$overlap(Node, (q, ln)) = \begin{cases} A1: (ln + r1) \geq l1 \text{ and } (ln + r2) \geq l2 \text{ and } (ln + l1) > r1.1 \text{ and } (ln + l2) > r2.1 \\ A2: (ln + r1) \geq l1 \text{ and } (ln + l2) > r2 \text{ and } (ln + l1) > r1.2 \text{ and } (ln + l2) > r2.2 \\ A3: (ln + r3) \geq l2 \text{ and } (ln + l1) > r1 \text{ and } (ln + l1) > r1.3 \text{ and } (ln + l2) > r2.3 \\ A4: (ln + l1) > r1 \text{ and } (ln + l2) > r2 \text{ and } (ln + l1) > r1.4 \text{ and } (ln + l2) > r2.4 \end{cases} \quad (11)$$

The developed modification MRDINMD by the authors replaces the classic rules of intersection of areas for MVP-tree in search algorithms and DIHPN (see Fig. 1).

The methods obtained during the research for reducing the number of calculated distances to improve the search efficiency in the MVP-tree have a theoretical basis [formulas (10) and (11)]. Thus, the purpose of the authors' practical research is to evaluate and analyze the performance gain of the modified search in the MVP-tree compared to the classical one, depending on the dimension of the space and the number of objects in the data.

## Experiments

We have carried out an experimental evaluation of the efficiency of the proposed algorithms in the context of the nearest neighbor search problem. During the experiments, we tested: 1) algorithm DIHPN; 2) combination of DIHPN and DDRPAN; 3) combination of DIHPN + DDRPAN + MRDINMD algorithms.

Computer characteristics as a part of experiments: 64-bit Windows 7 operating system, Intel i5-2500K processor, 16 GB RAM. We wrote the program in Java and compiled it using the NetBeans IDE. We use the following MVP-tree parameters: 1)  $U = 2$ ; 2)  $Y = 2$ ; 3)  $B = 4$ ; 4)  $K = 1$ ; 5)  $P = 0$ ;

Four data sets were selected as experimental data from repository ANN-Benchmarks [33]: 1) Last.fm (65 number of attributes); 2) SIFT (128 number of attributes); 3) MNIST (784 number of attributes); 4) GIST (960 number of attributes). Note that we have divided each dataset into two data categories: a set of SQ queries to find the nearest neighbor and the original set of objects.

Within the experiment, the MVP-tree has the following features: 1) we use the Euclidean distance as the distance metric; 2) we normalized each dimension of the data sets in the range  $[0, 1]$ ; 3) we use a random selection heuristic to select the vantage points. We repeated each experiment 3 times, after which we averaged the results.

A measure of the computational complexity of the tested algorithm is the absolute number of operations (NumberDistances) for calculating the distances between two points during a search for a SQ set. To simplify understanding, all the results of the tested algorithms were normalized relative to the basic search algorithm. Based on the stated goal of the research, we have chosen the classical search algorithm MVP-tree [29] as the basic algorithm for comparison. We calculate the effectiveness of the proposed algorithms using the formula:

$$Eff_{ALG} = \frac{\text{NumberDistances}_{ALG}}{\text{NumberDistances}_{MVP}} \times 100\%, \quad (12)$$

where  $Eff_{ALG}$  — the efficiency of the tested algorithm ALG;  $\text{NumberDistances}_{ALG}$  — the number of operations for calculating the distances between two points during the search for the SQ set of the algorithm ALG;  $\text{NumberDistances}_{MVP}$  — the number of calculated distances between two points for the set SQ using the classical search algorithm MVP-tree.

Thus, we obtained all the results of the work based on measuring and normalizing the absolute number of calculated distances between two objects (NumberDistances). Therefore, the evaluation of the efficiency of the tested algorithms does not

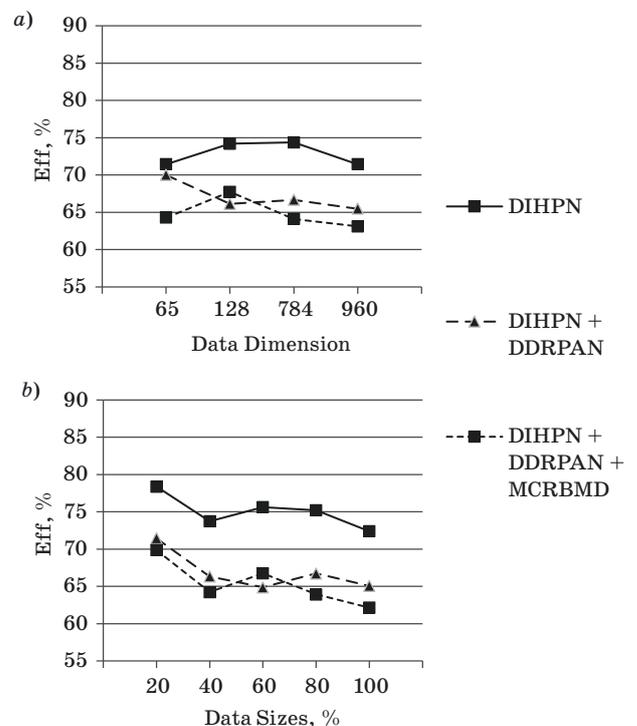
depend on the programming language, the development environment and technical characteristics of computing resources.

We used the following data set parameters for the experiment: 1) the original set of 60 000 objects, the number of attributes in the object is 65, 128, 784 and 960; 2) a corresponding set of SQ queries of 3000 objects.

We presented two summary graphs: 1) the graph of the computational complexity of search algorithms for different data dimensions (the number of attributes), shown in Fig. 3, a; 2) the graph of the computational complexity of search algorithms for different original data sizes (the number of objects), which is shown in Fig. 3, b.

Based on the analysis of Fig. 3, we were made the following conclusions for the algorithms developed by the authors to reduce the number of computations for the nearest neighbor search in the MVP-tree: 1) the proposed algorithms increase the search performance in both small and large space dimensions; 2) with the growth of the data dimension (the number of attributes), the efficiency of the developed search algorithms increases; 3) with the growth of the number objects of data, the efficiency of the applied algorithms increases.

Thus, the results of the experiment confirm the effectiveness of the author's solutions for nearest neighbors searching in big data with a significant number of attributes. The proposed theoretical al-



■ Fig. 3. Efficiency of different data dimensions (a) and sizes (b)

gorithms have found application in an applied classification system for solving pattern recognition problems based on precedent methods. The developed algorithms increase the computational performance of the nearest neighbor search procedure in the indexed set of original objects.

## Conclusion

We presented algorithmic solutions related to the adaptation of the classical search algorithm to the priority queue of node processing, sorted by increasing the distances from the query point to the nodes.

The proposed solutions allow to:

- eliminate additional calculations when using the priority queue for processing nodes by reducing the number of intersecting nodes of the same level;
- identify and generalize the theoretical dependence of the number of intersections on the complexity of the search calculations (the number of additional processed levels);

- improve the efficiency of the priority queue for processing nodes by more accurately determining the distances from the query point to the subtrees of the tree node;

- use additional minimum distances between the vantage points and the subtrees of the node, which improves the search algorithm;

- reduce search time by reducing the number of processed nodes based on the use of minimum distances in the procedure for determining the intersection of the search area with the area of the tree node.

The developed algorithms of search procedures in tree structures have demonstrated promising results of their application to the class of data indexing problems in metric space.

A limitation of the proposed algorithms is the heuristic of random selection of vantage points. The use of random selection heuristics does not guarantee stable results for the same data. As further improvements to the search algorithms, we are planning to use combinations of greedy heuristics when selecting vantage points.

## References

1. Fomin V. V., Duke V. A., Aleksandrov I. V. The use of machine learning methods for the determination of the fuel consumption of a gas turbine frigate. *Marine Intellectual Technologies*, 2019, vol. 4, no. 4(42), pp. 197–201.
2. Nielsen F., Piro P., Barlaud M. Bregman vantage point trees for efficient nearest Neighbor Queries. *2009 IEEE International Conference on Multimedia and Expo*, 2009, pp. 878–881. doi:10.1109/ICME.2009.5202635
3. Oliveira P. H., Traina C., Kaster D. S. CLAP, ACIR and SCOOP: Novel techniques for improving the performance of dynamic Metric Access Methods. *Information Systems*, 2017, vol. 72, pp. 117–135. doi:10.1016/j.is.2017.10.003
4. Zobot G. F., Cazzolato M. T., Scabora L. C., Traina A. J. M., Traina Jr. C. Efficient indexing of multiple metric spaces with spectra. *2019 IEEE International Symposium on Multimedia (ISM)*, 2019, pp. 169–1697. doi:10.1109/ISM46123.2019.00038
5. Calistru C. M., Ribeiro M. C., David G. T. Flexible descriptor indexing for multimedia databases. *Proceedings of the Fourth International Workshop on Content-Based Multimedia Indexing*, 2005, pp. 1–8.
6. Sehili Z., Rahm E. Speeding up privacy preserving record linkage for metric space similarity measures. *Datenbank-Spektrum*, 2016, vol. 16(3), pp. 227–236. doi:10.1007/s13222-016-0222-9
7. Chávez E., Ludueña V., Reyes N., Roggero P. Faster proximity searching with the distal SAT. *Information Systems*, 2016, vol. 59, pp. 15–47. doi:10.1007/978-3-319-11988-5\_6
8. Hanyf Y., Silkan H., Labani H. Criteria and technique to choose a good  $\rho$  parameter for the D-index. *Intelligent Systems and Computer Vision (ISCV)*, 2015, pp. 1–6. doi:10.1109/ISACV.2015.7106169
9. Almeida J., Torres R. D. S., Leite N. J. BP-tree: An efficient index for similarity search in high-dimensional metric spaces. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 2010, pp. 1365–1368. doi:10.1145/1871437.1871622
10. Pola I. R. V., Traina Jr. C., Traina A. J. M. The MM-tree: a memory-based metric tree without overlap between nodes. *Proceedings of the 11th East European Conference on Advances in Databases and Information Systems*, 2007, pp. 157–171. doi:10.1007/978-3-540-75185-4\_13
11. Carelo C. C. M., Pola I. R. V., Ciferri R. R., Traina A. J. M., Traina C. T., de Aguiar Ciferri C. D. Slicing the metric space to provide quick indexing of complex data in the main memory. *Information Systems*, 2011, vol. 36(1), pp. 79–98. doi:10.1016/j.is.2010.06.004
12. Guhlemann S., Petersohn U., Meyer-Wegener K. Reducing the distance calculations when searching an M-Tree. *Datenbank-Spektrum*, 2017, vol. 17(2), pp. 155–167. doi:10.1007/s13222-017-0258-5
13. Kokotinis I., Kendea M., Nodarakis N., Rapti A., Sioutas S., Tsakalidis A. K., Tsolis D., Panagis Y. NSM-tree: Efficient indexing on top of NoSQL databases. *International Workshop of Algorithmic Aspects of Cloud Computing*, 2016, pp. 3–14. doi:10.1007/978-3-319-57045-7\_1

14. Donkó I., Szalai-Gindl J. M., Gombos G., Kiss A. An implementation of the M-tree index structure for PostgreSQL using GiST. *2019 IEEE 15th International Scientific Conference on Informatics*, 2019, pp. 189–194. doi:10.1109/Informatics47936.2019.9119265
15. Chen L., Gao Y., Zheng B., Jensen C. S., Yang H., Yang K. Pivot-based metric indexing. *Proceedings of the VLDB Endowment*, 2017, vol. 10(10), pp. 1058–1069. doi:10.14778/3115404.3115411
16. Maliki N., Silkan H., Maghri M. Efficient indexing and similarity search using the Geometric Near-neighbor Access Tree (GNAT) for face-images data. *Procedia Computer Science*, 2019, vol. 148, pp. 600–609. doi:10.1016/j.procs.2019.01.033
17. Razente H. L., Sousa R. M. S., Barioni M. C. N. Metric indexing assisted by short-term memories. *Similarity Search and Applications*, 2018, vol. 11223, pp. 107–121. doi:10.1007/978-3-030-02224-2\_9
18. Chen L., Gao Y., Li X., Jensen C. S., Chen G. Efficient metric indexing for similarity search and similarity joins. *IEEE Transactions on Knowledge and Data Engineering*, 2017, vol. 29(3), pp. 556–571. doi:10.1109/TKDE.2015.2506556
19. Li S., Yu H., Yuan L. A Novel approach to remote sensing image retrieval with multi-feature VP-tree indexing and online feature selection. *IEEE Second International Conference on Multimedia Big Data*, 2016, vol. 1, pp. 133–136. doi:10.1109/BigMM.2016.11
20. Jiang D., Sun H., Yi J., Zhao X. The research on nearest neighbor search algorithm based on vantage point tree. *2017 8th IEEE International Conference on Software Engineering and Service Science*, 2017, pp. 354–357. doi:10.1109/ICSESS.2017.8342931
21. Su Y., Zhang Y., Wan C., Yu G. GDPC: A GPU-accelerated density peaks clustering algorithm. *Database Systems for Advanced Applications*, 2020, vol. 1, pp. 305–313. doi:10.1007/978-3-030-59410-7\_21
22. Cheng H., Yang W., Tang R., Mao J., Luo Q., Li C., Wang A. Distributed indexes design to accelerate similarity based images retrieval in airport video monitoring systems. *Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery*, 2015, pp. 1908–1912. doi:10.1109/FSKD.2015.7382239
23. Lin Y. Q., Fu Y. G., Su Q., Wang Y. M., Gong X. T. A rule activation method for extended belief rule base with VP-tree and MVP-tree. *Journal of Intelligent & Fuzzy Systems*, 2017, vol. 33(6), pp. 3695–3705. doi:10.3233/JIFS-17521
24. Figueroa K., Paredes R., Ibarrola J. A. C., Reyes N. Fixed height queries tree permutation index for proximity searching. In: *Pattern Recognition. MCP R 2017*. Carrasco-Ochoa J., Martínez-Trinidad J., Olivera-López J. (eds). Lecture Notes in Computer Science, Springer, Cham, 2017. Vol. 10267. Pp. 74–83. doi:10.1007/978-3-319-59226-8\_8
25. Mao R., Liu W., Iqbal Q., Miranker D. P. On index methods for an image database. *ACM International Workshop on Multimedia Databases*, 2003, pp. 1–9.
26. Boytsov L., Nyberg E. Pruning algorithms for low-dimensional non-metric k-NN search: A Case study. *Proceedings of the 12th International Conference on Similarity Search and Applications*, 2019, pp. 72–85. doi:10.1007/978-3-030-32047-8\_7
27. Pandey V., Renen A., Kipf A., Kemper A. How good are modern spatial libraries? *Data Science and Engineering*, 2021, vol. 6(11), pp. 192–208. doi:10.1007/s41019-020-00147-9
28. Liu S., Wei Y. Fast nearest neighbor searching based on improved VP-tree. *Pattern Recognition Letters*, 2015, vol. 60, pp. 8–15. doi:10.1016/j.patrec.2015.03.017
29. Bozkaya T., Ozsoyoglu M. Distance-based indexing for high-dimensional metric spaces. *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data*, 1997, pp. 357–368. doi:10.1145/253262.253345
30. Hjaltason G. R., Samet H. Index-driven similarity search in metric spaces. *ACM Transactions on Database Systems*, 2002, vol. 28, no. 4, pp. 517–580. doi:10.1145/958942.958948
31. Fomin V., Aleksandrov I., Gallyamov D., Kirichek R. Modified indexing algorithm based on priority queue in metric space for MVP tree. *The 4th International Conference on Future Networks and Distributed Systems (ICFNDS 2020)*, Saint-Petersburg, Russia, November 26–27, 2020, 8p. doi:10.1145/3440749.3442617
32. Humberto L. R., Maria C. N. B. Storing data once in M-tree and PM-tree. *12th International Conference on Similarity Search and Applications SISAP*, 2019, pp. 18–31. doi:10.1007/978-3-030-32047-8\_2
33. Erik Bernhardsson — *Ann-Benchmarks*. Available at: <https://github.com/erikbern/ann-benchmarks> (accessed 15 July 2021).

УДК 004.93

doi:10.31799/1684-8853-2021-4-18-27

**Алгоритм индексации, основанный на хранении дополнительных расстояний в метрическом пространстве, в рамках структуры дерева множества опорных точек**В. В. Фомин<sup>а</sup>, доктор техн. наук, профессор, orcid.org/0000-0001-7040-5386, v\_v\_fomin@mail.ruИ. В. Александров<sup>б</sup>, аспирант, orcid.org/0000-0002-6258-8191<sup>а</sup>Российский государственный педагогический университет им. А. И. Герцена, набережная р. Мойки, 48, Санкт-Петербург, 191186, РФ<sup>б</sup>Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича, Большевиков пр., 22-1, Санкт-Петербург, 193232, РФ

**Введение:** парадигма поиска по сходству применяется в различных вычислительных задачах, таких как классификация, интеллектуальный анализ данных, распознавание образов и др. В настоящее время среди алгоритмов поиска значительное место занимает технология древовидных метрических методов доступа. Классическая проблема сокращения времени поиска по сходству в метрическом пространстве является актуальной для современных систем при обработке больших сложных данных. Ввиду многоаспектности проблемы эффективности поисковых алгоритмов локальные исследования в этом направлении востребованы и продолжают приносить полезные результаты. **Цель:** снизить вычислительную сложность алгоритмов древовидного поиска в задачах, использующих метрическую близость. **Результаты:** разработан алгоритм поиска для структуры данных в виде дерева множества опорных точек, основанный на приоритетной очереди обработки узлов; математически формализованы проблемы дополнительных вычислений и способы их решения. Для повышения быстродействия поиска по сходству предложены процедуры формирования приоритетной очереди обработки узлов и уменьшения количества пересечений узлов одного уровня. Повышение эффективности происходит на основе изменения древовидной структуры данных и использования минимальных расстояний между опорными точками и поддеревьями узла. Уменьшение числа вычислений достигается за счет более точного определения расстояния до узлов от искомого объекта и факта пересечения области поиска с узлом дерева. **Практическая значимость:** полученным алгоритмам поиска требуется меньше времени для обработки информации за счет несущественного повышения требований к памяти. Снижение времени обработки информации расширяет границы применения древовидных метрических методов индексации в задачах поиска в больших массивах данных.

**Ключевые слова** — индексация сложных данных, древовидные структуры данных, метрические методы доступа, приоритетная очередь узлов, дерево множества опорных точек.

**Для цитирования:** Fomin V. V., Aleksandrov I. V. Indexing algorithm based on storing additional distances in metric space for multi-vantage-point tree. *Информационно-управляющие системы*, 2021, № 4, с. 18–27. doi:10.31799/1684-8853-2021-4-18-27

**For citation:** Fomin V. V., Aleksandrov I. V. Indexing algorithm based on storing additional distances in metric space for multi-vantage-point tree. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 18–27. doi:10.31799/1684-8853-2021-4-18-27

UDC 62.506.29.519

doi:10.31799/1684-8853-2021-4-28-36

## Planning polyphase behavior of autonomous intelligent mobile systems in uncertain environments

V. B. Melekhin<sup>a</sup>, Dr. Sc., Tech., Professor, [orcid.org/0000-0002-7614-2860](https://orcid.org/0000-0002-7614-2860), [pashka1602@rambler.ru](mailto:pashka1602@rambler.ru)

M. V. Khachumov<sup>b,c</sup>, PhD, Phys.-Math., Senior Researcher, [orcid.org/0000-0001-5117-384X](https://orcid.org/0000-0001-5117-384X)

<sup>a</sup>Dagestan State Technical University, Imam Shamil Ave., 70, 367015, Makhachkala, Russian Federation

<sup>b</sup>Federal Research Center "Computer Science and Control" of RAS, 60-Letiya Oktyabrya Pr., 9, 117313, Moscow, Russian Federation

<sup>c</sup>Program Systems Institute of the RAS, Petra Pervogo St., 4a, 152021, Veskovo, Russian Federation

**Introduction:** Modern ways to develop intelligent problem solvers have shortcomings in the efficiency of their application for planning purposeful behavior of autonomous intelligent mobile systems in a priori undescribed conditions of a problem environment. **Purpose:** Developing a model of knowledge representation and processing which would provide the ways to organize purposeful activity of autonomous intelligent mobile system in uncertain environment. **Methods:** Synthesis of frame-like behavior scenarios in the form of polyvariable conditionally dependent predicates whose structure includes complex variables as well as related variables of types "object", "event" and "relationship"; synthesis of heuristic rules for knowledge representation in the process of purposeful behavior planning. In order to represent complex variables in polyvariable conditionally dependent predicates, fuzzy semantic networks are used which can represent knowledge of variously purposed intelligent systems without regard to particular knowledge domains, being adaptable to a priori undescribed operational conditions. **Results:** We have proposed a structure of various polyvariable conditionally dependent predicates. On their base, an autonomous intelligent mobile system can organize various activities in a priori undescribed and unstable problem environments. Specially developed knowledge processing tools allow such a system to automatically plan its purposeful behavior in a space of subtasks during the fulfillment of tasks formulated for it. **Practical relevance:** The obtained results can be efficiently used in building intelligent problem solvers for autonomous intelligent mobile systems of various purpose, capable of performing complex tasks in a priori undescribed operational conditions.

**Keywords** – intelligent mobile system, a priori undescribed problem environment, polyvariable conditionally dependent predicates, formulated task, behavior planning.

**For citation:** Melekhin V. B., Khachumov M. V. Planning polyphase behavior of autonomous intelligent mobile systems in uncertain environments. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 28–36. doi:10.31799/1684-8853-2021-4-28-36

### Introduction

The development of autonomous intelligent mobile systems (AIMS) of various purposes, capable of solving complex problems in real a priori undescribed problem environments (PE) is crucial for the modern AI research. It requires a model of knowledge representation and processing which would help an AIMS to plan purposeful activity in uncertain environments, usually on the base of an airborne intelligent problem solver with limited resources. This model should provide the AIMS with the following features:

— Adaptation to the current operational conditions. Various situations possible in a PE are transformable according to certain rules which are usually defined in general terms, without regard to particular knowledge domains, but should be particularized. The objects in the PE should then be transformed into their specified state.

— Purposeful activity should be planned both in the space of subtasks and in the space of states, according to the form in which the specified activity purpose is represented, with the complexity

acceptable for the airborne computer. This endows the AIMS with fairly powerful functional abilities, making it capable of solving complex practical problems in all kinds of PE.

As for the first AIMS feature mentioned above, it should be noted that the modern models of knowledge representation and processing cannot provide its effective implementation because of the following main reasons.

1. There are certain limitations due to the use of network intelligence or neural networks in order to build computing units for intelligent problem solvers [1–3] which, after self-training or training, process visual data using various fish-swarm [4–6] or genetic algorithms [7–9]. This approach can efficiently organize only reflexive behaviour of an AIMS by modelling visual-action thinking of living systems [10]. In its turn, it allows the AIMS to solve a number of practical problems in uncertain environments [11], but does not provide ways to effectively plan its behavior when implementing complex tasks which demand time-consuming transformations of various situations in the AIMS by purposeful movement and manipulation of objects inside

the environment. Hence, in order to solve complex problems, the AIMS, in addition to its visual-action thinking, should be capable of visual-image and conceptual thinking. In other words, it should have inference tools based on the knowledge representation model and information from the PE [10].

2. In the field of algorithmic intelligence or applying the known models of knowledge representation and processing developed within the logical framework [12–14] and network models [13, 15], the main limitations are as follows. In the first place, an AIMS cannot afford the high inference complexity based on logical models of knowledge representation which requires a high-performance computer, often a bulky one. As for the known network models of knowledge representation (for example, semantic networks), they allow you to build a sufficiently expressive model of a given knowledge domain. However, the lack of effective inference procedures for semantic networks which could be the base for automated planning of AIMS behavior, and close association of these procedures with a particular knowledge domain considerably restrict the ways of using these models in intelligent problem solvers of AIMS which can purposefully function in PE.

In the second place, if, for example, the inference has a linear complexity, as stated by the authors who follow the MIVAR approach [16], its support will require a detailed PE model in a given knowledge domain. On the one hand, building such a model for a real PE without involving any bulky structures seems to be practically impossible [17]. On the other hand, the need of such a model does not allow us to use the logical approach for organizing purposeful AIMS behavior in uncertain environments. However, as follows from the above, the ability of an AIMS to perform complex tasks in uncertain PE is exactly what we want.

In order to endow an AIMS with the second feature mentioned above, associated with building a model of knowledge representation and processing without regard to particular knowledge domains, organizing inference in uncertain environments with a complexity affordable for an airborne computer, we can use frame-like scenarios of solving typical behavior subproblems [18] based on fuzzy semantic networks [19]. These scenarios allow the AIMS to build complex programs of purposeful activity in the space of state, with polynomial complexity of inference.

However, this approach is not optimal for problems associated with the need to plan polyphase behavior when at each stage of its implementation you have new operational conditions in the PE. Generally, polyphase behavior should be considered as AIMS activity which consists of several stages. At each stage, the intelligent system must automatically build and implement an activity plan en-

suring the achievement of its respective behavioral subgoal. After that, in most cases, the operational conditions of the AIMS fully change, and the system focuses on achieving the subgoal of the next stage. The achievement of the subgoals of all polyphase behavior stages means that the AIMS has accomplished its task.

In other words, the knowledge representation model proposed in [18] does not ensure that an AIMS can fulfil complex tasks whose implementation requires purposeful movement and manipulation of objects in various conditions of an a priori undescribed environment. In this connection, the necessity arises to develop standardized structures of a knowledge representation model which would help an AIMS to plan polyphase behavior in dynamically changing operational conditions.

In this work, we propose an approach which allows you to organize polyphase behavior of an AIMS in a priori undescribed PE conditions. Our approach assumes creating a model of knowledge representation and processing based on standardized structures which are polyvariable conditionally dependent predicates (PCDP). Generally, a PCDP structure, in contrast to conventional predicates [13], can include:

- complex variables described using active fuzzy semantic networks (AFSN) [19], which allow you to define various situations and subsituations of the environment specifying them in general terms, without regard to particular knowledge domains;

- miscellaneous variables (in fact, connected), such as “objects” of the problem environment, “events” happening in the PE, and “relationships” observed between the objects, events and AIMS, being descriptive for their influence on the current state on the intelligent system and its behavior.

## Statement of the problem

Let us consider an AIMS equipped with a manipulator, a transportation cart, and technical vision, capable of performing a set of various complex operations  $B_1 = \{b_{j_1}\}$ ,  $j_1 = \overline{1, m_1}$ . For example, it can move specified objects in the PE, find objects in a certain PE zone, transform the current environment situation when it is necessary for achieving one of its behavioral subgoals, etc. Each complex operation  $b_{j_1} \in B$  of this kind is implemented by the AIMS which performs a certain sequence  $B_{j_1}$  of elementary operations  $B_2 = \{b_{j_2}\}$ ,  $j_2 = \overline{1, m_2}$ . These can be operations such as “coming up to an object”, “seizing the object”, etc. Thus, AIMS is a system capable of moving and manipulating PE objects  $O = \{o_{j_3}(X_{j_3}, o_{j_3}^*)\}$ ,  $j_3 = \overline{1, m_3}$ , where  $X_{j_3}$  is a set of

markers providing the ways for the AIMS to identify  $j_3$  objects of the PE and perform certain elementary manipulations with them;  $o_{j_3}^*$  is the current state  $j_3$  of an object in the environment.

In general, an unstable PE can be characterized by its set of objects  $o_{j_3}(X_{j_3}, o_{j_3}^*) \in O$  and events  $Y = \{y_{j_4}(Y_{j_4})\}$ ,  $j_4 = \overline{1, m_4}$  which happen independently of the AIMS. The occurrence of such events leads to various changes in the PE which can prevent the AIMS from successfully performing certain elementary operations  $b_{j_2} \in B_2$ , where  $Y_{j_4}$  is a set of markers which determine the event  $j_4$ . Such an event can be, for example, the appearance of an object in the PE that blocks a passage.

Knowledge representation model, in accordance to the functional purpose of the AIMS, consists of a set of PCDP. Using their different combinations, an intelligent system can design a behavioral plan in order to solve various subproblems of its task. In each particular case, these PCDP serve for specifying an ordered sequence of elementary operations  $b_{j_2} \in B_2$  whose completion ensures that their respective complex operations  $b_{j_1} \in B_1$  are performed. During the inference, the values of different variables in a PCDP are determined more precisely, based on the target task formulated by the AIMS and visual data reflecting the current state of the PE.

The target task is formulated by the AIMS in a procedural representation form as a sequence of complex operations  $b_{j_1} \in B_1$  which have to be performed for its implementation. For example,  $b_{j_1}^1$ : locate the object  $o_{i_2}(X_{i_2}, o_{i_2}^*) \in O$  in a preset square of the PE;  $b_{j_1}^2$ : transfer this object into the desired state  $o_{j_3}^{**}$  taking into account the fact that the PE can produce certain events  $y_{i_3}(Y_{i_3}) \in Y$  keeping the AIMS from performing the respective elementary operations.

We have to develop:

— structure and content of variably purposed PCDP allowing the AIMS to automatically plan its purposeful behavior in order to fulfil the formulated tasks;

— knowledge processing procedures as a part of automated planning of AIMS behavior on the base of PCDP in order to build and solve ordered sequences of behavioral subproblems.

### Content and functional purpose of PCDP

Generally, an AIMS knowledge representation model is a specified set of standardized frame-like structures built on the base of the following PCDP types.

1. Dyadic polyvariable conditionally dependent predicates of the 1st type

“ $b_{j_1}^1 : P_{i_1}(S_{i_1}^*(i_1), x_{i_1}^*(X_{i_1}^*), o_{j_3}^{**}), i_1 = \overline{1, n_1}, n_1 \leq m_3$ ”

determining the following chains of elementary operations:

$$L_{i_1} : \left\langle \left\langle o_{j_3}^{*1} \& b_{j_2}^1, \right. \right. \\ \left. \left. b_{j_2}^2 \rightarrow \dots \rightarrow o_{j_3}^{*k} \& b_{j_1}^k \rightarrow \dots \rightarrow o_{j_3}^{*h} \& b_{j_1}^h \rightarrow o_{j_3}^{**} \right\rangle, \right. \\ \left. k = \overline{1, h_{j_3}}, \right.$$

which ensure that the AIMS performs complex operations  $b_{j_1}^1 \in B_1$  as their identifiers.

These chains of operations manipulate certain PE objects  $o_{j_3}(X_{j_3}, o_{j_3}^{*k}) \in O$  in order to transfer them from a random ( $k$ ) current state  $o_{j_3}^k$  into a target state  $o_{j_3}^{**}$  when, for various problem environment objects, the condition  $X_{i_1}^* \subseteq X_{j_3}$  is true, where  $h_{j_3}$  is the total number of permissible states  $j_3$  of the object;

$P_{i_1}$  is a dyadic predicate symbol with the following meaning: “Transfer the object  $x_{i_1}^*(X_{i_1}^*)$  from a random current state  $o_{j_3}^k$  into a target state  $o_{j_3}^{**}$  by launching the respective subchain of operations  $L_{i_1}^* : o_{j_3}^{*k} \& b_{j_2}^k \rightarrow \dots \rightarrow o_{j_3}^{*d} \& b_{j_2}^d \rightarrow o_{j_3}^{**}$ ,  $L_{i_1}^* \subseteq L_{i_1}$ , which brings result if the PE satisfies the conditions determined by the complex variable  $S_{i_1}^*(i_1)$ ;

$o_{j_3}^{*k} \& b_{j_2}^k \rightarrow o_{j_3}^{*k+1}$  is an elementary behavioral act, meaning that if the object  $o_{j_3}(X_{j_3}, o_{j_3}^{*k})$  is in the state  $o_{j_3}^{*k}$  then the AIMS needs to launch the operation  $b_{j_2}^k$  to transfer it into the state  $o_{j_3}^{*k+1}$ ;

$S_{i_1}(i_1)$  is a complex variable represented in the form of AFSN  $G_{i_1} = (V_{i_1}, E_{i_1})$ . The nodes  $V_{i_1}$  of the network  $G_{i_1}$  are marked with the following sorts of variables: “objects”  $x_{i_2}^*(X_{i_2}^*), i_2 = \overline{1, m_5}, m_5 < m_3$  and “events”  $d_{i_3}^*(X_{i_3}^*), i_3 = \overline{1, m_6}, m_6 < m_4$ . These variables are given their values in the current conditions by way of their respective substitution by objects  $o_{j_3}(X_{j_3}, o_{j_3}^{*k}) \in O$  and events  $d_{j_4}(X_{j_4}) \in D$  which the AIMS perceives from the PE when the conditions  $X_{i_1}^* \subseteq X_{j_3}$  and  $X_{i_3}^* \subseteq X_{j_4}$  are true for them. The ribs  $E_{i_1}$  of the AFSN  $G_{i_1}$  are determined by the following kinds of “relationship” variables:

— If a rib in the network  $G_{i_1}$  is incident to the nodes marked with object variables, or one of them is marked with an object variable and the second one is marked with the AIMS, then it is determined by a summarized state space relationship. It can be, for example, the distance between the AIMS and the object.

— In the case when a rib of the network  $G_{i_1}$  is incident to a node marked with a PE object and to another one marked with a PE event, it should be marked with a relationship which generally determines how this event affects the state of its adjacent object.

— If a rib in the network  $G_{i_1}$  is incident to a node marked with the AIMS and to a node marked with a PE event, then it is determined by a summarized relationship which shows how this event affects a certain aspect of the intelligent system activity: for example, interfering with the movement.

It should be noted that a generalized representation of various relationships in AFSN  $G_{i_1}$  is determined using the intervals of numerical values of their respective linguistic variable terms found by expertise [20].

It should also be noted that when the object variables  $x_{i_1}^*(X_{i_1}^*, x_{i_1}^{**})$  in predicates  $P_{i_1}(S_1^*(i_1), x_{i_1}^*(X_{i_1}^*), o_{j_3}^{**})$  receive values of random PE objects  $o_{j_3}(X_{j_3}, o_{j_3}^*) \in O$  and slots of complex variables  $S_1^*(i_1)$  receive values of appropriate objects, events, and relationships between them in the PE, we get statements in the format  $P_{i_1}(S_1(i_1), o_{j_3}(X_{j_3}, o_{j_3}^{*k}))$ . These statements are true if and only if for the current PE situation the following conditions are satisfied:  $X_{i_1}^* \subseteq X_{j_3}$ ,  $o_{j_3}^{*k} \neq o_{j_3}^{**}$  and  $S_1(i_1) \approx S_1^*(i_1)$ , where  $S_1(i_1) \approx S_1^*(i_1)$  means a fuzzy equality [19] between the passive fuzzy semantic network  $S_1(i_1)$  determining the current PE subsituation, and the network  $S_1^*(i_1)$  which corresponds to a complex variable of the desired PCDP. The truth of these conditions in the current PE situation is a prerequisite for the AIMS to successfully perform the respective subchains of elementary operations  $o_{j_3}^{*k} \& b \rightarrow \dots o_{j_3}^{*d} \& b_j^d \rightarrow x_{i_1}^{**}$ . The formal description of the current PE subsituation  $S_1(i_1)$  used by the AIMS during the inference is formed on the base of visual data in the intelligent problem solver via modification of the AFSN  $S_1(i_1)$  by substitution of particular objects, events and relationships between them in the PE.

## 2. Dyadic PCDP of the 2nd type

$$"b_{j_1}^2 : P_{i_2}(S_1^*(j_1, i_2), S_2^*(j_1, i_2)), i_2 = \overline{1, n_2} ",$$

where  $P_{i_2}$  is a predicate symbol with the following meaning: " $b_{j_1}^2$  : in order to transform the current PE subsituation corresponded by a complex variable  $S_1^*(j_1, i_2)$  into a subsituation determined by a complex variable  $S_2^*(j_1, i_2)$  a tuple of elementary operations  $\langle b_{j_2}^1, b_{j_2}^2, \dots, b_{j_2}^d \rangle$  should be performed".

## 3. Dyadic and polyadic PCDP of the 3rd type

$$"b_{j_1}^{3,1} : P_{i_3}(S_{3,1}^*(i_3), y_{i_3}^*(Y_{i_3}^*)), i_3 = \overline{1, m_3} "$$

$$\text{and } "b_{j_1}^{3,2} : P_{i_3}(S_{3,2}^*(i_3), y_{1,i_3}^*(Y_{1,i_3}^*), \dots, y_{k,i_3}^*(Y_{k,i_3}^*)),$$

determining tuples of elementary operations  $\langle b_{j_2}^1, b_{j_2}^2, \dots, b_{j_2}^d \rangle$  which are necessary for the

complex operation  $b_{j_1}^3$ , neutralizing the influence of separate  $y_{i_3}(Y_{i_3}) \in Y$  and mutually related PE events  $\langle y_{i_3}^1(Y_{i_3}^1), \dots, y_{i_3}^k(Y_{i_3}^k) \in Y \rangle$ , respectively, which negatively affect the purposeful activity of the AIMS. In the last expression,  $y_{i_3}^*(Y_{i_3}^*)$  are event variables which receive values of random events  $y_{i_3}(Y_{i_3}) \in Y$  observed by the AIMS in the PE when  $Y_{i_3}^* \subseteq Y_{j_4}$  is true.

$S_{3,1}^*(i_3)$  is a complex variable needed for checking certain PE conditions which provide that the AIMS successfully performs the operation tuple  $\langle b_{j_2}^1, b_{j_2}^2, \dots, b_{j_2}^d \rangle$  neutralizing the influence of an event observed in the environment to its further purposeful behavior.

Observing PE events  $y_{i_3}(Y_{i_3}) \in Y$  determined by the variables  $y_{i_5}^*(Y_{i_5}^*) \in Y^*$  stimulates AIMS unconditional reactions  $b_{j_2}^1, b_{j_2}^2, \dots, b_{j_2}^d$  in the case when the AIMS is about to perform operations that can be disturbed by these reactions. Hence, the standard elements in the knowledge representation model of the AIMS should be represented in the form of various combinations of the following PCDP pairs:

$$\begin{aligned} & \langle b_{j_1}^{3,1} : P_{i_3}(S_{3,1}^*(i_3), y_{i_3}^*(Y_{i_3}^*)), \\ & b_{j_1}^1 : P_{i_1}(S_1^*(i_1), x_{i_1}^*(X_{i_1}^*), o_{j_3}^{**}) \rangle; \\ & \langle b_{j_1}^{3,1} : P_{i_3}(S_{3,1}^*(i_3), y_{i_3}^*(Y_{i_3}^*)), \\ & b_{j_1}^2 : P_{i_2}(S_1^*(i_2), S_2^*(i_2)) \rangle. \end{aligned}$$

Such pairs allow you to recognize, by their first projection, whether the PE contains events that disturb the operations determined by the predicates which form their second projection. Note that similar pairs can also be formed on the base of k-adic predicates  $b_{j_1}^{3,2} : P_{i_3}(S_{3,2}^*(i_3), y_{1,i_3}^*(Y_{1,i_3}^*), \dots, y_{k,i_3}^*(Y_{k,i_3}^*))$ . This makes the AIMS capable of automatically recognize, by the characteristics of variables  $y_{i_3}^*(Y_{i_3}^*)$ , and eliminate in the PE various mutually related "braking" events  $y_{i_3}(Y_{i_3}) \in Y$  which disturb the successful performance of the operations determined by the PCDP of the 1st and 2nd types included in these pairs.

## 4. Dyadic PCDP of the 4th type

$$b_{j_1}^{4,z} : P_{i_4}(S_{i_4}^{*z}, RA(i_4)), z = \overline{1, 3} .$$

The AIMS uses them to determine a purposeful movement routing algorithm  $RA(i_4)$  according to the specified behavior conditions  $S_{i_4}^{*z}$  which can have one of the following meanings:

$S_{i_4}^1$  — the AIMS directly observes a PE object, access to which provides that the current subtask is

fulfilled. In other words, the system needs to make sure it takes such a position in the PE that the specified object is within its manipulator's working zone. In this case, a PCDP of the 4th type determines the conditions  $S_{i_4}^{*1}$  whose satisfaction launches the local optimal planning algorithms PA(1) which can lay a route of movement to the object in an environment with impediments [21];

$S_{i_4}^2$  — the AIMS needs to meet a target which is a moveable object  $o_{i_2}(X_{i_2}, o_{i_2}^*) \in O$  observed in the PE. In this case, on the base of a PCDP of the 4th type, the AIMS chooses a behavior planning algorithm PA(2) which can figure out the polar coordinates of the meeting point according to the object speed and current movement direction, as well as the probable speed of the AIMS itself. Then, based on the obtained data, the route is laid for the AIMS, ensuring it will meet the object [21];

$S_{i_4}^3$  — the AIMS needs to avoid colliding with a moveable object  $o_{i_2}(X_{i_2}, o_{i_2}^*) \in O$  observed in the PE. In this situation, on the base of a PCDP of the 4th type, the AIMS chooses a movement planning algorithm PA(3) which can figure out the polar coordinates of the probable meeting point and lay a route to it. The only difference from PA(2) is a correction making sure that the AIMS will not meet the object.

5. Polyvariable predicates of the 5th type  $b_{j_1}^5 : P_{i_5}(S_{i_5}^*, \sqrt{RA(4)})$  which endow the AIMS with a capability to choose a movement planning algorithm PA(4) for a PE with impediments when the coordinates of its current and target position in the given reference system are specified and expressed in the polyvariable  $S_{i_5}^*$ .

6. PCDP of the 6th type, determining complex operations  $b_{j_1}^6 \in B_1$  whose performance requires performing, in a cycle, their respective tuples  $\langle b_{j_2}^1, b_{j_2}^2, \dots, b_{j_2}^d \rangle$  of elementary operations. For example, it can be the complex operation  $b_{j_1}^6$ : load into the cart all the PE objects which satisfy the conditions of the object variable  $x_{j_1}^*(X_{j_1}^*)$ , etc., where  $X_{j_1}$  is a set of characteristics necessary for certain PE objects  $o_{j_3}(X_{j_3}, o_{j_3}^{*k}) \in O$  which can be seized by the manipulator's hand and lifted.

Generally, all PCDP of the 6th type have the following structure and content:  $b_{j_1}^6 : P_{i_6}(S_1^*(i_6), x_{j_1}^*(X_{j_1}^*), IF(S_2^*(i_6) \vee S_3^*(i_6)))$ . Here,  $P_{i_6}$  is a predicate symbol meaning the necessity to launch, in a cycle, a tuple of elementary operations  $\langle b_{j_2}^1, b_{j_2}^2, \dots, b_{j_2}^d \rangle$  in order to perform the complex operation  $b_{j_1}^6$  with a set of PE objects  $o_{j_3}(X_{j_3}, o_{j_3}^{*k}) \in O$  for which  $X_{j_1}^* \subseteq X_{j_1}$  is true.

$S_1^*(i_6)$  is an AFSN which can determine the conditions that should be satisfied in the current PE sit-

uation to ensure the successful performance of the tuple of elementary operations  $\langle b_{j_2}^1, b_{j_2}^2, \dots, b_{j_2}^d \rangle$ .

$IF(S_2^*(i_6) \vee S_3^*(i_6))$  is an operator of checking the conditions  $\langle b_{j_2}^1, b_{j_2}^2, \dots, b_{j_2}^d \rangle$ . For example, when  $S_2^*(i_6)$  is true, it means that all the PE objects  $o_{j_3}(X_{j_3}, o_{j_3}^{*k}) \in O$  are loaded into the cart, and when  $S_3^*(i_6)$  is true, it means that the total weight of the objects loaded into the cart has reached its bearing capacity;

$\vee$  is an operator showing that the cyclic performance of the operation tuple  $\langle b_{j_2}^1, b_{j_2}^2, \dots, b_{j_2}^d \rangle$  should be stopped if at least one of the conditions it is applied to is true.

7. PCDP of conditionally dependent predicates of the 7th type which determine the behavior planning tools associated with search in the PE and performing a certain tuple of elementary operations on found objects which satisfy the demands of the specified object variable  $x_{j_1}^*(X_{j_1}^*)$ . Generally, PCDP of this type can have, for example, the following structure and content:

$$b_{j_1}^{6,f} \left( \langle b_{j_1}^{4,1}, b_{j_1}^1 \rangle \right) : P_{i_6} (\downarrow A_{i_6} (5) \langle b_{j_1}^{4,1} : P_{i_4} (S_{i_4}^{*1}, RA(1)), b_{j_1}^1 : P_{i_1} (S_1^*(i_1), x_{i_1}^*(X_{i_1}^*), o_{j_3}^{**}) \rangle) \\ IF(search\ is\ over) \uparrow end), f = \overline{1, m_5},$$

where following the arrow means that if the condition  $IF$  is false, you should continue searching for required objects in the specified PE zone.

$A_{i_6} (5)$  is an algorithm of search for required PE objects. It splits the specified zone into segments, taking into account the resolving power of the AIMS machine vision, and scans each segment, looking for the objects. After an object is found in the current segment, the AIMS performs the elementary operations specified by the PCDP  $P_{i_4} (S_{i_4}^{*1}, RA(1))$  and  $A_{i_6} (5)$ .

$P_{i_4} (S_{i_4}^{*1}, RA(1))$  is a PCDP which launches the route planning algorithm in order to move toward the found object.

$P_{i_1} (S_1^*(i_1), x_{i_1}^*(X_{i_1}^*), o_{j_3}^{**})$  is a PCDP according to which the AIMS performs the operations providing that the object transfers from a random ( $k$ ) current state  $o_{j_3}^k$  into the specified state  $o_{j_3}^{**}$ . After that, we can pass to checking the condition determined by the  $IF$  operator.

### Knowledge processing procedures in automated planning of AIMS behavior in the space of subtasks

Generally, decision making in the process of automated planning of purposeful AIMS behavior on

the base of PCDP in the space of subtasks includes the following main stages of knowledge processing in the intelligent problem solver.

1. The received task is split into behavioral subtargets and their respective subtasks. Each subtask includes one complex operation  $b_{j_1} \in B_1$ .

2. From the knowledge representation model, we choose those PCDP which correspond to the selected operations  $b_{j_1} \in B_1$ . If the structure of the chosen PCDP contains "object" or "event" variables, we have to take into account the characteristics determining those variables of these predicates that are contained in the task formulated by the AIMS.

3. It has to be decided how independently different subtasks can be implemented. With this goal in mind, for each selected complex operation  $b_{j_1} \in B_1$  starting from the last one, we need to determine whether it is nested in the structure of another PCDP selected according to the task. For example, let the operation  $b_{j_1}^1, b_{j_1}^1 : P_{i_1}(S_1^*(i_1), x_{i_1}^*(X_{i_1}^*), o_{j_3}^{**})$  selected at the stage 2 be included in the structure of a PCDP of the 7th type found on the base of a complex operation  $b_{j_1}^{6,f} (< b_{j_1}^{4,1}, b_{j_1}^1 >)$ . In this case, a PCDP of the 1st type  $b_{j_1}^1 : P_{i_1}(S_1^*(i_1), x_{i_1}^*(X_{i_1}^*), o_{j_3}^{**})$  found in the knowledge representation model taking into account the characteristics of the object variable specified in the task received by the AIMS, is substituted into the structure of a polyvariable conditionally dependent predicate of the 7th type.

The process of searching for PCDP continues until having specified and corrected all the predicates which correspond to the independently performed complex operations  $b_{j_1} \in B_1$  in the structure of the task.

4. A plan of behavior is synthesized by concatenation of the PCDP which were found at the stage 3, in the order of their respective complex operations  $b_{j_1} \in B_1$  in the task.

5. When necessary, at various stage of the formed behavior plan implementation, a tuple of PCDP has to be built, which can help to form a sequence of elementary operations providing certain transformations of the situation perceived in the PE in order to fulfil the current subtask. For example, at some stage, the AIMS can face the necessity to ensure conditions determined by the polyvariable  $S_2^*(j_1, i_2)$  represented in AFSN form. In this case, based on visual data, the intelligent problem solver produces a formal description of the current operational conditions in the form of a passive fuzzy semantic network  $G_T=(V_T, E_T)$ . Its nodes  $V_T$  are marked with PE objects  $o_{j_3}(X_{j_3}, o_{j_3}^*) \in O$  and PE events  $y_{i_3}(Y_{i_3}) \in Y$  while its arcs  $E_T$  are determined by quantitative estimates of the relationships between the objects, events, and AIMS.

Now, if the AFSN corresponding to the subsituation  $S_1^*(j_1, i_2)$  is fuzzily equal to and nested in-

to a formal description of the current PE situation [19], then in order to provide the environment conditions determined by the subsituation  $S_2^*(j_1, i_2)$ , the AIMS performs a tuple of elementary operations  $\langle b_{j_2}^1, b_{j_2}^2, \dots, b_{j_2}^d \rangle$  found on the base of PCDP  $b_{j_1}^2 : P_{i_2}(S_1^*(j_1, i_2), S_2^*(j_1, i_2))$ . Otherwise, this tuple will consist of PCDP of the 2nd type whose elements satisfy the following conditions:

a) for the first PCDP  $P_{i_2}(S_1^{*1}(j_1, i_2), S_2^{*1}(j_1, i_2))$  in the structure of the tuple, the AFSN corresponding to the subsituation  $S_1^{*1}(j_1, i_2)$  should be fuzzily equal to and nested into the passive fuzzy semantic network  $G_T=(V_T, E_T)$ ;

b) for all pairs of adjacent PCDP, for example, for  $P_{i_2}^1(S_1^{*1}(j_1, i_2), S_2^{*1}(j_1, i_2))$  and  $P_{i_2}^2(S_1^{*2}(j_1, i_2), S_2^{*2}(j_1, i_2))$ , the  $S_2^{*1}(j_1, i_2) \approx S_1^{*2}(j_1, i_2)$  should be true;

c) for the sought predicate  $P_{i_2}(S_1^*(j_1, i_2), S_2^*(j_1, i_2))$  at the last position in the formed tuple and its left adjacent predicate  $P_{i_2}^2(S_1^{*2}(j_1, i_2), S_2^{*2}(j_1, i_2))$ , the condition  $S_2^{*k}(j_1, i_2) \approx S_1^{*k}(j_1, i_2)$  should be true.

In this case, performing, in the preset order, elementary operations specified by the PCDP in the tuple allows us, by means of consecutive purposeful transformation of the current operational conditions, to ensure that the PE conditions  $S_2^*(j_1, i_2)$  are true.

As an example, let us consider the process of AIMS behavior planning associated with the fulfilment of the following task: " $b_{j_1}^6 (< b_{j_1}^{4,1}, b_{j_1}^1 >)$ : in a given PE zone, find all objects  $o_{j_3}(X_{j_3}, o_{j_3}^{k*}) \in O$  possessing the subset of characteristics  $X_{j_3}^3; b_{j_1}^1$ ; transfer the found objects from a random current state  $o_{j_3}^{k*}$  into a specified state  $o_{j_3}^{**}; b_{j_1}^5$ ; after scanning the specified zone, proceed to the location point".

In order to fulfil this task, at the first step of building a behavior plan based on complex operations  $b_{j_1}^6 (\langle b_{j_1}^{4,1}, b_{j_1}^1 \rangle)$ ,  $b_{j_1}^1$  and  $b_{j_1}^5$ , we should pick up their respective PCDP from the memory and pass them to the intelligent problem solver. When choosing PCDP of the 1st type  $b_{j_1}^1 : P_{i_1}(S_1^*(i_1), x_{i_1}^*(X_{i_1}^*), o_{j_3}^{**})$ , we should remember that its object variable must be determined by a set of characteristics  $X_{j_3}^3$  specified in the task received by the AIMS.

At the next step of the planning, it is determined that the operation  $b_{j_1}^1$  is included in the structure of the complex operation  $b_{j_1}^6 (\langle b_{j_1}^{4,1}, b_{j_1}^1 \rangle)$ , i.e. its performance is not independent. After this, the predicate  $b_{j_1}^1 : P_{i_1}(S_1^*(i_1), x_{i_1}^*(X_{i_1}^3), o_{j_3}^{**})$  is substituted into the structure of a PCDP of the 7th type.

At the last step of the planning, a PCDP tuple is built, corresponding to independently performed complex operations  $b_{j_1}^6 (\langle b_{j_1}^{4,1}, b_{j_1}^1 \rangle)$  and  $b_{j_1}^5$ .

At the next stage of building an intellectual problem solver for the AIMS, we develop, based on the proposed knowledge representation and processing model, information technologies in order to organize automated planning of purposeful polyphase behavior in the PE. You can read about the methodology and peculiarities of developing hybrid or autonomous intelligent information technologies and systems in such works as [22].

To summarize, we can state that the proposed model of AIMS knowledge representation and processing, unlike the known algorithms of search for decisions in a space of subtasks [11–13], allows you to build procedures for automated planning of polyphase behavior during the implementation of PE tasks of various complexity by means of forming an inference tree without any empty alternatives. This is associated with the fact that generally, on the one hand, the search for a subproblem solution at each behavior planning step assumes choosing several effective operations at once; on the other hand, these operations are chosen in a purposeful way during the search. As for complex problem solution search in a space of subtasks on the base of the known inference procedures, generally it is reduced to brute-force search which, in its turn, leads to a higher dimension of the search space [11–13].

It should also be noted that in the framework of the proposed model, the main operation performed in the problem solver is comparison of different fuzzy semantic networks which are marked graphs. The complexity of this comparison has an order of  $Q[n^2]$ , where  $n$  is the number of nodes in the graphs. Hence, the purposeful behavior planning procedures built on their base will have polynomial complexity, which opens the ways to their efficient

implementation on airborne computers of various AIMS.

## Conclusion

1. The proposed standard structures in the form of variously purposed PCDP allow you to build a knowledge representation model for an AIMS without regard to particular knowledge domains, in accordance to its functional purpose. Thus, an intelligent mobile system can adapt to a PE with a priori undescribed conditions.

2. Generalized development of principles for purposeful behavior planning on the base of a specified set of PCDP provides the ways for its application in intelligent problem solvers of variously purposed AIMS, endowing them with the capability of performing tasks various in complexity and content, in uncertain environments.

3. The completeness of a PCDP set that allows an AIMS to perform variously purposed tasks in a PE with a priori undescribed conditions is determined based on its functional purpose or the class of subproblems that the system will have to solve during its purposeful behavior planning. A given PCDP set is an open model of knowledge representation which can be enriched as the system accumulates experience in various problem environments, different in complexity and content.

## Financial support

This study was supported by the Russian Science Foundation, Grant No. 21-71-10056 (<https://rscf.ru/en/project/21-71-10056/>).

## References

1. Yushenko A. S., Lebedev K. R., Zabihafar S. H. The adaptive neural network control of quadrotor helicopter. *Nauka i obrazovanie. MGTU im. N. E. Baumana*, 2017, no. 07, pp. 262–277. Available at: [https://www.elibrary.ru/download/elibrary\\_30585864\\_44197366.pdf](https://www.elibrary.ru/download/elibrary_30585864_44197366.pdf) (accessed 15 August 2021) (In Russian).
2. Korovin Ya. S., Khisamutdinov M. V., Ivanov D. Ya. Experimental researches of methods of neural network analysis of oilfield data. *Vestnik nauki i obrazovaniya*, 2019, no. 1-1, pp. 17–26 (In Russian).
3. Bai S., Wang C. Information aggregation and fusion in deep neural networks for object interaction exploration for semantic segmentation. *Knowledge-Based Systems*, 2021, vol. 218. doi:10.1016/j.knosys.2021.106843
4. Karpov V. E., Karpova I. P., Kulinich A. A. *Social'nye soobshchestva robotov* [Social communities of robots]. Moscow, LENAND Publ., 2019. 352 p. (In Russian).
5. Uriol R., Moran A. Mobile robot path planning in complex environments using ant colony optimization algorithm. *Proc. 3rd International Conference on Control, Automation and Robotics*, Nagoya, 2017, pp. 15–21. doi:10.1109/ICCAR.2017.7942653
6. Zhu B., Xie L., Han D., Meng X., Teo R. A survey on recent progress in control of swarm systems. *Science China Information Sciences*, 2017, vol. 60, no. 7, pp. 1–24. doi:10.1007/s11432-016-9088-2
7. Simon D. *Evolutionary Optimization Algorithms*. Wiley, 2013. 784 p.
8. Mogilev A. A., Kureichik V. M. Modified genetic project planning algorithm implemented with the use of cloud computing. *Izvestiya SFedU. Engineering*

- sciences, 2020, no. 2(212), pp. 157–159 (In Russian). doi:10.18522/2311-3103-2020-2-157-169
9. Dong H., Xu Y., Li X., Yang Z., Zou C. An improved ant lion optimizer with dynamic random walk and dynamic opposite learning. *Knowledge-Based Systems*, 2021, vol. 216. doi:10.1016/j.knosys.2021.106752
  10. Melekhin V. B., Khachumov M. V. Forms of thinking of autonomous intelligent agents: features and problems of their organization. *Marine Intelligent Technologies*, 2020, no 4–1(50), pp. 223–229. doi:10.37220/MIT.2020.50.4.031
  11. Folgheraiter M., Keldibek A., Aubakir B., Gini G., Franchi A. M., Bana M. A neuromorphic control architecture for a biped robot. *Robotics and Autonomous Systems*, 2019, vol. 120. doi:10.1016/j.robot.2019.07.014
  12. Broussard M. *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press, 2018. 246 p.
  13. Russell S., Norvig P. *Artificial Intelligence: A Modern Approach*. 4 ed. Pearson, 2020. 1216 p.
  14. Dushkin R. V. *Iskusstvennyj intellekt* [Artificial Intelligence]. Moscow, DMK Press Publ., 2019. 279 p. (In Russian).
  15. Ostrouh A. V., Surkova N. E. *Sistemy iskusstvennogo intellekta* [Artificial intelligence systems]. Saint-Petersburg, Lan' Publ., 2019. 227 p. (In Russian).
  16. Varlamov O. O., Aladdin D. V. On the application of mivar networks for intellectual planning of the behavior of robots in the space of conditions. *Izvestiya Kabardino-Balkarskogo nauchnogo centra RAN*, 2018, no. 6-2 (86), pp. 75–82 (In Russian).
  17. Ford M. *Architects of Intelligence: The Truth about AI from the People Building it*. Packt Publishing, 2018. 554 p.
  18. Melekhin V. B., Khachumov M. V. Partitioning complex tasks into subtasks in the process of planning goal-seeking behavior of an integral unmanned aerial vehicle in the state space. *Marine Intelligent Technologies*, 2020, no. 3–1 (50), pp. 237–244. doi:10.37220/MIT.2020.49.3.032
  19. Melekhin V. B., Khachumov M. V. Fuzzy semantic networks as an adaptive model of knowledge representation of autonomous intelligent systems. *Iskusstvennyj intellekt i prinyatie reshenij*, 2020, no. 3, pp. 61–72 (In Russian). doi:10.14357/20718594200306
  20. Flegontov A. V., Vilkov V. B., Chernyh A. K. *Modelirovanie zadach prinyatiya reshenij pri nechetkih iskhodnykh dannykh* [Modeling decision-making problems with fuzzy initial data]. Saint-Petersburg, Lan' Publ., 2020. 332 p. (In Russian).
  21. Melekhin V. B., Khachumov M. V. Path planning for purposeful flight of an autonomous aircraft at low altitude in uncertain environment. *Aviakosmicheskoe priborostroenie*, 2018, no. 1, pp. 18–27 (In Russian).
  22. Golenkov V. V., Guliakina N. A., Davydenko I. T., Shunkevich D. V., Ereemeev A. P. Ontological design of hybrid semantically compatible intelligent systems based on sense representation of knowledge. *Ontologiya proektirovaniya*, 2019, vol. 9, no. 1 (34), pp. 132–151 (In Russian). doi:10.18287/2223-9537-2019-9-1-132-151

УДК 62.506.29.519

doi:10.31799/1684-8853-2021-4-28-36

#### Планирование полифазного поведения автономных мобильных интеллектуальных систем в условиях неопределенности

В. Б. Мелехин<sup>а</sup>, доктор техн. наук, профессор, orcid.org/0000-0002-7614-2860, pashka1602@rambler.ru

М. В. Хачумов<sup>б,в</sup>, канд. физ.-мат. наук, старший научный сотрудник, orcid.org/0000-0001-5117-384X

<sup>а</sup>Дагестанский государственный технический университет, Имама Шамиля пр., 70, Махачкала, 367015, РФ

<sup>б</sup>Федеральный исследовательский центр «Информатика и управление» РАН, 60-летия Октября пр., 9, Москва, 117313, РФ

<sup>в</sup>Институт программных систем им. А. К. Айламазяна РАН, Петра Первого ул., 4а, с. Вельсково, 152021, РФ

**Введение:** проведен анализ современных принципов организации интеллектуальных решателей задач, позволивший обозначить их основные недостатки с точки зрения эффективности применения для планирования целенаправленного поведения автономных мобильных интеллектуальных систем в априори неопределенных условиях проблемной среды. **Цель:** разработка модели представления и обработки знаний, позволяющей организовать целенаправленную деятельность автономных мобильных интеллектуальных систем в условиях неопределенности. **Методы** исследования связаны с синтезом: 1) фреймоподобных сценариев поведения, представляющих собой полипеременные условно-зависимые предикаты, структура которых включает комплексные переменные, а также связанные переменные сорта «объекты», «события», происходящие в среде, и «отношения» между ними; 2) эвристических правил обработки знаний в процессе планирования целенаправленного поведения. Для представления комплексных переменных в полипеременных условно-зависимых предикатах применяются нечеткие семантические сети, позволяющие представлять знания различных по назначению интеллектуальных систем безотносительно к конкретной предметной области и на этой основе адаптироваться к априори неопределенным условиям функционирования. **Результаты:** предложена структура различных по функциональному назначению полипеременных условно-зависимых предикатов, позволяющих автономным мобильным интеллектуальным системам организовать различные виды деятельности в априори неопределенных нестабильных проблемных средах. Разработаны инструментальные средства обработки знаний, обеспечивающие автономным мобильным интеллектуальным системам возможность автоматически планировать целенаправленное поведение в пространстве подзадач в процессе выполнения

сформулированных им заданий. **Практическая значимость:** использование полученных результатов повышает эффективность построения интеллектуальных решателей задач для автономных мобильных интеллектуальных систем различного назначения, способных выполнять сложные задания в априори неописанных условиях функционирования.

**Ключевые слова** — интеллектуальная мобильная система, априори неописанная проблемная среда, полипеременные условно-зависимые предикаты, сформулированное задание, планирование поведения.

**Для цитирования:** Melekhin V. B., Khachumov M. V. Planning polyphase behavior of autonomous intelligent mobile systems in uncertain environments. *Информационно-управляющие системы*, 2021, № 4, с. 28–36. doi:10.31799/1684-8853-2021-4-28-36

**For citation:** Melekhin V. B., Khachumov M. V. Planning polyphase behavior of autonomous intelligent mobile systems in uncertain environments. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 28–36. doi:10.31799/1684-8853-2021-4-28-36

---

---

---

---

### УВАЖАЕМЫЕ АВТОРЫ!

Научные базы данных, включая SCOPUS и Web of Science, обрабатывают данные автоматически. С одной стороны, это ускоряет процесс обработки данных, с другой — различия в транслитерации ФИО, неточные данные о месте работы, области научного знания и т. д. приводят к тому, что в базах оказывается несколько авторских страниц для одного и того же человека. В результате для всех по отдельности считаются индексы цитирования, что снижает рейтинг ученого.

Для идентификации авторов в сетях Thomson Reuters проводит регистрацию с присвоением уникального индекса (ID) для каждого из авторов научных публикаций.

Процедура получения ID бесплатна и очень проста, есть возможность провести регистрацию на 12-ти языках, включая русский (чтобы выбрать язык, кликните на зеленое поле сверху справа на стартовой странице): <https://orcid.org>

---

---

# CubeSat formation architecture for small space debris surveillance and orbit determination

A. A. Afanasev<sup>a</sup>, Post-Graduate Student, [orcid.org/0000-0001-7503-9696](https://orcid.org/0000-0001-7503-9696), [anton.afanasev@skoltech.ru](mailto:anton.afanasev@skoltech.ru)

S. N. Biktimirov<sup>a</sup>, Post-Graduate Student, [orcid.org/0000-0003-0969-0944](https://orcid.org/0000-0003-0969-0944)

<sup>a</sup>Skolkovo Institute of Science and Technology, bld. 1, 30, Bolshoy Boulevard, 121205, Moscow, Russian Federation

**Introduction:** Satellites which face space debris cannot track it throughout the whole orbit due to natural limitations of their optical sensors, such as field of view, Earth occultation, or solar illumination. Besides, the time of continuous observations is usually very short. Therefore, we are trying to offer the most effective configuration of optical sensors in order to provide short-arc tracking of a target piece of debris, using a scalable Extended Information Filter. **Purpose:** The best scenario for short-arc tracking of a space debris orbit using multipoint optical sensors. **Results:** We have found optimal configurations for groups of satellites with optical sensors which move along a sun-synchronous orbit. Debris orbit determination using an Extended Information Filter and measurements from multipoint sensors was simulated, and mean squared errors of the target's position were calculated. Based on the simulation results for various configurations, inter-satellite distances and measurement time, the most reliable scenario (four satellites in tetrahedral configuration) was found and recommended for practical use in short-arc debris tracking.

**Keywords** – space debris, short-arc tracking, sun-synchronous orbit, relative motion, extended information filter.

**For citation:** Afanasev A. A., Biktimirov S. N. CubeSat formation architecture for small space debris surveillance and orbit determination. *Informatsionno-upravlyaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 37–46. doi:10.31799/1684-8853-2021-4-37-46

## Introduction

Space Situational Awareness (SSA) is the knowledge and characterization of space objects and their operational environment to support safe, stable, and sustainable space activities. The problems arising in the context of SSA have become a significant concern for the space community, because of expanding growth in numbers of satellite systems and operators in Earth orbits, corresponding growth in the existing space debris population, and interaction between space systems and space debris populations, which necessitates collision risk assessment and general measures to mitigate the debris problem before it develops into what is termed Kessler syndrome [1]. Presently, space debris are monitored and catalogued using ground-based optical telescopes and radars, which allow obtaining comprehensive knowledge of all objects larger than 10 cm in size and their orbits. However, it is estimated that only about 4 % of the objects in low Earth orbit (LEO) sized one centimeter and larger are contained in the public space object catalog [2]. In particular, it is reported that the space debris density is very high in sun-synchronous orbits (SSO), which is very popular among satellite systems operators.

This study considers a number of formation-flying architectures to perform space-based optical observations for on-request orbit determination of smaller (1 to 10 cm in size) space debris objects in SSO. Prior research indicates that the on-board sen-

sors can be advantageous in SSA programs, because they are not interrupted by the daylight, and they are not characterized by the scattering, diffractions, aberrations and turbulences in the atmosphere [3]. In recent years, a number of studies have been carried out towards analyzing different aspects of spaceborne debris surveillance missions. [4] proposed space-based optical instruments and discussed how the existing knowledge gap in the space debris population in the millimetre and centimetre regime may be closed by means of a passive optical instrument. [5] discusses how a MEO satellite can potentially detect debris objects in GEO. [6] proposes a method for accurate orbit determination and prediction using short-arc tracking data. Various techniques for space debris orbit determination based on short-arc data are also discussed in [7, 8]. A multi-spacecraft formation approach to space-debris monitoring is developed in [3, 9], these papers propose a networking system of spacecraft that dynamically processes distributed measurements and recovers the debris object's orbits. Such systems allow processing multipoint measurements and exhibit collective behavior are discussed in other space applications [10], and their usage becomes indispensable for debris tracking or orbit acquisition tasks, because they cannot be carried out by a single device. The idea of cooperative object tracking using space-based optical sensors is also developed in [11, 12]. Further studies [13] go as far as proposing large constellations of spacecraft for global space debris monitoring.

Understanding the inevitable use of satellite formations for accurate small debris object tracking and their orbit determination, we carry out a comparative analysis of different formation architectures, comprising different number of spacecraft (two to four), and assuming different orbital configurations (from the simplest train-like formation, through moderately complex general-circular-orbit (GCO) formation, and ending up with the most complex tetrahedral formation). We assume that the formations are deployed into a circular sun-synchronous orbit at 700 km altitude and each spacecraft carries an optical payload to track the debris objects. We discuss the observation conditions and select an orbit close to the terminator plane. We then analyze a construction of an extended information filter to process the multipoint measurements obtained by the formation spacecraft and analyze its orbit determination accuracy taking into account the payload constraints and depending on such systems parameters as the formation type, number of spacecraft in a formation, and intersatellite distances.

### Mission design

Let us consider a circular Sun-synchronous orbit passing near to the Earth terminator line as a reference orbit for space debris observation with the aid of a small formation flying satellites. We use these kinds of reference orbits because it allows observing space debris using optical cameras at any point on orbit. For numerical simulations of the proposed algorithms let us consider the circular SSO orbits with 700 km altitude. Thus, the inclination  $i$  of the reference orbit can be found from the expression for secular rate of right ascension of ascending node  $\dot{\Omega}$ :

$$\dot{\Omega} = -\frac{3nR_{\oplus}^2 J_2}{2R_{\text{orb}}^2} \cos i, \quad (1)$$

where  $R_{\oplus} \approx 6371$  km is the radius of the Earth;  $J_2 \approx 0.00108263$  is the second order zonal harmonic of the Earth gravitational potential;  $R_{\text{orb}}$  is the

reference orbit radius,  $n = \sqrt{\frac{\mu}{R_{\text{orb}}^3}}$  is the orbit mean

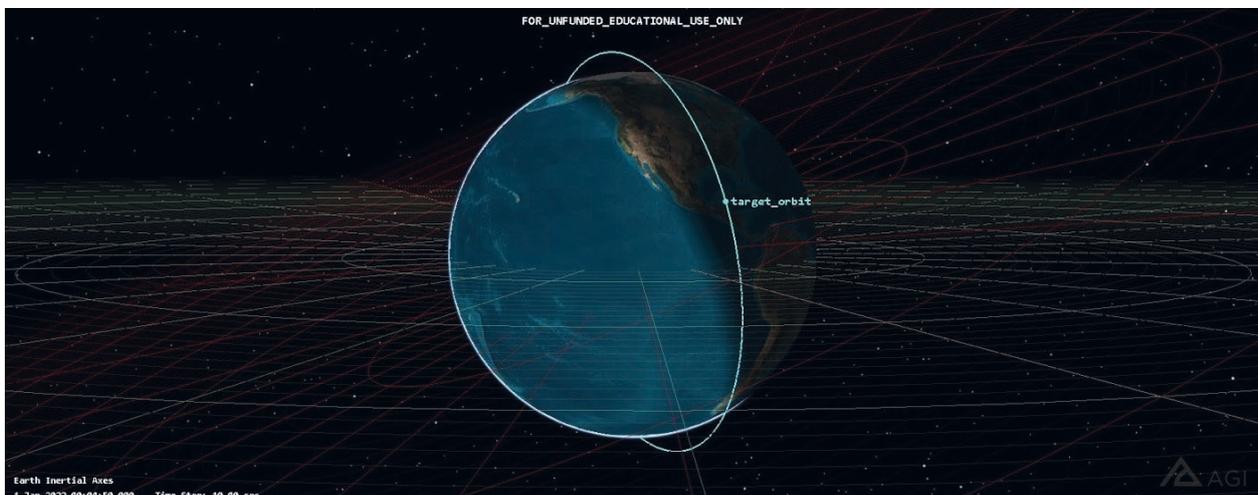
motion,  $\mu \approx 3.986 \times 10^{14}$  m<sup>3</sup>/s<sup>2</sup> is the standard gravitational parameter of the Earth. From the expression (1) the inclination of the corresponding SSO orbit is equal to 98.18 degrees.

Finally, we should find such a right ascension of ascending node (RAAN)  $\Omega$  that will orient the orbit close to the terminator line. Let us define the RAAN  $\Omega$  in way that will align projection of the orbit's unit angular momentum vector  $\mathbf{e}_h$  onto the  $xy$ -plane along the projection of the Sun position vector  $\mathbf{R}_{\text{Sun}}$  onto the  $xy$ -plane. Let us consider the simulation start epoch 1st January, 2022. The corresponding Sun position vector, given in Earth-Centered Inertial coordinates, is equal to  $\mathbf{R}_{\text{Sun}} = [26127801, -132825709.3, -57579560.5]^T$  m. This yields RAAN  $\Omega = 11.13^\circ$ . Fig. 1 depicts the reference orbit, built by Systems Toolkit software.

Reference orbit parameters at epoch Jan 1, 2022, 00:00:00 UTC:

- SMA, km — 7071;
- Eccentricity — 0;
- Inclination, deg — 98.18;
- RAAN, deg — 11.13;
- Argument of perigee (AOP), deg — 0;
- True anomaly (TA), deg — 0.

To demonstrate the proposed algorithms performance, we consider 3 common satellite formation configuration types such as train, GCO, and



■ Fig. 1. Systems Toolkit simulation of the target orbit

tetrahedron formation. For trade-off analysis we will also test each orbital configuration with different triangulation bases. In order to design the formations' orbital configurations, we employ the Hill — Clohessy — Wiltshire (HCW) equations that describe linearized relative motion dynamics for a couple of satellites at near-circular orbits. The HCW equations can be written as follows:

$$\begin{aligned} \ddot{x} + 2n\dot{z} &= \mathbf{u}_x; \\ \ddot{y} + n^2y &= \mathbf{u}_y; \\ \ddot{z} - 2n\dot{x} - 3n^2z &= \mathbf{u}_z, \end{aligned} \quad (2)$$

where  $\mathbf{u}$  is unit force vector acting on a satellite;  $x, y, z$  are satellite position vector components given in the orbital reference frame (along track, normal and radial respectively).

The HCW equations (2) admit periodic analytical solutions in case of free motion that can be written in the following form [14, 15]:

$$\begin{aligned} x(t) &= c_1 \cos(nt + \alpha) + c_3; \\ y(t) &= c_2 \sin(nt + \beta); \\ z(t) &= \frac{c_1}{2} \sin(nt + \alpha), \end{aligned} \quad (3)$$

where constants  $c_1, c_2, c_3, \alpha, \beta$  are used to define the initial conditions of the relative motion.

The orbital configurations used in the study are defined in Table 1. Parameter  $\rho$  defines a triangulation base. In the numerical simulation we use three

different values for  $\rho = 1, 5, 10$  km [10]. Fig. 2, *a-c* depict resulting orbital configurations from (3).

### Optical system dynamics

Let's consider the Earth-Centered Inertial reference frame and express the orbital dynamics of the target debris with radius-vector  $\mathbf{r}$  and velocity  $\mathbf{v}$  in this frame [16]:

$$\begin{pmatrix} \dot{\mathbf{r}} \\ \dot{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} \mathbf{v} \\ \mathbf{a}_c + \mathbf{a}_{J_2} \end{pmatrix} + \begin{pmatrix} \eta_{\mathbf{v}} \\ \eta_{\mathbf{a}} \end{pmatrix}, \quad (4)$$

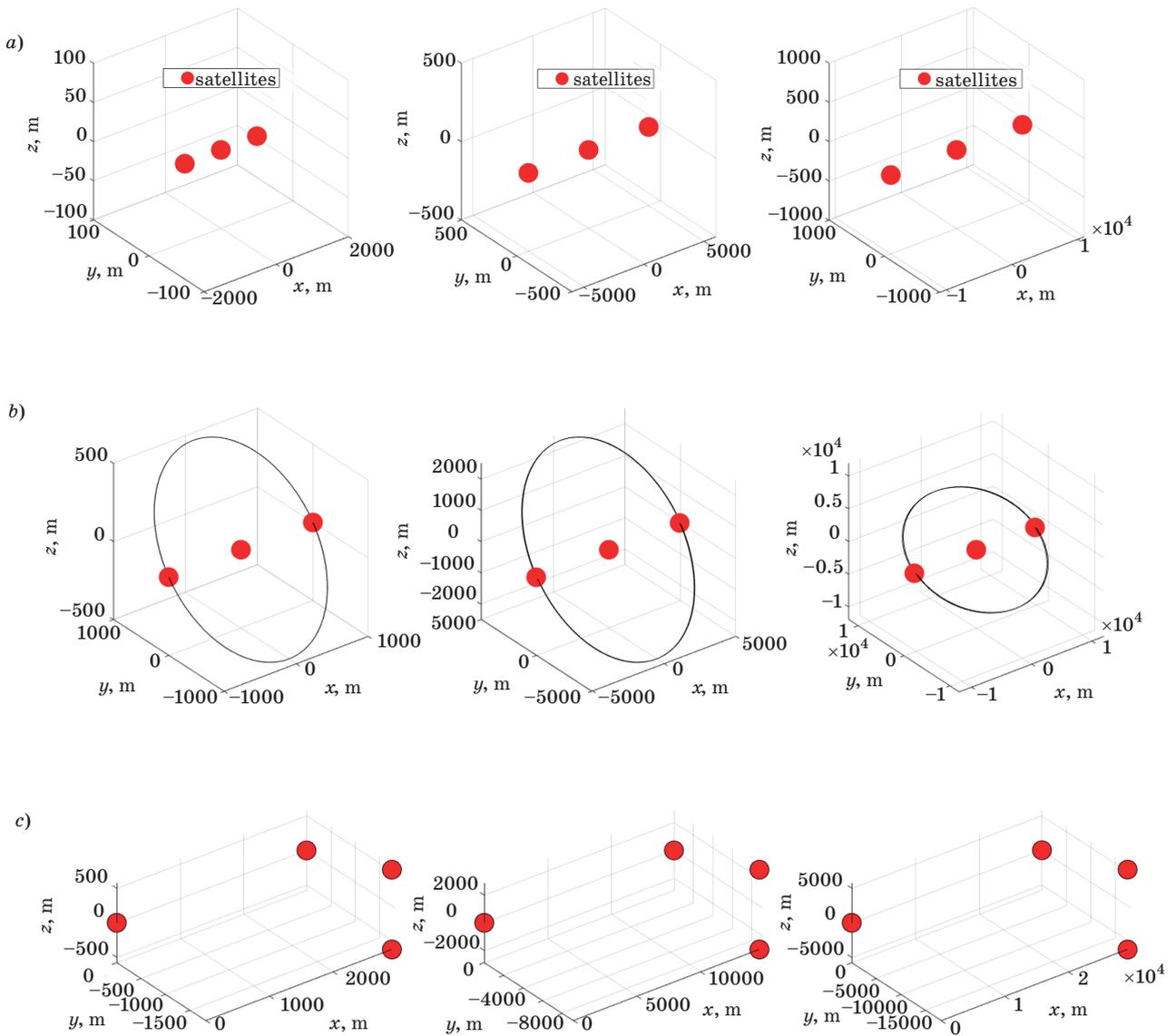
where

$$\mathbf{a}_c = -\mu \frac{\mathbf{r}}{r^3}; \quad \mathbf{a}_{J_2} = \frac{3}{2} J_2 \left( \frac{R_{\oplus}}{r} \right)^2 \begin{pmatrix} \mu \\ r^3 \end{pmatrix} \begin{pmatrix} x \left( 5 \frac{z^2}{r^2} - 1 \right) \\ y \left( 5 \frac{z^2}{r^2} - 1 \right) \\ z \left( 5 \frac{z^2}{r^2} - 3 \right) \end{pmatrix},$$

in which  $\mathbf{r} = (x \ y \ z)^T$ ;  $r = |\mathbf{r}|$ ;  $\mathbf{a}_c$  is the central body acceleration;  $\mathbf{a}_{J_2}$  is the acceleration from  $J_2$  non-spherical perturbation, and process Gaussian noises  $\eta_{\mathbf{v}} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{v}}^2 \mathbf{I}_{3 \times 3})$  and  $\eta_{\mathbf{a}} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{a}}^2 \mathbf{I}_{3 \times 3})$

■ Table 1. Initial conditions

Satellite №	$c_1$	$c_2$	$c_3$	$\alpha$	$\beta$
<b>Train formation</b>					
1	0	0	0	0	0
2	0	0	$\rho$	0	0
3	0	0	$-\rho$	0	0
<b>GCO formation</b>					
1	0	0	0	0	0
2	$\rho$	$\sqrt{3}\rho/2$	0	0	0
3	$\rho$	$\sqrt{3}\rho/2$	0	$\pi$	$\pi$
<b>Tetrahedron</b>					
1	0	0	0	0	0
2	$2\rho/5$	0	$2\rho\sqrt{5/3}$	0	0
3	$2\rho$	$\rho\sqrt{5}$	$\rho\sqrt{5/3}$	$-\arctan(1/\sqrt{2})$	$\arctan(\sqrt{2}) - \pi$
4	$2\rho$	$\rho\sqrt{5}$	$\rho\sqrt{5/3}$	$\arctan(1/\sqrt{2})$	$-\arctan(\sqrt{2})$



■ **Fig. 2.** Train (a), GCO (b) and tetrahedron (c) formations: left —  $\rho = 1$  km; centered —  $\rho = 5$  km; on the right —  $\rho = 10$  km

of the velocity and acceleration with zero mean and corresponding standard deviations  $\sigma_v$  and  $\sigma_a$  respectively,  $\mathbf{I}_{3 \times 3}$  is the identity matrix of size 3.

The position of the target is measured with two angles — elevation  $\theta$  and azimuth  $\varphi$ , which then converted into primary parameters of the debris orbit, according to the data about background stars on the image. The model of observations is following:

$$\varphi = \arctan\left(\frac{y - y_i}{x - x_i}\right) + \eta_\varphi;$$

$$\theta = \arctan\left(\frac{z - z_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2}}\right) + \eta_\theta, \quad (5)$$

where  $(x_i, y_i, z_i)^T$  is the position of the  $i$ -th sensor in Earth-Centered Inertial;  $\eta_\varphi, \eta_\theta \sim \mathcal{N}(\mathbf{0}, \sigma_{\text{meas}}^2)$  are measurement Gaussian noises with standard deviation  $\sigma_{\text{meas}}$  and zero mean [17].

### Target position estimation with Extended Information Filter

The state-space vector is  $\mathbf{X} = (\mathbf{r} \ \mathbf{v})^T$ , its covariance matrix is  $\mathbf{P}$ , measurement vector is  $\mathbf{z} = (\varphi \ \theta)^T$ . The prediction for estimating the value of state-space vector in the next time-step is given by the equation (4):  $\mathbf{X}_k = f(\mathbf{X}_{k-1}) + (\eta_v \ \eta_a)^T$  with noise covariance matrix  $\mathbf{Q}$  (identical for all  $k$ ).

Taking Jacobian of the function  $f$  gives evolution matrix:

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ F_1^c + F_2^c(x, x) + F_1^{J_2}(x) & F_2^c(x, y) + F_3^{J_2} & F_2^c(x, z) + F_4^{J_2}(x) & 0 & 0 & 0 \\ F_2^c(y, x) + F_3^{J_2} & F_1^c + F_2^c(y, y) + F_1^{J_2}(y) & F_2^c(y, z) + F_4^{J_2}(y) & 0 & 0 & 0 \\ F_2^c(z, x) + F_4^{J_2}(x) & F_2^c(z, y) + F_4^{J_2}(y) & F_1^c + F_2^c(z, z) + F_2^{J_2} & 0 & 0 & 0 \end{pmatrix}, \quad (6)$$

where

$$\begin{aligned} F_1^c &= -\frac{\mu}{r^3}; \quad F_2^c(\text{coord}_1, \text{coord}_2) = \frac{3\mu \cdot \text{coord}_1 \cdot \text{coord}_2}{r^5}; \\ F_1^{J_2}(\text{coord}) &= \frac{3J_2 R_{\oplus}^2 \mu}{2r^5} \left( \frac{5}{r^2} (\text{coord}^2 + z^2) - 1 - \frac{35 \cdot \text{coord}^2 \cdot z^2}{r^4} \right); \\ F_2^{J_2} &= \frac{3J_2 R_{\oplus}^2 \mu}{2r^5} \left( \frac{30z^2}{r^2} - 3 - \frac{35z^4}{r^4} \right); \quad F_3^{J_2} = \frac{15J_2 R_{\oplus}^2 \mu xy}{2r^7} \left( 1 - \frac{7z^2}{r^2} \right); \\ F_4^{J_2}(\text{coord}) &= \frac{15J_2 R_{\oplus}^2 \mu \cdot \text{coord} \cdot z}{2r^7} \left( 3 - \frac{7z^2}{r^2} \right). \end{aligned}$$

Transition matrix  $\Phi_{k,k-1}$  is acquired from  $\mathbf{F}_{k-1}$  (matrix  $\mathbf{F}$  from (6) in  $k-1$  time-step) as

$$\Phi_{k,k-1} = \exp(\mathbf{F}_{k-1} \cdot T), \quad (7)$$

where  $T$  is the period of movement.

Similarly, measurement vector in the current time-stop for  $i$ -th sensor is acquired from (5):  $\mathbf{z}_{k,i} = h(\mathbf{X}_{k,i}) + (\eta_{\phi} \ \eta_{\theta})^T$  with noise covariance matrix  $\Sigma$  (identical for all  $k$  and  $i$ ).

Observation matrix is the Jacobian of the function  $h$ :

$$\mathbf{H}_i = \begin{pmatrix} H_{\phi,x}^i & H_{\phi,y}^i & H_{\phi,z}^i & 0 & 0 & 0 \\ H_{\theta,x}^i & H_{\theta,y}^i & H_{\theta,z}^i & 0 & 0 & 0 \end{pmatrix}, \quad (8)$$

where

$$\begin{aligned} H_{\phi,x}^i &= \frac{y_i - y}{(x_i - x)^2 + (y_i - y)^2}; \quad H_{\phi,y}^i = \frac{x - x_i}{(x_i - x)^2 + (y_i - y)^2}; \\ H_{\phi,z}^i &= 0; \quad H_{\theta,z}^i = \frac{\sqrt{(x_i - x)^2 + (y_i - y)^2}}{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}; \\ H_{\theta,x}^i &= \frac{(x_i - x)(z - z_i)}{\sqrt{(x_i - x)^2 + (y_i - y)^2} \left( (x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2 \right)}; \\ H_{\theta,y}^i &= \frac{(y_i - y)(z - z_i)}{\sqrt{(x_i - x)^2 + (y_i - y)^2} \left( (x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2 \right)}. \end{aligned}$$

Using Extended Information Filter [18–21], state-space vector, transition (7), covariance matrix, information state  $\mathbf{y}$  and information matrix  $\mathbf{Y}$  are calculated for the predicted state:

$$\begin{aligned}
 \mathbf{X}_{k|k-1} &= f(\mathbf{X}_{k-1}); \\
 \mathbf{P}_{k|k-1} &= \Phi_{k,k-1} \mathbf{P}_{k-1} \Phi_{k,k-1}^T + \mathbf{Q}; \\
 \mathbf{y}_{k|k-1} &= \mathbf{P}_{k|k-1}^{-1} \mathbf{X}_{k|k-1}; \\
 \mathbf{Y}_{k|k-1} &= \mathbf{P}_{k|k-1}^{-1}.
 \end{aligned} \quad (9)$$

Information step contribution  $\mathbf{i}_i$  and information matrix contributions  $\mathbf{I}_i$  for each sensor from the measurements (8) update is following:

$$\begin{aligned}
 \mathbf{i}_{k,i} &= \mathbf{H}_{k,i}^T \Sigma^{-1} \left( \mathbf{z}_{k,i} - h(\mathbf{X}_{k|k-1}) + \mathbf{H}_{k,i} \mathbf{X}_{k|k-1} \right); \\
 \mathbf{I}_{k,i} &= \mathbf{H}_{k,i}^T \Sigma^{-1} \mathbf{H}_{k,i}.
 \end{aligned} \quad (10)$$

Update of the information state and matrix from contributions (10):

$$\begin{aligned}
 \mathbf{y}_k &= \mathbf{y}_{k|k-1} + \sum_{i=1}^{N_{\text{sens}}} \mathbf{i}_{i,k}; \\
 \mathbf{Y}_k &= \mathbf{Y}_{k|k-1} + \sum_{i=1}^{N_{\text{sens}}} \mathbf{I}_{i,k},
 \end{aligned} \quad (11)$$

where  $N_{\text{sens}}$  is the number of CubeSats with optical sensors.

The procedure (9)–(11) is repeated over the time-frame of the simulation for the given target and sensors orbits.

### Constraints on the optical sensors

It's clear that sensors will not be able to track the target debris in the majority of time due to various constraints, especially the limited field of view. Let's consider them and apply to the scenario.

1. The target should be bright enough to be observed:

$$\begin{aligned}
 -26.74 - 2.5 \lg \left( \frac{2}{3\pi^2} A \xi (\sin \varphi_i - \varphi_i \cos \varphi_i) \right) + \\
 + 5 \lg R_i \leq m_{\text{max}},
 \end{aligned} \quad (12)$$

where  $A$  is the target's cross-section in motion direction;  $\xi$  is the target's albedo;  $\varphi_i$  is the angle of the  $i$ -th sensor, adjacent to the solar phase angle (sun-target-sensor);  $R_i$  is the distance from the  $i$ -th sensor to the target;  $m_{\text{max}}$  is the maximal detectable magnitude

2. Optical sensors should be oriented along with the sun:

$$\cos \varphi_i \leq 0. \quad (13)$$

3. Sensors and target should be in the direct sight line above the Earth's atmosphere:

$$\sqrt{r_i^2 - (R_{\oplus} + h)^2} + \sqrt{r^2 - (R_{\oplus} + h)^2} > R_i, \quad (14)$$

where  $r_i$  and  $r$  are the absolute values of the radius-vectors of  $i$ -th sensor and target respectively;  $h \approx 100$  km is the Earth's atmosphere thickness.

4. Limit of the optical sensor is represented by the sight cone with Full Width Half Maximum (FWHM) of the sensor. Thus, the target should be inside this cone:

$$\arccos \left( \frac{\mathbf{R}_i}{R_i} \cdot \mathbf{e}_{\text{cone}} \right) < \text{FWHM}, \quad (15)$$

where  $\mathbf{R}_i$  is the vector from sensor to target;  $\mathbf{e}_{\text{cone}}$  is the normalized vector, defining cone's axis.

Applying constraints (12)–(15) drastically decreases the amount of time when target debris is detectable from satellites optical sensors configuration.

### Simulations

We now consider the simulations for all configurations of the satellites with optical sensors with following parameters: measurement period  $T = 1$  s, cone angle  $\text{FWHM} = 10^\circ$ , measurement error  $\sigma_{\text{meas}} = 5$  arcsec, maximal detectable magnitude  $m_{\text{max}} = 18$ . Orbit parameters are represented in the section "Mission design".

An average detectable target debris is given following parameters: albedo  $\xi = 0.3$ , cross-section  $A = 0.01$  m<sup>2</sup>, velocity error  $\sigma_v = 10^{-4}$  m/s, acceleration error  $\sigma_a = 10^{-6}$  m/s<sup>2</sup>. The orbit is chosen as the common debris-polluted LEO with parameters: SMA  $a_{\text{targ}} = 7177$  km, eccentricity  $e_{\text{targ}} = 10^{-6}$ , inclination  $i_{\text{targ}} = 85.4^\circ$ , RAAN  $\Omega_{\text{targ}} = 136.6^\circ$ , AOP  $\omega_{\text{targ}} = 0$ , TA  $\nu_{\text{targ}} = 300^\circ$ .

■ **Table 2.** RMSE of the target debris position for sensor configurations with  $\rho = 1$  km

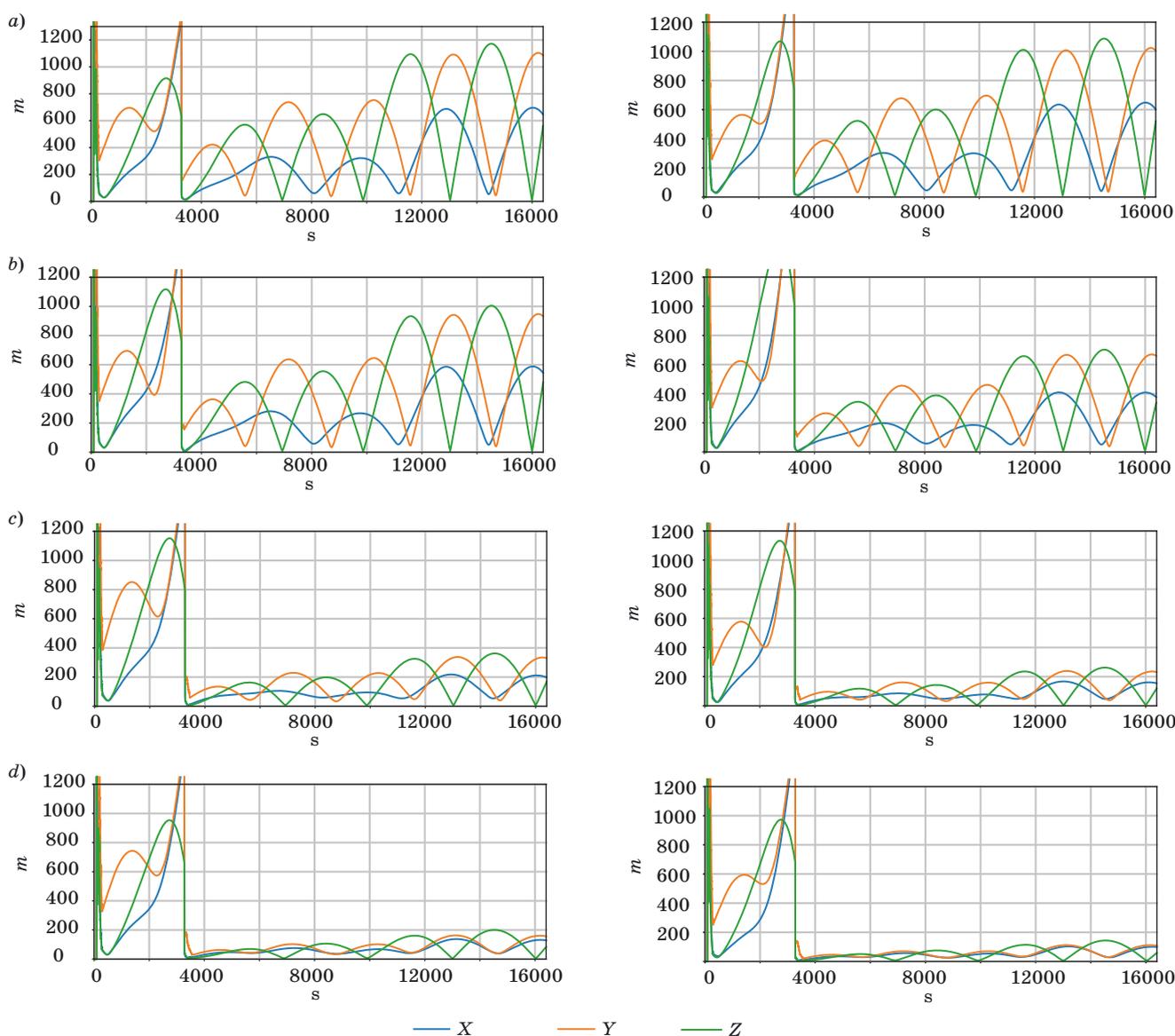
Config-uration	Number of satel-lites	RMSE, m			
		$t = 50$ s	$t = 100$ s	$t = 200$ s	$t = 300$ s
Train	2	1200	1050	280	200
	3	1070	730	220	190
GCO	2	1150	1000	380	200
	3	1100	700	220	160
Tetra-hedron	4	820	660	210	200

■ **Table 3.** RMSE of the target debris position for sensor configurations with  $\rho = 5$  km

Configuration	Number of satellites	RMSE, m			
		$t = 50$ s	$t = 100$ s	$t = 200$ s	$t = 300$ s
Train	2	1300	970	300	180
	3	1080	900	300	170
GCO	2	1200	920	230	210
	3	900	550	220	170
Tetrahedron	4	800	790	230	170

■ **Table 4.** RMSE of the target debris position for sensor configurations with  $\rho = 10$  km

Configuration	Number of satellites	RMSE, m			
		$t = 50$ s	$t = 100$ s	$t = 200$ s	$t = 300$ s
Train	2	1250	1050	360	200
	3	1110	820	300	190
GCO	2	1310	1100	260	180
	3	980	680	260	170
Tetrahedron	4	990	720	340	220



■ **Fig. 3.** RMSE of the target debris position determination vs time on orbit from 2 (left) and 3 (on right) optical sensors within satellites in GCO configuration with  $\rho = 1$  km and measurement time of  $t = 50$  s (a);  $t = 100$  s (b);  $t = 200$  s (c);  $t = 300$  s (d)

Simulations are held for  $N = 18\,000$  timesteps, with initial state covariance matrix for Extended Information Filter

$$\mathbf{P}_0 = \text{diag}(\sigma_{r_0}^2, \sigma_{r_0}^2, \sigma_{r_0}^2, \sigma_{v_0}^2, \sigma_{v_0}^2, \sigma_{v_0}^2),$$

where  $\sigma_{r_0} = 10$  km and  $\sigma_{v_0} = 10$  m/s. Initial state of the state-space vector  $\mathbf{X}_0$  coincides with the true value of this vector (that's why the initial error on the graphs will be 0). Simulations are repeated  $M = 200$  times to calculate the round mean-squared error (RMSE) of the estimated position of the target debris.

The result of all simulations can be seen in tables 2–4. RMSEs are taken for each projection of the radius-vector of the target. It's clear, that's the biggest error is located in the  $Z$ -coordinate of the target position in all cases. Thus, we are putting it in the tables as the most appropriate one to show the effectiveness of the filter.

The example of RMSE simulations is given on Fig. 3, *a–d*.

Conclusions from these simulations are following:

1. The time of measurement dramatically decrease the error of the position determination for any sensors configuration in any size. For instance, the error for  $t = 50$  s can reach up to 1.3 km, while at  $t = 300$  s the error decreases to 160 m.

2. An amount of satellites also decreases the error of position estimation. So, for 2 satellites in small times of measurement RMSEs are actually worst. The situation flattens out as the time of observation increases — for times of 300 s errors for all different number of satellites are basically same, near 180 m.

3. The type configuration seems to have rather low impact on the RMSE of the target position determination, but the results of simulations imply that GCO type of formation gives better results than the train type. Tetrahedral sensor composition is actually incomparable with that of GCO and train because of the different amount of satellites in those groups, but since this configuration is direct extension of GCO on the 4-point measurement sys-

tem, we can claim that the tetrahedral formation is the best fit for our purposes. Thus, we recommend to use tetrahedral configuration in the problem of short-arc tracking of target debris.

## Conclusions

The problem of space debris detection is highly relevant and focused by the modern community of researchers and engineers. One of the high priority problems in this area is to determine the orbit of the passing by piece of space debris with short-track measurements from the multipoint formation of the optical sensors, located in the group of satellites mainly on the sun-synchronous orbit. We managed to construct viable configurations for these sensors and compare the effectiveness in scenarios with different number of points of view (satellites) and measurement times. We implemented the extended information filter to make the simulations of the measurement of the target position by several optical sensors and calculated respective round mean-squared errors. The error diminishes when the number of sensors increases and also when the time of continuous measurement grows. The most important part is actually that the GCO configuration outstrips the train one in cases of 2 and 3 observation points, which means that the tetrahedral configuration, which is natural extension of GCO onto the case of 4 satellites, is the best use-case for determining the target debris orbit using short-arc optical measurements. The next step of this research is to collect enough statistics on the performance of this filter and configuration on the different orbits, different from sun-synchronous, and test whether it will be good enough without constant position near the terminator. The undeniable advantage of such research would be full coverage of LEO debris.

## Financial support

The reported study was funded by RFBR, project 19-38-90278.

## References

1. Kessler D. J., Cour-Palais B. G. Collision frequency of artificial satellites: the creation of a debris belt. *J. Geophys. Res.*, 1978, vol. 83, A6, pp. 2637–2646. doi:10.1029/JA083iA06p02637
2. Oltrogge D. L., Alfano S. The technical challenges of better Space Situational Awareness and Space Traffic Management. *Journal of Space Safety Engineering*, 2019, vol. 6, iss. 2, pp. 72–79. doi:10.1016/j.jsse.2019.05.004
3. Felicetti L., Emami M. R. A multi-spacecraft formation approach to space debris surveillance. *Acta Astronautica*, 2016, vol. 127, pp. 491–504. doi:10.1016/j.actaastro.2016.05.040
4. Flohrer T., Krag H., Klinkrad H., Schildknecht T. Feasibility of performing space surveillance tasks with a proposed space-based optical architecture, *Ad-*

- vances in Space Research, 2011, vol. 47, iss. 6, pp. 1029–1042. doi:10.1016/j.asr.2010.11.021
5. Jianli Du, Xiangxu Lei, Jizhang Sang. A space surveillance satellite for cataloging high-altitude small debris. *Acta Astronautica*, 2019, vol. 157, pp. 268–275. doi:10.1016/j.actaastro.2019.01.003
  6. Bin Li, Jizhang Sang, Junyu Chen. Achievable orbit determination and prediction accuracy using short-arc space-based observations of space debris. *Advances in Space Research*, 2018, vol. 62, iss. 11, pp. 3065–3077. doi:10.1016/j.asr.2018.08.038
  7. Ansalone L., Curti F. A genetic algorithm for initial orbit determination from a too short arc optical observation. *Advances in Space Research*, 2013, vol. 52, no. 3, pp. 477–489. doi:10.1016/j.asr.2013.04.004
  8. Sciré G., Santoni F., and Piergentili F. Analysis of orbit determination for space based optical space surveillance systems. *Advances in Space Research*, 2015, vol. 56, no. 3, pp. 421–428. doi:10.1016/j.asr.2015.02.031
  9. Felicetti L., Emami M. R. Attitude coordination of multiple spacecraft for space debris surveillance. *Advances in Space Research*, 2017, vol. 59, no. 5, pp. 1270–1288. doi:10.1016/j.asr.2016.12.012
  10. Afanasev A., Shavin M., Ivanov A., Pritykin D. Tetrahedral satellite formation: Geomagnetic measurements exchange and interpolation. *Advances in Space Research*, 2021, vol. 67, iss. 10, pp. 3294–3307. doi:10.1016/j.asr.2021.02.012
  11. Jia B., Pham K. D., Blasch E., Shen D., Wang Z., and Chen G. Cooperative space object tracking using space-based optical sensors via consensus-based filters. *IEEE Transactions on Aerospace and Electronic Systems*, 1936, vol. 52, no. 4, pp. 1908–2016. doi:10.1109/TAES.2016.140506
  12. Sharma J. Space-based visible space surveillance performance. *Journal of Guidance, Control, and Dynamics*, 2000, vol. 23, no. 1, pp. 153–158. doi:10.2514/2.4503
  13. Du J., Chen J., Li B., and Sang J. Tentative design of SBSS constellations for LEO debris catalog maintenance. *Acta Astronautica*, 2019, vol. 155, pp. 379–388. doi:10.1016/j.actaastro.2018.06.054
  14. Biktimirov S., Ivanov D., Sadretdinov T., Omran B., Pritykin D. A multi-satellite mission to illuminate the earth: Formation control based on impulsive maneuvers. *Advances in the Astronautical Sciences*, 2020, vol. 173, pp. 463–474.
  15. Ivanov D., Biktimirov S., et al. Writing with sunlight: cubesat formation control using aerodynamic forces. *Proceedings of the International Astronautical Congress, IAC 2019*, October, 2019.
  16. Battin R. H. *An Introduction to the Mathematics and Methods of Astrodynamics*. AIAA, 1999.
  17. Stokes G. H., Braun C. V., Sridharan R., and Sharma J. The space-based visible program. *Lincoln Laboratory Journal*, 1998, vol. 11, no. 2, pp. 205–238. doi:10.2514/6.2000-5334
  18. Jia B., Ming X., and Cheng Y. Multiple sensor estimation using the sparse Gauss — Hermite quadrature information filter. *2012 American Control Conference (ACC)*, Montreal, QC, Canada, June 2012, pp. 5544–5549, doi:10.1109/ACC.2012.6315385
  19. Chang L., Qin F., Li A., and Hu B. Unscented type Kalman filter: limitation and combination. *IET Signal Processing*, 2013, vol. 7, no. 3, pp. 167–176. doi:10.1049/iet-spr.2012.0330
  20. Wang Y., Zheng W., Sun S., and Li L. Robust information filter based on maximum correntropy criterion. *Journal of Guidance, Control, and Dynamics*, 2016, vol. 39, no. 5, pp. 1126–1131. doi:10.2514/1.G001576
  21. Simon D. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. John Wiley and Sons, 2006. 552 p.

УДК 629.783

doi:10.31799/1684-8853-2021-4-37-46

#### Архитектура формаций кубсатов для наблюдения за малым космическим мусором и определения его орбиты

А. А. Афанасьев<sup>а</sup>, аспирант, orcid.org/0000-0001-7503-9696, anton.afanasev@skoltech.ru

Ш. Н. Биктимиров<sup>а</sup>, аспирант, orcid.org/0000-0003-0969-0944

<sup>а</sup>Сколковский институт науки и технологий, Большой б-р, 30, стр. 1, Москва, 121205, РФ

**Введение:** спутники с сенсорами, которые встречают мусор, не могут отслеживать его на всей орбите из-за естественных ограничений оптических датчиков: поля зрения, затмения Земли, освещенности Солнцем. К тому же время непрерывных наблюдений обычно очень невелико. В связи с этим представляется актуальным создание наиболее эффективной конфигурации оптических датчиков для обеспечения отслеживания короткой дуги орбиты целевого куска мусора с использованием масштабируемого обобщенного информационного фильтра. **Цель:** определить наилучший сценарий отслеживания орбиты космического мусора по короткой дуге с использованием многоточечных оптических датчиков. **Результаты:** выявлены оптимальные конфигурации группы спутников с оптическими датчиками, движущихся по солнечно-синхронной орбите. Смоделирован процесс определения орбиты мусора с помощью расширенного информационного фильтра с использованием измерений от многоточечных датчиков и вычислены среднеквадратические ошибки местоположения цели. На основе результатов моделирования для различных конфигураций, межспутниковых расстояний и времени измерения найден наиболее надежный сценарий из четырех спутников в тетраэдральной конфигурации, который рекомендуется использовать на практике отслеживания мусора по короткой дуге.

**Ключевые слова** — космический мусор, слежение на короткой дуге, солнечно-синхронная орбита, относительное движение, обобщенный информационный фильтр.

Для цитирования: Afanasev A. A., Biktimirov S. N. CubeSats formation architecture for small space debris surveillance and orbit determination. *Информационно-управляющие системы*, 2021, № 4, с. 37–46. doi:10.31799/1684-8853-2021-4-37-46

For citation: Afanasev A. A., Biktimirov S. N. CubeSat formation architecture for small space debris surveillance and orbit determination. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 37–46. doi:10.31799/1684-8853-2021-4-37-46

### Финансовая поддержка

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта 19-38-90278.

### Уважаемые авторы!

При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

Статьи должны содержать изложение новых научных результатов. Название статьи должно быть кратким, но информативным. В названии недопустимо использование сокращений, кроме самых общепринятых (РАН, РФ, САПР и т. п.).

Текст рукописи должен быть оригинальным, а цитирование и самоцитирование корректно оформлено.

Объем статьи (текст, таблицы, иллюстрации и библиография) не должен превышать эквивалента в 20 страниц, напечатанных на бумаге формата А4 на одной стороне через 1,5 интервала Word шрифтом Times New Roman размером 13, поля не менее двух сантиметров.

Обязательными элементами оформления статьи являются: индекс УДК, заглавие, инициалы и фамилия автора (авторов), ученая степень, звание (при отсутствии — должность), полное название организации, аннотация и ключевые слова на русском и английском языках, ORCID и электронный адрес одного из авторов. При написании аннотации не используйте аббревиатур и не делайте ссылок на источники в списке литературы. Предоставляйте подрисовочные подписи и названия таблиц на русском и английском языках.

Статьи авторов, не имеющих ученой степени, рекомендуется публиковать в соавторстве с научным руководителем, наличие подписи научного руководителя на рукописи обязательно; в случае самостоятельной публикации обязательно предоставляйте заверенную по месту работы рекомендацию научного руководителя с указанием его фамилии, имени, отчества, места работы, должности, ученого звания, ученой степени.

Простые формулы набирайте в Word, сложные с помощью редактора Mathtype или Equation. Для набора одной формулы не используйте два редактора; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта в Mathtype никогда не пользуйтесь вкладкой Other, Smaller, Larger, используйте заводские установки редактора, не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст; пробелы в формуле ставьте только после запятой при перечислении с помощью Ctrl+Shift+Space (пробел); не отделяйте пробелами знаки: + – ×, а также пространство внутри скобок; для выделения греческих символов в Mathtype полужирным начертанием используйте Style → Other → bold.

Для набора формул в Word никогда не используйте вкладки: «Уравнение», «Конструктор», «Формула» (на верхней панели: «Вставка» — «Уравнение»), так как этот ресурс предназначен только для внутреннего использования в Word и не поддерживается программами, предназначенными для изготовления оригинал-макета журнала.

При наборе символов в тексте помните, что символы, обозначаемые латинскими буквами, набираются светлым курсивом, русскими и греческими — светлым прямым, векторы и матрицы — прямым полужирным шрифтом.

Подробнее см. pdf-файл «Правила подготовки рукописей» (стр. 11) на сайте <https://guar.ru/ric>

#### Иллюстрации:

— рисунки, графики, диаграммы, блок-схемы предоставляйте в виде отдельных исходных файлов, поддающихся редактированию, используя векторные программы: Visio (\*.vsd, \*.vsdx); Adobe Illustrator (\*.ai); Coreldraw (\*.cdr, версия не выше 15); Excel (\*.xls); Word (\*.docx); AutoCad, Matlab (экспорт в PDF, EPS, SVG, WMF, EMF); Компас (экспорт в PDF), веб-портал DRAW. IO (экспорт в PDF);

— фото и растровые — в формате \*.tif, \*.png с максимальным разрешением (не менее 300 pixels/inch).

Наличие подрисовочных подписей и названий таблиц на русском и английском языках обязательно (желательно не повторяющих дословно комментарии к рисункам в тексте статьи).

#### В редакцию предоставляются:

— сведения об авторе (фамилия, имя, отчество, место работы, должность, ученое звание, учебное заведение и год его окончания, ученая степень и год защиты диссертации, область научных интересов, количество научных публикаций, домашний и служебный адреса и телефоны, e-mail), фото авторов: анфас, в темной одежде на белом фоне, должны быть видны плечи и грудь, высокая степень четкости изображения без теней и отблесков на лице, фото можно представить в электронном виде в формате \*.tif, \*.png, \*.jpg с максимальным разрешением — не менее 300 pixels/inch при минимальном размере фото 40×55 мм;

— экспертное заключение.

#### Список литературы

составляется по порядку ссылок в тексте и оформляется следующим образом:

— для книг и сборников — фамилия и инициалы авторов, полное название книги (сборника), город, издательство, год, общее количество страниц, doi;

— для журнальных статей — фамилия и инициалы авторов, полное название статьи, название журнала, год издания, номер журнала, номера страниц, doi;

— ссылки на иностранную литературу следует давать на языке оригинала без сокращений;

— при использовании web-материалов указывайте адрес сайта и дату обращения.

Список литературы оформляйте двумя отдельными блоками по образцам lit.dot на сайте журнала (<http://i-us.ru/paperrules/>):

Литература и References.

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Руководство для авторов».

#### Контакты

Куда: 190000, Санкт-Петербург,  
Б. Морская ул., д. 67, ГУАП, РИЦ

Кому: Редакция журнала «Информационно-управляющие системы»

Тел.: (812) 494-70-02

Эл. почта: [i-us.spb@gmail.com](mailto:i-us.spb@gmail.com)

Сайт: [www.i-us.ru](http://www.i-us.ru)

## Базовый сценарий интеллектуальной поддержки принятия решений на основе моделей жизни пользователей в цифровой среде

**А. В. Смирнов<sup>а</sup>**, доктор техн. наук, профессор, главный научный сотрудник, [orcid.org/0000-0001-8364-073X](https://orcid.org/0000-0001-8364-073X), [smir@iias.spb.su](mailto:smir@iias.spb.su)

**Т. В. Левашова<sup>а</sup>**, канд. техн. наук, старший научный сотрудник, [orcid.org/00000002-1962-7044](https://orcid.org/00000002-1962-7044)

**М. В. Петров<sup>а</sup>**, научный сотрудник, [orcid.org/0000-0001-7403-5036](https://orcid.org/0000-0001-7403-5036)

<sup>а</sup>Санкт-Петербургский федеральный исследовательский центр РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ

**Введение:** в области поддержки принятия решений практика использования информации, оставленной пользователями в виде цифровых следов, пока не является широко распространенной. Ранее авторами данной работы была предложена концепция интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде для формирования рекомендаций. Задачами описываемых исследований является создание базового сценария, реализующего эту концепцию, и подхода к формированию групп пользователей с однотипными предпочтениями и поведением при принятии решений. **Цель:** прогнозирование рекомендательных решений на основе информации, которая содержится в цифровых следах пользователей. **Результаты:** разработаны 1) базовый сценарий интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде, предназначенный для рекомендации пользователю решения на основе знаний о его типе как лица, принимающего решения, о задаче принятия решений и проблемной области, и 2) подход к формированию групп пользователей с однотипными предпочтениями и поведением при принятии решений на основе информации о сегментах пользователей, о правилах поведенческой сегментации и о действиях, представленных в моделях жизни пользователей в цифровой среде. **Практическая значимость:** результаты исследований целесообразно использовать при построении современных рекомендующих систем, в которых закладываются требования к отслеживанию цифровых следов.

**Ключевые слова** – поддержка принятия решений, цифровые следы, модель жизни пользователя в цифровой среде, обобщенный паттерн, онтология лица, принимающего решения, сценарий поддержки принятия решений.

**Для цитирования:** Смирнов А. В., Левашова Т. В., Петров М. В. Базовый сценарий интеллектуальной поддержки принятия решений на основе моделей жизни пользователей в цифровой среде. *Информационно-управляющие системы*, 2021, № 4, с. 47–60. doi:10.31799/1684-8853-2021-4-47-60

**For citation:** Smirnov A. V., Levashova T. V., Petrov M. V. Scenario model of intelligent decision support based on user's digital life models. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 47–60 (In Russian). doi:10.31799/1684-8853-2021-4-47-60

### Введение

Использование информации, оставленной пользователями в виде цифровых следов, становится обычной практикой современных информационных систем. Для систем поддержки принятия решений (СППР) такая информация предоставляет возможности и может быть основой для формирования прогностических решений [1–3]. Однако выполненный авторами данной работы анализ проектов и моделей, в которых используется концепция жизни человека в цифровой среде, показал, что широкого применения в СППР эта концепция пока не нашла [4]. В данной работе жизнь человека в цифровой среде понимается в соответствии с определением, которое предложено в работе Digital Era Framework («Концепция цифровой эры») [5]. Опираясь на эту работу, термин «жизнь человека в цифровой среде» можно сопоставить с термином “digital life” («цифровая жизнь») и определить как способ су-

ществования, при котором цифровые технологии являются неотъемлемой частью всех аспектов жизни.

В работе предлагается подход, нацеленный на формирование рекомендательных решений на основе информации, которая содержится в цифровых следах пользователей. Неструктурированный контент цифровых следов систематизируется при помощи модели жизни пользователя в цифровой среде. Для работы представляет интерес аспект жизни, связанный с задачей принятия решений. В СППР модель жизни пользователя в цифровой среде служит источником информации для получения представления о задаче принятия решений, проблемной области, поведении пользователей, принятых ими решениях и для формирования групп пользователей с однотипными предпочтениями и поведением при принятии решений. Предпочтения и поведение групп пользователей описываются обобщенным паттерном группы. Для пользователя, обратившегося к СППР, опре-

деляется группа, к которой он принадлежит, и на основе предпочтений пользователей этой группы и ранее принятых ими решений формируется решение, рекомендуемое в текущей ситуации. Задача принадлежности пользователя группе решается онтологической машиной вывода как задача классификации [6–8].

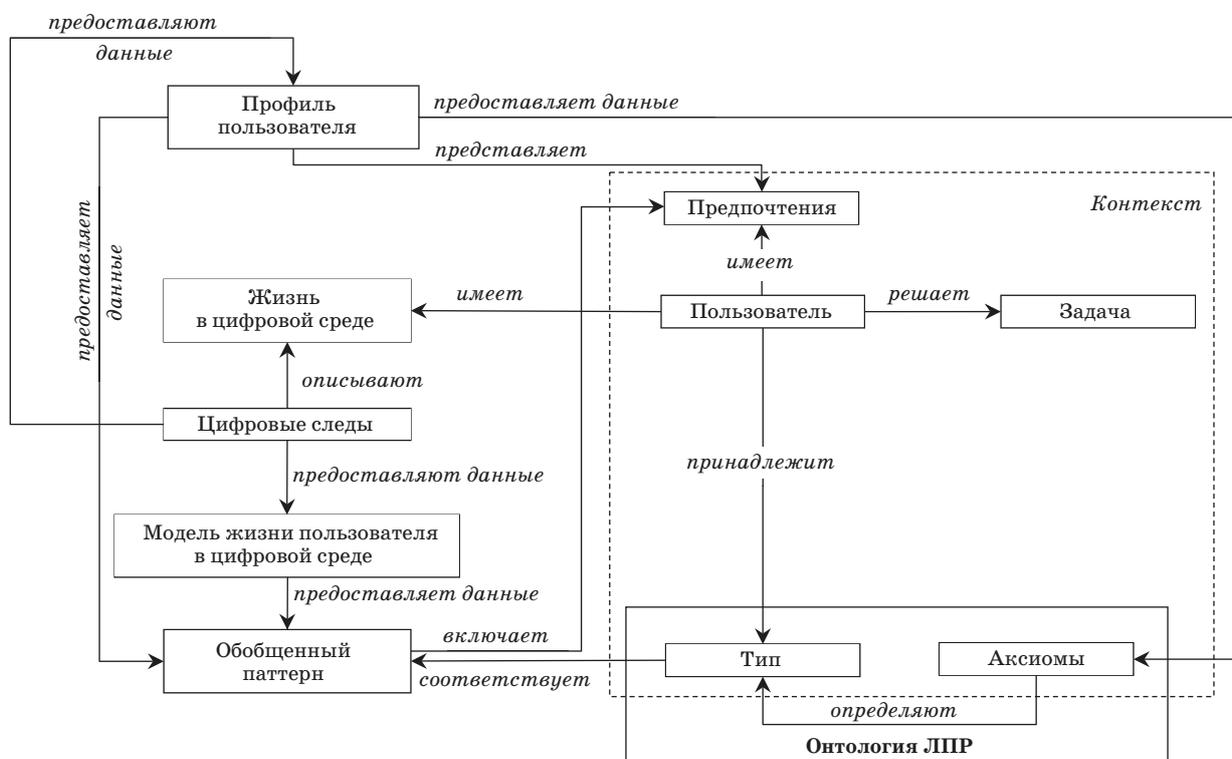
Основными задачами исследований, представленных в данной работе, являются разработка базового сценария интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде и подхода к формированию групп пользователей с однотипными предпочтениями и поведением при принятии решений.

### Концептуальная модель интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде

Концептуальная модель интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде (рис. 1) предназначена для рекомендации решения, которое пользователь принял бы в текущей ситуации (контексте). Основными блоками концептуаль-

ной модели являются профиль пользователя, модель жизни пользователя в цифровой среде, обобщенный паттерн пользователей с однотипными предпочтениями и поведением при принятии решений, онтология лица, принимающего решения (ЛПР), и контекст [4].

Профиль пользователя содержит информацию о его контекстно-зависимых и контекстно-независимых характеристиках [9]. Модель жизни пользователя в цифровой среде является комплексным формализованным описанием действий пользователя при его взаимодействии с различными сайтами, информационными средами и системами в целях решения интересующих его задач. Информация, входящая в данную модель, либо явно содержится в цифровых следах, либо выявляется из них специальными методами. Обобщенный паттерн — это формальное экстенциональное описание группы пользователей, отличающихся однотипным поведением при принятии решений и предпочтениями. Онтология ЛПР формализует аксиомы, на основании которых может быть решена задача классификации для определения, какому типу ЛПР принадлежит пользователь. Контекст представляет информацию, которая характеризует пользователя и решаемую задачу. Он складывается из информации о типе пользователя как ЛПР, его



■ **Рис. 1.** Концептуальная модель интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде [4]  
 ■ **Fig. 1.** Conceptual framework of intelligent decision support based on model of user digital life [4]

предпочтениях, задаче принятия решений и проблемной области, к которой относится задача.

Интеллектуальная поддержка принятия решений согласно концептуальной модели осуществляется следующим образом. Когда пользователь сталкивается с задачей принятия решений, информация об этой задаче и проблемной области передается из модели жизни пользователя в цифровой среде в модель контекста. Информация о пользователе из его профиля передается в онтологию ЛПП, и там определяется, какому типу ЛПП принадлежит данный пользователь в рассматриваемом контексте. На основании соответствий между типом ЛПП и обобщенными паттернами выбирается обобщенный паттерн  $GP$  и на основе информации о предпочтениях группы, описываемой этим паттерном, определяются предпочтения данного пользователя. Эти предпочтения также передаются в модель контекста. Контекстная информация и информация о решениях пользователей, описываемых обобщенным паттерном  $GP$ , являются основанием для формирования решения, которое активный пользователь принял бы в рассматриваемом контексте. Решения пользователей представлены в моделях их жизни в цифровой среде.

### Разработка базового сценария интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде

Для разработки базового сценария требуется идентифицировать процессы, которые должна поддерживать СППР, чтобы реализовать заложенную в концептуальной модели функциональность, и определить последовательность информационного взаимодействия блоков концептуальной модели при выполнении этих процессов.

### Определение последовательности информационного взаимодействия блоков концептуальной модели

Исходными данными для рассматриваемой задачи являются:

— профиль пользователя ( $PU$ ):

$$PU = \langle User\_ID, P\_out, P\_in(C) \rangle, \\ P\_in(C) = P_c(C) \cup Pr(C), \quad (1)$$

где  $User\_ID$  — уникальное имя пользователя;  $P\_out$  — множество контекстно-независимых характеристик пользователя;  $P\_in(C)$  — множество контекстно-зависимых характеристик пользователя в контексте  $C(T)$ ;  $P_c(C)$  — прочие контекстно-зависимые характеристики (текущее место-

положение пользователя, время в зоне этого местоположения и др.);  $Pr(C)$  — множество предпочтений пользователя в контексте  $C(T)$ ;  $T$  — период существования контекста  $C$ ;

— модель жизни пользователя в цифровой среде ( $DL$ ):

$$DL = \langle User\_ID, Problem(t_0, t_n), Domain, \\ \{Action(t_a^-, t_a^+)\}, Decision(t_n), R_1, R_2, R_3 \rangle,$$

$$R_1 \in Problem \times Domain, R_2 \in Problem \times Decision,$$

$$R_3 \in Action(t_a^-, t_a^+) \times Problem(t_0, t_n), \quad (2)$$

где  $Problem$  — вид решаемой пользователем задачи;  $t_0$  — время начала решения задачи;  $t_n$  — время принятия решения;  $Domain$  — проблемная область, которой принадлежит задача;  $Action(t_a^-, t_a^+)$  — действие, выполняемое на интервале  $t_a^-, t_a^+$  ( $t_0 \leq t_a^-, t_a^- < t_a^+, t_a^+ < t_n$ ) в процессе решения задачи  $Problem$ ;  $Decision$  — принятое решение;

— обобщенный паттерн группы пользователей с однотипными предпочтениями и поведением при принятии решений ( $GP$ ):

$$GP = \langle group, behaviour\_type_g, Pr_g \rangle, \quad (3)$$

где  $group$  — группа пользователей;  $behaviour\_type_g$  — вид поведения, типичный для пользователей группы  $group$ ;  $Pr_g$  — множество предпочтений, типичных для пользователей группы  $group$ ;

— онтология ЛПП ( $O_{DM}$ ):

$$O_{DM} = (Cl, Rel, A), Cl = Cl_o \cup DM\_Type, \\ A = A_o \cup A_{DM\_Type}, \quad (4)$$

где  $Cl$  — множество классов онтологии;  $Rel$  — множество отношений между классами ( $Rel \rightarrow Cl \times Cl$ );  $DM\_Type$  — класс, представляющий типы ЛПП;  $Cl_o = Cl \setminus DM\_Type$ ;  $A$  — множество аксиом;  $A_{DM\_Type}$  — множество аксиом, которые специфицируют принадлежность ЛПП типам, представленным в классе  $DM\_Type$ ;  $A_o = A \setminus A_{DM\_Type}$ ;

— множество отображений между типами ЛПП и группами пользователей с однотипными предпочтениями и поведением:  $DM\_Type \rightarrow group$ ;

— модель контекста ( $C$ ):

$$C(T) = \langle user\_ID, user\_type(T), domain(T), \\ problem(T), Pr_u(T), R_4 \rangle,$$

$$\begin{aligned}
 &user\_ID \rightarrow User\_ID, domain(T) \rightarrow Domain, \\
 &problem(T) \rightarrow Problem, Pr_u(T) \subseteq Pr, \\
 &R_4 \in domain(T) \times Pr, \quad (5)
 \end{aligned}$$

где  $T = [t_0, t_n]$ ;  $Pr_u(T)$  — множество предпочтений пользователя в контексте  $C(T)$ ;  $Pr$  — множество предпочтений в проблемной области  $domain(T)$ .

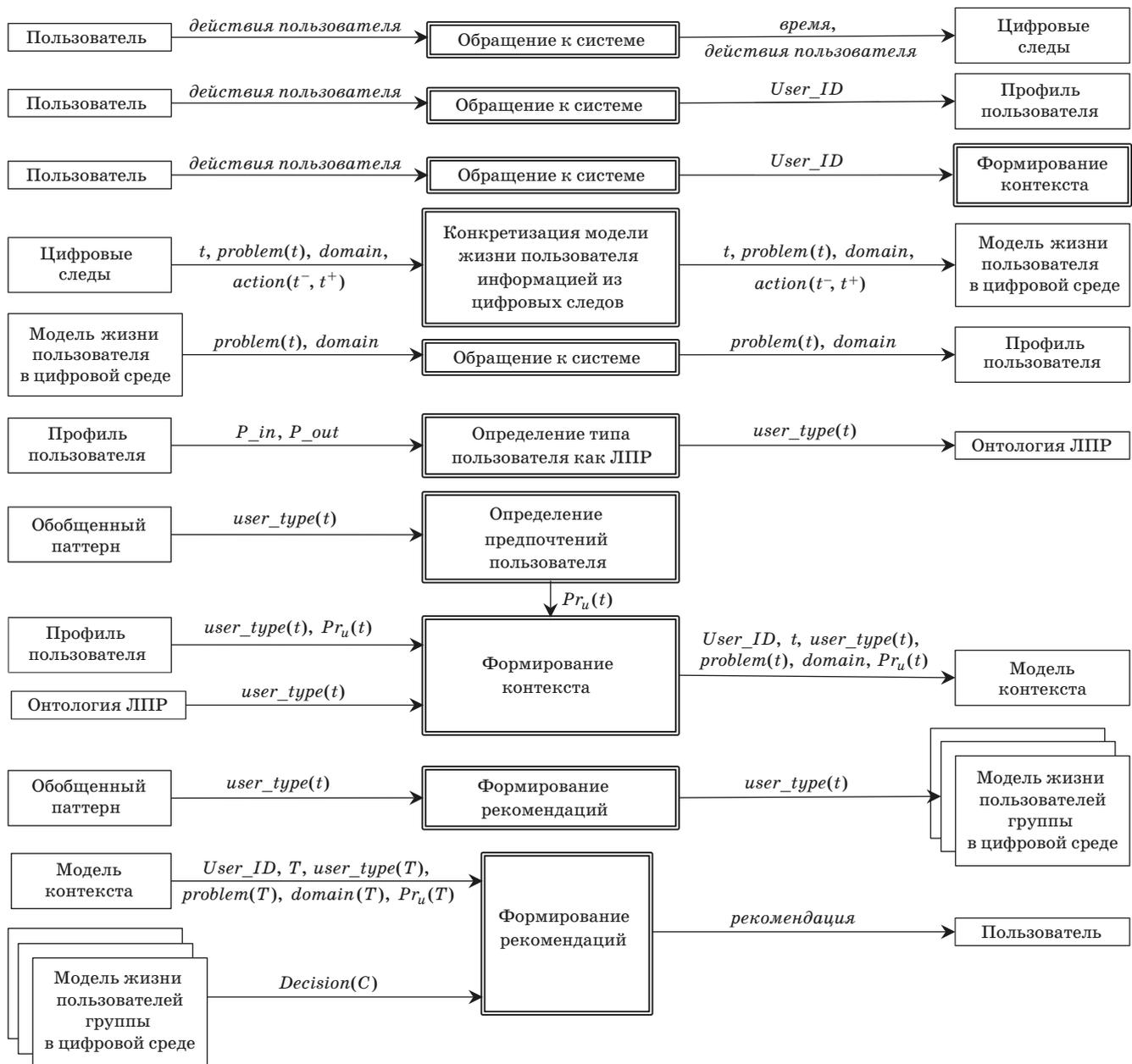
Требуется определить последовательность информационного взаимодействия блоков концептуальной модели, которая позволяет реализовать предполагаемую модель функциональности

СППР. В соответствии с этой последовательностью требуется разработать базовый сценарий интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде.

Концептуальная модель интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде предполагает следующие процессы:

— обращение пользователя к сайту, сервису или информационной системе с задачей;

— запоминание действий, выполняемых пользователем, в цифровых следах (процесс выполняется автоматически);



■ **Рис. 2.** Входные и выходные информационные потоки в СППР  
 ■ **Fig. 2.** Input & output information flows in decision support system

— конкретизация модели жизни пользователя информацией из цифровых следов (время обращения и выполнения действий, выполняемые действия, вид решаемой задачи, проблемная область);

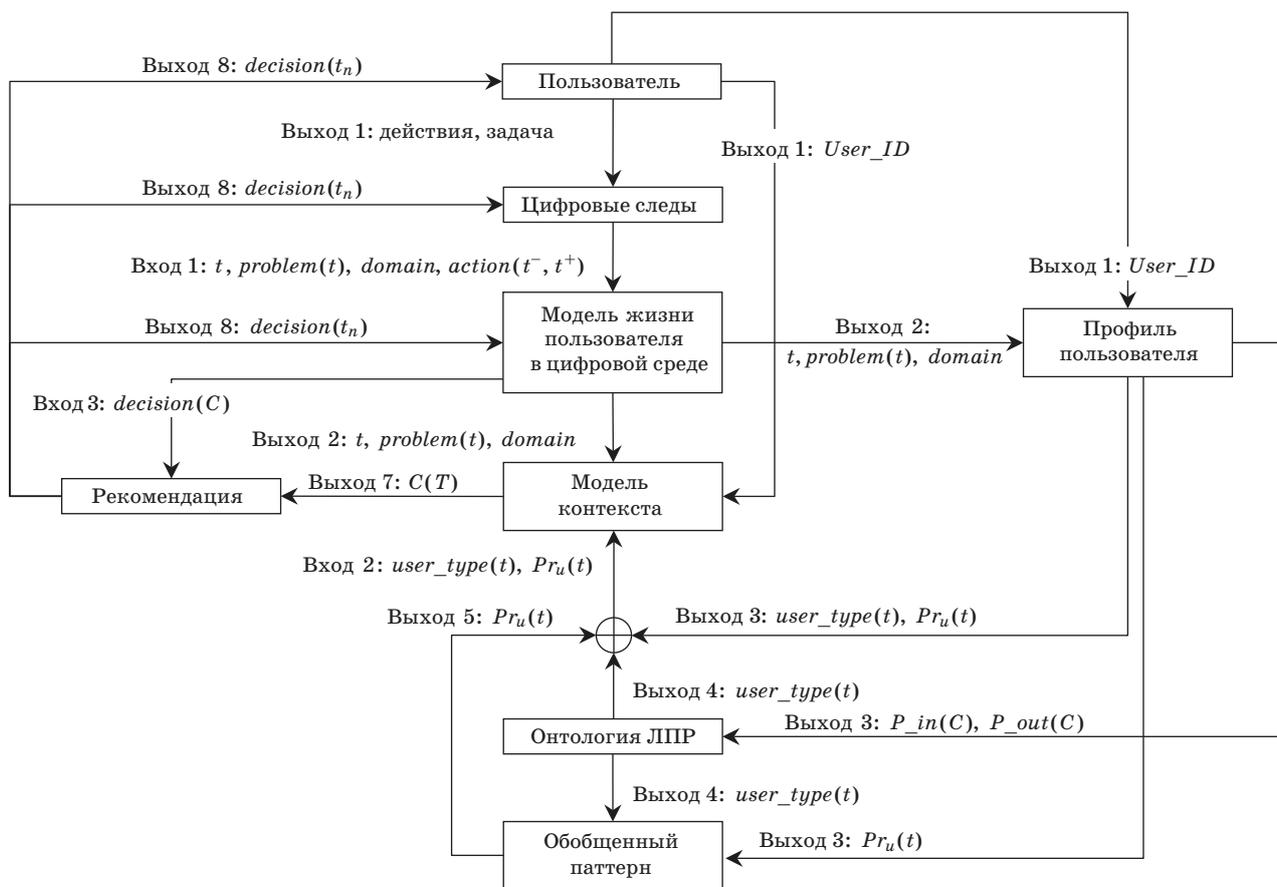
- определение типа пользователя как ЛПР;
- определение предпочтений пользователя;
- формирование контекста;
- формирование рекомендаций.

Входные и выходные информационные потоки между блоками концептуальной модели и перечисленными выше процессами показаны на рис. 2 (блоки, представляющие процессы, на рисунке обведены двойными линиями). На основании этих информационных потоков определены информационные потоки между основными блоками концептуальной модели (рис. 3). Обмен информацией между этими блоками происходит следующим образом.

Когда пользователь сталкивается с задачей принятия решений, он обращается к сайту, сервису или информационной системе с этой задачей. Уникальный идентификатор пользователя ( $User\_ID$ )

$ID$ ) передается в модель контекста, и устанавливается связь с профилем данного пользователя (профиль привязан к идентификатору). В цифровых следах пользователя появляется информация о времени его обращения к системе ( $t$ ) и о выполняемых им действиях ( $action(t^-, t^+)$ ) (фиксируется время начала и окончания выполнения каждого действия). Информация о действиях пользователя включает в себя информацию о решаемой задаче (задаче принятия решений) ( $problem(t)$ ) и проблемной области ( $domain$ ). Вся информация, связанная с действиями  $action(t^-, t^+)$ , заносится в модель жизни пользователя в цифровой среде. Процесс выявления информации о задаче и проблемной области не предполагается концептуальной моделью. Предопределено, что модель жизни пользователя содержит требуемую информацию.

Из модели жизни пользователя в цифровой среде информация о задаче и проблемной области передается в его профиль для связи контекстно-зависимых характеристик пользователя с контекстом, который создается в концептуальной модели.



■ Рис. 3. Информационные потоки между основными блоками концептуальной модели интеллектуальной поддержки принятия решений

■ Fig. 3. Information flows between main blocks of conceptual framework of intelligent decision support based on model of user digital life

Передача информации из профиля пользователя зависит от того, какая именно информация там представлена. Если есть информация о типе пользователя ( $user\_type(t)$ ) и его предпочтениях ( $Pr_u(t)$ ), то эта информация передается в модель контекста. Если нет информации о типе пользователя (вне зависимости от наличия информации о предпочтениях), то передаются значения его контекстно-зависимых ( $P\_in(C)$ ) и контекстно-независимых ( $P\_out(C)$ )

характеристик в онтологию ЛПП. Если есть информация о типе пользователя, но неизвестны его предпочтения, то тип пользователя отображается в обобщенный паттерн. На рис. 3 указанные альтернативы представлены блоком суммирования, который обозначает альтернативные варианты передачи информации в модель контекста (*выход 2*): из профиля пользователя (*выход 3*), онтологии ЛПП (*выход 4*) или обобщенного паттерна (*выход 5*).



■ **Рис. 4.** Базовый сценарий интеллектуальной поддержки принятия решений  
 ■ **Fig. 4.** Basic scenario of intelligent decision support

Из онтологии ЛПП информация о типе пользователя ( $user\_type(t)$ ) передается в обобщенный паттерн и в модель контекста. Из обобщенного паттерна информация о предпочтениях пользователя ( $Pr_u(t)$ ) передается в модель контекста. СППР формирует рекомендуемое решение (рекомендацию) на основании модели контекста и информации о решениях, принимаемых пользователями типа  $user\_type(t)$ , которые представлены в моделях жизней пользователей этого типа (вход 3). Рекомендуемое решение предлагается пользователю, отображается в цифровых следах и в модели его жизни в цифровой среде.

### Базовый сценарий

Базовый сценарий интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде приведен на рис. 4. В сценарии представлены ситуации, обрабатываемые СППР с момента обращения пользователя к системе до момента формирования рекомендаций. С точки зрения лица, получающего рекомендации, пользователь является ЛПП, так как он может принять или отклонить рекомендуемое решение. Рассматриваемые ситуации соответствуют этапам сценария. Для каждого этапа определены цель, входные данные и результаты (выходные данные).

*Этап 1*, когда пользователь столкнулся с задачей принятия решений, связан с определением типа задачи принятия решений и проблемной области, к которой относится данная задача. Как правило, тип задачи нетрудно идентифицировать в цифровых следах, в то время как для выявления информации о проблемной области могут потребоваться подходы к контекстуализации цифровых следов (например, [10–13]). Такие подходы относятся к области обработки больших данных.

Из цифровых следов тип задачи принятия решений и проблемной области заносится в модель жизни пользователя в цифровой среде ( $DL$ ) (2). Базовый сценарий получает информацию о задаче и проблемной области из этой модели. Поскольку процесс выявления информации о задаче и проблемной области не предполагается концептуальной моделью, этап извлечения информации из цифровых следов в сценарии не предусмотрен.

Для того чтобы СППР могла обрабатывать задачу, ее постановка должна быть представлена в виде, который «понятен» системе, т. е. в формальном виде. Формальная постановка задачи во многом зависит от технологии обработки задач, которая используется в системе. В данной работе речь идет об интеллектуальной поддержке принятия решений, что предполагает использование базы знаний. Такие СППР при решении задач оперируют фактами, правилами и процедурами.

Применительно к вышеописанной концептуальной модели интеллектуальной поддержки принятия решений задача должна быть представлена в виде, согласующемся с онтологическим представлением. Это обусловлено тем, что в основе концептуальной модели лежат принципы интеллектуальной поддержки принятия решений на основе моделей жизни пользователей в цифровой среде [4], в соответствии с которыми представление модели жизни пользователя (задача является частью этой модели) должно быть совместимо с онтологической моделью знаний, которая используется в СППР. В онтолого-ориентированном виде задача описывается множеством фактов (аксиом) на одном из языков представления онтологий. Тип задачи принятия решений, ее формальная постановка и проблемная область, к которой относится данная задача, являются выходными данными первого этапа.

Этап получения представления о пользователе (*этап 2*) нацелен на выбор аксиом онтологии ЛПП, которые будут использоваться для вывода типа пользователя как ЛПП. Аксиомы онтологии могут быть представлены утверждениями и правилами. Аксиомы, которые описывают принадлежность пользователя типу ЛПП, оперируют знаниями, которые связаны с характеристиками пользователя. Полностью или частично значения этих характеристик представлены в профиле пользователя. Выбираются аксиомы, которые после подстановки в них значений характеристик из профиля становятся, пользуясь терминологией OWL, утверждениями об индивидах [14], иными словами, аксиомы, для которых есть значения характеристик пользователя в его профиле.

Утверждения об индивидах (или конкретизированные аксиомы) являются входной информацией для *этапа 3*. Они используются для решения задачи классификации, в результате чего выводится тип пользователя как ЛПП —  $dm_u$  ( $dm_u \subset DM\_Type$ ).

Следует отметить, что так как некоторые характеристики пользователя являются контекстно-зависимыми, то и тип пользователя становится контекстно-зависимым, при этом один и тот же пользователь в разных контекстах может быть отнесен к разным типам ЛПП.

Следующий этап (*этап 4*) связан с выбором группы пользователей, поведение которых при принятии решений можно рассматривать как сходное по типу с поведением пользователя, взаимодействующего с СППР (активного пользователя). Такая группа определяется на основании отображения  $DM\_Type \rightarrow group$ , из которого следует, что пользователь типа  $dm_u \subset DM\_Type$  принадлежит группе  $Gr$  ( $Gr \rightarrow group$ ). Результатом этапа является вид группы, которой принадлежит активный пользователь, и соответствующий

этой группе обобщенный паттерн поведения и предпочтений.

На *этапе 5* формируется контекст (5), который представляет информацию о типе пользователя как ЛПР, его предпочтениях, задаче принятия решений и проблемной области, к которой относится задача. Информация о задаче принятия решений и проблемной области является результатом выполнения этапа 1. Информация о типе пользователя является результатом выполнения этапа 3. На рассматриваемом этапе определяются предпочтения пользователя и создается контекст. Предпочтения пользователя соответствуют предпочтениям пользователей, образующих группу *Gr*, и описаны в обобщенном паттерне этой группы. Модель контекста (5), которая конкретизирована перечисленной информацией, является результатом этапа 5.

На *этапе 6* информация о пользователе, выявленная на различных этапах базового сценария интеллектуальной поддержки принятия решений, заносится в его профиль. Данная информация касается типа пользователя как ЛПР и его предпочтений применительно к конкретной проблемной области.

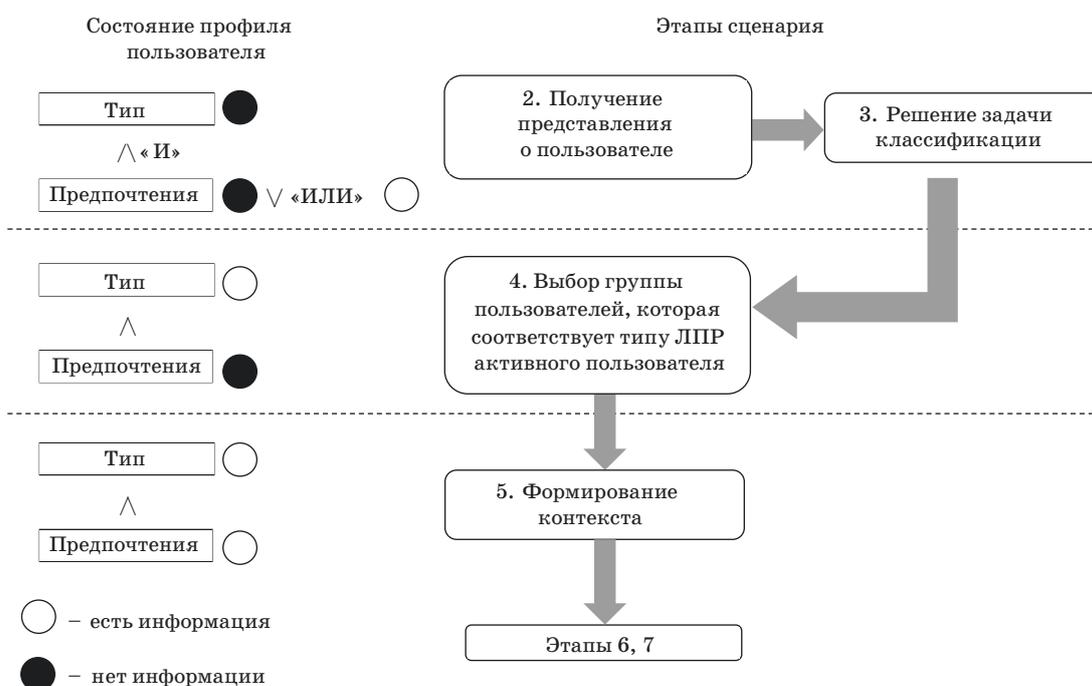
Заключительный этап сценария (*этап 7*) связан с формированием решения, которое пользователь принял бы в текущей ситуации, т. е. в контексте, сформированном на этапе 5. На основании моделей жизней пользователей в цифровой

среде, входящих в группу *Gr*, определяется, какие решения  $Decision_p$  ( $Decision_p \rightarrow problem(T) \times \times Decision$ ) пользователи этой группы принимали при рассмотрении задачи  $problem(T)$  в проблемной области  $domain(T)$ , и формируется соответствующая рекомендация.

Базовый сценарий (см. рис. 4) описывает все этапы работы СППР. Однако ввиду наличия альтернативных вариантов передачи информации в модель контекста (см. рис. 3) некоторые этапы могут не потребоваться. Необходимость выполнения этапов зависит от информации, представленной в профиле пользователя, а именно от наличия информации о его типе как ЛПР и его предпочтениях (рис. 5). Такая информация может содержаться в профиле, если:

- пользователь ранее обращался к СППР и тип данного пользователя и его предпочтения для данной проблемной области были сохранены системой в его профиле;
- профиль пользователя содержит информацию о его предпочтениях применительно к рассматриваемой проблемной области, выявленную различными сайтами, средами и системами, с которыми данный пользователь когда-либо взаимодействовал.

Для оценки базового сценария интеллектуальной поддержки принятия решений на основе модели жизни пользователя в цифровой среде использовался сравнительный метод. Для сравнения выбраны следующие подходы:



■ **Рис. 5.** Выполнение этапов в зависимости от состояния профиля пользователя  
 ■ **Fig. 5.** Phases depending on information available in user profile

а) формирование рекомендаций без использования модели жизни пользователя в цифровой среде;

б) формирование рекомендаций на основе модели жизни пользователя в цифровой среде и обобщенных паттернов поведения и предпочтений;

в) формирование рекомендаций на основе модели жизни пользователя в цифровой среде и онтологии ЛПР;

г) базовый сценарий: формирование рекомендаций на основе модели жизни пользователя в цифровой среде, обобщенных паттернов поведения и предпочтений, а также онтологии ЛПР.

В качестве критериев оценки сложности процесса формирования рекомендации и качество решений. Сложность процесса формирования рекомендации оценивалась посредством оценки количества и сложности процессов, требующихся для реализации базового сценария. Процессы, связанные с определением предпочтений пользователя, при оценке не учитывались, поскольку отличительной особенностью предлагаемого сценария для формирования персонализированных рекомендаций является учет типов поведения пользователя как ЛПР. Для оценки сложности процесса использовались константы  $\delta_1, \dots, \delta_n$ , где  $\delta_{i+1} > \delta_i$  (чем больше значение индекса у  $\delta$ , тем сложнее процесс);  $n$  — количество используемых констант. При оценке качества решений использовалась относительная шкала оценок от низкого до хорошего.

Процессы, требующиеся для реализации базового сценария:

$P_1$ : определение вида задачи принятия решений и проблемной области;

$P_2$ : отслеживание выполняемых пользователем действий;

$P_3$ : архивирование принятых пользователем решений;

$P_4$ : формирование групп пользователей с однотипным поведением (и предпочтениями);

$P_5$ : определение типа пользователя как ЛПР;

$P_6$ : определение группы пользователей с однотипным поведением (и предпочтениями), к которой относится активный пользователь.

Оцениваемая сложность процессов в подходе «а»: сложность процесса  $P_1$  зависит от наличия в СППР классификатора видов задач (применительно к проблемным областям). Если классификатор есть, то  $P_1 = \delta_2$ , в противном случае  $P_1 = \delta_5$ . Сложность остальных процессов оценивается следующим образом:  $P_2 = \delta_3$ ;  $P_3 = \delta_1$ ;  $P_4 = \delta_4$  (группы формируются из пользователей, когда-либо обращавшихся к СППР);  $P_5 = \delta_4$ ;  $P_6 = \delta_2$ . Подход обеспечивает хорошее качество решений для известных пользователей и известных задач, но низкое для нового пользователя в связи с от-

сутствием накопленных о нем данных. В целом качество этого подхода оценивается как ниже среднего.

Оцениваемая сложность процессов в подходе «б»: процессы  $P_1, P_2, P_3, P_4$  не требуются;  $P_5 = \delta_5$ ;  $P_6 = \delta_2$ . Подход обеспечивает одинаково хорошее качество решений как для известных и новых пользователей, так и для повторяющихся и новых задач.

Оцениваемая сложность процессов в подходе «в»: процессы  $P_1, P_2, P_3$  не требуются;  $P_4 = \delta_5$ ;  $P_5 = \delta_2$ ;  $P_6 = \delta_4$ . Подход обеспечивает хорошее качество решений при наличии однозначного соответствия между логически выведенным типом ЛПР и группой пользователей с однотипным поведением и предпочтениями. В противном случае качество решений оценивается как среднее. В целом качество этого подхода оценивается как среднее.

Оцениваемая сложность процессов в подходе «г»: процессы  $P_1, P_2, P_3, P_4$  не требуются;  $P_5 = \delta_5$ ;  $P_6 = \delta_3$ . Подход обеспечивает одинаково хорошее качество решений как для известных и новых пользователей, так и для повторяющихся и новых задач.

Результаты сравнения подходов приведены в таблице. Из таблицы видно, что самая высокая сложность процесса формирования рекомендаций в подходе «а». В то же время этот подход не гарантирует хорошего качества решений. Подход «б» обеспечивает хорошее качество решений при меньшей сложности процесса, чем подход «а». Подход «в» не является гарантией хорошего качества решений при достаточно высокой сложности (меньше, чем в «а», но больше, чем в «б»). Подход «г» — базовый сценарий — обеспечивает хорошее качество решений, позволяя при этом уменьшить сложность процесса формирования рекомендаций.

- Сравнение подходов к реализации базового сценария
- Evaluation of approaches to the basic scenario implementation

Подход	Сложность процесса формирования рекомендаций	Качество решений
а	$\delta_1 + 2\delta_2 + \delta_3 + 2\delta_4$ или $\delta_1 + \delta_2 + \delta_3 + 2\delta_4 + \delta_5$	Ниже среднего
б	$\delta_2 + \delta_5$	Хорошее
в	$\delta_2 + \delta_4 + \delta_5$	Среднее
г (базовый сценарий)	$\delta_2 + \delta_3$	Хорошее

### Формирование групп пользователей с однотипными предпочтениями и сценариями поведения

В концептуальной модели интеллектуальной поддержки принятия решений введено понятие группового паттерна поведения и предпочтений, которые типичны для группы пользователей *GP* (3). Источниками информации для выявления таких групп являются модели жизни пользователей в цифровой среде *DL* (2).

В качестве основы для определения видов групп пользователей с однотипными предпочтениями и поведением предлагается использовать виды сегментов пользователей на основе их поведения, которые выделяются в различных проблемных областях, а также правила, применяемые для определения принадлежности пользователей этим сегментам (рис. 6). Поскольку в различных проблемных областях интерес представляют различные поведенческие типы пользователей, в каждой проблемной области используются собственные способы и правила сегментации [15]. Например, сегменты пользователей в области электронной торговли соответствуют различным типам покупателей (осторожные, рациональные, зависимые и т. д.), сегменты пользователей в области туризма соответствуют различным типам туристов (нацеленные на развлечения, деловой туризм, приверженцы определенных культурных центров и т. д.).

Выявленные правила сегментации приводят к общему виду, который выбирается разработчиками системы. Основные требования, предъявляемые к виду правил: а) правила осуществляют логический вывод на основе информации, которая может быть предоставлена моделями жизни пользователей в цифровой среде; б) правила семантически совместимы с онтологией ЛПП. Последнее требование связано с тем, что совместимость позволяет пополнять онтологию новыми знаниями, полученными в ходе логического вывода. В частности, уточнять классификацию типов ЛПП. Примерами правил, удовлетворяющих перечисленным требованиям, являются RuleML (Rule Markup Language) [16], RIF (Rule Interchange Format) [17], SWRL (Semantic Web Rule Language) [18] и др. [19].

Работа по идентификации видов сегментов, извлечению правил сегментации и их преобразованию выполняется экспертами.

Применение правил сегментации к действиям пользователя, которые представлены в модели его жизни в цифровой среде ( $Action(t_a^-, t_a^+)$ ), позволяет получить группы пользователей с однотипным поведением и предпочтениями в заданной проблемной области. Действия  $Action(t_a^-, t_a^+)$ ,



■ **Рис. 6.** Схема формирования групп пользователей с однотипными предпочтениями и сценариями поведения

■ **Fig. 6.** Scheme of creation of groups of users with the same kinds of preferences and decision-making behaviour

помимо прочих, включают в себя действия, связанные с указанием предпочитаемых критериев оценки альтернатив при выборе решения, т. е. с предпочтениями пользователя.

С учетом вышеизложенного модель обобщенного паттерна *GP* (3) дополняется информацией о проблемной области:

$$GP^* = (group, behaviour\_type_g \rightarrow behaviour\_type, Pr_g \subseteq Pr_{dr}, R_5, R_6),$$

$$R_5 \in Domain \times Behaviour\_Type,$$

$$R_6 \in Domain \times Pr,$$

где  $GP^*$  — уточненная модель обобщенного паттерна *GP*;  $behaviour\_type \in Behaviour\_Type$ ,  $Behaviour\_Type$  — множество сегментов на основе поведения пользователей в проблемной области *Domain*;  $Pr_{dr}$  — множество предпочтений, используемых в правилах сегментации для определения принадлежности пользователей сегментам  $Behaviour\_Type$ .

#### Пример «Интеллектуальная поддержка принятия решений для рекомендации пользователю книги в библиотеке»

Пользователь *u* обращается в библиотеку. Модель жизни пользователя *u* ( $DL_u$ ):

$User\_ID = Alex$   
 $Problem(t_0) = \text{Поиск}(2020-11-19\ 19:55:16.057)$   
 $Domain = \text{Библиотека}$

$$\left\{ Action(t_a^-, t_a^+) \right\} = \left\{ \begin{array}{l} Press.MenuItem.Search(2020-11-19\ 19:55:17.0648, 2020-11-19\ 19:55:17.926), \\ Chose.Option.Title(2020-11-19\ 19:55:18.0256, 2020-11-19\ 19:55:18.873), \\ Enter.Title."Progtramming in Java"(2020-11-19\ 19:55:19.203, 2020-11-19\ 19:55:19.936) \end{array} \right\}.$$

Интерпретация модели: 19 ноября 2020 года в 19:55 пользователь Alex обратился в библиотеку с задачей поиска. Он выполнял следующие действия: выбрал пункт меню “Search” (искать), выбрал опцию “Title” (заголовок) и ввел название книги “Programming in Java”. На момент начала работы пользователя в библиотеке время принятия решения  $t_n$  и само решение  $decision(t_n)$  не известны.

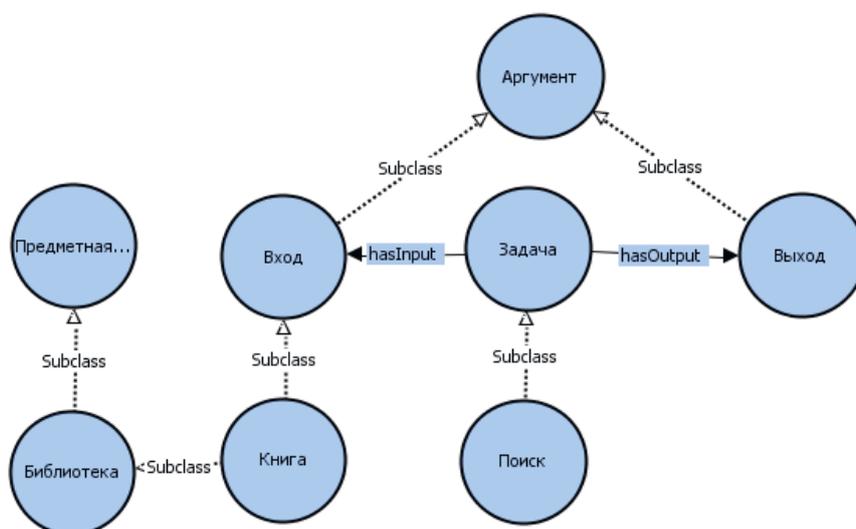
Графический пример формальной постановки рассматриваемой задачи в онтолого-ориентированном виде на языке OWL [20] приведен на рис. 7. Изображение выполнено при помощи средства визуализации OWL-онтологий — VOWL (Visual Notation for OWL Ontologies) [21]. В используемой версии VOWL возможности для представления индивидов не предусмотрены, поэтому название книги на рисунке не показано.

Значения характеристик, которые представлены в профиле пользователя и могут быть подставлены в аксиомы онтологии:

- род деятельности: библиотекарь;
- количество анализируемых критериев: не более четырех;
- набор анализируемых критериев: предопределен;
- предпочитаемые годы публикаций: современные;
- отношение предпочтения между формальными и содержательными критериями: формальные критерии на 80 % предпочтительнее, чем содержательные.

Утверждения об индивидах, полученные после подстановки значений в аксиомы, классифицировали пользователя как принадлежащему типу ЛПР-аналитик на среднем количестве известных критериев с общим представлением содержимого. В онтологии этот тип представлен классом «Аналитик\_В». В настоящей статье сама онтология не приводится из-за ограниченного объема.

Для данного типа ЛПР в проблемной области «Библиотека» существуют отображения в различные группы пользователей, которые характеризуются однотипным поведением и предпочтениями. Для установки однозначного отображения требуется дополнительный вывод. Это достигается посредством применения правил, которые связаны с полученным типом ЛПР. В частности, одним из таких правил является правило, что если в модели жизни пользователя есть действия, связанные с получением заказов



■ **Рис. 7.** Онтолого-ориентированная постановка задачи поиска книги  
 ■ **Fig. 7.** Ontology-based representation for task of book search

на библиографию, или пользователь делится библиографиями, то типу этого пользователя в рассматриваемой проблемной области соответствует группа консультантов. На языке SWRL это правило представляется множеством правил R1–R3 (данный язык не поддерживает операцию логического «ИЛИ»):

```
R1: User(?u) ^ tracedAction(?x,?y) ^
Request(?y) ^ requestedItem(?y,?x) ^
Bibliography(?z) -> User_type(Консультант)
R2: User(?u) ^ tracedAction(?x,?y) ^
Send(?y) ^ sentItem(?y,?x) ^ Bibliography(?z) ->
User_type(Консультант)
R3: User(?u) ^ tracedAction(?x,?y) ^
Makes_Shared(?y) ^ sharedItem(?y,?x) ^
Bibliography(?z) -> User_type(Консультант),
```

где *tracedAction* — отношение между классом *Action* (действий пользователя) и классами *Request* (запрос), *Send* (отправка) и *Makes\_Shared* (делиться (чем-либо)); *requestedItem* — отношение между классом *Request* и объектом запроса, в рассматриваемом случае *Bibliography* (библиография); *sentItem* — отношение между классом *Send* и объектом отправки (*Bibliography*); *sharedItem* — отношение между классом *Makes\_Shared* и объектом, к которому организован доступ (*Bibliography*).

В модели жизни активного пользователя обнаружены указанные действия, поэтому для него получено отображение: Аналитик *B* → Консультант. Обобщенный паттерн группы *lk*:

$$GP_{lk} = (\text{Консультант}, behaviour_{lk}, Pr_{lk}),$$

где *behaviour<sub>lk</sub>* — распространение библиографий; *Pr<sub>lk</sub>* — множество предпочтений группы.

Чтобы не усложнять пример, множество предпочтений не раскрыто. Примерами предпочтений являются: при наличии переизданий предпочтение отдается последнему; при наличии переводов предпочтение отдается переводу на язык консультанта; при поиске профильных публикаций предпочтение отдается профильным изданиям; при наличии электронных публикаций предпочтение отдается бесплатным и т. п.

Информация о типе пользователя как ЛПР при поиске книг в библиотеке и его поведенческих предпочтениях заносится в его профиль.

Контекст для задачи обращения пользователя в библиотеку:

$$C(T) = (Alex, \text{Аналитик\_}B(T), \text{Библиотека}(T), \text{Поиск}(T), Pr_{lk}(T)),$$

где  $T = [2020-11-19\ 19:55:16.057, t_c)$ ,  $t_c$  — текущий момент времени в СППР.

Анализ моделей жизней пользователей группы Консультант показал, что они при поиске книги “Programming in Java” выбирали книгу:

Java SE 9 : полное руководство: [пер. с англ.] / Герберт Шилдт. — Москва [и др.]: Вильямс, 2018. — 1488 с. — Пер. изд.: Java SE 9 / Herbert Schildt. — 500 экз. — ISBN 978-5-6040043-6-4 (в пер.).

Указанная книга рекомендуется пользователю. Если бы в моделях жизней пользователей не нашлось решения для задачи поиска именно этой книги, то СППР предложила бы книгу, найденную на основе предпочтений группы.

Если пользователь удовлетворен рекомендацией, то предложенное решение появляется в цифровых следах и, соответственно, в модели его жизни в цифровой среде.

## Заключение

Предложен базовый сценарий интеллектуальной поддержки принятия решений на основе моделей жизни пользователей в цифровой среде, который реализует ранее предложенную авторами данной работы концепцию интеллектуальной поддержки принятия решений на основе моделей жизни пользователей в цифровой среде. Эта концепция нацелена на формирование рекомендательных решений на основе информации, которая содержится в цифровых следах пользователей. Сценарий описывает ситуации, обрабатываемые СППР, с момента обращения пользователя к системе до момента формирования рекомендаций. Сценарий позволяет обрабатывать запросы с неполной информацией за счет учета предпочтений и анализа решений пользователей, которые имеют такой же тип поведения при принятии решений и такие же предпочтения, как и активный пользователь.

Предложен способ формирования групп пользователей с однотипными поведением и предпочтениями при принятии решений с использованием информации о сегментах пользователей, созданных в различных проблемных областях на основе поведенческих признаков этих пользователей, о правилах поведенческой сегментации и о действиях, представленных в моделях жизни пользователей в цифровой среде.

## Финансовая поддержка

Концептуальная модель, базовый сценарий интеллектуальной поддержки принятия решений на основе моделей жизни пользователей в цифровой среде, способ формирования групп пользователей с однотипными поведением и предпочте-

ниями при принятии решений разрабатывались при финансовой поддержке РФФИ в рамках научного проекта № 20-07-00455. Онтология ЛПП разрабатывалась при финансовой поддержке РФФИ в части классификации типов ЛПП в рамках научного проекта № 19-37-90094 и в части

аксиоматизации типов ЛПП в рамках научного проекта № 20-07-00490. Контекстно-зависимая типизация пользователей является частью исследований по контекстно-ориентированному поведению пользователей, которые выполнялись в рамках бюджетной темы № 0073-2019-0005.

## Литература

1. Araujo T., Helberger N., Kruikemeier S., de Vreese C. H. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, 2020, vol. 35, pp. 611–623. doi: <https://doi.org/10.1007/s00146-019-00931-w>
2. Han M. L., Kwak B. H., Kim H. K. CBR-based decision support methodology for cybercrime investigation: Focused on the data-driven website defacement analysis. *Security and Communication Networks*, 2019, vol. 2019, Article ID 1901548. doi: [10.1155/2019/1901548](https://doi.org/10.1155/2019/1901548)
3. Asniar Surendro K. Predictive analytics for predicting customer behavior. *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT)*, IEEE, 2019, pp. 230–233. doi: [10.1109/ICAIIIT.2019.8834571](https://doi.org/10.1109/ICAIIIT.2019.8834571)
4. Смирнов А. В., Левашова Т. В. Контекстно-управляемый подход к интеллектуальной поддержке принятия решений на основе цифровых следов пользователей. *Информатика и автоматизация*, 2020, т. 19, № 5, с. 915–941. doi: [10.15622/ia.2020.19.5.1](https://doi.org/10.15622/ia.2020.19.5.1)
5. Lengsfeld J. *Digital Era Framework*. Bad Waldsee, Dr. Jörn Lengsfeld, 2019. 306 p.
6. Dentler K., Cornet R., ten Teije A., de Keizer N. Comparison of reasoners for large ontologies in the OWL 2 EL Profile. *Semantic Web*, 2011, vol. 2, no. 2, pp. 71–87. doi: [10.3233/SW-2011-0034](https://doi.org/10.3233/SW-2011-0034)
7. Abburu S. A Survey on ontology reasoners and comparison. *International Journal of Computer Applications*, 2012, vol. 57, no. 17, pp. 33–39. doi: [10.5120/9208-3748](https://doi.org/10.5120/9208-3748)
8. Parsia B., Matentzoglou N., Gonçalves R. S., Glimm B., Steigmiller A. The OWL Reasoner Evaluation (ORE) 2015 Competition Report. *Journal of Automated Reasoning*, 2017, vol. 59, no. 4, pp. 455–482. doi: [10.1007/s10817-017-9406-8](https://doi.org/10.1007/s10817-017-9406-8)
9. Городецкий В. И., Тушканова О. Н. Онтологии и персонализация профиля пользователя в рекомендующих системах третьего поколения. *Онтология проектирования*, 2014, т. 13, № 3, с. 7–31.
10. Vuong T., Jacucci G., Ruotsalo T. Watching inside the Screen. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2017, vol. 1, no. 3, Article no. 109, pp. 1–23. doi: [10.1145/3130974](https://doi.org/10.1145/3130974)
11. Breiter A., Hepp A. The Complexity of Datafication: Putting Digital Traces in Context. In: *Communicative Figurations*. Cham, Springer International Publishing, 2018. Pp. 387–405. doi: [10.1007/978-3-319-65584-0\\_16](https://doi.org/10.1007/978-3-319-65584-0_16)
12. IJOC Special Section “Digital Traces In Context”. *International Journal of Communication* / ed. by A. Hepp, T. N. Friemel, A. Breiter. Los Angeles: USC Annenberg Press, 2018, vol. 12, pp. 439–705.
13. Pentland B. T., Recker J., Wolf J., Wyner G. Bringing context inside process research with digital trace data. *Journal of the Association for Information Systems*, 2020, vol. 21, no. 5, pp. 1214–1236. doi: [10.17705/1jais.00635](https://doi.org/10.17705/1jais.00635)
14. Glimm B., Horrocks I., Motik B., Shearer R., Stoilos G. A novel approach to ontology classification. *Journal of Web Semantics*, 2012, vol. 14, pp. 84–101. doi: [10.1016/j.websem.2011.12.007](https://doi.org/10.1016/j.websem.2011.12.007)
15. Bayer J. Customer segmentation in the telecommunications industry. *Journal of Database Marketing & Customer Strategy Management*, 2010, vol. 17, no. 3–4, pp. 247–256. doi: [10.1057/dbm.2010.21](https://doi.org/10.1057/dbm.2010.21)
16. Specification of RuleML. [http://wiki.ruleml.org/index.php/Specification\\_of\\_RuleML](http://wiki.ruleml.org/index.php/Specification_of_RuleML) (дата обращения: 31.05.2021).
17. Kifer M., Boley H. *RIF Overview*. Second Edition. W3C Working Group Note. W3C, 2013. <https://www.w3.org/TR/rif-overview/> (дата обращения: 31.05.2021)
18. Horrocks I., Patel-Schneider P. F., Boley H., Tabet S., Grosz B., Dean M. *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*. W3C Member Submission, 2004. <https://www.w3.org/Submission/SWRL/> (дата обращения: 31.05.2021).
19. Mehla S., Jain S. Rule Languages for the Semantic Web. In: *Emerging Technologies in Data Mining and Information Security*. Ed. by A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, S. Dutta. Springer Singapore. Advances in Intelligent Systems and Computing, 2019. Vol. 755. Pp. 825–834. doi: [10.1007/978-981-13-1951-8\\_73](https://doi.org/10.1007/978-981-13-1951-8_73)
20. OWL 2 Web Ontology Language Document Overview. W3C Recommendation. Ed. W3C OWL Working Group. W3C, 2012. <https://www.w3.org/TR/owl2-overview/> (дата обращения: 31.05.2021).
21. Negru S., Lohmann S., Haag F. *VOWL: Visual Notation for OWL Ontologies*. Specification of Version 2.0. VisualDataWeb.org, 2014. <http://vowl.visualdataweb.org/v2/> (дата обращения: 31.05.2021).

UDC 004.89:004.822

doi:10.31799/1684-8853-2021-4-47-60

Scenario model of intelligent decision support based on user's digital life models

A. V. Smirnov<sup>a</sup>, Dr. Sc., Tech., Professor, Chief Researcher, orcid.org/0000-0001-8364-073X, smir@iiias.spb.su

T. V. Levashova<sup>a</sup>, PhD, Tech., Senior Researcher, orcid.org/00000002-1962-7044

M. V. Petrov<sup>a</sup>, Researcher, orcid.org/0000-0001-7403-5036

<sup>a</sup>St. Petersburg Federal Research Center of the RAS, 39, 14th Line, 199178, Saint-Petersburg, Russian Federation

**Introduction:** In the decision support domain, the practice of using information from users' digital traces has not been widespread so far. Earlier, the authors of this paper developed a conceptual framework of intelligent decision support to recommend decisions based on the user's digital life model. The present research is aiming at the development of a basic scenario that implements this framework, and the creation of an approach to grouping users with similar preferences and decision-making behaviours. **Purpose:** Predictive modeling of advisory decisions based on the information contained in the user's digital traces. **Results:** A basic scenario has been developed for intelligent decision support based on a user's digital life model. The scenario is intended to provide the user with advisory decisions based on the knowledge about his or her decision-maker type, decision support problem, and problem domain. An approach has been proposed to group users with the same kinds of preferences and decision-making behaviors, based on the information about the segments of users, behavioral segmentation rules, and actions represented in the users' digital life models. **Practical relevance:** The research results are beneficial for the development of advanced recommendation systems which follow certain standards in tracking digital traces.

**Keywords** — decision support, digital traces, user's digital life model, group pattern, decision maker ontology, decision support scenario.

**For citation:** Smirnov A. V., Levashova T. V., Petrov M. V. Scenario model of intelligent decision support based on user's digital life models. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 47–60 (In Russian). doi:10.31799/1684-8853-2021-4-47-60

References

1. Araujo T., Helberger N., Kruike-meier S., de Vreese C. H. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, 2020, vol. 35, pp. 611–623. doi:<https://doi.org/10.1007/s00146-019-00931-w>
2. Han M. L., Kwak B. Il., Kim H. K. CBR-based decision support methodology for cybercrime investigation: Focused on the data-driven website defacement analysis. *Security and Communication Networks*, 2019, Article ID 1901548. doi:10.1155/2019/1901548
3. Asniar Surendro K. Predictive analytics for predicting customer behavior. *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT)*, IEEE, 2019, pp. 230–233. doi:10.1109/ICAIIIT.2019.8834571
4. Smirnov A. V., Levashova T. V. Context-aware approach to intelligent decision support based on user digital traces. *Informatics and Automation*, 2020, vol. 19, no. 5, pp. 915–941 (In Russian). doi:10.15622/ia.2020.19.5.1
5. Lengsfeld J. *Digital Era Framework*. Bad Waldsee, Dr. Jörn Lengsfeld, 2019. 306 p.
6. Dentler K., Cornet R., ten Teije A., de Keizer N. Comparison of reasoners for large ontologies in the OWL 2 EL Profile. *Semantic Web*, 2011, vol. 2, no. 2, pp. 71–87. doi:10.3233/SW-2011-0034
7. Abburu S. A survey on ontology reasoners and comparison. *International Journal of Computer Applications*, 2012, vol. 57, no. 17, pp. 33–39. doi:10.5120/9208-3748
8. Parsia B., Matentzoglou N., Gonçalves R. S., Glimm B., Steigmiller A. The OWL Reasoner Evaluation (ORE) 2015 Competition Report. *Journal of Automated Reasoning*, 2017, vol. 59, no. 4, pp. 455–482. doi:10.1007/s10817-017-9406-8
9. Gorodetsky V. I., Tushkanova O. N. Ontology-based user profile personification in 3G recommender systems. *Ontology of Designing*, 2014, vol. 13, no. 3, pp. 7–31 (In Russian).
10. Vuong T., Jacucci G., Ruotsalo T. Watching inside the Screen. *Proceedings of ACM Interactive, Mobile, Wearable Ubiquitous Technologies*, 2017, vol. 1, no. 3, Article no. 109, pp. 1–23. doi:10.1145/3130974
11. Breiter A., Hepp A. *The Complexity of Datafication: Putting Digital Traces in Context*. In: *Communicative Figurations*. Cham, Springer International Publishing, 2018. Pp. 387–405. doi:10.1007/978-3-319-65584-0\_16
12. IJOC special section “Digital Traces in Context”. *International Journal of Communication*. Eds. A. Hepp, T. N. Friemel, A. Breiter. Los Angeles, USC Annenberg Press, 2018, vol. 12, pp. 439–705.
13. Pentland B. T., Recker J., Wolf J., Wyner G. Bringing context inside process research with digital trace data. *Journal of the Association for Information Systems*, 2020, vol. 21, no. 5, pp. 1214–1236. doi:10.17705/1jais.00635
14. Glimm B., Horrocks I., Motik B., Shearer R., Stoilos G. A novel approach to ontology classification. *Journal of Web Semantics*, 2012, vol. 14, pp. 84–101. doi:10.1016/j.websem.2011.12.007
15. Bayer J. Customer segmentation in the telecommunications industry. *Journal of Database Marketing & Customer Strategy Management*, 2010, vol. 17, no. 3–4, pp. 247–256. doi:10.1057/dbm.2010.21
16. *Specification of RuleML*. Available at: [http://wiki.ruleml.org/index.php/Specification\\_of\\_RuleML](http://wiki.ruleml.org/index.php/Specification_of_RuleML) (accessed 31 May 2021).
17. Kifer M., Boley H. *RIF Overview*. Second Ed. W3C Working Group Note, W3C, 2013. Available at: <https://www.w3.org/TR/rif-overview/> (accessed 31 May 2021).
18. Horrocks I., Patel-Schneider P. F., Boley H., Tabet S., Groszof B., Dean M. *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*. W3C Member Submission, 2004. Available at: <https://www.w3.org/Submission/SWRL/> (accessed 31 May 2021).
19. Mehla S., Jain S. *Rule Languages for the Semantic Web*. In: *Emerging Technologies in Data Mining and Information Security*. Eds. A. Abraham, P. Dutta, J. K. Mandal, A. Bhat-tacharya, S. Dutta. Springer Singapore, Advances in Intelligent Systems and Computing, 2019. Vol. 755. Pp. 825–834. doi:10.1007/978-981-13-1951-8\_73
20. *OWL 2 Web Ontology Language Document Overview*. W3C Recommendation. Ed. W3C OWL Working Group. W3C, 2012. Available at: <https://www.w3.org/TR/owl2-overview/> (accessed 31 May 2021).
21. Negru S., Lohmann S., Haag F. *VOWL: Visual Notation for OWL Ontologies*. Specification of Version 2.0. VisualData-Web.org, 2014. Available at: <http://vowl.visualdataweb.org/v2/> (accessed 31 May 2021).

## Формирование обучающей выборки в задачах машинного обучения. Обзор

А. В. Парасич<sup>а</sup>, инженер-программист, [orcid.org/0000-0003-2728-0893](https://orcid.org/0000-0003-2728-0893) [parasichav@yandex.ru](mailto:parasichav@yandex.ru)

В. А. Парасич<sup>б</sup>, канд. техн. наук, доцент, [orcid.org/0000-0003-3593-2345](https://orcid.org/0000-0003-3593-2345)

И. В. Парасич<sup>б</sup>, канд. техн. наук, доцент, [orcid.org/0000-0003-1965-8737](https://orcid.org/0000-0003-1965-8737)

<sup>а</sup>ООО «ТРИДИВИ», Ленина пр., 64, Челябинск, 454080, РФ

<sup>б</sup>Южно-Уральский государственный университет, Ленина пр., 76, Челябинск, 454080, РФ

**Введение:** правильное формирование обучающей выборки является ключевым фактором при решении задач машинного обучения. При этом в реальных обучающих выборках часто встречаются те или иные трудности и ошибки формирования выборки, оказывающие критическое влияние на результат обучения. Проблема формирования обучающей выборки возникает во всех задачах машинного обучения, поэтому знание возможных вопросов при формировании обучающей выборки будет полезно при решении любой задачи машинного обучения. **Цель:** обзор возможных проблем формирования обучающей выборки с целью облегчить их обнаружение и устранение при работе с реальными обучающими выборками. Анализ влияния этих проблем на результат обучения. **Результаты:** проведен обзор возможных ошибок формирования обучающей выборки, таких как отсутствие данных, разбалансировка, ложные внутривыборочные закономерности, формирование выборки из ограниченного набора источников, изменение генеральной совокупности во времени и др. Исследовано влияние этих ошибок на результат обучения, а также на формирование тестовой выборки и измерение качества алгоритма обучения. Pseudo-labeling, data augmentation, hard samples mining рассматриваются как наиболее эффективные способы расширения обучающей выборки. Предложены практические рекомендации по формированию обучающей и тестовой выборок. Приведены примеры из практики соревнований Kaggle. Рассмотрена проблема cross-dataset generalization. Предложен алгоритм решения проблемы cross-dataset generalization при обучении нейронных сетей, названный Cross-Dataset Machine, простой в реализации и позволяющий получить выигрыш в cross-dataset обобщении. **Практическая значимость:** материалы статьи могут использоваться в качестве практического руководства при решении задач машинного обучения.

**Ключевые слова** – машинное обучение, обучающая выборка, Kaggle, глубокие нейронные сети, деревья решений, ImageNet.

**Для цитирования:** Парасич А. В., Парасич В. А., Парасич И. В. Формирование обучающей выборки в задачах машинного обучения. Обзор. *Информационно-управляющие системы*, 2021, № 4, с. 61–70. doi:10.31799/1684-8853-2021-4-61-70

**For citation:** Parasich A. V., Parasich V. A., Parasich I. V. Training set formation in machine learning tasks. Survey. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 61–70 (In Russian). doi:10.31799/1684-8853-2021-4-61-70

### Введение

Принципиальным недостатком большинства теоретических построений из области машинного обучения и статистики [1] (в том числе рассуждений о переобучении и измерении качества работы алгоритма) является заложенное в них предположение о том, что обучающая выборка представляет идеальную независимую репрезентативную выборку из генеральной совокупности, отражающую все ее свойства. В действительности подавляющее большинство реальных обучающих выборок содержат те или иные погрешности в формировании выборки, начиная от отсутствия данных определенного вида, заканчивая ложными закономерностями внутри данной выборки. И эти проблемы формирования выборки часто оказывают решающее влияние на качество алгоритма. При этом с точки зрения успеха в решении задач машинного обучения качество данных, как правило, намного важнее качества алгоритма обучения. Именно устранение проблем формирования обучающей выборки часто явля-

ется ключом к успешному решению задачи распознавания. Рассмотрим наиболее распространенные проблемы и ошибки при формировании обучающей выборки [2, 3].

### Проблемы формирования обучающей выборки

**Отсутствие данных определенного типа.** Самая простая, самая распространенная и самая опасная проблема. Если данных определенного типа (или из определенного участка пространства признаков) в обучающей выборке  $X^L = (x_i, y_i)_{i \in 1..L}$  нет (например, в задаче распознавания лиц в выборке нет людей в очках), то алгоритм не сможет научиться правильно работать на этих данных. Как правило, очень трудно гарантировать, что в выборке есть данные всех возможных типов, поэтому с проблемой отсутствия данных приходится постоянно бороться. Следовательно, одной из главных ценностей при решении задач машинного обучения является изучение обучающей выборки.

**Слишком мало данных определенного типа.** Если данных определенного типа в обучающей выборке  $X^L = (x_i, y_i)_{i \in 1..L}$  слишком мало (например, один объект), то крайне высока вероятность, что алгоритм не сможет выучить по этим данным правильные закономерности  $y = f_k(x_i)$ , а вместо этого выучит первую попавшуюся закономерность (наподобие того, что если левый верхний пиксель изображения — красный, то это изображение собаки, иначе — изображение кошки). При малом числе данных в обучающей выборке для их классификации внутри обучающей выборки подойдет практически любое (первое попавшееся) решающее правило. Поэтому данных в выборке должно быть достаточно в той мере, чтобы ошибочные закономерности случайно не позволили правильно классифицировать эти данные внутри обучающей выборки.

**Разбалансировка.** Нарушение равномерного количественного соотношения данных разных типов. Простейший пример — разное количество обучающих примеров разных классов  $C_1..C_n$  в задаче классификации [4] (например, в задаче распознавания пола человека по изображению лица в выборке мужчин гораздо больше, чем женщин, тогда алгоритму может быть выгоднее всегда говорить, что на изображении мужчина). Более сложный пример — разбалансировка по некоторой оси разнообразия данных (в задаче распознавания пола человека по изображению лица в выборке гораздо больше дневных снимков, чем ночных, из-за чего может пострадать качество распознавания ночных снимков).

Разбалансировка возникает очень часто, поскольку трудно гарантировать соблюдение равномерных пропорций для всех возможных осей вариативности данных. При этом может очень негативно сказаться на качестве обучения, приводя к необоснованному с семантической точки зрения завышению влияния одних примеров и занижению влияния или полному игнорированию других примеров. Так происходит потому, что в алгоритмах обучения в процессе обучения оптимизируются метрики качества, представляющие собой простую сумму ошибки по всем обучающим примерам, и эта сумма, понятное дело, будет зависеть от соотношения количеств обучающих примеров разных типов. И те примеры, которых в выборке больше, будут больше влиять на эту ошибку, а значит, и на результат обучения. В предельном случае при резкой разбалансировке слабо представленный класс (тип примера) будет полностью проигнорирован и никогда не будет распознаваться (этот класс никогда не будет возвращаться алгоритмом в качестве ответа). Иногда для этого достаточно 10-кратной разбалансировки.

Проблема разбалансировки решается простым перевзвешиванием обучающих примеров

(повышением веса слабо представленным примерам либо понижением веса сильно представленным примерам). Если интерфейс алгоритма обучения не поддерживает веса примеров, то недостаточно представленные примеры можно просто продублировать нужное число раз. Если доступно добавление новых данных, то недостающих данных можно добавить. Сложность заключается в том, что разбалансировку не всегда удается обнаружить. На практике встречаются весьма сложные и нестандартные виды разбалансировок, которые оказывают большое влияние на качество алгоритмов.

Если число примеров разных классов сильно несбалансировано, и при этом известно, что при реальном использовании модели распределение числа примеров разных классов будет таким же несбалансированным, то возникает дилемма. Если сбалансировать выборку, то у алгоритма будет неверное представление о вероятности появления примеров разных классов, из-за чего распределение ответов алгоритма будет смещено относительно истинного распределения классов, что может привести к существенному увеличению числа ошибок. Если же не сбалансировать выборку, то алгоритм может просто проигнорировать слабо представленные классы и не научиться их распознавать. Для обучения глубоких нейронных сетей [5] существует следующий прием: на первых эпохах обучения баланс классов выравнивают, на последних эпохах обучения возвращают тот баланс, который был изначально. Это позволяет улучшить сходимость сети (за счет того, что слабо представленные классы не игнорируются), и в то же время у сети не возникает искаженного представления о вероятности появления примеров разных классов. Также при обучении нейронных сетей для решения проблемы разбалансировки можно использовать *focal loss* [6]. Классическими методами борьбы с разбалансировкой являются *SMOTE* [7] и *ADASYN* [8]. Проблема разбалансировки подробно исследуется в работах [9–13].

**Ложные внутривыборочные закономерности.** Ложные зависимости, которые существуют только внутри данной обучающей выборки  $X^L = (x_i, y_i)_{i \in 1..L}$  и являются следствием неправильного формирования выборки, из-за чего алгоритм может выучить эти ложные зависимости вместо реальных зависимостей  $y = f_k(x_i)$  (если эти ложные зависимости будет проще выучить, чем реальные зависимости) и будет некорректно работать за пределами обучающей выборки. Например, в задаче распознавания пола человека по изображению лица в выборке все изображения мужчин сняты в дневное время, а все изображения женщин — в ночное. Тогда алгоритм может научиться определять пол человека исключи-

тельно на основе уровня освещенности изображения и не будет работать в реальных условиях.

Ложные внутривыборочные закономерности могут проявляться в виде ложной внутривыборочной разбалансировки, т. е. жесткой зависимости (все изображения мужчин сняты только в дневное время, а все изображения женщин — только в ночное) внутри обучающей выборки нет, но есть сильный количественный перекос в сторону такой зависимости, из-за чего алгоритм выучит данную ложную закономерность.

**Ложные внутривыборочные закономерности второго порядка.** Внешняя зависимость, которая не является ложной и реально существует в данных, но не является прямым отражением природы данных, из-за чего качество работы алгоритма в нестандартных условиях может пострадать при заучивании такой зависимости. Например, в задаче детекции пешеходов [14] в абсолютном большинстве случаев пешеход идет по земле. Эта закономерность не является ложной и реально существует в данных. Но если алгоритм выучит, что пешеход всегда идет по земле, то он не будет детектировать прыгающих пешеходов. Поэтому лучше учить те закономерности, которые являются прямым следствием природы данных.

**Формирование выборки из ограниченного набора источников («микроисточников»).** Очень часто обучающие выборки формируются не путем независимого выбора объектов из генеральной совокупности, а набираются из ограниченного набора источников данных  $S_1 \dots S_n$ , что создает определенные риски при обучении. Алгоритм может выучить конкретные источники данных вместо истинной зависимости  $y = f_k(x_i)$ . Например, в задаче распознавания спама, если обучающая выборка писем была собрана из ограниченного набора спамерских и «чистых» ящиков и среди признаков есть признак «имя почтового ящика», то алгоритм может выучить конкретные имена спамерских почтовых ящиков и будет некорректно работать для любых имен спамерских ящиков, не представленных в обучающей выборке.

Поэтому в обучении нельзя использовать признаки, все значения которых принципиально не могут быть представлены в обучающей выборке (фамилии, города, названия почтовых ящиков и т. д.).

Формирование выборки из ограниченного набора источников также может стать причиной появления ложных внутривыборочных закономерностей (например, в задаче распознавания пола человека по изображению лица мужчин снимали хорошей камерой, а женщин — плохой, тогда вместо реальных зависимостей алгоритм может научиться определять пол по качеству изображения). Поэтому при таком способе формирования

выборки нужно быть предельно внимательным к возможным негативным последствиям.

**Разное распределение значений признаков.** В более общем виде не все возможные значения признаков представлены в обучающей выборке. Например, в задаче предсказания поведения покупателя в магазине в обучающей выборке только люди молодого возраста, а в тестовой выборке — только люди пожилого возраста. Частный случай проблемы отсутствия данных определенного вида.

**Недозаполненность признакового пространства.** Пространство признаков  $f_1, \dots, f_p$  порождает некоторое разбиение обучающей выборки  $X^L = (x_i, y_i)_{i \in 1..L}$  на группы примеров  $C_1 \dots C_n$  в соответствии с тем, какие значения признаков  $f_k(x_i)$  у данных примеров. При этом в зависимости от того, какие признаки используются, в некоторые части множества возможных значений признаков может не попасть ни один обучающий пример либо попасть слишком мало обучающих примеров.

В качестве примера можно привести использование гистограммы ориентированных градиентов [15] в задаче детекции объекта. При использовании слишком большого числа угловых ячеек (например, 100 ячеек) может получиться так, что некоторым ячейкам не будет соответствовать ни один объект из обучающей выборки, поэтому алгоритм не сможет научиться детектировать объекты с такой ориентацией. При использовании меньшего числа ячеек (например, 50 ячеек) подобной проблемы может не возникнуть.

Чем более сложные признаки используются, тем выше вероятность того, что некоторые части признакового пространства останутся недозаполненными примерами из обучающей выборки. Поэтому при переходе к более сложным признакам может потребоваться расширение обучающей выборки, чтобы добавление более сложных признаков приводило к росту качества, а не к переобучению [16].

**Сбор одинаковых данных при формировании выборки.** Часто встречающаяся на практике ошибка. При формировании обучающей выборки собираются слишком одинаковые данные либо варьируются не все степени свободы данных (например, собраны данные не для всех возможных диапазонов расстояний от объекта до камеры), что приводит к неработоспособности алгоритма в тех условиях, данные для которых не были собраны.

**Сбор данных не в тех условиях, в которых будет использоваться система.** Другая важная с практической точки зрения ошибка. Сбор данных не в тех условиях, в которых будет использоваться система, может, с одной стороны, привести к появлению в данных тех степеней свободы,

которых не будет при реальном использовании системы (что затруднит обучение), а, с другой стороны, может привести к отсутствию в выборке некоторых разновидностей данных, на которых система должна будет работать, из-за чего алгоритм не сможет обучиться правильно работать на этих данных. Подобная ошибка опасна еще и тем, что может увести всю разработку системы в неверном направлении (разработка не будет сконцентрирована на решении тех проблем, которые наиболее важны при реальном использовании системы, и стратегические решения о выборе направления развития системы могут быть приняты неверно). Поэтому с самого начала разработки любой системы распознавания рекомендуется собирать данные в условиях, максимально приближенных к тем, в которых эта система будет использоваться.

**Неправильные значения целевой переменной.** Предельный случай проблем формирования выборки, тем не менее встречающийся в реальных выборках. Может возникать в результате ошибок разметки.

**Смещение малых подвыборок.** Даже если обучающая выборка не содержит в себе ни одной из вышеперечисленных проблем, вроде отсутствия данных или присутствия ложных закономерностей, это не значит, что таких проблем нет в ее малых подвыборках. Малые подвыборки образуются при обучении нижних уровней деревьев решений (так как при обучении деревьев решений при обучении каждой следующей вершины ее множество данных делится между ее левым и правым сыном для дальнейшего обучения). При обучении нейронных сетей малой подвыборкой можно считать мини-батч, либо множество обучающих примеров с высокой ошибкой в конце обучения, сильнее всего влияющих на обучение сети на поздних этапах, либо множество примеров, вызывающих активацию определенного нейрона, либо множество примеров, соответствующих определенному внутреннему состоянию сети. При этом наличие рассмотренных проблем в малой подвыборке гораздо более вероятно, чем во всей выборке, следовательно, высок риск неправильного обучения. Это может приводить к трудно диагностируемому, но опасным формам переобучения.

**Изменение генеральной совокупности во времени.** Даже если у нас есть идеально правильная обучающая выборка  $X^L = (x_i, y_i)_{i \in 1..L}$ , не содержащая никаких ошибок и проблем формирования выборки (допустим, мы решаем задачу машинного обучения для крупной интернет-компании, используя для обучения датасет из данных о нескольких миллионах пользователей, где проблемы отсутствия данных какого-либо вида не может быть в принципе), это не гарантирует нам

полное отсутствие проблем, связанных с формированием выборки. Генеральная совокупность может меняться во времени (если алгоритм работает в онлайн-режиме). Могут появляться новые разновидности данных (новые источники данных), происходить сезонные колебания или локальные всплески определенного рода активности. В данном случае может помочь правильная схема валидации. Например, рекомендуется делать валидацию по времени (данные из последнего месяца — в тестовую выборку, остальные — в обучающую).

### Проблемы формирования тестовой выборки

Те же самые проблемы (отсутствие данных, недостаточное количество данных, разбалансировка, ложные зависимости) могут присутствовать и в тестовой выборке, по которой измеряется качество работы алгоритма, из-за чего оценка качества работы алгоритма может оказаться некорректной.

Стандартная процедура тестирования (разбиение множества данных  $D$  на обучающую  $T$  и тестовую  $V$  выборки случайным образом) и кросс-валидация [17] не позволяют обнаружить проблемы формирования выборки.

Если в исходной выборке  $D$  отсутствуют данные некоторого вида, то этих данных не будет ни в обучающей, ни в тестовой выборке, поэтому качество работы на этих данных измерено не будет, и отсутствие данных никак не проявится при тестировании.

Если в исходной выборке  $D$  присутствует ложная внутривыборочная закономерность, то она будет существовать и в обучающей, и в тестовой выборках, порожденных из выборки  $D$ . И если алгоритм выучит эту ложную внутривыборочную закономерность, то на тестовой выборке он сможет отработать корректно, т. е. проблема никак не проявится. Если в выборке нет примеров определенного типа, то и качество работы на них измерено не будет, и эта проблема тоже никак не проявится при тестировании.

Разбалансировка также оказывает влияние на результат тестирования алгоритма. При разной сложности распознавания данных разных типов может получиться так, что качество работы алгоритма на тестовой выборке будет в первую очередь зависеть от баланса количества сложных данных в выборке по отношению к простым данным, нежели от качества алгоритма.

Если в тестовой выборке присутствуют неправильные значения целевой переменной, то качество алгоритма будет просто неправильно измерено, независимо от предположений о пра-

вильности / неправильности закономерностей. Данный пример наглядно показывает, что не стоит слепо полагаться на результаты тестирования, не анализируя суть происходящего и возникающие ошибки.

Например, если выборка составлена из данных из заранее известного набора источников  $D_1, D_2, \dots, D_n$  и известно, что в ходе использования системы могут появляться новые источники  $D_i$ , то для проверки обобщающей способности алгоритма выборку следует разбивать на обучающую и тестовую по источникам (например, данные из источников  $D_1, D_2, \dots, D_k$  — в обучение, а данные из источников  $D_{k+1}, D_{k+2}, \dots, D_n$  — в тест).

### Эффективные способы расширения обучающей выборки

Добавление данных в обучающую выборку является, как правило, самым эффективным способом повышения качества обучения. При этом не стоит забывать о том, что лучше всего добавлять именно те данные, которых не хватает в обучающей выборке и с распознаванием которых алгоритм испытывает проблемы. И не создать дополнительную разбалансировку либо ложные внутривыборочные закономерности. Рассмотрим наиболее распространенные способы добавления данных в обучающую выборку.

**Pseudo-labeling. Noisy Student.** В последнее время в соревновательной практике широко используется *pseudo-labeling* [18] (метод *self-supervised learning* [19], когда алгоритм сначала обучается на коллекции размеченных данных, затем ответы обученного алгоритма используются для разметки набора неразмеченных данных, и далее полученные данные используются для дообучения алгоритма, при этом используются только те данные, в ответах на которые алгоритм уверен). Разновидности *pseudo-labelling* — алгоритмы *Noisy Student* [20] и *Meta Pseudo Labels* [21] — показали одни из лучших результатов на *ImageNet* [22].

**Data augmentation.** Важный этап обучения нейронных сетей, состоящий в модификации обучающих изображений (поворот, масштабирование, зеркальное отражение и т. д.) по определенному правилу с целью расширить обучающую выборку и повысить ее разнообразие. Рассмотрим наиболее эффективные виды аугментаций.

**Аугментации цвета и контраста.** Случайное изменение компонент  $R, G, B$  цвета пикселей изображения. Один из самых эффективных методов аугментации данных, потому что нейросети без этой аугментации имеют тенденцию к заучиванию фич вида «сумма цветов пикселей в области».

**Аугментации масштаба, random cropping.** Практически всегда приводят к улучшению качества. В работе [23] показано повышение качества детекции объектов на 10 % на датасете *COCO* [24] благодаря добавлению аугментаций масштаба. Эффективность данного вида аугментаций объясняется тем, что сверточные нейронные сети по своей природе не инвариантны к масштабу, а изменение масштаба изображения значительно повышает разнообразие данных с точки зрения нейросети.

Еще одним эффективным видом аугментации являются *CutOut* [25] и *Random Erasing* [26] (закрашивание случайных прямоугольников на картинке, чтобы нейросеть не могла научиться распознавать объект по одной конкретной детали внешнего вида, например, распознавать машину по колесу).

Одним из эффективных приемов обучения нейросетей является постепенное уменьшение интенсивности аугментаций по ходу обучения. Это позволяет нейросети лучше адаптироваться к исходному распределению, в то же время улучшает сходимость и устойчивость сети за счет повышения разнообразия данных. Также существуют методы автоматического подбора наиболее эффективных аугментаций под заданную выборку [27]. Эффективность различных схем аугментации исследуется в работах [28–30].

**Hard Samples Mining.** Классическая проблема при обучении детектора объектов — сбор в обучающую выборку *hard negative examples* [31, 32] (фрагментов изображений, внешне похожих на детектируемый объект, но не являющихся детектируемым объектом). Такие объекты нужны, для того чтобы научиться отличать объект от похожих объектов фона. Сложность в том, что в естественных условиях такие объекты встречаются редко, поэтому для эффективного обучения требуются специальные методы майнинга таких объектов. Одним из эффективных приемов майнинга *hard negative examples* из соревновательной практики *Kaggle* [33] является использование в качестве *hard negative examples* изображений, вызвавших ложноположительные срабатывания недоученной версии детектора (детектора, полученного после небольшого числа эпох обучения). Данный прием использовался в решении, занявшем первое место в соревновании [34].

**Generative Adversarial Networks (GAN).** *Generative Adversarial Networks* [35] могут использоваться для генерации изображений [36] либо для стилизации (*Style Transfer* [37]) изображений под новые условия. Современные GAN часто генерируют некорректные изображения, также существуют большие проблемы с их сходимостью, поэтому GAN не всегда подходят для генерации самих распознаваемых объектов, од-

нако их вполне можно использовать для генерации фона [38] либо для адаптации имеющихся изображений к другим условиям (например, для переделывания дневных изображений в ночные [39–41]).

**Имитация добавления данных.** Широко известный метод регуляризации обучения нейронных сетей *Dropout* [42] можно рассматривать как имитацию добавления данных. С точки зрения  $i$ -го слоя нейронной сети нет разницы между тем, изменяются ли входные данные или изменяются значения выходов  $i - 1$ -го слоя нейронной сети. Другой реализацией данного принципа является *Shake-Shake regularization* [43] — метод, некоторое время являвшийся *State-of-the-Art* результатом на *CIFAR-10* [44] — датасете, содержащем малое число изображений (около 10 000), из-за чего имитация добавления данных для этого датасета представляется целесообразной. Развитием идеи *Shake-Shake regularization* является *Shake-Drop regularization* [45].

### Проблема cross-dataset generalization. Алгоритм Cross-Dataset Machine

Одной из проблем, связанных с формированием выборки, является проблема обобщения алгоритма на данные тех же классов из другого датасета — *cross-dataset generalization* [46]. То есть в качестве тестовой выборки используется не выборка из того же датасета, на котором шло обучение, а выборка из другого датасета (допустим, в задаче распознавания машины обучение производилось на ImageNet [22], а валидация — на PASCAL VOC [47]). Cross-dataset обобщающая способность обычно существенно хуже простой обобщающей способности из-за того, что датасеты могут сильно отличаться друг от друга, а в процессе обучения не было обеспечено обобщение на данные, сильно не похожие на обучающие. Cross-dataset тестирование алгоритмов является более надежным в плане определения качества работы алгоритма, так как позволяет исключить влияние на результат тестирования части проблем формирования выборки, таких как ложные внутривыборочные закономерности.

Рассмотрим один из возможных алгоритмов повышения cross-dataset generalization при обучении нейронных сетей (назовем этот алгоритм *Cross-Dataset Machine*). Допустим, есть три датасета (для одной и той же задачи)  $D_1, D_2, D_3$ , обучение производится на датасетах  $D_1$  и  $D_2$ , тестирование — на датасете  $D_3$ . Будем при обучении в четных батчах подавать на вход сети только объекты из датасета  $D_1$ , в нечетных — только объекты из датасета  $D_2$ . Если для обучения используется  $n$  датасетов  $D_1, D_2, \dots, D_n$ , то разобьем

эти датасеты на два непересекающихся множества  $M_1$  и  $M_2$  ( $\{D_1 \dots D_m\} \in M_1, \{D_{m+1} \dots D_n\} \in M_2$ ) и будем при обучении в четных батчах подавать на вход сети только объекты из множества  $M_1$ , в нечетных — только объекты из множества  $M_2$ . В качестве датасетов могут выступать не датасеты целиком, а отдельные источники данных, из которых был собран датасет (такая схема чаще всего и используется на практике). Идея алгоритма в том, что при обучении сети на батче из датасета  $D_1$  сеть выучивает два типа фич: специфичные для датасета  $D_1$  и не обобщаемые на другие датасеты (назовем такие фичи  $F_p$ ) и обобщаемые между датасетами ( $F_g$ ). При подаче на вход сети батча из датасета  $D_2$  необобщаемые фичи  $F_p$  будут разрушаться (не будут выживать), а обобщаемые фичи  $F_g$  будут выживать и развиваться дальше. Таким образом, после нескольких итераций алгоритма в сети должны остаться и обучиться только хорошо обобщаемые между датасетами фичи.

Проверим алгоритм на задаче соревнования *State Farm Distracted Driver Detection* [48], проводившегося на платформе Kaggle. По изображению водителя изнутри машины надо определить факт отвлечения водителя (курит, ест, разговаривает по телефону). Датасет соревнования составлен из групп изображений, где каждая группа — это изображения одного и того же водителя в одной и той же машине (всего в датасете 26 водителей). Поэтому в качестве датасетов (источников данных)  $D_1, D_2, \dots, D_n$  будем использовать группы изображений одного и того же водителя (все изображения одного водителя  $i$  — это отдельный источник данных  $D_i$ ). Разделим их на два непересекающихся множества (по 11 водителей в каждом), пять водителей оставим для валидации. В таблице приводятся результаты сравнения работы данного алгоритма и стандартного алгоритма обучения нейросети. Мы видим, что Cross-Dataset Machine показывает ошибку на валидации меньшую, чем у стандартного алгоритма.

Возможной модификацией алгоритма Cross-Dataset Machine является 3-стадийный Cross-Dataset Machine. В этом алгоритме одна итерация обучения состоит из трех стадий: на первой стадии на вход сети подаются обучающие примеры из подмножества датасетов  $M_1$ , на второй стадии — из подмножества датасетов  $M_2$ , на третьей стадии — обучающие примеры из всей обучающей выборки вперемешку. Полный цикл обучения сети состоит из  $k$  таких итераций. Это позволяет сети лучше адаптироваться ко всей генеральной совокупности, в то же время сохраняются преимущества исходного алгоритма, заключающиеся в выживании хорошо обобщаемых между датасетами фич и невыживании плохо обобщаемых фич. Как мы видим из таблицы, данный алгоритм позволяет получить вы-

- Результаты работы алгоритма Cross-Dataset Machine на задаче State Farm Distracted Driver Detection
- Cross-Dataset Machine results on State Farm Distracted Driver Detection task

Ошибка	Стандартный алгоритм	Cross-Dataset Machine	3-стадийный Cross-Dataset Machine
Минимальная	1,78	1,70	1,66
В конце обучения	1,85	1,78	1,71
Средняя	2,05	2,04	2,01

игрыш относительно стандартного Cross-Dataset Machine.

Проверим алгоритм на задаче соревнования *HuVMAP — Hacking the Kidney* [49]. В этой задаче требуется отсегментировать клетки функциональных единиц ткани. Всего дано семь больших цельных изображений функциональных единиц ткани, обучение и распознавание происходит по маленьким фрагментам больших изображений (всего 11 473 фрагмента). Поэтому в качестве да-

тасетов (источников данных)  $D_1, D_2, \dots, D_n$  будем использовать цельные изображения (одно цельное изображение  $i$  — это отдельный источник данных  $D_i$ ). В качестве backbone использовался EfficientNet-B4 [50]. В результате доля корректно отсегментированных пикселей на задаче HuVMAP — Hacking the Kidney следующая:

- стандартный алгоритм — 0,848;
- Cross-Dataset Machine — 0,855.

И в этой задаче Cross-Dataset Machine позволяет получить выигрыш в качестве.

## Заключение

В статье сделан обзор возможных проблем формирования обучающей выборки, проведен анализ их влияния на результат обучения, даны рекомендации по их устранению. Рассмотрены наиболее эффективные способы расширения обучающей выборки. Исследована проблема cross-dataset generalization. Предложен алгоритм Cross-Dataset Machine, позволяющий получить выигрыш в cross-dataset generalization.

## Литература

1. Vapnik V. N., Chervonenkis A. Y. On the uniform convergence of relative frequencies of events to their probabilities. In: *Measures of complexity*. Springer, Cham, 2015. Pp. 11–30.
2. Gonzalez-Diaz R., Gutiérrez-Naranjo A., Paluzo-Hidalgo E. Representative datasets: the perceptron case. *arXiv preprint arXiv:1903.08519*, 2019.
3. Roh Y., Heo G., Whang S. E. A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 2019. <https://doi.org/10.1109/TKDE.2019.2946162>
4. Kotsiantis S. B., Zaharakis I., Pintelas P. Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 2007, vol. 160, no. 1, pp. 3–24.
5. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, vol. 60, no. 6, pp. 84–90.
6. Lin T. Y., Goyal P., Girshick R., He K., Dollár P. Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.
7. Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 2002, vol. 16, pp. 321–357.
8. He H., Bai Y., Garcia E. A., Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1322–1328.
9. Longadge R., Dongre S. Class imbalance problem in data mining review. *arXiv preprint arXiv:1305.1707*, 2013.
10. Japkowicz N., Stephen S. The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 2002, vol. 6, no. 5, pp. 429–449.
11. Buda M., Maki A., Mazurowski M. A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 2018, vol. 106, pp. 249–259.
12. Johnson J. M., Khoshgoftaar T. M. Survey on deep learning with class imbalance. *Journal of Big Data*, 2019, vol. 6, no. 1, pp. 1–54.
13. Liu X. Y., Wu J., Zhou Z. H. Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2008, vol. 39, no. 2, pp. 539–550.
14. Enzweiler M., Gavrilu D. M. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, vol. 31, no. 12, pp. 2179–2195.
15. Dalal N., Triggs B. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 886–893.

16. Ying X. An overview of overfitting and its solutions. *Journal of Physics: Conference Series, IOP Publishing*, 2019, vol. 1168, no. 2, pp. 022022.
17. Browne M. W. Cross-validation methods. *Journal of Mathematical Psychology*, 2000, vol. 44, no. 1, pp. 108–132.
18. Lee D. H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. *Workshop on Challenges in Representation Learning, ICML*, 2013, vol. 3, no. 2, pp. 896.
19. Hendrycks D., Mazeika M., Kadavath S., Song D. Using self-supervised learning can improve model robustness and uncertainty. *arXiv preprint arXiv:1906.12340*, 2019.
20. Xie Q. Self-training with noisy student improves imagenet classification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10687–10698.
21. Pham H., Dai Z., Xie Q., Luong M. T., Le Q. V. Meta pseudo labels. *arXiv preprint arXiv:2003.10580*, 2020.
22. ImageNet Dataset. <http://www.image-net.org/> (дата обращения: 05.01.2021).
23. Cai Z., Vasconcelos N. Cascade R-CNN: high quality object detection and instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. <https://doi.org/10.1109/TPAMI.2019.2956516>
24. Common Objects in Context (COCO) Dataset. <https://cocodataset.org/#home>. (дата обращения: 05.01.2021).
25. DeVries T., Taylor G. W. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.
26. Zhong Z., Zheng L., Kang G., Li S., Yang Y. Random erasing data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 07, pp. 13001–13008.
27. Cubuk E. D., Zoph B., Mane D., Vasudevan V., Le Q. V. Autoaugment: Learning augmentation strategies from data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 113–123.
28. Perez L., Wang J. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.
29. Shorten C., Khoshgoftaar T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019, vol. 6, no. 1, pp. 1–48.
30. O’Gara S., McGuinness K. Comparing data augmentation strategies for deep image classification. *Irish Machine Vision and Image Processing Conference (IMVIP)*, 2019. <https://doi.org/10.21427/148B-AR75>
31. Canavet O., Fleuret F. Efficient sample mining for object detection. *Proceedings of the Asian Conference on Machine Learning (ACML)*, 2014, pp. 48–63.
32. Jin S. Y., RoyChowdhury A., Jiang H., Singh A., Prasad A., Chakraborty D., Learned-Miller E. Unsupervised hard example mining from videos for improved object detection. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 307–324.
33. Kaggle: You Machine Learning and Data Science Community. <https://www.kaggle.com/> (дата обращения: 05.06.2021).
34. NFL 1st and Future — Impact Detection. <https://www.kaggle.com/c/nfl-impact-detection/discussion/209403> (дата обращения: 05.03.2021).
35. Creswell A., White T., Dumoulin V., Arulkumaran K., Sengupta B., Bharath A. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 2018, vol. 35, no. 1, pp. 53–65.
36. Antoniou A., Storkey A., Edwards H. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017.
37. Xu Z., Wilber M., Fang C., Hertzmann A., Jin H. Learning from multi-domain artistic images for arbitrary style transfer. *arXiv preprint arXiv:1805.09987*, 2018.
38. Ma Y., Liu K., Guan Z., Xu X., Qian X., Bao H. Background augmentation generative adversarial networks (BAGANs): Effective data generation based on GAN-augmented 3D synthesizing. *Symmetry*, 2018, vol. 10, no. 12, pp. 734.
39. Meng Y., Kong D., Zhu Z., Zhao Y. From night to day: GANs based low quality image enhancement. *Neural Processing Letters*, 2019, vol. 50, no. 1, pp. 799–814.
40. Ardiyanto I., Soesanti I., Qairawan D. C. Night-to-day road scene translation using generative adversarial network with structural similarity loss for night driving safety. In: *Deep Learning and Big Data for Intelligent Transportation: Enabling Technologies and Future Trends*, 2021. Pp. 119–133.
41. Xie H., Xiao J., Lei J., Xie W., Klette R. Image Scene Conversion Algorithm Based on Generative Adversarial Networks. In: *Asian Conference on Pattern Recognition*. Springer, Singapore, 2019. Pp. 29–36.
42. Srivastava N., Hinton G. E., Krizhevsky A., Sutskever I., Salakhutdinov R. R. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 2014, vol. 15, no. 1, pp. 1929–1958.
43. Gastaldi X. Shake-shake regularization. *arXiv preprint arXiv:1705.07485*, 2017.
44. CIFAR-10 Dataset. <https://www.cs.toronto.edu/~kriz/cifar.html> (дата обращения: 05.02.2021).
45. Yamada Y., Iwamura M., Akiba T., Kise K. Shakedrop regularization for deep residual learning. *IEEE Access*, 2019, vol. 7, pp. 186126–186136.
46. Torralba A., Efros A. A. Unbiased look at dataset bias. *CVPR*, 2011, pp. 1521–1528.
47. PASCAL VOC Dataset. <http://host.robots.ox.ac.uk/pascal/VOC/> (дата обращения: 05.01.2021).
48. State Farm Distracted Driver Detection. <https://www.kaggle.com/c/state-farm-distracted-driver-detection> (дата обращения: 05.06.2021).

49. HuBMAP — Hacking the Kidney. <https://www.kaggle.com/c/hubmap-kidney-segmentation> (дата обращения: 05.06.2021).

50. Tan M., Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning, PMLR*, 2019, pp. 6105–6114.

UDC 004.855.5

doi:10.31799/1684-8853-2021-4-61-70

### Training set formation in machine learning tasks. Survey

A. V. Parasich<sup>a</sup>, Programmer Engineer, [orcid.org/0000-0003-2728-0893](https://orcid.org/0000-0003-2728-0893), [parasichav@yandex.ru](mailto:parasichav@yandex.ru)

V. A. Parasich<sup>b</sup>, PhD, Tech., Associate Professor, [orcid.org/0000-0003-3593-2345](https://orcid.org/0000-0003-3593-2345)

I. V. Parasich<sup>b</sup>, PhD, Tech., Associate Professor, [orcid.org/0000-0003-1965-8737](https://orcid.org/0000-0003-1965-8737)

<sup>a</sup>3DiVi Inc, 64, Lenina Pr., 454080, Chelyabinsk, Russian Federation

<sup>b</sup>South Ural State University, 76, Lenina Pr., 454080, Chelyabinsk, Russian Federation

**Introduction:** Proper training set formation is the key factor in solving machine learning tasks. At the same time, in real training sets, there are often some problems and errors that have a critical impact on the training result. The training set formation problem arises in all machine learning problems; therefore, knowledge of the possible problems of forming a training set will be useful when solving any machine learning problem. **Purpose:** Make an overview of possible problems in the formation of a training set, in order to facilitate their detection and elimination when working with real training sets. Analyze the impact of these problems on learning. **Results:** The article makes an overview of possible errors in the formation of a training set, such as lack of data, imbalance, false patterns, sampling from a limited set of sources, change in the general population over time, and others. The influence of these errors on the learning result is considered. The influence of the same problems on the formation of a test set and measurement of the quality of the learning algorithm is considered. The pseudo-labeling, data augmentation, hard samples mining are considered as the most effective ways to expand the training set. Practical recommendations for the formation of training and test set are offered. Practical recommendations for the formation of training and test set are offered. Examples from the practice of Kaggle competitions are given. The problem of cross-dataset generalization is considered. An algorithm for solving the problem of cross-dataset generalization in training neural networks, called the Cross-Dataset Machine, is proposed, which is very simple to implement and allows you to get a gain in cross-dataset generalization. **Practical relevance:** The materials of the article can be used as a practical guide in solving machine learning problems.

**Keywords** — machine learning, training set, Kaggle, deep neural networks, decision trees, ImageNet.

**For citation:** Parasich A. V., Parasich V. A., Parasich I. V. Training set formation in machine learning tasks. Survey. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 61–70 (In Russian). doi:10.31799/1684-8853-2021-4-61-70

### References

- Vapnik V. N., Chervonenkis A. Y. *On the uniform convergence of relative frequencies of events to their probabilities*. In: *Measures of complexity*. Springer, Cham, 2015. Pp. 11–30.
- Gonzalez-Diaz R., Gutiérrez-Naranjo A., Paluzo-Hidalgo E. Representative datasets: the perceptron case. *arXiv preprint arXiv:1903.08519*, 2019.
- Roh Y., Heo G., Whang S. E. A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 2019. <https://doi.org/10.1109/TKDE.2019.2946162>
- Kotsiantis S. B., Zaharakis I., Pintelas P. Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 2007, vol. 160, no. 1, pp. 3–24.
- Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, vol. 60, no. 6, pp. 84–90.
- Lin T. Y., Goyal P., Girshick R., He K., Dollár P. Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.
- Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 2002, vol. 16, pp. 321–357.
- He H., Bai Y., Garcia E. A., Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1322–1328.
- Longadge R., Dongre S. Class imbalance problem in data mining review. *arXiv preprint arXiv:1305.1707*, 2013.
- Japkowicz N., Stephen S. The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 2002, vol. 6, no. 5, pp. 429–449.
- Buda M., Maki A., Mazurowski M. A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 2018, vol. 106, pp. 249–259.
- Johnson J. M., Khoshgoftaar T. M. Survey on deep learning with class imbalance. *Journal of Big Data*, 2019, vol. 6, no. 1, pp. 1–54.
- Liu X. Y., Wu J., Zhou Z. H. Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2008, vol. 39, no. 2, pp. 539–550.
- Enzweiler M., Gavrilu D. M. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, vol. 31, no. 12, pp. 2179–2195.
- Dalal N., Triggs B. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 886–893.
- Ying X. An overview of overfitting and its solutions. *Journal of Physics: Conference Series, IOP Publishing*, 2019, vol. 1168, no. 2, pp. 022022.
- Browne M. W. Cross-validation methods. *Journal of Mathematical Psychology*, 2000, vol. 44, no. 1, pp. 108–132.
- Lee D. H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. *Workshop on Challenges in Representation Learning, ICML*, 2013, vol. 3, no. 2, pp. 896.
- Hendrycks D., Mazeika M., Kadavath S., Song D. Using self-supervised learning can improve model robustness and uncertainty. *arXiv preprint arXiv:1906.12340*, 2019.
- Xie Q. Self-training with noisy student improves imagenet classification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10687–10698.

21. Pham H., Dai Z., Xie Q., Luong M. T., Le Q. V. Meta pseudo labels. *arXiv preprint arXiv:2003.10530*, 2020.
22. *ImageNet Dataset*. Available at: <http://www.image-net.org/> (accessed 5 January 2021).
23. Cai Z., Vasconcelos N. Cascade R-CNN: high quality object detection and instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. <https://doi.org/10.1109/TPAMI.2019.2956516>
24. *Common Objects in Context (COCO) Dataset*. Available at: <https://cocodataset.org/#home> (accessed 5 January 2021).
25. DeVries T., Taylor G. W. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.
26. Zhong Z., Zheng L., Kang G., Li S., Yang Y. Random erasing data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 07, pp. 13001–13008.
27. Cubuk E. D., Zoph B., Mane D., Vasudevan V., Le Q. V. Autoaugment: Learning augmentation strategies from data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 113–123.
28. Perez L., Wang J. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.
29. Shorten C., Khoshgoftaar T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019, vol. 6, no. 1, pp. 1–48.
30. O’Gara S., McGuinness K. Comparing data augmentation strategies for deep image classification. *Irish Machine Vision and Image Processing Conference (IMVIP)*, 2019. <https://doi.org/10.21427/148B-AR75>
31. Canavet O., Fleuret F. Efficient sample mining for object detection. *Proceedings of the Asian Conference on Machine Learning (ACML)*, 2014, pp. 48–63.
32. Jin S. Y., RoyChowdhury A., Jiang H., Singh A., Prasad A., Chakraborty D., Learned-Miller E. Unsupervised hard example mining from videos for improved object detection. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 307–324.
33. *Kaggle: You Machine Learning and Data Science Community*. Available at: <https://www.kaggle.com/> (accessed 5 June 2021).
34. *NFL 1st and Future — Impact Detection*. Available at: <https://www.kaggle.com/c/nfl-impact-detection/discussion/209403> (accessed 5 March 2021).
35. Creswell A., White T., Dumoulin V., Arulkumaran K., Sengupta B., Bharath A. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 2018, vol. 35, no. 1, pp. 53–65.
36. Antoniou A., Storkey A., Edwards H. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017.
37. Xu Z., Wilber M., Fang C., Hertzmann A., Jin H. Learning from multi-domain artistic images for arbitrary style transfer. *arXiv preprint arXiv:1805.09987*, 2018.
38. Ma Y., Liu K., Guan Z., Xu X., Qian X., Bao H. Background augmentation generative adversarial networks (BAGANs): Effective data generation based on GAN-augmented 3D synthesizing. *Symmetry*, 2018, vol. 10, no. 12, pp. 734.
39. Meng Y., Kong D., Zhu Z., Zhao Y. From night to day: GANs based low quality image enhancement. *Neural Processing Letters*, 2019, vol. 50, no. 1, pp. 799–814.
40. Ardiyanto I., Soesanti I., Qairawan D. C. *Night-to-day road scene translation using generative adversarial network with structural similarity loss for night driving safety*. In: *Deep Learning and Big Data for Intelligent Transportation: Enabling Technologies and Future Trends*, 2021. Pp. 119–133.
41. Xie H., Xiao J., Lei J., Xie W., Klette R. *Image scene conversion algorithm based on generative adversarial networks*. In: *Asian Conference on Pattern Recognition*. Springer, Singapore, 2019. Pp. 29–36.
42. Srivastava N., Hinton G. E., Krizhevsky A., Sutskever I., Salakhutdinov R. R. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 2014, vol. 15, no. 1, pp. 1929–1958.
43. Gastaldi X. Shake-shake regularization. *arXiv preprint arXiv:1705.07485*, 2017.
44. *CIFAR-10 Dataset*. Available at: <https://www.cs.toronto.edu/~kriz/cifar.html> (accessed 5 February 2021).
45. Yamada Y., Iwamura M., Akiba T., Kise K. Shakedown regularization for deep residual learning. *IEEE Access*, 2019, vol. 7, pp. 186126–186136.
46. Torralba A., Efros A. A. Unbiased look at dataset bias. *CVPR*, 2011, pp. 1521–1528.
47. *PASCAL VOC Dataset*. Available at: <http://host.robots.ox.ac.uk/pascal/VOC/> (accessed 5 January 2021).
48. *State Farm Distracted Driver Detection*. Available at: <https://www.kaggle.com/c/state-farm-distracted-driver-detection> (accessed 5 June 2021).
49. *HuBMAP — Hacking the Kidney*. Available at: <https://www.kaggle.com/c/hubmap-kidney-segmentation> (accessed 5 June 2021).
50. Tan M., Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning, PMLR*, 2019, pp. 6105–6114.

## Вычисление аддитивной границы вероятности ошибки декодирования с использованием характеристических функций

А. Н. Трофимов<sup>а</sup>, канд. техн. наук, доцент, [orcid.org/0000-0003-1233-5222](https://orcid.org/0000-0003-1233-5222), [andrei.trofimov@k36.org](mailto:andrei.trofimov@k36.org)  
Ф. А. Таубин<sup>а</sup>, доктор техн. наук, профессор, [orcid.org/0000-0002-8781-9531](https://orcid.org/0000-0002-8781-9531)

<sup>а</sup>Санкт-Петербургский государственный университет аэрокосмического приборостроения,  
Б. Морская ул., 67, Санкт-Петербург, 190000, РФ

**Введение:** точное значение вероятности ошибки декодирования обычно не удается вычислить, поэтому стандартный подход при получении верхней границы вероятности ошибки декодирования по максимуму правдоподобия базируется на использовании аддитивной границы и границы Чернова и (или) ее модификаций. Для многих ситуаций этот подход не обеспечивает достаточной точности. **Цель:** разработка метода точного вычисления аддитивной границы вероятности ошибочного декодирования для широкого класса кодов и моделей каналов без памяти. **Методы:** использование характеристических функций логарифма отношения правдоподобия для произвольной пары кодовых слов, решетчатого представления кодов и численного интегрирования. **Результаты:** полученная точная аддитивная граница вероятности ошибочного декодирования основана на сочетании использования характеристических функций и произведения решетчатых диаграмм для рассматриваемого кода, благодаря чему можно получить итоговое выражение в интегральной форме, удобной для численного интегрирования. Важной особенностью предложенной процедуры является то, что она позволяет точно вычислить аддитивную границу с помощью подхода, основанного на использовании передаточных (производящих) функций. При таком подходе метки ветвей в произведении решетчатых диаграмм для рассматриваемого кода заменяются соответствующими характеристическими функциями. Полученное итоговое выражение позволяет с использованием стандартных методов численного интегрирования вычислять значения аддитивной границы вероятности ошибочного декодирования с требуемой точностью. **Практическая значимость:** по итогам проведенной работы можно заметно улучшить качество оценивания вероятности ошибочного декодирования и тем самым повысить эффективность решений, связанных с применением конкретных схем кодирования для широкого класса каналов связи.

**Ключевые слова** – декодирование по максимуму правдоподобия, аддитивная граница вероятности ошибки, попарная вероятность ошибки, характеристическая функция, канал с ID-AGN-шумом, канал с аддитивным лапласовым шумом.

**Для цитирования:** Трофимов А. Н., Таубин Ф. А. Вычисление аддитивной границы вероятности ошибки декодирования с использованием характеристических функций. *Информационно-управляющие системы*, 2021, № 4, с. 71–85. doi:10.31799/1684-8853-2021-4-71-85

**For citation:** Trofimov A. N., Taubin F. A. Evaluation of the union bound for the decoding error probability using characteristic functions. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 71–85 (In Russian). doi:10.31799/1684-8853-2021-4-71-85

### Введение

Широкое использование кодов с исправлением ошибок для хранения, обработки и передачи данных диктует интерес к разработке новых, более точных методов оценки вероятности ошибки, достигаемой при введении помехоустойчивого кодирования. Эффективный метод вычисления вероятности ошибки декодирования может значительно снизить затраты, связанные с анализом и реализацией процедуры кодирования, как за счет рационального выбора параметров схемы кодирования, так и благодаря тому, что не требуется прибегать к компьютерному моделированию.

Поскольку точное значение вероятности ошибки декодирования обычно не удается вычислить, стандартный подход при получении верхней границы вероятности ошибки как для декодирования по максимуму правдоподобия (МП), так и для декодирования по максимуму апостериорной

вероятности (МАН) базируется на использовании аддитивной границы [1, 2]. Хотя эти процедуры декодирования, как правило, являются чрезмерно сложными для длинных кодов, получение подобного рода верхних границ для вероятности ошибки декодирования представляет интерес для коротких и умеренно длинных кодов. При использовании аддитивной границы точная вероятность ошибки декодирования ограничивается суммой попарных вероятностей. Попарная вероятность ошибки представляет собой вероятность такого события для пары кодовых слов, при котором значение правдоподобия (или апостериорная вероятность) для правильного (переданного) кодового слова меньше, чем для неправильного кодового слова. Для канала без памяти и двоичного кода эта вероятность представляет собой вероятность того, что сумма некоторых случайных величин превосходит определенный порог. В ряде ситуаций аддитивная граница позволяет достаточно

точно оценить истинное значение вероятности ошибочного декодирования.

Наряду с аддитивной границей, для оценки вероятности ошибки декодирования могут быть использованы и другие, более точные верхние границы (см., например, обзор [3]), которые обычно подразделяют на две категории (хотя между ними существует весьма тесная связь). К первой категории относятся граница Галлагера — Фано и ее вариации, такие как граница Думана — Салехи (Duman — Salehi bound) и граница Шульмана — Федера (Shulman — Feder bound), а ко второй — так называемые тангенциальные границы (tangential bounds), восходящие к тангенциальной границе Берлекемпа и включающие, в частности, тангенциальную сферическую границу (tangential-sphere bound — TSB) Полтырева, границу Дивсалара (Divsalar bound) и границу Энгдала — Зигангирова (Engdahl — Zigangirov bound) для сверточных кодов. Однако сфера применения указанных границ довольно ограничена, что связано со следующими факторами. Во-первых, улучшенные верхние границы получены, как правило, для простейших моделей канала (в частности, для канала с аддитивным белым гауссовым шумом) и частных вариантов сигнальных созвездий. Во-вторых, для коротких и умеренно длинных кодов (длиной до 100) улучшенные верхние границы практически совпадают с аддитивной границей при вероятности ошибки декодирования порядка  $10^{-6}$  и меньше. Заметный же выигрыш в точности наблюдается для турбокодов (с большой величиной перемежения) и длинных низкоплотностных кодов, имеющих достаточно большую сложность декодирования. И, наконец, вычисление улучшенных верхних границ оказывается существенно более сложным, чем аддитивной границы. Поэтому аддитивная граница остается доминирующим инструментом, особенно в плане прикладных исследований для практически значимых моделей каналов, при анализе помехоустойчивости конкретных кодовых схем на основе коротких и умеренно длинных кодов.

Для произвольных кодов и каналов без памяти вычисление как попарных вероятностей, так и аддитивной границы в целом оказывается весьма сложным, поэтому обычно используется следующий подход. Попарная вероятность вначале оценивается сверху с использованием границы Чернова, позволяющей представить ее в виде произведения ряда сомножителей, а затем с помощью передаточной (производящей) функции кода учитываются и суммируются все возможные попарные вероятности, обеспечивая окончательный результат для аддитивной границы в замкнутой форме. Некоторые результаты по уточнению границы Чернова для ряда моделей каналов без памяти можно найти в работе [4].

Точное вычисление попарных вероятностей в замкнутой форме возможно только для относительно простых моделей каналов без памяти. Среди известных подходов к получению в замкнутой форме выражений для попарных вероятностей [5–9] следует выделить подход, предложенный в работе W. Turin: Union Bound on Viterbi Algorithm Performance, *AT&T Tech. Journal*, 1985, основанный на использовании аппарата характеристических функций. Модификация этого подхода с ориентацией на произвольные модели каналов без памяти и использование компьютерных методов, таких как квадратурные формулы Гаусса — Чебышева и численное интегрирование, оказывается более предпочтительным подходом по сравнению с подходами, основанными на получении аппроксимации аддитивной границы в замкнутой форме.

Также важно, что численное интегрирование, а также квадратурные формулы Гаусса — Чебышева могут применяться для вычисления аддитивной границы, даже если передаточная функция не задана аналитически.

В данной работе предоставлен метод точного вычисления аддитивной границы вероятности ошибочного декодирования для широкого класса кодов и произвольных моделей каналов без памяти. Предложенный метод включает получение точного выражения для попарной вероятности, используемой при получении аддитивной границы. Получение точного выражения для попарной вероятности достигается благодаря использованию характеристических функций (х.ф.) логарифма отношения правдоподобия для рассматриваемой пары кодовых слов.

### Представление аддитивной границы с использованием характеристических функций

Решение при декодировании некоторого кода по МП принимается по правилу  $\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p_{y|\mathbf{x}}(\mathbf{y} | \mathbf{x})$ , где  $p_{y|\mathbf{x}}(\mathbf{y} | \mathbf{x})$  — условная вероятность (для канала с дискретным выходом), или  $n$ -мерная условная функция плотности вероятности (ф.п.в.), задающая распределение выходных значений  $\mathbf{y} = (y^{(1)}, \dots, y^{(n)})$  канала с непрерывным выходом;  $\mathbf{x}$  — переданное кодовое слово длиной  $n$ . Для канала без памяти

$$p_{y|\mathbf{x}}(\mathbf{y} | \mathbf{x}) = \prod_{l=1}^n p_{y^{(l)}|x^{(l)}}(y^{(l)} | x^{(l)}), \quad (1)$$

где  $p_{y^{(l)}|x^{(l)}}(y | x)$  — одномерная условная вероятность или условная ф.п.в. для канала с дискретным и непрерывным выходом соответственно. Вероятность ошибки при передаче слова  $\mathbf{x}$  может быть

ограничена сверху с использованием аддитивного неравенства как

$$P_e(\mathbf{x}) \leq \sum_{\mathbf{x}' \neq \mathbf{x}} P_e(\mathbf{x} \rightarrow \mathbf{x}' | \mathbf{x}), \quad (2)$$

где

$$P_e(\mathbf{x} \rightarrow \mathbf{x}' | \mathbf{x}) = \Pr[p(\mathbf{y} | \mathbf{x}') > p(\mathbf{y} | \mathbf{x}) | \mathbf{x}] \quad (3)$$

— вероятность ошибки декодирования для кода из двух слов  $\mathbf{x}$  и  $\mathbf{x}'$  при условии, что было передано слово  $\mathbf{x}$ , а суммирование по  $\mathbf{x}'$  здесь и далее понимается как суммирование по словам кода. Вероятность (3) с использованием равенства (1) может быть записана в эквивалентной форме  $P_e(\mathbf{x} \rightarrow \mathbf{x}' | \mathbf{x}) = \Pr[Z(\mathbf{y} | \mathbf{x}, \mathbf{x}') > 0 | \mathbf{x}]$ , где  $Z(\mathbf{y} | \mathbf{x}, \mathbf{x}') = \sum_{l=1}^n z(y^{(l)} | x^{(l)}, x'^{(l)})$  и

$$z(y | x, x') = \ln \frac{p_{y|x}(y | x')}{p_{y|x}(y | x)}. \quad (4)$$

Вычисление аддитивной границы для вероятности ошибки декодирования  $p_e$  базируется на использовании следующего результата [10]:

«Утверждение. Пусть  $Z$  — вещественная случайная величина и  $w_Z(\cdot)$  — ее ф.п.в. Пусть  $C(\omega)$  — х.ф. случайной величины  $Z$ , т. е.

$$C_Z(\omega) = \overline{e^{j\omega Z}} = \int_{-\infty}^{\infty} e^{j\omega Z} w_Z(x) dx, \quad (5)$$

черта сверху здесь и далее обозначает усреднение. Тогда

$$\Pr[Z > 0] = \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \frac{C_Z(\alpha - j\beta)}{\beta + j\alpha} d\alpha, \quad 0 < \beta < \beta_0,$$

где  $\beta_0$  — максимальное значение  $\operatorname{Im}\omega$ , при котором сходится интеграл (5), который может быть записан в виде

$$\begin{aligned} C_Z(\omega) \Big|_{\omega=\alpha-j\beta} &= C_Z(\alpha - j\beta) = \\ &= \overline{e^{j(\alpha-j\beta)Z}} = \int_{-\infty}^{\infty} e^{j\alpha x} e^{\beta x} w_Z(x) dx = \\ &= \int_{-\infty}^{\infty} e^{\beta x} \cos(\alpha x) w_Z(x) dx + j \int_{-\infty}^{\infty} e^{\beta x} \sin(\alpha x) w_Z(x) dx. \end{aligned} \quad (6)$$

Заметим, что обычно х.ф. определяется как функция вещественного аргумента, но здесь рассмотрено ее обобщение как функции комплексного аргумента  $\omega$ .

Используя это утверждение, а также то обстоятельство, что канал описывается моделью без памяти [см. (1)], можно записать, что

$$P_e(\mathbf{x} \rightarrow \mathbf{x}' | \mathbf{x}) = \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \frac{\prod_{l=1}^n c_z(\alpha - j\beta; x^{(l)}, x'^{(l)})}{\beta + j\alpha} d\alpha$$

при  $0 < \beta < \beta_0$ , (7)

где

$$c_z(\omega; x, x') = \overline{\exp(j\omega z(y | x, x'))} \quad (8)$$

— х.ф. случайной величины  $z(y | x, x')$ , определенной равенством (4), а верхняя граница  $\beta_0$  следует из условия сходимости интеграла (6).

Вероятность ошибки декодирования  $p_e$  оценивается сверху с использованием аддитивного неравенства (2)

$$\begin{aligned} p_e &\leq \sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}} P_e(\mathbf{x} \rightarrow \mathbf{x}' | \mathbf{x}) P(\mathbf{x}) = \\ &= \frac{1}{\pi M} \sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}_0} \int_0^{\infty} \operatorname{Re} \frac{\prod_{l=1}^n c_z(\alpha - j\beta; x^{(l)}, x'^{(l)})}{\beta + j\alpha} d\alpha = \\ &= \frac{1}{\pi M} \int_0^{\infty} \operatorname{Re} \frac{\sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}} \prod_{l=1}^n c_z(\alpha - j\beta; x^{(l)}, x'^{(l)})}{\beta + j\alpha} d\alpha, \end{aligned} \quad (9)$$

где  $P(\mathbf{x})$  — вероятность использования последовательности  $\mathbf{x}$ . При записи равенства в выражении (9) учтено предположение о равновероятном использовании последовательностей  $\mathbf{x}$ , т. е. что  $P(\mathbf{x}) = 1/M$ , где  $M$  — объем кода.

Поскольку  $c_z(\omega; x, x) = 1$  для любого значения аргумента  $\omega$ , то

$$\begin{aligned} \sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}} \prod_{l=1}^n c_z(\omega; x^{(l)}, x'^{(l)}) &= \\ &= \sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}} \prod_{l=1}^n c_z(\omega; x^{(l)}, x'^{(l)}) - \sum_{\mathbf{x}} \prod_{l=1}^n c_z(\omega; x^{(l)}, x^{(l)}) = \\ &= \sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}} \prod_{l=1}^n c_z(\omega; x^{(l)}, x'^{(l)}) - M. \end{aligned}$$

Поэтому можно записать, что

$$p_e \leq \frac{1}{\pi M} \int_0^{\infty} \operatorname{Re} \frac{D(\alpha - j\beta) - M}{\beta + j\alpha} d\alpha, \quad (10)$$

где

$$D(\omega) = \sum_{\mathbf{x}} \sum_{\mathbf{x}' \neq \mathbf{x}} \prod_{l=1}^n c_z(\omega; x^{(l)}, x'^{(l)}). \quad (11)$$

Задача вычисления границы (10) состоит из трех этапов: получение выражения для функции  $D(\omega)$ , получение выражений для х.ф.  $c_z(\omega; x, x')$

и вычисление интеграла (10). Первые две задачи могут быть решены аналитически, а для решения последней задачи обычно требуется численное интегрирование.

### Вывод выражения для функции $D(\omega)$

Функция  $D(\omega)$ , определенная равенством (11), зависит от структуры кода (включающей канальный алфавит) и от вида х.ф. и их значений, определяемых моделью рассматриваемого канала без памяти. В частном случае — для геометрически равномерных кодов (geometrically uniform codes) — функция  $D(\omega)$  может быть вычислена с использованием спектра кода. Основным примером геометрически равномерных кодов являются линейные коды в совокупности с равномерной фазовой модуляцией. В общем же случае для произвольных кодов (как блочных, так и сверточных) и произвольных канальных алфавитов функция  $D(\omega)$  может быть вычислена с использованием подходящей решетчатой диаграммы. Рассмотрим сначала случай блочного кодирования. Для описания структуры кода воспользуемся понятием решетчатой диаграммы, или кодовой решетки блочного кода. При этом описании используются терминология и определения из работы [11].

Решетчатая диаграмма представляет собой размеченный направленный граф  $\Omega = (V, E, A)$ , где  $V, E, A$  — множество вершин, множество ребер и множество меток ребер (алфавит) соответственно. Множество ребер  $E$  состоит из троек  $e = (v, v', a)$ ,  $v, v' \in V, a \in A$ . Множество вершин  $V$  решетчатой диаграммы (кодовой решетки) разделяется на непересекающиеся подмножества как

$$V = V^{(0)} \cup V^{(1)} \cup \dots \cup V^{(n-1)} \cup V^{(n)}, \quad (12)$$

где каждое множество  $V^{(i)}$  образует вершины  $i$ -го уровня диаграммы,  $i = 0, 1, \dots, n-1$ . При этом каждое ребро  $e \in E$  начинается в вершине из  $V^{(i)}$  и заканчивается в вершине из  $V^{(i+1)}$ , и каждая вершина  $v \in V$  лежит хотя бы на одном из путей от вершины из множества начальных вершин  $V^{(0)}$  до вершины из множества конечных вершин  $V^{(n)}$ . Представление (12) множества вершин  $V$  приводит к тому, что множество ребер может быть задано как  $E = E^{(1)} \cup E^{(2)} \cup \dots \cup E^{(n-1)} \cup E^{(n)}$ , где  $E^{(i)}$  — непересекающиеся подмножества, содержащие ребра, которые завершаются в вершинах из  $V^{(i)}$ . Аналогично можно записать, что  $A = A^{(1)} \cup A^{(2)} \cup \dots \cup A^{(n-1)} \cup A^{(n)}$ , где подмножества  $A^{(i)}$  образуют множества меток ребер из  $E^{(i)}$ . Множества  $A^{(i)}$  не обязательно должны быть дизъюнктивными, т. е. не имеющими общих элементов, и во многих случаях  $A^{(1)} = A^{(2)} = \dots = A^{(n)} = A$ . Последовательность меток вдоль пути

в диаграмме, начинающегося в вершине из  $V^{(0)}$  и приходящего в вершину из  $V^{(n)}$ , представляет собой кодовое слово блочного кода. Очевидно, что номер последнего уровня  $n$  равен длине блочного кода. В решетчатых диаграммах многих блочных кодов  $|V^{(0)}| = |V^{(n)}| = 1$ . Примерами блочных кодов, в решетках которых  $|V^{(0)}| > 1, |V^{(n)}| > 1$ , могут служить циклически усеченные сверточные коды (tail-biting convolutional codes). Обозначим  $\Sigma^{(i)} = \{0, 1, \dots, |V^{(i)}| - 1\}$ . Будем считать, что зафиксировано некоторое взаимно-однозначное соответствие  $V^{(i)} \leftrightarrow \Sigma^{(i)}$ , обозначаемое как  $\Sigma(v) = k, v \in V^{(i)}, 0 \leq k \leq |V^{(i)}| - 1$ . Тогда элементы множества  $\Sigma^{(i)}$  могут рассматриваться как индексы вершин  $i$ -го уровня. Множество  $\Sigma = \Sigma^{(1)} \cup \Sigma^{(2)} \cup \dots \cup \Sigma^{(n-1)} \cup \Sigma^{(n)}$  образует множество состояний решетчатой диаграммы и  $S_\Omega = \max_i |\Sigma^{(i)}| = \max_i |V^{(i)}|$  представляет собой максимальное число состояний диаграммы, которое некоторым образом характеризует ее сложность.

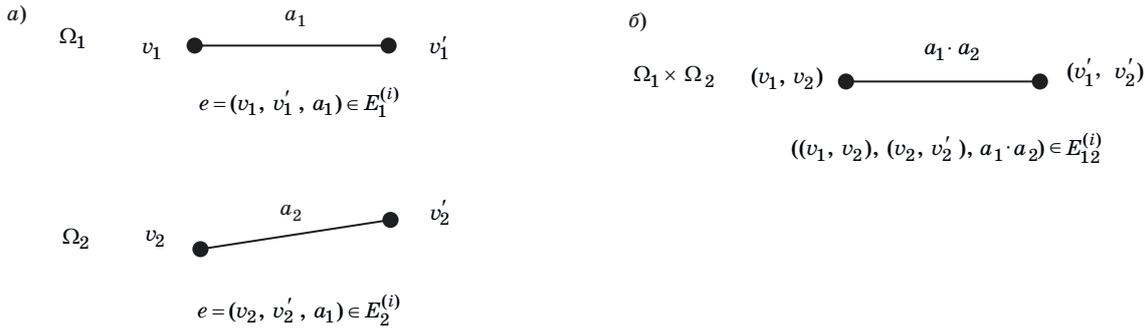
Отметим, что в литературе представлены формализованные процедуры построения кодовых решеток для различных блочных кодов [11]. В частности, для любого линейного двоичного блочного кода кодовая решетка с максимальным числом состояний диаграммы  $S_\Omega \leq \min(M, 2^n/M)$  легко строится по проверочной матрице кода. Также легко могут быть построены кодовые решетки для каскадных кодов над не двоичным алфавитом, состоящие из двоичных компонентных кодов (некоторые примеры приведены в работах [10, 12]).

Вычисление функции  $D(\omega)$  базируется на использовании произведения кодовых решеток. Произведение решеток  $\Omega_1 = (V_1, E_1, A_1)$  и  $\Omega_2 = (V_2, E_2, A_2)$ , где  $V_1 = V_1^{(0)} \cup V_1^{(1)} \cup \dots \cup V_1^{(n-1)} \cup V_1^{(n)}$  и  $V_2 = V_2^{(0)} \cup V_2^{(1)} \cup \dots \cup V_2^{(n-1)} \cup V_2^{(n)}$ , обозначается как  $\Omega_1 \times \Omega_2$  и представляет собой решетку вида  $\Omega_1 \times \Omega_2 = \{V_{12}, E_{12}, A_{12}\}$ , где

$$V_{12} = (V_1^{(0)} \times V_2^{(0)}) \cup (V_1^{(1)} \times V_2^{(1)}) \cup \dots \cup (V_1^{(n)} \times V_2^{(n)}).$$

Ребро  $e = ((v_1, v_2), (v'_1, v'_2), (a_1 \cdot a_2)) \in E_{12}$  для вершин  $(v_1, v_2) \in V_1^{(i)} \times V_2^{(i)}$  и  $(v'_1, v'_2) \in V_1^{(i+1)} \times V_2^{(i+1)}$  существует в решетке  $\Omega_1 \times \Omega_2$ , если существуют ребра  $e_1 = (v_1, v'_1, a_1)$  и  $e_2 = (v_2, v'_2, a_2)$  в решетках  $\Omega_1$  и  $\Omega_2$  соответственно для пар вершин  $v_1 \in V_1^{(i)}, v'_1 \in V_1^{(i+1)}$  и  $v_2 \in V_2^{(i)}, v'_2 \in V_2^{(i+1)}$ . Символ « $\cdot$ » обозначает здесь некоторую композицию символов  $a_1$  и  $a_2$ , а алфавит произведения решеток определяется как  $A_{12} = \{a_1 \cdot a_2 | a_1 \in A_1, a_2 \in A_2\}$ . На рис. 1 приведена поясняющая иллюстрация.

Для получения выражения для функции  $D(\omega)$ , определенной равенством (11), применяется ранее использовавшийся подход [10, 12]. Чтобы построить функцию  $D(\omega)$ , необходимо рассмотреть все пары кодовых слов  $x, x'$  или все



■ **Рис. 1.** Вид ребер в исходных решетках (а) и в произведении  $\Omega_1 \times \Omega_2$  (б)  
 ■ **Fig. 1.** The form of edges in the original trellises (а) and in the product  $\Omega_1 \times \Omega_2$  (б)

возможные пары путей в решетчатой диаграмме кода  $\Omega = (V, E, X)$ ,  $E = E^{(1)} \cup E^{(2)} \cup \dots \cup E^{(n)}$ ,  $X$  — кодовый алфавит. Суммирование по  $x$  и  $x'$  в (11) соответствует суммированию по всем путям в решетке  $\Omega^2 = \Omega \times \Omega = \{V \times V, E_2, \{c_z(\omega; x, x')\}\}$ , в которой любое ребро имеет вид  $((v_1, v_2), (v'_1, v'_2), c_z(\omega; x, x'))$ , если в решетке  $\Omega$  соответствующие ребра на одном уровне решетки имеют вид  $(v_1, v_2, x) \in E^{(i)}$  и  $(v'_1, v'_2, x') \in E^{(i)}$ ,  $x, x' \in X$ . Иными словами, здесь композиция меток ребер задается как  $x \cdot x' = c_z(\omega; x, x')$ .

Рассмотрим сначала простой случай, когда кодовая решетка  $\Omega$  имеет одну начальную и одну конечную вершину. В этом случае и решетка  $\Omega^2$  имеет одну начальную и одну конечную вершину. Пусть  $E_2^{(i)}$  — множество ребер решетки  $\Omega^2$  на  $i$ -м уровне,  $i = 1, 2, \dots, n$ . Для уровня  $i$  определим величины

$$P_{kl}^{(i)}(\omega) = \sum_{e(v,v',c_z(\omega;x,x')) \in E_2^{(i)}} c_z(\omega; x, x'),$$

где  $k = \Sigma(v)$ ,  $l = \Sigma(v')$  — индексы вершин  $v$  и  $v'$ , или состояния решетки  $\Omega^2$ , т. е.  $P_{kl}^{(i)}(\omega)$  вычисляется как сумма меток параллельных ребер, связывающих состояния  $k$  и  $l$  в решетке  $\Omega^2$ . Определим матрицы

$$P^{(i)}(\omega) = [P_{kl}^{(i)}(\omega)], \tag{13}$$

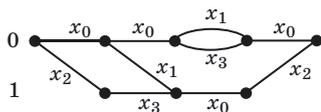
$k = 0, 1, \dots, |V^{(i-1)}|^2 - 1$ ,  $l = 0, 1, 2, \dots, |V^{(i)}|^2 - 1$ . Тогда нетрудно заметить, что

$$D(\omega) = P^{(1)}(\omega)P^{(2)}(\omega)\dots P^{(n)}(\omega). \tag{14}$$

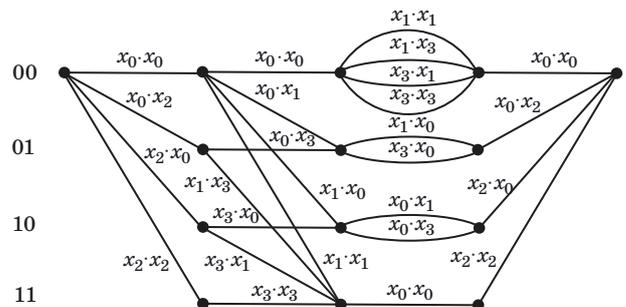
Следующий пример иллюстрирует процесс получения выражения для функции  $D(\omega)$  для простого нелинейного недвоичного блочного кода.

**Пример.** Пусть решетка  $\Omega = (V, E, A)$ , задающая некоторый блочный код, имеет вид, показанный на рис. 2.

Этот код, очевидно, имеет длину  $n = 4$ , построен над алфавитом  $A = \{x_0, x_1, x_2, x_3\}$  и содержит четыре кодовых слова:  $(x_0 x_0 x_1 x_0)$ ,  $(x_0 x_0 x_3 x_0)$ ,  $(x_0 x_1 x_0 x_2)$ ,  $(x_2 x_3 x_0 x_2)$ , т. е. для него  $M = 4$ . На рис. 3 показана решетка  $\Omega^2 = \Omega \times \Omega$ .



■ **Рис. 2.** Пример решетки  $\Omega = (V, E, A)$   
 ■ **Fig. 2.** Example of trellis  $\Omega = (V, E, A)$



■ **Рис. 3.** Решетка  $\Omega^2 = \Omega \times \Omega$   
 ■ **Fig. 3.** Trellis  $\Omega^2 = \Omega \times \Omega$

Далее под композицией  $x \cdot x'$ , представляющей собой метку ребра в решетке  $\Omega^2$ , будем понимать х.ф.  $c_z(\omega; x, x')$ . Поэтому для матриц  $\mathbf{P}^{(1)}(\omega)$ ,  $\mathbf{P}^{(2)}(\omega)$ ,  $\mathbf{P}^{(3)}(\omega)$  и  $\mathbf{P}^{(4)}(\omega)$  с учетом того, что  $c_z(\omega; x, x) = 1$ , можно записать следующие выражения:

$$\begin{aligned} \mathbf{P}^{(1)}(\omega) &= [1 \quad c_z(\omega; x_0, x_2) \quad c_z(\omega; x_2, x_0) \quad 1]; \\ \mathbf{P}^{(2)}(\omega) &= \begin{bmatrix} 1 & c_z(\omega; x_0, x_1) & c_z(\omega; x_1, x_0) & 1 \\ 0 & c_z(\omega; x_0, x_3) & 0 & c_z(\omega; x_1, x_3) \\ 0 & 0 & c_z(\omega; x_3, x_0) & c_z(\omega; x_3, x_1) \\ 0 & 0 & 0 & 1 \end{bmatrix}; \\ \mathbf{P}^{(3)}(\omega) &= \begin{bmatrix} 2+c_z(\omega; x_1, x_3)+c_z(\omega; x_3, x_1) & 0 & 0 & 0 \\ 0 & c_z(\omega; x_1, x_0)+c_z(\omega; x_3, x_0) & 0 & 0 \\ 0 & 0 & 0 & c_z(\omega; x_0, x_1)+c_z(\omega; x_0, x_3) \\ 0 & 0 & 0 & 1 \end{bmatrix}; \\ \mathbf{P}^{(4)}(\omega) &= [1 \quad c_z(\omega; x_0, x_2) \quad c_z(\omega; x_2, x_0) \quad 1]^T. \end{aligned}$$

Вычисляя выражение для  $D(\omega)$ , получаем, что

$$\begin{aligned} D(\omega) &= \mathbf{P}^{(1)}(\omega)\mathbf{P}^{(2)}(\omega)\mathbf{P}^{(3)}(\omega)\mathbf{P}^{(4)}(\omega) = \\ &= c_z(\omega; x_1, x_3) + c_z(\omega; x_3, x_1) + c_z(\omega; x_0, x_2)c_z(\omega; x_1, x_3) + c_z(\omega; x_2, x_0)c_z(\omega; x_3, x_1) + \\ &+ c_z(\omega; x_0, x_1)c_z(\omega; x_0, x_2)c_z(\omega; x_1, x_0) + c_z(\omega; x_0, x_1)c_z(\omega; x_1, x_0)c_z(\omega; x_2, x_0) + \\ &+ c_z(\omega; x_0, x_1)c_z(\omega; x_0, x_2)c_z(\omega; x_3, x_0) + c_z(\omega; x_0, x_3)c_z(\omega; x_1, x_0)c_z(\omega; x_2, x_0) + \\ &+ c_z(\omega; x_0, x_2)^2 c_z(\omega; x_0, x_3)c_z(\omega; x_1, x_0) + c_z(\omega; x_0, x_2)^2 c_z(\omega; x_0, x_3)c_z(\omega; x_3, x_0) + \\ &+ c_z(\omega; x_0, x_1)c_z(\omega; x_2, x_0)^2 c_z(\omega; x_3, x_0) + c_z(\omega; x_0, x_3)c_z(\omega; x_2, x_0)^2 c_z(\omega; x_3, x_0) + 4. \end{aligned}$$

Это выражение содержит  $M^2 = 16$  слагаемых вида  $\prod_{l=1}^4 c(\omega; x^{(l)}, x'^{(l)})$  по числу пар кодовых слов  $(\mathbf{x}, \mathbf{x}')$ , из них  $M(M - 1) = 12$  нетривиальных слагаемых по числу пар  $(\mathbf{x}, \mathbf{x}')$ ,  $\mathbf{x} \neq \mathbf{x}'$  и  $M = 4$  тривиальных слагаемых, равных единице, по числу пар  $(\mathbf{x}, \mathbf{x}')$ ,  $\mathbf{x} = \mathbf{x}'$ .

Если решетка  $\Omega$  имеет несколько начальных и столько же конечных состояний, т. е.  $|V^{(0)}| = |V^{(n)}| = S$ , то вычисление функции  $D(\omega)$  становится чуть сложнее. Обозначим через  $\Omega_s$  подрешетку в решетчатой диаграмме  $\Omega = (V, E, X)$ , в которой есть только одно начальное состояние  $s = \Sigma(v_s)$ ,  $v_s \in V^{(0)}$ , и только одно конечное состояние  $s = \Sigma(v_s)$ ,  $v_s \in V^{(n)}$ ,  $s = 0, 1, \dots, S - 1$ , т. е.  $\Omega_s = (V_s, E_s, X_s)$ , где  $V_s = \{v_s\} \cup V^{(1)} \cup \dots \cup V^{(n-1)} \cup \{v_s\}$ . Очевидно, решетчатая диаграмма  $\Omega$  может быть представлена как набор из  $S$  таких подрешеток  $\Omega_s$  (малых решеток) с идентичными начальными и конечными состояниями. Обозначим  $\Omega_s \times \Omega_{s'} = \{V_s \times V_{s'}, E_{2_{ss'}}, \{c_z(\omega; x, x')\}\}$  решетку, в которой каждое ребро имеет вид  $((v_1, v_2), (v'_1, v'_2), c_z(\omega; x, x'))$ , если в решетках  $\Omega_s$  и  $\Omega_{s'}$  соответствующие ребра на одном уровне решетки имеют вид  $(v_1, v_2, x) \in E_s^{(i)}$  и  $(v'_1, v'_2, x') \in E_{s'}^{(i)}$ ,  $x, x' \in X$ ,  $s, s' = 0, 1, \dots, S - 1$ . С учетом такого представления функция  $D(\omega)$  может быть построена как

$$D(\omega) = \sum_{s=0}^{S-1} \sum_{s'=0}^{S-1} D_{ss'}(\omega), \tag{15}$$

где

$$D_{ss'}(\omega) = \mathbf{P}_{ss'}^{(1)}(\omega)\mathbf{P}_{ss'}^{(2)}(\omega)\dots\mathbf{P}_{ss'}^{(n)}(\omega), \tag{16}$$

а матрицы  $\mathbf{P}_{ss'}^{(i)}(\omega)$  строятся для решеток  $\Omega_s \times \Omega_{s'}$  по той же процедуре, по какой были построены матрицы  $\mathbf{P}^{(i)}(\omega)$ ,  $i = 1, 2, \dots, n$ , для решетки  $\Omega \times \Omega$  для случая с одной начальной и одной конечной вершинами. Очевидно, что выражение (14) — это частный случай формул (15) и (16) для случая  $S = |V^{(0)}| = |V^{(n)}| = 1$ .

Можно показать, что при  $\beta = 1/2$ , т. е. когда  $\omega = \alpha - j/2$ , выполняется равенство  $D_{ss'}(\omega) = D_{s's}(\omega)$ . Поэтому равенство (15) может быть представлено как

$$D(\alpha - j/2) = \sum_{s=0}^{S-1} D_{ss}(\alpha - j/2) + 2 \sum_{s=0}^{S-1} \sum_{s'=s+1}^{S-1} D_{ss'}(\alpha - j/2).$$

Это свойство используется для уменьшения сложности вычисления верхней границы вероятности  $p_e$ . Окончательное выражение для вероятности ошибки декодирования, в котором используются функции (16), имеет вид

$$p_e \leq \frac{1}{\pi M} \int_0^{\infty} \operatorname{Re} \frac{\sum_{s=0}^{S-1} D_{ss}(\alpha - j/2) + 2 \sum_{s=0}^{S-1} \sum_{s'=s+1}^{S-1} D_{ss'}(\alpha - j/2) - M}{1/2 + j\alpha} d\alpha. \quad (17)$$

Отметим, что для вычисления правой части (17) нужно вычислить  $S(S+1)/2$  функций  $D_{ss'}(\cdot)$ , а не  $S^2$ , как это следует из формулы (15).

### Упрощение для линейных кодов и аддитивных каналов

Для линейных кодов и для каналов, для которых выполняется следующее свойство для х.ф.:

$$c_z(\omega; x, x') = c_z(\omega; 0, x' - x), \quad (18)$$

вычисление границы (9) может быть упрощено. Заметим, что определение разности  $x' - x$  в этом выражении зависит от модели канала и может вычисляться по-разному для различных случаев. Примерами каналов, для которых выполняется это равенство, могут служить дискретные и непрерывные каналы с аддитивным шумом, в частности, двоичный симметричный канал и двоичный канал с аддитивным белым гауссовым шумом. Для двоичного симметричного канала разность  $x' - x$  понимается как разность в  $GF(2)$ , т. е.  $x' \oplus x$ . Для канала с аддитивным белым гауссовым шумом и модуляционного отображения  $s: GF(2) \rightarrow \{A_0, A_1\}$ ,  $A_0, A_1 \in \mathbb{R}$ , эта разность также понимается как  $x' \oplus x$ , а при вычислении значений х.ф. используются условные гауссовы ф.п.в. со средними  $s(0)$  и  $s(x' \oplus x)$ . В этом случае двойная сумма по словам кода в выражении (9) может быть представлена как

$$\sum_x \sum_{x' \neq x} \prod_{l=1}^n c_z(\omega; x^{(l)}, x'^{(l)}) = \sum_x \sum_{x' \neq x} \prod_{l=1}^n c_z(\omega; 0, x^{(l)} - x'^{(l)}) =$$

$$= \sum_x \sum_{x' \neq x} \prod_{l=1}^n c_z(\omega; 0, x^{(l)} - x'^{(l)}) = \quad (19)$$

$$= M \sum_{x \neq 0} \prod_{l=1}^n c_z(\omega; 0, x^{(l)}) = M \left( \sum_x \prod_{l=1}^n c_z(\omega; 0, x^{(l)}) - 1 \right), \quad (20)$$

где в переходе от (19) к (20) использовано свойство линейности кода. Для представления суммы по  $x$  в (20) рассмотрим решетку вида  $\Omega_* \times \Omega$ , где  $\Omega_* = \{\{v_0\} \cup \{v_0\} \cup \dots \cup \{v_0\}, E, \{0\}\}$  — решетка, содержащая лишь одну вершину на каждом уровне и только один путь из начальной вершины в конечную, отмеченный нулевыми символами. Тогда сумму по кодовым словам  $x$  в (20) можно представить как

$$D_1(\omega) = \sum_x \prod_{l=1}^n c_z(\omega; 0, x^{(l)}) = P_1^{(1)}(\omega) P_1^{(2)}(\omega) \dots P_1^{(n)}(\omega), \quad (21)$$

где матрицы  $P_1^{(i)}(\omega)$  строятся на основе решетки  $\Omega_* \times \Omega$  аналогично тому, как были построены матрицы (13) на основе решетки  $\Omega \times \Omega$ . То есть  $P_1^{(i)}(\omega) = [P_{1\ kl}^{(i)}(\omega)]$ , где

$$P_{1\ kl}^{(i)}(\omega) = \sum_{(v, v', c_z(\omega; 0, x)) \in E^{(i)}} c_z(\omega; 0, x),$$

$k = \Sigma(v)$ ,  $l = \Sigma(v')$  — индексы вершин  $v$  и  $v'$ , или состояния решетки  $\Omega_* \times \Omega$ , а  $E^{(i)}$  — множество ребер  $i$ -го уровня в решетке  $\Omega_* \times \Omega$ . Иными словами,  $P_{1\ kl}^{(i)}(\omega)$  вычисляется как сумма меток параллельных ребер, связывающих состояния  $k$  и  $l$  в решетке  $\Omega_* \times \Omega$ . Граница для  $p_e$  в окончательном виде выглядит следующим образом:

$$p_e \leq \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \frac{D_1(\alpha - j\beta) - 1}{\beta + j\alpha} d\alpha. \quad (22)$$

Граница (22), в которой используется функция  $D_1(\omega)$ , вычисляемая по формуле (21), совпадает с обычной аддитивной границей ошибки декодирования для линейного кода и канала с аддитивным шумом (дискретным или непрерывным). При этом сложность ее вычисления определяется сложностью решетки  $\Omega$ , а не решетки  $\Omega \times \Omega$ , как в общем случае.

Предложенный подход с использованием стандартной техники [2] может быть легко распространен на случай решетчатых и сверточных кодов. Тогда нетрудно показать, что, например, для двоичного сверточного кода со скоростью

$R = k/n$  с функцией нумератора весов  $T(D, N)$  [2] границы для вероятности ошибочного события  $P_E$  и вероятности ошибки на бит  $P_B$  принимают вид

$$P_E \leq \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \frac{T(c_z(\alpha - j\beta; 0, 1), 1)}{\beta + j\alpha} d\alpha;$$

$$P_B \leq \frac{1}{\pi k} \int_0^{\infty} \operatorname{Re} \left( \frac{1}{\beta + j\alpha} \frac{dT(c_z(\alpha - j\beta; 0, 1), N)}{dN} \Big|_{N=1} \right) d\alpha. \quad (23)$$

Заметим, что обычная граница вероятностей  $P_E$  и  $P_B$  для сверточных кодов имеет вид  $P_E \leq T(D_0, 1)$ ,  $P_B \leq (1/k)dT(D_0, N)/dN|_{N=1}$ , где  $D_0 = \min_{\beta > 0} e^{\beta z(y|x, x')}$  — показатель границы Чернова. В некоторых случаях эти границы могут быть улучшены за счет введения коэффициента  $K(d) \sim 1/\sqrt{d}$ ; конкретный вид функции  $K(d)$  зависит от модели канала [2, 4]. Границы в этом случае принимают вид  $P_E \leq K(d_f)T(D_0, 1)$ ,  $P_B \leq K(d_f)(1/k)dT(D_0, N)/dN|_{N=1}$ , где  $d_f$  — свободное расстояние сверточного кода.

### Примеры применения

В этом разделе приведены примеры применения границы (10) для некоторых каналов. Для простых случаев — двоичный симметричный канал и гауссов канал с дискретным временем — предложенный подход приводит, как и ожидалось, к известным результатам. Для более сложных моделей канала, рассмотренных в примерах 3–5, использование границы (10) позволяет получить новые результаты.

**Пример 1. Двоичный симметричный канал (ДСК).** В этом случае  $x, y = 0, 1$  и  $p_{x|y}(x|y) = 1 - \varepsilon$ , если  $y = x$ , и  $p_{x|y}(x|y) = \varepsilon$ , если  $y \neq x$ , где  $\varepsilon$  — вероятность ошибки в ДСК,  $\varepsilon < 1/2$ . Характеристическая функция (8) имеет в этом случае вид

$$c_z(\omega; x, x') = \begin{cases} (1 - \varepsilon)^{1-j\omega} \varepsilon^{j\omega} + \varepsilon^{1-j\omega} (1 - \varepsilon)^{j\omega}, & x \neq x', \\ 1, & x = x'. \end{cases} \quad (24)$$

Граница для вероятности  $p_e$  вычисляется в общем виде согласно выражению (10). Если выделить и рассмотреть одно слагаемое в сумме по  $x$  в (11), то получим выражение для вероятности  $P_e(x \rightarrow x' | x)$  [см. (3)] следующего вида:

$$P_e(x \rightarrow x' | x) = \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \frac{\prod_{l=1}^n c_z(\alpha - j\beta; x^{(l)}, x'^{(l)})}{\beta + j\alpha} d\alpha =$$

$$= \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \frac{f(\varepsilon, \alpha - j\beta)^{d_H(x, x')}}{\beta + j\alpha} d\alpha, \quad (25)$$

где  $f(\varepsilon, \omega) = (1 - \varepsilon)^{1-j\omega} \varepsilon^{j\omega} + \varepsilon^{1-j\omega} (1 - \varepsilon)^{j\omega}$ ;  $d_H(x, x')$  — расстояние Хэмминга между словами  $x$  и  $x'$ . Обозначим  $d = d_H(x, x')$  и  $u = \ln((1 - \varepsilon)/\varepsilon)$ ; тогда можно записать, что  $f(\varepsilon, \omega) = (1 - \varepsilon)e^{-j\omega u} + \varepsilon e^{j\omega u}$  и далее легко показать, что

$$f(\varepsilon, \alpha - j\beta)^{d_H(x, x')} = \sum_{l=0}^d C_d^l \varepsilon^{d-l} (1 - \varepsilon)^l e^{ju(d-2l)\alpha} e^{\beta u(d-2l)}.$$

Отсюда и из (25) следует, что

$$P_e(x \rightarrow x' | x) = \sum_{l=0}^d C_d^l \varepsilon^{d-l} (1 - \varepsilon)^l e^{\beta u(d-2l)} I(d-2l),$$

где обозначение  $I(m)$  используется для следующего интеграла:

$$I(m) = \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \frac{e^{jma}}{\beta - j\alpha} d\alpha = \frac{1}{\pi} \int_0^{\infty} \frac{\beta \cos m\alpha}{\beta^2 + \alpha^2} d\alpha + \frac{1}{\pi} \int_0^{\infty} \frac{\alpha \sin m\alpha}{\beta^2 + \alpha^2} d\alpha.$$

Для вычисления этих величин используем табличные тождества

$$\int_0^{\infty} \frac{\cos m\alpha}{\beta^2 + \alpha^2} d\alpha = \begin{cases} \frac{\pi}{2\beta} e^{-m\beta}, & m \geq 0, \\ \frac{\pi}{2\beta} e^{m\beta}, & m < 0; \end{cases}$$

$$\int_0^{\infty} \frac{\alpha \sin m\alpha}{\beta^2 + \alpha^2} d\alpha = \begin{cases} \frac{\pi}{2} e^{-m\beta}, & m > 0, \\ 0, & m = 0, \\ -\frac{\pi}{2} e^{m\beta}, & m < 0, \end{cases}$$

где  $\beta > 0$ , и тогда получим, что

$$I(d-2l) = \begin{cases} e^{-\beta(d-2l)}, & d > 2l, \\ \frac{1}{2}, & d = 2l, \\ 0, & d < 2l, \end{cases}$$

и поэтому

$$P_e(x \rightarrow x' | x) = \begin{cases} \sum_{l=(d+1)/2}^d C_d^l \varepsilon^l (1 - \varepsilon)^{d-l}, & d \text{ — нечетное,} \\ \frac{1}{2} C_d^{d/2} \varepsilon^{d/2} (1 - \varepsilon)^{d/2} + \\ + \sum_{l=d/2+1}^d C_d^l \varepsilon^l (1 - \varepsilon)^{d-l}, & d \text{ — четное.} \end{cases}$$

Последнее равенство означает, что граница (10) для этого случая совпадает с аддитивной границей, в которой каждое слагаемое вычисляется точно.

Для х.ф. (24) выполняется условие (18). Поэтому граница для вероятности  $P_e$  для линейного кода, которая может быть найдена согласно выражению (22), дает в точности тот же результат, что и граница (10), но вычисляется значительно проще.

**Пример 2.** Гауссов канал с дискретным временем. Для этого примера выход канала имеет вид  $y^{(i)} = s(x^{(i)}) + \xi^{(i)}$ , где  $s(x^{(i)})$  — переданная сигнальная точка,  $s(x^{(i)})$ ,  $y^{(i)}$ ,  $\xi^{(i)} \in \mathbb{R}^m$ , где  $m$  — размерность сигнального пространства,  $\xi^{(i)}$  —  $m$ -мерный гауссов случайный вектор с независимыми одинаково распределенными компонентами с параметрами  $(0, \sigma^2)$ . Условная ф.п.в., определяющая канал, задается поэтому следующим образом:

$$P_{y|x}(y|s(x)) = (\sqrt{2\pi}\sigma)^{-m} \exp(-\|y - s(x)\|^2 / 2\sigma^2).$$

Нетрудно показать, что х.ф. (8) имеет в этом случае вид

$$c_z(\omega; x, x') = \exp\left(-\frac{j\omega(1-j\omega)\|s(x) - s(x')\|^2}{2\sigma^2}\right). \quad (26)$$

Тогда для двух кодовых слов  $\mathbf{x} = (x^{(1)}, \dots, x^{(n)})$ ,  $\mathbf{x}' = (x'^{(1)}, \dots, x'^{(n)})$  и их модуляционных отображений  $\mathbf{s}(\mathbf{x}) = (s(x^{(1)}), \dots, s(x^{(n)}))$  и  $\mathbf{s}(\mathbf{x}') = (s(x'^{(1)}), \dots, s(x'^{(n)}))$  можно записать, что

$$\prod_{l=1}^n c_z(\omega; x^{(l)}, x'^{(l)}) = \exp\left(-\frac{j\omega(1-j\omega)d^2}{2\sigma^2}\right)_{\omega=\alpha-j\beta} = e^{-\frac{\alpha^2 d^2}{2\sigma^2}} e^{-\frac{j\alpha(1-2\beta)d^2}{2\sigma^2}} e^{-\frac{\beta(1-\beta)d^2}{2\sigma^2}},$$

где

$$d^2 = d_E^2(\mathbf{s}(\mathbf{x}), \mathbf{s}(\mathbf{x}')) = \sum_{l=1}^n \|s(x^{(l)}) - s(x'^{(l)})\|^2.$$

Если выделить и рассмотреть одно слагаемое в сумме по  $\mathbf{x}$  в (11), то получим выражение для вероятности  $P_e(\mathbf{x} \rightarrow \mathbf{x}' | \mathbf{x})$  [см. (3)] вида

$$P_e(\mathbf{x} \rightarrow \mathbf{x}' | \mathbf{x}) = \frac{1}{\pi} e^{-\frac{\beta(1-\beta)d^2}{2\sigma^2}} \int_0^\infty \operatorname{Re} \frac{\exp(-(\alpha^2 + j\alpha(1-\beta))d^2 / 2\sigma^2)}{\beta + j\alpha} d\alpha.$$

Для сокращения записи введем обозначения

$$a = d^2 / 2\sigma^2, \quad u = (1-2\beta)d^2 / 2\sigma^2. \quad (27)$$

Тогда для вероятности  $P_e(\mathbf{x} \rightarrow \mathbf{x}' | \mathbf{x})$  можно записать выражение

$$P_e(\mathbf{x} \rightarrow \mathbf{x}' | \mathbf{x}) = e^{-a\beta(1-\beta)} I(\alpha, u), \quad (28)$$

где

$$\begin{aligned} I(\alpha, u) &= \frac{1}{\pi} \int_0^\infty \operatorname{Re} \frac{e^{-\alpha a^2 + j\alpha u}}{\beta + j\alpha} d\alpha = \\ &= \frac{1}{\pi} \int_0^\infty \operatorname{Re} \frac{e^{-\alpha a^2} (\cos u\alpha - j \sin u\alpha)}{\beta + j\alpha} d\alpha = \\ &= \frac{1}{\pi} \beta \int_0^\infty \frac{e^{-\alpha a^2} \cos u\alpha}{\beta^2 + \alpha^2} d\alpha - \frac{1}{\pi} \int_0^\infty \frac{\alpha e^{-\alpha a^2} \sin u\alpha}{\beta^2 + \alpha^2} d\alpha = \\ &= \frac{1}{2} e^{a\beta^2 + \beta u} \operatorname{erfc}(\sqrt{a}\beta + u / (2\sqrt{a})). \end{aligned} \quad (29)$$

Последнее выражение в (29) получено с использованием табличных интегралов для условий  $a > 0, \beta > 0, u > 0$ :

$$\begin{aligned} \int_0^\infty \frac{e^{-ax^2} \cos xy}{\beta^2 + x^2} dx &= \frac{\pi}{4\beta} e^{a\beta^2} (e^{-\beta y} \operatorname{erfc}(t_1) + e^{\beta y} \operatorname{erfc}(t_2)); \\ \int_0^\infty \frac{x e^{-ax^2} \sin xy}{\beta^2 + x^2} dx &= \frac{\pi}{4} e^{a\beta^2} (e^{-\beta y} \operatorname{erfc}(t_1) - e^{\beta y} \operatorname{erfc}(t_2)), \end{aligned}$$

где  $t_1 = \sqrt{a}\beta - y / (2\sqrt{a})$ ;  $t_2 = \sqrt{a}\beta + y / (2\sqrt{a})$ . Собирая вместе выражения (28), (29) и (27), получаем в итоге, что

$$\begin{aligned} P_e(\mathbf{x} \rightarrow \mathbf{x}' | \mathbf{x}) &= \frac{1}{2} \operatorname{erfc}\left(\frac{d_E(\mathbf{s}(\mathbf{x}), \mathbf{s}(\mathbf{x}'))}{2\sqrt{2}\sigma}\right) = \\ &= Q\left(\frac{d_E(\mathbf{s}(\mathbf{x}), \mathbf{s}(\mathbf{x}'))}{2\sigma}\right). \end{aligned}$$

Это значит, что граница (10) для этого случая совпадает с аддитивной границей, в которой каждое слагаемое вычисляется точно.

Если для модуляционного отображения  $\mathbf{s}(\mathbf{x})$  выполняется свойство  $\|\mathbf{s}(\mathbf{x}) - \mathbf{s}(\mathbf{x}')\|^2 = \|\mathbf{s}(\mathbf{0}) - \mathbf{s}(\mathbf{x}' - \mathbf{x})\|^2$ , а код обладает свойством линейности, т. е.  $\mathbf{x} - \mathbf{x}'$  принадлежит коду, то граница для вероятности  $P_e$ , которая может быть найдена согласно выражению (22), дает в точности тот же результат, что и граница (10), но вычисляется гораздо проще.

Далее рассматриваются примеры, которые не сводятся к ранее известным аддитивным границам.

**Пример 3.** Канал с гауссовым шумом, дисперсия которого зависит от переданного значения (Input dependent additive Gaussian noise — ID-AGN). Эта модель используется при анализе систем записи/считывания данных в многоуров-

невой флеш-памяти, беспроводных систем связи в видимом оптическом диапазоне, волоконно-оптических систем связи, систем связи с неортогональным множественным доступом [13–21].

Первым среди далее приводимых вариантов мы рассматриваем тот, который использовался для описания модели ячейки многоуровневой флеш-памяти (гауссова аппроксимация [13, 14]).

Входные уровни каждой ячейки принимают некоторые фиксированные значения  $x_0, x_1, \dots, x_{q-1}$ , а выходные значения представляют собой случайные величины. Распределения этих случайных величин описываются условными ф.п.в.  $p_{y|x}(y|x), -\infty < y < \infty, x = x_0, x_1, \dots, x_{q-1}$ .

В публикациях [13–15] перечисляются факторы, определяющие конкретный вид условных ф.п.в.  $p_{y|x}(y|x), -\infty < y < \infty, x = x_0, x_1, \dots, x_{q-1}$ . Будем полагать далее, что модель ячейки определяется ф.п.в.  $p_{y|x}(y|x_i), i = 0, 1, \dots, q-1$ , которые могут быть аппроксимированы гауссовыми плотностями, т. е. будем считать, что

$$p_{y|x}(y|x_i) = \frac{1}{\sqrt{2\pi}\sigma(x_i)} \exp\left(-\frac{(y-x_i)^2}{2\sigma^2(x_i)}\right).$$

Эта модель известна также как ID-AGN-модель [22]. Пример с  $q = 4$  часто используется для описания канала записи флеш-памяти, который в этом случае представлен как набор одиночных многоуровневых ячеек памяти. Эта модель допускает обобщение на большее число входных уровней, в частности на шесть, восемь и двенадцать уровней [22]. С возрастанием числа циклов перезаписи и увеличением времени хранения записанных данных значения уровней записи  $x_i$  (кроме  $x_0$ ) уменьшаются, тогда как значения стандартного отклонения  $\sigma(x_i)$  (кроме  $\sigma(x_0)$ ) увеличиваются. Такое изменение параметров  $x_i$  и  $\sigma(x_i)$ , задающих канал, описывает процесс его постепенного ухудшения (деградации) в зависимости от времени и условий эксплуатации. Детали описания зависимости величин  $x_i$  и  $\sigma(x_i), i = 0, 1, \dots, q-1$ , от числа циклов перезаписи и времени хранения записанных данных могут быть найдены в работах [13, 14, 23].

В этом случае можно показать, что при  $\sigma(x) \neq \sigma(x')$

$$c_z(\omega; x, x') = \frac{\sigma(x)^{j\omega} \sigma(x')^{1-j\omega}}{\sqrt{j\omega\sigma^2(x) + (1-j\omega)\sigma^2(x')}} \times \exp\left(-\frac{j\omega(1-j\omega)(x-x')^2}{2(j\omega\sigma^2(x) + (1-j\omega)\sigma^2(x'))}\right) \quad (30)$$

при условиях  $\text{Im}\omega < \sigma^2(x')/(\sigma^2(x) - \sigma^2(x'))$ , если  $\sigma(x) > \sigma(x')$ , и  $-\text{Im}\omega < \sigma^2(x')/(\sigma^2(x) - \sigma^2(x'))$ , если  $\sigma(x) < \sigma(x')$ . Граница для вероятности  $p_e$  вычис-

ляется для этой модели с использованием общего выражения (10). Ряд результатов, относящихся к каналу записи флеш-памяти и полученных на основе описанного здесь подхода, может быть найден в работах [10, 12].

Модель с ID-AGN, используемая для описания взаимосвязи входного сигнала  $x$  и выходного сигнала  $y$  в беспроводной системе связи в видимом оптическом диапазоне и волоконно-оптической системе связи, имеет следующий вид [16, 17]:

$$y = gx + \sqrt{k_0 + k_1 x} \xi_I + \xi, \quad (31)$$

где  $g$  — коэффициент передачи канала;  $k_0, k_1 > 0$  — коэффициенты, учитывающие вклад ID-AGN;  $\xi_I$  и  $\xi$  — независимые гауссовы случайные величины с нулевым средним и дисперсиями  $\sigma_I^2$  и  $\sigma$  соответственно. Первая составляющая  $\sqrt{k_0 + k_1 x} \xi_I$  аддитивного шума возникает вследствие случайности механизма эмиссии фотонов светодиоидом. Вторая составляющая  $\xi$  аддитивного шума представляет собой обычный тепловой шум. Передаваемый сигнал  $x$  в (31) представляет собой случайную величину, которая в АМ, простейшем варианте модуляции, характеризует интенсивность излучаемого оптического сигнала. Очевидно, при использовании АМ  $x$  является неотрицательной случайной величиной, ограниченной сверху максимальной интенсивностью светодиода. Значения х.ф.  $c_z(\omega; x, x')$  для этой модели вычисляются по формуле (30), в которой зависимость  $\sigma(x)$  имеет вид  $\sigma(x) = \sqrt{(k_0 + k_1 x)\sigma_I^2 + \sigma^2}$ .

**Пример 4.** Канал с релейскими замираниями и непрерывным выходом. В этой модели входные величины  $x \in \{x_0, x_1\}$ , а выход канала (выходное значение демодулятора с мягкими решениями) может быть представлен как  $y = y_0 - y_1$ , где

$$y_0 = \begin{cases} (\zeta_c \sqrt{2\gamma} + \xi_{c0})^2 + (\zeta_s \sqrt{2\gamma} + \xi_{s0})^2, & \text{если } x = x_0, \\ \xi_{c0}^2 + \xi_{s0}^2, & \text{если } x = x_1; \end{cases}$$

$$y_1 = \begin{cases} \xi_{c1}^2 + \xi_{s1}^2, & \text{если } x = x_0, \\ (\zeta_c \sqrt{2\gamma} + \xi_{c1})^2 + (\zeta_s \sqrt{2\gamma} + \xi_{s1})^2, & \text{если } x = x_1, \end{cases}$$

и  $\zeta_c, \zeta_s, \xi_{c0}, \xi_{s0}, \xi_{c1}, \xi_{s1}$  — независимые гауссовы случайные величины с нулевыми средними. Величины  $\xi_{c0}, \xi_{s0}, \xi_{c1}, \xi_{s1}$  возникают из-за влияния аддитивного белого гауссова шума, а  $\zeta_c$  и  $\zeta_s$  представляют собой квадратурные компоненты релейского коэффициента передачи канала.

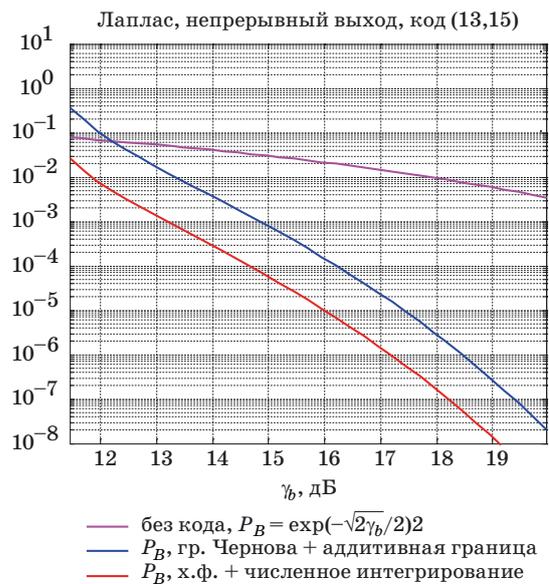
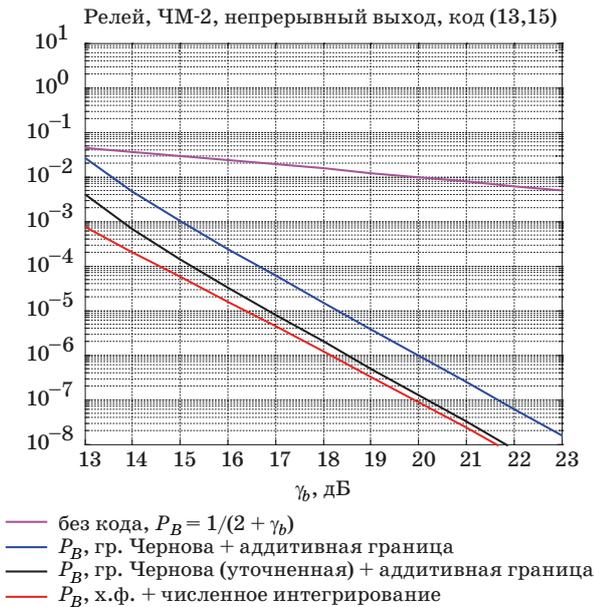
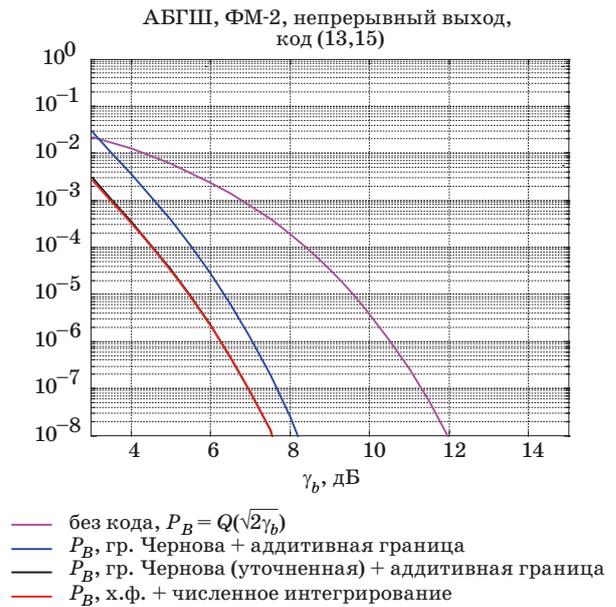
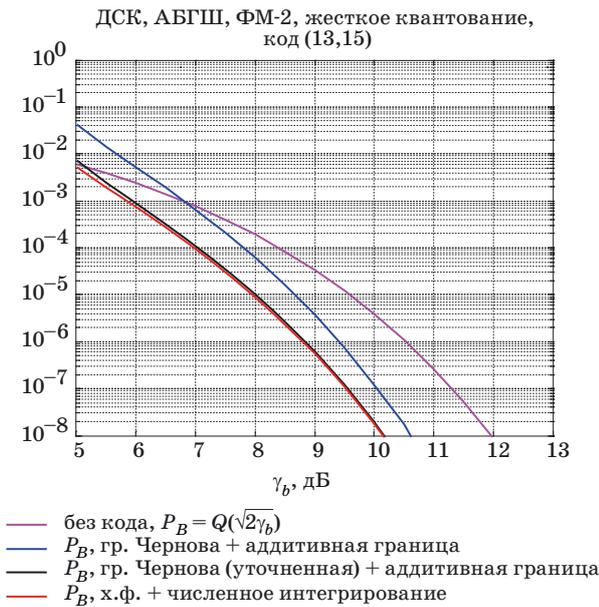
При соответствующей нормировке  $\zeta_c^2 = \zeta_s^2 = 1/2, \xi_{c0}^2 = \xi_{s0}^2 = \xi_{c1}^2 = \xi_{s1}^2 = 1$ , а величина  $\gamma$  имеет смысл среднего значения отношения сигнал/шум. Такая модель возникает при рассмотрении некоге-

рентного приема ортогональных сигналов ЧМ-2, переданных по каналу с релеевскими замираниями. Для условных ф.п.в. в этом случае справедливости выражения

$$P_{y|x}(y|x_0) = \begin{cases} \frac{1}{2(\gamma+2)} \exp\left(-\frac{y}{2(\gamma+1)}\right), & \text{если } y > 0, \\ \frac{1}{2(\gamma+2)} \exp\left(\frac{y}{2}\right), & \text{если } y \leq 0; \end{cases} \quad (32)$$

$$P_{y|x}(y|x_1) = P_{y|x}(-y|x_0) = \begin{cases} \frac{1}{2(\gamma+2)} \exp\left(-\frac{y}{2}\right), & \text{если } y \geq 0, \\ \frac{1}{2(\gamma+2)} \exp\left(\frac{y}{2(\gamma+1)}\right), & \text{если } y < 0. \end{cases} \quad (33)$$

Подстановка выражений (32) и (33) в (8) приводит в итоге к выражению



■ **Рис. 4.** Границы вероятности ошибки на бит  $P_B$  в зависимости от отношения сигнал/шум на бит  $\gamma_b$  для рассмотренных моделей каналов

■ **Fig. 4.** Bounds on the bit error probability  $P_B$  vs signal-to-noise ratio per bit  $\gamma_b$  for the considered channel models

$$c_z(\omega; x_0, x_1) = \frac{\gamma + 1}{\gamma + 2} \left( \frac{1}{1 + (1 - j\omega)\gamma} + \frac{1}{1 + j\omega\gamma} \right), \quad (34)$$

при этом должны выполняться условия  $1/\gamma > \text{Im}\omega > -(1 + \gamma)/\gamma$ . В дальнейшем потребуется значение правой части (34) при  $\omega = \alpha - j\beta$ . Из условий для  $\text{Im}\omega$  следует, что  $-1/\gamma < \beta < (1 + \gamma)/\gamma$ . Кроме того, имеется ограничение  $\beta > 0$ . Поэтому окончательно имеем границы для параметра  $\beta$ :  $0 < \beta < (1 + \gamma)/\gamma$ . Тогда для  $\beta = 1/2$  получаем

$$c_z(\alpha - j/2; x_0, x_1) = \frac{4(\gamma + 1)}{(\gamma + 2)^2 + 4\alpha^2\gamma^2}. \quad (35)$$

**Пример 5.** Канал с аддитивным шумом с распределением Лапласа. Эта модель является частным случаем модели канала с аддитивным шумом, имеющим обобщенное гауссово распределение. Аддитивный шум, имеющий распределение Лапласа, широко используется как а) модель интерференционного шума в системах связи миллиметрового диапазона; б) модель интерференционного шума в дуплексных радиосетях при наличии замираний; в) модель шума в оптических линиях [19–21]. В этой модели предполагается, что  $x \in \{x_0, x_1\}$ ,  $x_0, x_1 \in \mathbb{R}$ , и условные ф.п.в. задаются равенствами  $p_{y|x}(y|x_0) = \exp(-|y - x_0|/b)/(2b)$  и  $p_{y|x}(y|x_1) = \exp(-|y - x_1|/b)/(2b)$ , где  $b = \sigma/\sqrt{2}$ ,  $\sigma$  — среднеквадратическое отклонение. Подстановка этих выражений для ф.п.в. приводит после ряда преобразований к следующему выражению для х.ф.:

$$c_z(\omega; x_0, x_1) = \left( (1 - j\omega)e^{-j\omega|x_0 - x_1|/b} - j\omega e^{-(1 - j\omega)|x_0 - x_1|/b} \right) / (1 - 2j\omega).$$

Более удобным для практических вычислений оказывается функция  $c_z(\omega; x, x')|_{\omega = \alpha - j/2}$ , которая может быть приведена к виду

$$c_z(\alpha - j/2; x_0, x_1) = e^{-\frac{|x_0 - x_1|}{2b}} \left( \frac{\sin(\alpha|x_0 - x_1|/b)}{2\alpha} + \cos(\alpha(x_0 - x_1)/b) \right). \quad (36)$$

В приведенных ниже численных примерах под значением отношения сигнал/шум понимается величина  $|x_0 - x_1|/b$ .

**Численные результаты.** В качестве иллюстративного примера рассмотрим применение разработанного подхода для двоичного сверточного кода с  $R = 1/2$  с многочленной порождающей матрицей  $\mathbf{G}(z) = [1 + z + z^3 \quad 1 + z^2 + z^3]$ , или в восьмеричной компактной записи  $\mathbf{G}(z) = [13, 15]$ . Этот код имеет кодовое ограничение  $v = 3$ ,

свободное расстояние  $d_f = 6$  и функцию нумератора весов

$$T(D, N) = \frac{D^6 N^3 - D^{10} N^3 + D^6 N}{1 - 2D^2 N^3 - 3D^2 N + 2D^6 N^3}.$$

На рис. 4 представлены графики границы вероятности  $P_B$ , вычисленные согласно выражению (23) для рассмотренных примеров каналов, исключая модель с ID-AGN [см. (24), (26), (34) и (36)]. Результаты, относящиеся к модели с ID-AGN, полученные с использованием описанного здесь подхода, могут быть найдены в работах [10, 12].

Из приведенных графиков следует, что предложенный универсальный подход (красная линия) дает существенное улучшение границы вероятности ошибки для конкретных кодов по сравнению с обычной аддитивной границей в комбинации с границей Чернова (синяя линия): так, при вероятности ошибки  $10^{-5}$  оценка требуемых энергетических затрат снижается на величину от 0,7 до 2,0 дБ в зависимости от модели канала. Что касается уточненных вариантов границы Чернова [4] (черная линия), то предложенный подход тоже улучшает известные границы для этих моделей, хотя и очень незначительно. Заметим, что важнейшая положительная черта предложенного подхода состоит в его универсальности, что позволяет применить его к любому каналу без памяти и обеспечить при этом высокую точность оценивания при малых вычислительных затратах.

## Заключение

В данной работе представлен метод точного вычисления аддитивной границы вероятности ошибочного декодирования для широкого класса кодов и моделей каналов без памяти. Разработанный метод базируется на получении точного выражения для попарной вероятности ошибки с использованием аппарата характеристических функций. Получение аддитивной границы вероятности ошибочного декодирования основано на использовании произведения решетчатых диаграмм для рассматриваемого кода, позволяющего получить итоговое выражение в интегральной форме, удобной для численного интегрирования. Важной особенностью предложенной процедуры является то, что она позволяет точно вычислить аддитивную границу с помощью подхода, основанного на использовании передаточных (производящих) функций. При таком подходе метки ветвей в произведении решетчатых диаграмм для рассматриваемого кода заменяются соответствующими характеристическими функциями.

ческими функциями. Удобное для численного интегрирования итоговое выражение позволяет вычислять значения аддитивной границы вероятности ошибочного декодирования с желаемой точностью. В целом, как показывают конкретные примеры, вычислительные затраты при использовании предложенного подхода к вычислению аддитивной границы вероятности ошибочного декодирования оказываются вполне приемлемыми для кодов, имеющих сравнительно небольшую длину.

Результаты были частично представлены на XXIV международной конференции «Волновая электроника и инфокоммуникационные систе-

мы (WECONF-2021)», Санкт-Петербург, ГУАП, 2021 г.

### Финансовая поддержка

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации, соглашение № FSRF-2020-0004 «Научные основы построения архитектур и систем связи бортовых информационно-вычислительных комплексов нового поколения для авиационных, космических систем и беспилотных транспортных средств».

### Литература

1. Proakis J. G., M. Salehi M. *Digital Communications*. Fifth Ed. McGraw-Hill, 2008. 1151 p.
2. Gallager R. G. *Principles of Digital Communications*. Cambridge University Press, 2012. 407 p.
3. I. Sason I., Shamai S. Performance analysis of linear codes under maximum-likelihood decoding: A tutorial. *Foundations and Trends in Communications and Information Theory*, 2006, vol. 3, no. 1–2, pp. 1–225.
4. Trofimov A. N. Modified Chernoff Bound and Some Applications. In: *Modulation and Coding Techniques in Wireless Communications*. Ed. E. Krouk and S. Semenov. Wiley, 2011. Pp. 206–220.
5. Biglieri E. *Coding for Wireless Channels*. Springer US, 2006. 420 p.
6. Goldsmith A. *Wireless Communications*. Cambridge University Press, 2012. 644 p.
7. Abedi A., Khandani A. K. An analytical method for approximate performance evaluation of binary linear block codes. *IEEE Transactions on Communications*, 2004, vol. 52, no. 2, pp. 228–235.
8. Ali S. A., Kambo N. S., Ince E. A. Exact expression and tight bound on pairwise error probability for performance analysis of turbo codes over Nakagami-m fading channels. *IEEE Communications Letters*, 2007, vol. 11, no. 5, pp. 399–401.
9. Gupta B., Saini D. S. Moment generating function-based pairwise error probability analysis of concatenated low density parity check codes with Alamouti coded multiple input multiple output orthogonal frequency division multiplexing systems. *Communications IET*, 2014, vol. 8, no. 3, pp. 399–412.
10. Таубин Ф. А., Трофимов А. Н. Каскадное кодирование на основе многомерных решеток и кодов Рида — Соломона для многоуровневой флэш-памяти. *Труды СПИИРАН*, 2018, вып. 2(57), с. 75–103. doi:10.15622/sp.57.4
11. Vardy A. Trellis Structures of Codes. In: *Handbook of Coding Theory*. Ed. V. S. Pless and W. C. Huffman. Elsevier, Amsterdam, 1998. Vol. 2. 1106 p.
12. Таубин Ф. А., Трофимов А. Н. Каскадное кодирование для многоуровневой флэш-памяти с исправлением ошибок малой кратности во внешней ступени. *Труды СПИИРАН*, 2019, вып. 18(5), с. 1149–1181. doi:10.15622/sp.2019.18.5.1149-1181
13. Wang X., Dong G., Pan L., Zhou R. Error correction codes and signal processing in flash memory. In: *Flash Memories*. Ed. Igor Stievano. <http://www.intechopen.com/books/flash-memories/error-correction-codes-and-signal-processing-in-flash-memory> (дата обращения: 26.10.2014).
14. Dong G., Pan Y., Xie N., Varanasi C., Zhang T. Estimating information-theoretical NAND flash memory storage capacity and its implication to memory system design space exploration. *IEEE Trans. Very Large Scale Integration (VLSI) Systems*, 2012, vol. 20, no. 9, pp. 1705–1714.
15. Huang X., Kavcic A., Ma X., Dong G., Zhang T. Multilevel flash memories: Channel modeling, capacities and optimal coding rates. *International Journal on Advances in Systems and Measurement*, 2013, vol. 6, no. 3, 4, pp. 364–373. [http://www.iariajournals.org/systems\\_and\\_measurements/](http://www.iariajournals.org/systems_and_measurements/) (дата обращения: 26.10.2014).
16. Moser S. M. Capacity results of an optical intensity channel with input-dependent Gaussian noise. *IEEE Transactions on Information Theory*, 2012, vol. 58, no. 1, pp. 207–223.
17. Yuan M., Sha X., Liang X., Jiang M., Wang J., Zhao C. Coding performance for signal dependent channels in visible light communication system. *Signal and Information Processing (GlobalSIP) 2015 IEEE Global Conference*, pp. 1037–1041.
18. Gao Q., Hu S., Gong C., Xu Z. Modulation design for multi-carrier visible light communications with signal-dependent noise. *Communication Systems (ICCS) 2016 IEEE International Conference*, 2016, pp. 1–6.
19. Dytso A., Bustin R., Poor H. V., Shitz S. S. On additive channels with generalized Gaussian noise. *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, 2017, pp. 426–430.

20. Ding Z., Yang Z., Fan P., Poor H. V. On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Process. Letters*, 2014, vol. 21, no. 12, pp. 1501–1505.
21. Bariah L., Muhaidat S., Sofotasios P. C., Gurugopinath S., Hamouda W., Yanikomeroğlu H. Non-orthogonal multiple access in the presence of additive generalized Gaussian noise. *IEEE Communications Letters*, 2020, vol. 24, no. 10, pp. 2137–2141.
22. Sun F., Rose K., Zhang T. On the use of strong BCH codes for improving multilevel NAND flash memory storage capacity. [http://www.researchgate.net/publication/254376882\\_On\\_the\\_Use\\_of\\_Strong\\_BCH\\_Codes\\_for\\_Improving\\_Multilevel\\_NAND\\_Flash\\_Memory\\_Storage\\_Capacity](http://www.researchgate.net/publication/254376882_On_the_Use_of_Strong_BCH_Codes_for_Improving_Multilevel_NAND_Flash_Memory_Storage_Capacity) (дата обращения: 26.10.2014).
23. Трофимов А. Н., Таубин Ф. А. Теоретико-информационный анализ многоуровневой flash-памяти. Часть 1: Модель канала и границы случайного кодирования. *Информационно-управляющие системы*, 2016, № 2, с. 56–67. doi.org/10.15217/issn1684-8853.2016.2.56

UDC 621.391

doi:10.31799/1684-8853-2021-4-71-85

**Evaluation of the union bound for the decoding error probability using characteristic functions**A. N. Trofimov<sup>a</sup>, PhD, Tech., Associate Professor, orcid.org//0000-0003-1233-5222, andrei.trofimov@k36.orgF. A. Taubin<sup>a</sup>, Dr. Sc., Tech., Professor, orcid.org/0000-0002-8781-9531<sup>a</sup>Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation

**Introduction:** Since the exact value of a decoding error probability cannot usually be calculated, an upper bounding technique is used. The standard approach for obtaining the upper bound on the maximum likelihood decoding error probability is based on the use of the union bound and the Chernoff bound, as well as its modifications. For many situations, this approach is not accurate enough. **Purpose:** Development of a method for exact calculation of the union bound for a decoding error probability, for a wide class of codes and memoryless channels. **Methods:** Use of characteristic functions of logarithm of the likelihood ratio for an arbitrary pair of codewords, trellis representation of codes and numerical integration. **Results:** The resulting exact union bound on the decoding error probability is based on a combination of the use of characteristic functions and the product of trellis diagrams for the code, which allows to obtain the final expression in an integral form convenient for numerical integration. An important feature of the proposed procedure is that it allows one to accurately calculate the union bound using an approach based on the use of transfer (generating) functions. With this approach, the edge labels in the product of trellis diagrams for the code are replaced by their corresponding characteristic functions. The final expression allows, using the standard methods of numerical integration, to calculate the values of the union bound on the decoding error probability with the required accuracy. **Practical relevance:** The results presented in this article make it possible to significantly improve the accuracy of the bound of the error decoding probability, and thereby increase the efficiency of technical solutions in the design of specific coding schemes for a wide class of communication channels.

**Keywords** — maximum likelihood decoding, union bound on the error probability, pairwise error probability, characteristic function, ID-AGN channel, additive Laplacian noise channel.

**For citation:** Trofimov A. N., Taubin F. A. Evaluation of the union bound for the decoding error probability using characteristic functions. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 4, pp. 71–85 (In Russian). doi:10.31799/1684-8853-2021-4-71-85

**References**

- Proakis J. G., M. Salehi M. *Digital Communications*. Fifth Ed. McGraw-Hill, 2008. 1151 p.
- Gallager R. G. *Principles of Digital Communications*. Cambridge University Press, 2012. 407 p.
- I. Sason I., Shamai S. Performance analysis of linear codes under maximum-likelihood decoding: A tutorial. *Foundations and Trends in Communications and Information Theory*, 2006, vol. 3, no. 1–2, pp. 1–225.
- Trofimov A. N. *Modified Chernoff Bound and Some Applications*. In: *Modulation and Coding Techniques in Wireless Communications*. Ed. E. Krouk and S. Semenov. Wiley, 2011. Pp. 206–220.
- Biglieri E. *Coding for Wireless Channels*. Springer US, 2006. 420 p.
- Goldsmith A. *Wireless Communications*. Cambridge University Press, 2012. 644 p.
- Abedi A., Khandani A. K. An analytical method for approximate performance evaluation of binary linear block codes. *IEEE Transactions on Communications*, 2004, vol. 52, no. 2, pp. 228–235.
- Ali S. A., Kambo N. S., Ince E. A Exact expression and tight bound on pairwise error probability for performance analysis of turbo codes over Nakagami-m fading channels. *IEEE Communications Letters*, 2007, vol. 11, no. 5, pp. 399–401.
- Gupta B., Saini D. S. Moment generating function-based pairwise error probability analysis of concatenated low density parity check codes with alamouti coded multiple input multiple output orthogonal frequency division multiplexing systems. *Communications IET*, 2014, vol. 8, no. 3, pp. 399–412.
- Taubin F. A., Trofimov A. N. Concatenated Reed–Solomon/lattice coding for multilevel flash memory. *SPIIRAS Proceedings*, 2018, no. 2(57), pp. 75–101 (In Russian). doi: 10.15622/sp.57.4
- Vardy A. *Trellis Structures of Codes*. In: *Handbook of Coding Theory*. Ed. V. S. Pless and W. C. Huffman. Elsevier, Amsterdam, 1998. Vol. 2. 1106 p.
- Taubin F. A., Trofimov A. N. Concatenated coding for multilevel flash memory with low error correction capabilities in outer stage. *SPIIRAS Proceedings*, 2019, vol. 18, no. 5, pp. 1149–1181 (In Russian). doi:10.15622/sp.2019.18.5.1149-1181
- Wang X., Dong G., Pan L., Zhou R. *Error correction codes and signal processing in flash memor*. In: *Flash Memories*. Ed. Igor Stievano. Available at: <http://www.intechopen.com/books/flash-memories/error-correction-codes-and-signal-processing-in-flash-memory> (accessed 26 October 2014).

14. Dong G., Pan Y., Xie N., Varanasi C., Zhang T. Estimating information-theoretical NAND flash memory storage capacity and its implication to memory system design space exploration. *IEEE Trans. Very Large Scale Integration (VLSI) Systems*, 2012, vol. 20, no. 9, pp. 1705–1714.
15. Huang X., Kavcic A., Ma X., Dong G., Zhang T. Multilevel flash memories: Channel modeling, capacities and optimal coding rates. *International Journal on Advances in Systems and Measurement*, 2013, vol. 6, no. 3, 4, pp. 364–373. Available at: [http://www.iariajournals.org/systems\\_and\\_measurements/](http://www.iariajournals.org/systems_and_measurements/) (accessed 26 October 2014).
16. Moser S. M. Capacity results of an optical intensity channel with input-dependent Gaussian noise. *IEEE Transactions on Information Theory*, 2012, vol. 58, no. 1, pp. 207–223.
17. Yuan M., Sha X., Liang X., Jiang M., Wang J., Zhao C. Coding performance for signal dependent channels in visible light communication system. *Signal and Information Processing (GlobalSIP) 2015 IEEE Global Conference*, pp. 1037–1041.
18. Gao Q., Hu S., Gong C., Xu Z. Modulation design for multi-carrier visible light communications with signal-dependent noise. *Communication Systems (ICCS) 2016 IEEE International Conference*, 2016, pp. 1–6.
19. Dytso A., Bustin R., Poor H. V., Shitz S. S. On additive channels with generalized Gaussian noise. *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, 2017, pp. 426–430.
20. Ding Z., Yang Z., Fan P., Poor H. V. On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Process. Letters*, 2014, vol. 21, no. 12, pp. 1501–1505.
21. Bariah L., Muhaidat S., Sofotasios P. C., Gurugopinath S., Hamouda W. Yanikomeroglu H. Non-orthogonal multiple access in the presence of additive generalized Gaussian noise. *IEEE Communications Letters*, 2020, vol. 24, no. 10, pp. 2137–2141.
22. Sun F., Rose K., Zhang T. On the use of strong BCH codes for improving multilevel NAND flash memory storage capacity. Available at: [http://www.researchgate.net/publication/254376882\\_On\\_the\\_Use\\_of\\_Strong\\_BCH\\_Codes\\_for\\_Improving\\_Multilevel\\_NAND\\_Flash\\_Memory\\_Storage\\_Capacity](http://www.researchgate.net/publication/254376882_On_the_Use_of_Strong_BCH_Codes_for_Improving_Multilevel_NAND_Flash_Memory_Storage_Capacity) (accessed 26 October 2014).
23. Trofimov A. N., Taubin F. A. Information theory analysis of multilevel flash memory. Part 1: Channel model and random coding bounds. *Informatsionno-upravliaiushchie sistemy [Information and Control Systems]*, 2016, no. 2, pp. 56–67 (In Russian). doi.org/10.15217/issn1684-8853.2016.2.56

## ПАМЯТКА ДЛЯ АВТОРОВ

*Поступающие в редакцию статьи проходят обязательное рецензирование.*

При наличии положительной рецензии статья рассматривается редакционной коллегией. Принятая в печать статья направляется автору для согласования редакторских правок. После согласования автор представляет в редакцию окончательный вариант текста статьи.

Процедуры согласования текста статьи могут осуществляться как непосредственно в редакции, так и по e-mail ([ius.spb@gmail.com](mailto:ius.spb@gmail.com)).

При отклонении статьи редакция представляет автору мотивированное заключение и рецензию, при необходимости доработать статью — рецензию.

*Редакция журнала напоминает, что ответственность за достоверность и точность рекламных материалов несут рекламодатели.*

### **АЛЕКСАНДРОВ Игорь Васильевич**



Аспирант кафедры программной инженерии и вычислительной техники Санкт-Петербургского государственного университета телекоммуникаций им. проф. М. А. Бонч-Бруевича.

В 2014 году окончил Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича по специальности «Программное обеспечение вычислительной техники и автоматизированных систем».

Является автором десяти научных публикаций и одного свидетельства об интеллектуальной собственности.

Область научных интересов — системный анализ, управление и обработка информации, распознавание образов, индексация данных, параллельные вычисления.

Эл. адрес:  
chronos555@yandex.ru

### **АФАНАСЬЕВ Антон Андреевич**



Аспирант Космического центра Сколковского института науки и технологий.

В 2017 году окончил Московский физико-технический институт (Государственный университет) по специальности «Прикладная математика и физика» и Сколковский институт науки и технологий по специальности «Информационные системы и технологии».

Является автором пяти научных публикаций.

Область научных интересов — малые космические аппараты, теория управления, магнитные системы управления, статистические методы.

Эл. адрес:  
anton.afanasev@skolkovotech.ru

### **БАЛОНИН Николай Алексеевич**



Профессор кафедры вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения.

В 1982 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Автоматика и телемеханика».

В 2008 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 100 научных публикаций, в том числе трех монографий.

Область научных интересов — теория динамических систем, теория идентификации, теория операторов, теория матриц, вычислительные методы, интернет-робототехника, интернет-книги с исполняемыми алгоритмами, научные социальные сети.

Эл. адрес: korbendfs@mail.ru

### **БИКТИМІРОВ Шамиль Насимович**



Аспирант Сколковского института науки и технологий.

В 2018 году окончил магистратуру Сколковского института науки и технологий по специальности «Информационные системы и технологии».

Является автором девяти научных публикаций.

Область научных интересов — орбитальная механика, численные симуляции, аэрокосмические исследования, кибернетика 2.0.

Эл. адрес:  
shamil.biktimirov@skoltech.ru

### **ЛЕВАШОВА Татьяна Викторовна**



Старший научный сотрудник лаборатории интегрированных систем автоматизации Санкт-Петербургского института информатики и автоматизации РАН.

В 1986 году окончила Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Электронные вычислительные машины».

В 2009 году защитила диссертацию на соискание ученой степени кандидата технических наук. Является автором более 200 научных публикаций.

Область научных интересов — управление знаниями, инженерия онтологий, управление онтологиями, управление контекстом, социоконвергентные системы.

Эл. адрес:  
tatiana.levashova@iias.spb.ru

### **МЕЛЕХИН Владимир Борисович**



Профессор кафедры программного обеспечения вычислительной техники и автоматизированных систем Дагестанского государственного технического университета, Махачкала. Почетный работник науки и техники Российской Федерации.

В 1976 году окончил Дагестанский политехнический институт по специальности «Автоматика и телемеханика».

В 1996 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором 213 научных публикаций, в том числе пяти монографий.

Область научных интересов — искусственный интеллект, модели представления и обработки знаний автономных интеллектуальных мобильных систем, управление экономическими системами.

Эл. адрес:  
pashka160@rambler.ru

**ПАРАСИЧ  
Андрей  
Викторович**



Инженер-программист ООО «ТРИДИВИ», Челябинск. В 2014 году окончил Южно-Уральский государственный университет по специальности «Вычислительные машины, комплексы, системы и сети». Является автором десяти научных публикаций. Область научных интересов — машинное обучение, компьютерное зрение, распознавание позы человека, распознавание лиц, распознавание по радужной оболочке глаза. Эл. адрес: parasichav@yandex.ru

**ПАРАСИЧ  
Виктор  
Александрович**



Доцент кафедры электронных вычислительных машин Южно-Уральского государственного университета, Челябинск. В 1976 году окончил Южно-Уральский государственный университет по специальности «Электронные вычислительные машины». В 1987 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 23 научных публикаций. Область научных интересов — машинное обучение, программирование. Эл. адрес: pva16@yandex.ru

**ПАРАСИЧ  
Ирина  
Васильевна**



Доцент кафедры математического и компьютерного моделирования Южно-Уральского государственного университета, Челябинск. В 1981 году окончила Южно-Уральский государственный университет по специальности «Прикладная математика». В 1991 году защитила диссертацию на соискание ученой степени кандидата технических наук. Является автором 26 научных публикаций и одного патента на изобретение. Область научных интересов — машинное обучение, многомерный статистический анализ. Эл. адрес: parasichiv@susu.ru

**ПЕТРОВ  
Михаил  
Владимирович**



Научный сотрудник лаборатории интегрированных систем автоматизации Санкт-Петербургского федерального исследовательского центра РАН. В 2017 году окончил магистратуру Университета ИТМО по специальности «Бизнес-информатика». Является автором 17 научных публикаций. Область научных интересов — управление компетенциями, информационных технологии. Эл. адрес: mikhaill.petrov@iias.spb.su

**СЕРГЕЕВ  
Александр  
Михайлович**



Доцент кафедры вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения. В 2004 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Вычислительные машины, комплексы, системы и сети». В 2020 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 35 научных публикаций. Область научных интересов — численные методы, теория вычислительных процессов, проектирование специализированных процессоров. Эл. адрес: asklab@mail.ru

**СИНИЦЫНА  
Ольга  
Игоревна**



Аспирант кафедры вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения. В 2008 году окончила Санкт-Петербургский университет телекоммуникаций им. проф. М. А. Бонч-Бруевича по специальности «Автоматизация технологических процессов и производств». Является автором двух научных публикаций. Область научных интересов — теория динамических систем, теория матриц, вычислительные методы. Эл. адрес: libra18@yandex.ru

**СМИРНОВ  
Александр  
Викторович**



Профессор, главный научный сотрудник, заведующий лабораторией интегрированных систем автоматизации Санкт-Петербургского института информатики и автоматизации РАН.

В 1979 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Системы автоматического управления».

В 1994 году защитил диссертацию на соискание ученой степени доктора технических наук. Является автором более 350 научных публикаций.

Область научных интересов — концептуальное моделирование, управление знаниями, управление контекстом, конфигурирование систем, системы поддержки принятия решений, социоконвергентные системы.

Эл. адрес: [smir@iias.spb.su](mailto:smir@iias.spb.su)

**ТАУБИН  
Феликс  
Александрович**



Профессор кафедры информационных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения.

В 1968 году окончил Ленинградский институт авиационного приборостроения по специальности «Радиоэлектронные устройства систем управления».

В 1992 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 80 научных публикаций, девяти патентов и двух изобретений.

Область научных интересов — цифровые системы связи, методы помехоустойчивого кодирования, широкополосные системы, беспроводные сети.

Эл. адрес: [ftaubin@yahoo.com](mailto:ftaubin@yahoo.com)

**ТРОФИМОВ  
Андрей  
Николаевич**



Доцент кафедры инфокоммуникационных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения.

В 1976 году окончил Ленинградский институт авиационного приборостроения по специальности «Автоматизированные системы управления».

В 1982 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 41 научной публикации и шести патентов на изобретения.

Область научных интересов — теория передачи дискретных сообщений, теория информации, теория кодирования.

Эл. адрес: [andrei.trofimov@vu.spb.ru](mailto:andrei.trofimov@vu.spb.ru)

**ФОМИН  
Владимир  
Владимирович**



Профессор кафедры информационных систем Российского государственного педагогического университета им. А. И. Герцена, Санкт-Петербург.

В 1988 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Автоматизированные системы управления».

В 2001 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором 74 научных публикаций и десяти свидетельств об интеллектуальной собственности.

Область научных интересов — системная инженерия, проектирование информационных систем, технологии интеллектуального анализа данных, семитическое моделирование.

Эл. адрес: [v\\_v\\_fomin@mail.ru](mailto:v_v_fomin@mail.ru)

**ХАЧУМОВ  
Михаил  
Вячеславович**



Старший научный сотрудник Института программных систем им. А. К. Айламазяна РАН, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН, Москва.

В 2009 году окончил Российский университет дружбы народов по специальности «Прикладная математика и информатика».

В 2012 году защитил диссертацию на соискание ученой степени кандидата физико-математических наук.

Является автором более 120 научных публикаций, в том числе двух монографий.

Область научных интересов — искусственный интеллект, интеллектуальные системы управления, модели представления и обработки знаний интеллектуальных мобильных систем.

Эл. адрес: [khmike@inbox.ru](mailto:khmike@inbox.ru)