

ISSN 2305-9052 (Print)  
ISSN 2410-7034 (Online)

# ВЕСТНИК

ЮЖНО-УРАЛЬСКОГО  
ГОСУДАРСТВЕННОГО  
УНИВЕРСИТЕТА

# BULLETIN

OF THE SOUTH URAL  
STATE UNIVERSITY

**СЕРИЯ**

**ВЫЧИСЛИТЕЛЬНАЯ  
МАТЕМАТИКА  
И ИНФОРМАТИКА**

**2022, том 11, № 2**

**SERIES**

**COMPUTATIONAL  
MATHEMATICS  
AND SOFTWARE ENGINEERING**

**2022, volume 11, no. 2**



# ВЕСТНИК



ЮЖНО-УРАЛЬСКОГО  
ГОСУДАРСТВЕННОГО  
УНИВЕРСИТЕТА

2022  
Т. 11, № 2

ISSN 2305-9052 (Print)  
ISSN 2410-7034 (Online)

СЕРИЯ

## «ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА И ИНФОРМАТИКА»

Решением ВАК включен в Перечень научных изданий,  
в которых должны быть опубликованы результаты диссертаций  
на соискание ученых степеней кандидата и доктора наук

Учредитель — Федеральное государственное автономное образовательное учреждение  
высшего образования «Южно-Уральский государственный университет  
(национальный исследовательский университет)»

Тематика журнала:

- Вычислительная математика и численные методы
- Математическое программирование
- Распознавание образов
- Вычислительные методы линейной алгебры
- Решение обратных и некорректно поставленных задач
- Доказательные вычисления
- Численное решение дифференциальных и интегральных уравнений
- Исследование операций
- Теория игр
- Теория аппроксимации
- Информатика
- Искусственный интеллект и машинное обучение
- Системное программирование
- Перспективные многопроцессорные архитектуры
- Облачные вычисления
- Технология программирования
- Машинная графика
- Интернет-технологии
- Системы электронного обучения
- Технологии обработки баз данных и знаний
- Интеллектуальный анализ данных

### Редакционная коллегия

**Л.Б. Соколинский**, д.ф.-м.н., проф., *гл. редактор*  
**В.П. Танана**, д.ф.-м.н., проф., *зам. гл. редактора*  
**М.Л. Цымблер**, д.ф.-м.н., доц., *отв. секретарь*  
**Г.И. Радченко**, к.ф.-м.н., доц. (Австрия)  
**Я.А. Краева**, *техн. секретарь*

### Редакционный совет

**С.М. Абдуллаев**, д.г.н., профессор  
**А. Андряк**, PhD, профессор (Германия)  
**В.И. Бердышев**, д.ф.-м.н., акад. РАН, *председатель*  
**В.В. Воеводин**, д.ф.-м.н., чл.-кор. РАН

**Дж. Донгарра**, PhD, профессор (США)  
**С.В. Зыкин**, д.т.н., профессор  
**И.М. Куликов**, д.ф.-м.н.  
**Д. Маллманн**, PhD, профессор (Германия)  
**А.В. Панюков**, д.ф.-м.н., профессор  
**Р. Продан**, PhD, профессор (Австрия)  
**В.И. Ухоботов**, д.ф.-м.н., профессор  
**В.Н. Ушаков**, д.ф.-м.н., чл.-кор. РАН  
**М.Ю. Хачай**, д.ф.-м.н., профессор  
**А. Черных**, PhD, профессор (Мексика)  
**П. Шумяцкий**, PhD, профессор (Бразилия)



# BULLETIN

OF THE SOUTH URAL  
STATE UNIVERSITY

2022

Vol. 11, no. 2

SERIES

“COMPUTATIONAL  
MATHEMATICS AND SOFTWARE  
ENGINEERING”

ISSN 2305-9052 (Print)  
ISSN 2410-7034 (Online)

---

Vestnik Yuzhno-Ural'skogo Gosudarstvennogo Universiteta.  
Seriya “Vychislitel'naya Matematika i Informatika”

---

## South Ural State University

The scope of the journal:

- Numerical analysis and methods
- Mathematical optimization
- Pattern recognition
- Numerical methods of linear algebra
- Reverse and ill-posed problems solution
- Computer-assisted proofs
- Numerical solutions of differential and integral equations
- Operations research
- Game theory
- Approximation theory
- Computer science
- Artificial intelligence and machine learning
- System software
- Advanced multiprocessor architectures
- Cloud computing
- Software engineering
- Computer graphics
- Internet technologies
- E-learning
- Database processing
- Data mining

### Editorial Board

**L.B. Sokolinsky**, South Ural State University (Chelyabinsk, Russia)  
**V.P. Tanana**, South Ural State University (Chelyabinsk, Russia)  
**M.L. Zymbler**, South Ural State University (Chelyabinsk, Russia)  
**G.I. Radchenko**, Silicon Austria Labs (Graz, Austria)  
**Ya.A. Kraeva**, South Ural State University (Chelyabinsk, Russia)

### Editorial Council

**S.M. Abdullaev**, South Ural State University (Chelyabinsk, Russia)  
**A. Andrzejak**, Heidelberg University (Germany)  
**V.I. Berdyshev**, Institute of Mathematics and Mechanics, Ural Branch of the RAS (Yekaterinburg, Russia)  
**J. Dongarra**, University of Tennessee (USA)  
**M.Yu. Khachay**, Institute of Mathematics and Mechanics, Ural Branch of the RAS (Yekaterinburg, Russia)  
**I.M. Kulikov**, Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of RAS (Novosibirsk, Russia)  
**D. Mallmann**, Julich Supercomputing Centre (Germany)  
**A.V. Panyukov**, South Ural State University (Chelyabinsk, Russia)  
**R. Prodan**, Alpen-Adria-Universität Klagenfurt (Austria)  
**P. Shumyatsky**, University of Brasilia (Brazil)  
**A. Tchernykh**, CICESE Research Center (Mexico)  
**V.I. Ukhobotov**, Chelyabinsk State University (Chelyabinsk, Russia)  
**V.N. Ushakov**, Institute of Mathematics and Mechanics, Ural Branch of the RAS (Yekaterinburg, Russia)  
**V.V. Voevodin**, Lomonosov Moscow State University (Moscow, Russia)  
**S.V. Zykin**, Sobolev Institute of Mathematics, Siberian Branch of the RAS (Omsk, Russia)

## Содержание

РЕСУРСОНЕЗАВИСИМОЕ ОПИСАНИЕ ИНФОРМАЦИОННЫХ ГРАФОВ С ДИСТРИБУТИВНЫМИ ОПЕРАЦИЯМИ НА ЯЗЫКЕ ПРОГРАММИРОВАНИЯ SET@L И.И. Левин, И.В. Писаренко, Д.В. Михайлов, А.К. Мельников, А.И. Дордопуло .....	5
СРАВНЕНИЕ ЦЕЛЕВЫХ ФУНКЦИЙ В ЗАДАЧЕ ПРОНИ ДЛЯ АППРОКСИМАЦИИ ДАННЫХ ИЗМЕРЕНИЙ А.А. Ломов, Е.А. Русинова .....	18
ПРИМЕНЕНИЕ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ ДЛЯ АННОТИРОВАНИЯ СЕНСОРНЫХ ДАННЫХ А.И. Гоглачев, М.Л. Цымблер .....	30
МЕТОД ПОДАВЛЕНИЯ АКУСТИЧЕСКОГО ЭХА НА ОСНОВЕ РЕКУРРЕНТНОЙ НЕЙРОННОЙ СЕТИ И АЛГОРИТМА КЛАСТЕРИЗАЦИИ Д.М. Шаход, О.Л. Ибряева .....	43
ДИАГНОСТИКА НЕИСПРАВНОСТЕЙ ПОДШИПНИКОВ КАЧЕНИЯ С ИСПОЛЬЗОВАНИЕМ ПИКОВ СПЕКТРА И НЕЙРОННЫХ СЕТЕЙ О.Л. Ибряева, М.Н. Мохаммад .....	59

## Contents

RESOURCE-INDEPENDENT DESCRIPTION OF INFORMATION GRAPHS WITH DISTRIBUTIVE OPERATIONS IN THE SET@L PROGRAMMING LANGUAGE I.I. Levin, I.V. Pisarenko, D.V. Mikhailov, A.K. Melnikov, A.I. Dordopulo .....	5
COMPARISON OF THE TARGET FUNCTIONS IN THE PRONY'S PROBLEM OF MEASUREMENT DATA APPROXIMATION A.A. Lomov, E.A. Rusinova .....	18
APPLICATION OF PARALLEL COMPUTING FOR SENSOR DATA ANNOTATION A.I. Goglachev, M.L. Zymbler .....	30
METHOD OF AN ACOUSTIC ECHO SUPPRESSION BASED ON RECURRENT NEURAL NETWORK AND CLUSTERING Gh.M. Shahoud, O.L. Ibryaeva .....	43
FAULT DIAGNOSIS OF ROLLING BEARINGS USING SPECTRAL PEAKS AND NEURAL NETWORKS O.L. Ibryaeva, M.N. Mohammad .....	59



This issue is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

# РЕСУРСОНЕЗАВИСИМОЕ ОПИСАНИЕ ИНФОРМАЦИОННЫХ ГРАФОВ С ДИСТРИБУТИВНЫМИ ОПЕРАЦИЯМИ НА ЯЗЫКЕ ПРОГРАММИРОВАНИЯ SET@L

© 2022 И.И. Левин<sup>1</sup>, И.В. Писаренко<sup>2</sup>, Д.В. Михайлов<sup>2</sup>,  
А.К. Мельников<sup>3</sup>, А.И. Дордопуло<sup>2</sup>

<sup>1</sup> Южный федеральный университет (347928 Таганрог, пер. Некрасовский, д. 44),

<sup>2</sup> «НИЦ супер-ЭВМ и нейрокомпьютеров» (347922 Таганрог, пер. Итальянский, д. 106),

<sup>3</sup> АО «Вычислительные решения» (117587 Москва, Варшавское ш., д. 125с17)

E-mail: [iilevin@sfnedu.ru](mailto:iilevin@sfnedu.ru), [pisarenko@superevm.ru](mailto:pisarenko@superevm.ru), [mikhailow.den@gmail.com](mailto:mikhailow.den@gmail.com),

[ak@comp-sol.ru](mailto:ak@comp-sol.ru), [dordopulo@superevm.ru](mailto:dordopulo@superevm.ru)

Поступила в редакцию: 05.05.2022

В данной работе предлагается преобразовать стандартную последовательную топологию информационного графа с дистрибутивными операциями к комбинированному варианту с последовательными и параллельными фрагментами, что позволяет эффективно описать реализацию вычислений в ресурсонезависимой форме. Конечная топология зависит от доступного вычислительного ресурса реконфигурируемой системы и обеспечивает повышение удельной производительности в сравнении с исходным вариантом. Разработанный алгоритм преобразования линейной структуры в различные комбинированные топологии в зависимости от конфигурации вычислительной системы описан на языке программирования Set@L.

*Ключевые слова:* дистрибутивные операции, ресурсонезависимое программирование, реконфигурируемые вычислительные системы, редукция производительности, Set@L, признаки «разбиение пополам» и «голова/хвост».

## ОБРАЗЕЦ ЦИТИРОВАНИЯ

Левин И.И., Писаренко И.В., Михайлов Д.В., Мельников А.К., Дордопуло А.И. Ресурсонезависимое описание информационных графов с дистрибутивными операциями на языке программирования Set@L // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2022. Т. 11, № 2. С. 5–17. DOI: 10.14529/cmse220201.

## Введение

Информационные графы многих прикладных задач имеют последовательную (линейную) топологию, поэтому реализация их распараллеливания (ускорения) в общем виде затруднительна. Однако, если операции графа обладают определенными свойствами, такими как дистрибутивность и ассоциативность, то целесообразно преобразовать их к виду, для которого допустимо параллельное (ускоренное) выполнение. Например, в работе [1] показано, что за счет перестановки вершин в информационном графе с ассоциативными операциями и дальнейшей оптимизации вычислительной структуры время решения задачи может быть сокращено в число раз, соответствующее латентности операционной вершины. В данной статье представлен другой пример оптимизации последовательного информационного графа с дистрибутивными операциями для реализации на реконфигурируемых ВС с различными конфигурациями. Подобные графы широко применяются для решения многих задач цифровой обработки сигналов [2, 3]. Для практического применения предложенных преобразований необходимы эффективные методы и средства представления параллельных программ в ресурсонезависимом виде. В работах [4–6] предложен язык архитектурно-

независимого параллельного программирования Set@l (Set Aspect-Oriented Language), основанный на парадигме аспектно-ориентированного программирования (АОП) и теоретико-множественном представлении исходного кода программы. Описав базовую топологию графа с дистрибутивными операциями и принципы ее преобразования в виде специальных признаков метода обработки на языке Set@l, возможно синтезировать множество вариантов топологий и переходить между ними путем изменения типов и разбиений совокупностей. Приведение графа к оптимальному виду при заданном объеме вычислительного ресурса реконфигурируемой ВС может осуществляться без изменения исходного кода программы.

## 1. Преобразование информационных графов с дистрибутивными операциями

Рассмотрим последовательный информационный граф, топология которого изображена на рис. 1. Представленный граф состоит из чередующихся операционных вершин  $\alpha$  и  $\beta$ , причем обе базовые операции обладают свойством ассоциативности, а операция  $\beta$  является дистрибутивной по отношению к операции  $\alpha$ . Граф на рис. 1 позволяет вычислить последовательность коэффициентов  $W_1, W_2, \dots, W_n$ , определяемых следующей рекуррентной формулой:

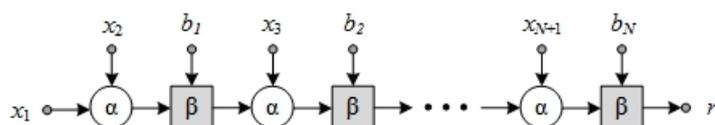


Рис. 1. Последовательный информационный граф с дистрибутивными операциями  $\alpha$  и  $\beta$

Граф на рис. 1 позволяет вычислить последовательность коэффициентов  $W_1, W_2, \dots, W_n$ , определяемых следующей рекуррентной формулой:

$$W_0 = x_1; \quad W_i = (W_{i-1} \alpha x_{i+1}) \beta b_i, \quad i = 1, 2, \dots, N, \quad (1)$$

где  $N$  — мощность множества, включающего все коэффициенты  $\beta_i$ . Время выполнения всех операций графа составляет

$$T_G = \tau \cdot N \cdot (l_\alpha + l_\beta), \quad (2)$$

где  $\tau$  — длительность такта в секундах;  $l_\alpha, l_\beta$  — латентности операционных вершин  $\alpha$  и  $\beta$  в тактах.

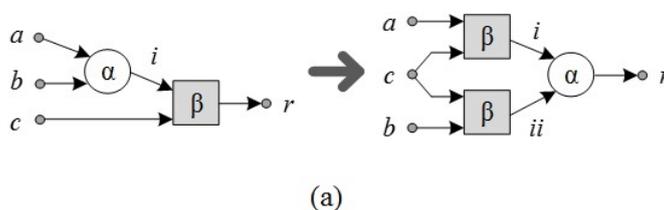
Предположим, что объем доступного вычислительного ресурса  $R$  реконфигурируемой ВС не ограничен. Тогда для сокращения времени решения задачи  $T_G$  возможно модифицировать последовательную топологию на рис. 1, используя свойства дистрибутивности и ассоциативности операций  $\alpha$  и  $\beta$ . Введем элементарные преобразования пар вершин, показанные на рис. 2, а и б. Преобразование на рис. 2, а перегруппирует пару последовательно соединенных вершин  $\alpha$  и  $\beta$  в соответствии с законом дистрибутивности:

$$(\alpha(a, b, i), \beta(i, c, r)) = (\beta(a, c, i), \beta(b, c, ii), \alpha(i, ii, r)), \quad (3)$$

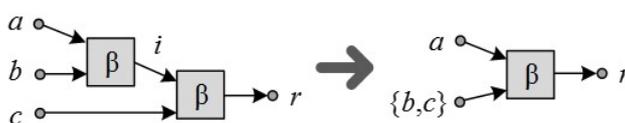
где  $a, b, c$  — входные вершины;  $r$  — выходная вершина;  $i, ii$  — дуги, соответствующие промежуточным результатам вычислений. Преобразование на рис. 2, б представляет со-

бой слияние двух ассоциативных вершин  $\beta$  в одну вершину  $\beta$ , на вход которой подается множество исходных операндов:

$$(\beta(a, b, i), \alpha(i, c, r)) = \beta(a, \{b, c\}, r). \tag{4}$$



(а)

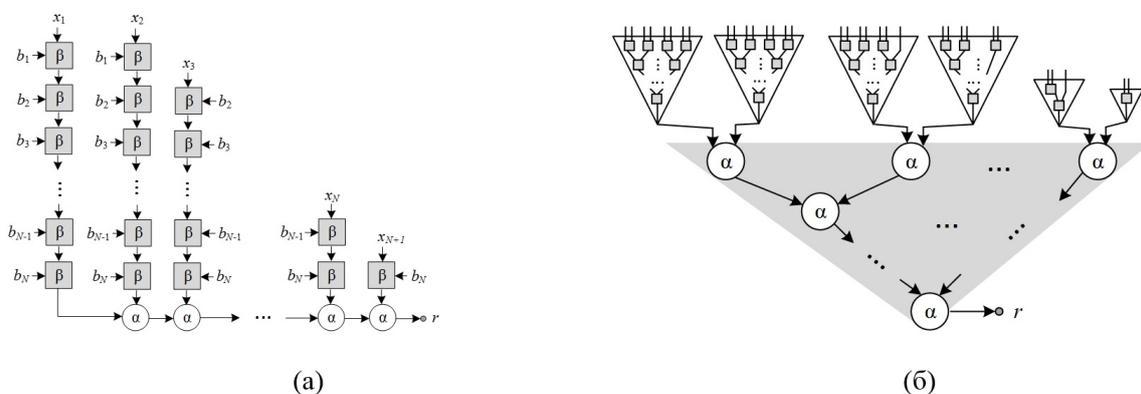


(б)

**Рис. 2.** Преобразование пары дистрибутивных вершин  $\alpha$  и  $\beta$  (а) и слияние пары ассоциативных вершин (б), показанное на примере операции  $\beta$

Последовательно применяя преобразования на рис. 2, а и б к исходному линейному графу (см. рис. 1), возможно получить модифицированную топологию с множеством подграфов-ветвей, одна из которых состоит только из вершин  $\alpha$ , а остальные – только из вершин  $\beta$ , как показано на рис. 3, а.

Теперь, когда граф разбит на фрагменты, содержащие ассоциативные операционные вершины только одного типа  $\alpha$  или  $\beta$  (см. рис. 3, а), возможно распараллелить вычисления, преобразовав каждый из последовательных подграфов к пирамидальному виду [7]. В результате будет получена топология, изображенная на рис. 3, б и включающая нижний пирамидальный подграф из вершин  $\alpha$ , входы которого соединены с выходами множества верхних пирамидальных подграфов, составленных из вершин  $\beta$ . Размер пирамид из вершин  $\beta$  будет меняться от  $N$  вершин (для двух крайних левых подграфов на рис. 3, б) до одной вершины (крайне правая пирамида на рис. 3, б).



(а)

(б)

**Рис. 3.** Модифицированные топологии информационного графа с дистрибутивными операциями  $\alpha$  и  $\beta$  с последовательными (а) и параллельными (б) подграфами-ветвями

Время выполнения всех операций информационного графа на рис. 3, б является минимальным и составляет

$$T_G = (l_\alpha + l_\beta) \cdot \lceil \log_2 N \rceil \cdot \tau, \quad (5)$$

где  $\lceil \cdot \rceil$  — округление до целого в большую сторону. В формуле (5) округление множителя  $\log_2 N$  учитывает случаи, когда размерность пирамидальных подграфов необходимо дополнить до степени 2. В отличие от исходной топологии (см. рис. 1) информационные графы на рис. 3 характеризуются существенно большим объемом аппаратного ресурса, занимаемого соответствующими им вычислительными структурами: количество вершин  $\beta$  возрастает до  $(N^2 + N)/2$ , тогда как число вершин  $\alpha$  остается равным  $N$ . Например, если в исходном графе была 1000 вершин  $\beta$ , то после преобразования их количество возрастет более чем в 500 раз.

На практике при решении задач даже небольшой размерности может возникнуть ситуация, когда объем аппаратного ресурса реконфигурируемой ВС недостаточен для реализации топологий, изображенных на рис. 3. Анализируя информационный граф на рис. 3, а, выделим следующие особенности:

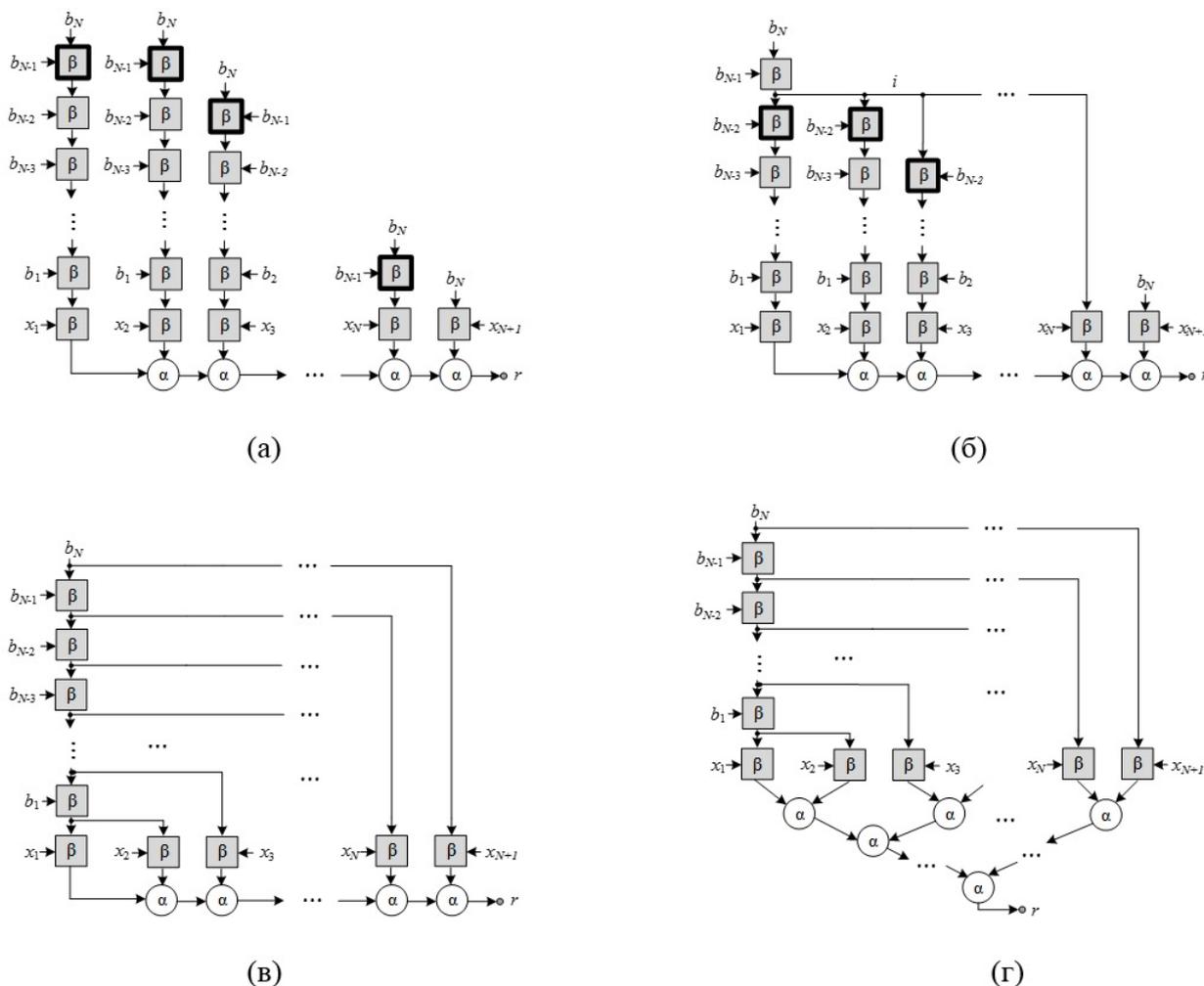
- каждая ветвь из вершин  $\beta$  использует только одну из переменных  $x_i$ ;
- на две левые ветви из вершин  $\beta$  подаются все коэффициенты  $\{b_1 \dots b_N\}$ , а на каждую последующую ветвь — на один коэффициент меньше, то есть  $m$ -я ветвь оперирует коэффициентами  $\{b_{m-1} \dots b_N\}$ .

Поскольку операция  $\beta$  ассоциативна, перегруппируем вершины в ветвях таким образом, чтобы блоки с входами  $\{x_1 \dots x_{N+1}\}$  располагались вплотную к подграфу из вершин  $\alpha$  (см. рис. 4, а). В результате такого преобразования каждая ветвь, кроме крайней правой, будет содержать вершину  $\beta$  с входами  $b_{N-1}$  и  $b_N$  (выделены на рис. 4, а). Оставим одну вершину  $\beta(b_{N-1}, b_N, i)$  в крайней левой ветви и удалим повторяющиеся вершины, подав на входы остальных ветвей данные с выхода  $i$  (см. рис. 4, б). В полученном графе возникнет новое множество идентичных вершин  $\beta(i, b_{N-2}, ii)$ , которое выделено на рис. 4, б и может быть заменено одной вершиной аналогичным образом. Каждый шаг преобразования уменьшает количество вершин  $\beta$  и формирует промежуточную топологию, реализуемую на ВС с определенным объемом вычислительного ресурса. В предельном случае, показанном на рис. 4, в, процедура удаления одинаковых вершин повторялась до тех пор, пока во всех ветвях, кроме крайней левой, не осталось по одной вершине  $\beta$ . Количество вершин  $\beta$  в такой топологии равно  $2 * N - 1$  вместо  $(N^2 + N)/2$  в исходном графе (см. рис. 4, а). Линейку вершин  $\alpha$  в графе на рис. 4 в целесообразно преобразовать в пирамиду (см. рис. 4, г), обеспечив снижение времени выполнения всех операций до

$$T_G = (l_\alpha \cdot \lceil \log_2 N \rceil + l_\beta \cdot N) \cdot \tau. \quad (6)$$

Дальнейшее уменьшение аппаратных затрат возможно при разбиении исходного графа (см. рис. 1) на линейные изоморфные подграфы, содержащие максимальное реализуемое число вершин, однако при таком подходе время решения задачи  $T_G$  будет не оптимальным. Более эффективный подход предполагает преобразование каждого подграфа к виду, аналогичному рис. 4, г.

Если количество операционных вершин в графе с дистрибутивными операциями после модификаций, представленных на рис. 3 и 5, все еще велико для реализации соответствующей вычислительной структуры на имеющемся аппаратном ресурсе, то для формирования плотного потока данных необходимо выполнить следующие преобразования. Вернемся к по-



**Рис. 4.** Модификация информационного графа с дистрибутивными операциями: перестановка вершин  $\beta$  в ветвях (а), однократное удаление повторяющихся вершин  $\beta$  (б), граф с одной полной ветвью (в) и перегруппировка вершин  $\alpha$  в пирамидальную структуру (г)

следовательной топологии (см. рис. 1) и разобьем исходный граф на изоморфные подграфы, содержащие по  $L$  линейно соединенных пар вершин  $\alpha$  и  $\beta$ . Далее, используя предложенную выше методику, приведем каждый выделенный подграф к виду «полная линейка  $\beta$  — пирамида  $\alpha$ », показанному на рис. 5, а. В соответствующей полученному графу вычислительной структуре обратная связь будет охватывать полную линейку операций  $\beta$  и пирамиду операций  $\alpha$ , поэтому интервал обработки данных будет достаточно высоким. Для его снижения перегруппируем вершины таким образом, чтобы обратная связь охватывала только пару вершин  $\alpha$  и  $\beta$ . Поскольку операция  $\alpha$  ассоциативна, полную пирамиду размерности  $N$  можно заменить на пирамиду меньшей размерности  $(N-1)$  и одиночную операционную вершину  $\alpha$ , как показано на рис. 5, б. Вместе с вершиной  $\alpha$  из линейки операций  $\beta$  будет вынесена одна ближайшая к пирамиде вершина  $\beta$ , поэтому в конечной вычислительной структуре на рис. 5, б обратной связью будет охвачено всего две вершины, по одной каждого типа. Чтобы обеспечить минимальный интервал обработки данных, размерность выделенных подграфов должна соотноситься с латентностями операционных вершин:  $L = (l_\alpha + l_\beta)$ .

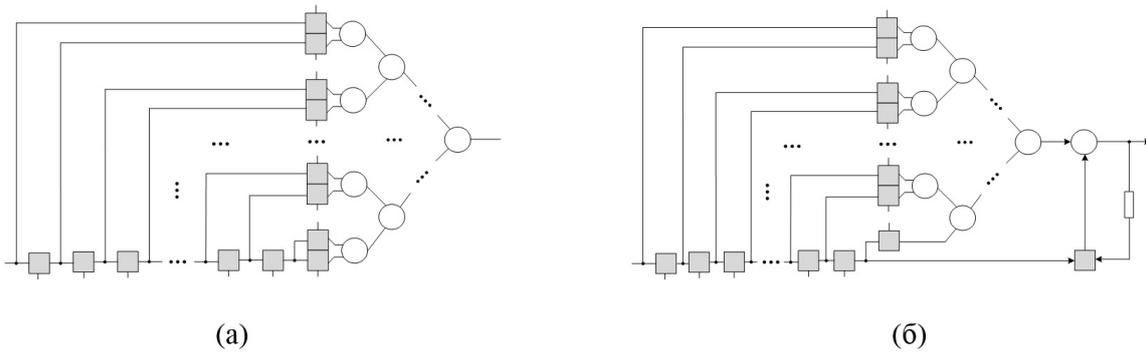
После описанного преобразования время решения задачи составит:

$$T_G = \left( L + \left\lceil \frac{N}{l_\alpha + l_\beta} \right\rceil \right) \cdot \tau, \quad (7)$$

где  $L = (l_\alpha + l_\beta) \cdot l_\beta + \lceil \log_2(l_\alpha + l_\beta) \rceil \cdot l_\alpha$  — длина критического пути в тактах для вычислительной структуры на рис. 5, б.

Если аппаратный ресурс реконфигурируемой ВС недостаточен даже для реализации вычислительной структуры на рис. 5, б, то целесообразно использовать минимальную структуру с парой вершин  $\alpha$  и  $\beta$ . В таком случае время выполнения всех операций графа составит

$$T_G = (l_\alpha + l_\beta) \cdot \tau \cdot N. \quad (8)$$



**Рис. 5.** Преобразованный фрагмент исходного информационного графа (а) и вычислительная структура (б), получаемая при вынесении пары вершин  $\alpha$  и  $\beta$

Таким образом, в зависимости от соотношения объемов вычислительного ресурса  $R_\alpha$  и  $R_\beta$ , доступных для реализации вершин соответствующих типов, их латентностей  $l_\alpha$  и  $l_\beta$ , а также размерности обрабатываемых массивов данных  $N$  возможно выделить несколько способов построения вычислительной структуры, обеспечивающих разные скорости решения задачи. Ниже приведено составное выражение, позволяющее рассчитать время выполнения всех операций графа для каждого из этих случаев:

$$T_G = \begin{cases} (l_\alpha + l_\beta) \cdot \lceil \log_2 N \rceil \cdot \tau, & \text{если } R_\alpha \geq N, R_\beta \geq \frac{N^2+N}{2}; \\ (l_\alpha \cdot \lceil \log_2 N \rceil + l_\beta \cdot N) \cdot \tau, & \text{если } R_\alpha \geq N, 2N \leq R_\beta < \frac{N^2+N}{2}; \\ \left\{ (l_\alpha + l_\beta - 1) \cdot l_\beta + \lceil \log_2(l_\alpha + l_\beta) \rceil \cdot l_\alpha + \left\lceil \frac{N}{l_\alpha + l_\beta} \right\rceil \right\} \cdot \tau \cdot (l_\alpha + l_\beta), & \\ \text{если } l_\alpha + l_\beta \leq R_\alpha < N, 2 \cdot (l_\alpha + l_\beta) \leq R_\beta < 2N; \\ \left\{ (k - 1) \cdot l_\beta + \lceil \log_2 k \rceil \cdot l_\alpha + \frac{N}{k} \right\} \cdot \tau \cdot (l_\alpha + l_\beta), & \\ \text{если } R_\alpha < l_\alpha + l_\beta, R_\beta < 2 \cdot (l_\alpha + l_\beta), \end{cases} \quad (9)$$

где  $(k - 1)$  — число автоподстановок [8].

Первый вариант описывает предельный случай, когда аппаратного ресурса достаточно для реализации вычислительной структуры, соответствующей полному графу с пирамидальными подграфами на рис. 3, б, что позволяет выполнить все операции за наименьшее время. Второй вариант описывает ситуацию, когда аппаратного ресурса  $R_\beta$  не хватает на построение множества пирамидальных структур, но хватает на построение линейки, использование которой позволяет удалить повторяющиеся вершины (см. топологию на

рис. 4, г). Третий вариант описывает ситуацию, когда доступного ресурса не хватает на реализацию всего графа целиком, но достаточно для реализации подсистемы, охваченной обратной связью, включающей в себя  $l_\alpha + l_\beta$  блоков  $\alpha$  и  $2 \cdot (l_\alpha + l_\beta)$  блоков  $\beta$  (см. вычислительную структуру на рис. 5, б). Наконец, последний вариант описывает случай, когда оборудования не хватает для того, чтобы реализовать даже описанную выше подсистему. В таком случае автоподстановка реализуется столько раз, сколько позволяет объем свободного аппаратного ресурса. В предельном случае  $k = l_\alpha + l_\beta$  рассматриваемое выражение переходит в третье, а при  $k = 1$  — в формулу (8) для минимальной вычислительной структуры с парой вершин  $\alpha$  и  $\beta$ , охваченных обратной связью.

Если исходный последовательный граф состоит из чередующихся вершин, содержащих ассоциативные операции двух типов, одна из которых является дистрибутивной по отношению к другой, мы можем значительно ускорить выполнение всех операций графа. Степень этого ускорения зависит от того, сколько вычислительных блоков возможно реализовать на доступном нам оборудовании и как именно будет преобразован исходный граф. При этом в случае, если мы реализуем вычислительную систему с обратной связью, глубина которой и размер системы определяется суммарной латентностью блоков  $\alpha$  и  $\beta$ , то мы получим лучший результат, чем при реализации всего графа целиком, при условии, что у нас не хватает ресурса для представления всех блоков  $\beta$  в виде пирамидальной структуры.

## 2. Ресурснезависимая программа на языке Set@l

Традиционные методы параллельного программирования, как правило, оперируют информационным графом с фиксированной структурой, поэтому их использование для описания преобразований топологии графа с дистрибутивными операциями (см. рис. 1–5) достаточно трудоемко и малоэффективно. Код ресурснезависимой программы, которая реализует рассмотренные преобразования, будет состоять из множества подпрограмм, связанных условными операторами и задающих отдельные варианты топологий. В отличие от классических языков параллельного программирования, возможности языка архитектурно-независимого программирования Set@l позволяют описать принципы построения и модификации графов с дистрибутивными операциями в виде множеств типов «голова/хвост» и «разбиение пополам» и правил их декомпозиции. В таком случае исходный код программы описывает не отдельные реализации, а множество возможных топологий графа для заданной размерности задачи. Аспекты выбирают определенный вариант топологии, исходя из конкретных значений параметров конфигурации. Для изменения структуры информационного графа достаточно изменить тип и разбиение множеств, тогда как исходный код программы, описывающий математическую сущность алгоритма, остается неизменным.

Исходный код программы на языке Set@l, приведенный на рис. 6, описывает базовую последовательную топологию графа с дистрибутивными операциями (см. рис. 1). Отметим, что вершины  $\alpha$  (**alpha**) и  $\beta$  (**beta**) могут соответствовать как простым операциям над числами с фиксированной запятой, так и более сложным операциям сложения и умножения комплексных чисел, векторов и матриц.

Полный граф описывается рекурсивно (Rec, см. строку (1) на рис. 6) как отношение  $F$  между множествами переменных  $X$  и коэффициентов  $B$  и результирующим элементом  $r$ . Поскольку граф строится по последовательному принципу «голова/хвост», в условии завершения рекурсии (**break**, строки (2)–(4)) и правил раскрытия итераций (**main**, строки (5), (6)) используются соответствующие признаки **Head** («голова»: выделение первого

```

(1)  Rec[F(X,B,r) | set(X,B) && element(r)]:
(2)    break[cap(B)=1:
(3)      F(X,B,r)=conc{alpha(Head(Tail(X)),Head(X),i),
(4)                          beta(i,Head(B),r) | element(i)}];
(5)    main[F(X,B,r)=conc{F(Tail(X),Tail(B),e),alpha(e,Head(X),i),
(6)                          beta(i,Head(B),r) | element(e,i)}];
(7)  end(F);

```

**Рис. 6.** Исходный код программы на языке Set@l, описывающий последовательную топологию информационного графа с дистрибутивными операциями (см. рис. 1)

элемента множества) и **Tail** («хвост»: выделение подмножества, содержащего все элементы исходного множества, кроме первого). Альтернативный принцип построения графов, рассматриваемый в данной работе, — «разбиение пополам» — предполагает формирование параллельной пирамидальной структуры.

На каждой итерации (см. строки (5), (6) на рис. 6) линейку дистрибутивных операций **F** над множествами **X** и **B** можно представить как аналогичную линейку операций над «хвостами» этих множеств и пару связанных через элемент **i** вершин  $\alpha$  и  $\beta$ , на входы которых подаются «головы» множеств **X** и **B** и промежуточный результат **e** последовательного выполнения ассоциативных операций **F** над **Tail(X)**, **Tail(B)**, а выход **r** является конечным или промежуточным результатом вычислений. Так как исходный последовательный граф не привязан к какой-либо вычислительной архитектуре, все его элементы связаны между собой по параллельно-зависимому принципу **conc** [6]. При выполнении условия завершения рекурсии (строки (2)–(4)) последняя линейка операций **F** преобразуется в особую пару вершин, на входы которой подаются оставшиеся элементы множеств **X** и **B**.

На рис. 7 представлен код на языке программирования Set@l, который описывает базовое преобразование пары дистрибутивных вершин  $\alpha$  и  $\beta$  (признак **distr**, строки (1)–(4), см. рис. 2, а) и слияние пары ассоциативных вершин любого типа (признак **mrg**, строки (5)–(8), см. рис. 2, б), на основе которых осуществляется дальнейшая модификация последовательного информационного графа. Преобразование **mrg** позволяет передать тип Н/Т («голова/хвост») или DIV2 («разбиение пополам») формируемому множеству **k**, подаваемому на вход блока **f**.

```

(1)  attribute(distr(a1,a2) | a1=alpha(a,b,i) && a2=beta(i,c,r)) :
(2)    conc(a1,a2)=conc(beta(a,c,i),beta(b,c,ii),
(3)                          alpha(i,ii,r) | element(ii));
(4)  end(distr);
(5)  attribute(mrg(a1,a2,z) | a1=f(a,b,c) && a2=f(c,d,e) &&
(6)  z=type('Н/Т' || 'DIV2')) :
(7)    conc(a1,a2)=[f(a,k,e) | k=union(b,d) && type(k)=z];
(8)  end(mrg);

```

**Рис. 7.** Код базовых признаков, описывающих преобразование пары дистрибутивных вершин (**distr**) и слияние пары ассоциативных вершин (**mrg**), на языке программирования Set@l

Фрагмент программы на языке Set@l изображенный на рис. 8, определяет процедуру преобразования **Tr** линейного информационного графа с дистрибутивными операциями

(см. топологию на рис. 1 и код на рис. 6) к топологии с подграфами-ветвями (см. рис. 3, а). Движение по итерациям начинается от пары вершин  $\alpha$  и  $\beta$ , соединенной с выходом  $r$ .

На каждой итерации (**main**, см. строки (6)–(8) на рис. 8) последовательно (**seq**) реализуются одно преобразование пары дистрибутивных вершин (**distr**, строка (6)) и два отдельных слияния (**mrq**, строки (6)–(8)) групп ассоциативных вершин  $\alpha$  и  $\beta$ . В итоге граф представляется как объединение этих операций с аналогичным подграфом, вершины которого входят в множество  $\text{Tail}^2(G) = \text{Tail}(\text{Tail}(G))$ . Преобразования **mrq** осуществляются над группами вершин, связанных общими элементами  $s$ , а характер параллелизма объединяемых вершин определяется передаваемыми типами **talpha** и **tbeta** (H/T или DIV2). Строки (10)–(14) описывают параллелизм ветвей из вершин  $\beta$ , свойственный для графа с топологией на рис. 3, а.

Переход от топологии с последовательными подграфами-ветвями (см. рис. 3, а) к его параллельной реализации (см. рис. 3, б) может осуществляться за счет передачи типов **talpha**='DIV2' и **tbeta**='DIV2' при вызове процедуры **Tr**.

```
(1) Rec[Tr(G, talpha, tbeta) | set(G), type(talpha, tbeta)]:
(2)   break [cap(G)=2:
(3)     Tr(G)=seq[distr(Head(G), Head(Tail(G))),
(4)     mrq(k, l, m | k, l in G && k=alpha(a, b, c) && l=alpha(c, d, e))]];
(6)   main[Tr(G)=seq[Tr(Tail^2(G)), distr(Head(G), Head(Tail(G))),
(7)     mrq(e, f, m | e, f in G && e=alpha(a, b, c) && f=alpha(c, d, e)),
(8)     mrq(x, y, z | x, y in G && y=beta(c, d, e) && z=tbeta)]];
(9) end(Tr);
(10) attribute(Tpar(G) | set(G)):
(11)   set(K);
(12)   K=(a | a in G && a=beta(c, d, e));
(13)   type(K)='par';
(14) end(Tpar);
```

**Рис. 8.** Код преобразования исходного информационного графа (см. рис. 1) в структуру с последовательными подграфами-ветвями (см. рис. 3, а) на языке Set@l

На следующем этапе топология с подграфами-ветвями (см. рис. 3, а) оптимизируется путем удаления повторяющихся вершин  $\beta$  (см. рис. 8). Соответствующая процедура **Gcut** задается фрагментом кода на языке программирования Set@l, представленном на рис. 9. Признак **vcut** (см. строки (1)–(3)) выделяет пару вершин  $\beta$  с одинаковыми входами  $s$ ,  $f$  и разными выходами  $e$ ,  $ee$  и заменяет их на одну вершину с объединенным выходом  $k$ , в котором обе связи реализуются параллельно (**par**). Процедура **scut** (см. строки (4)–(7)) последовательно выделяет все повторяющиеся вершины в текущем графе  $V$  и применяет к ним преобразование **vcut**. Процесс продолжается до тех пор, пока из исходной группы не останется одна уникальная вершина (см. рис. 4, б). Процедура **Gcut** многократно повторяет преобразование **scut** и приводит граф к виду, показанному на рис. 4, в: все ветви из вершин  $\beta$ , кроме первой, содержат по одной вершине, а повторяющиеся вершины полностью удалены.

После оптимизации топологии графа (см. код на рис. 9) необходимо синтезировать вычислительную структуру **CS**, показанную на рис. 5, б. Для этого множества входных данных  $XD$  и  $VD$  типа **pipe** (конвейерная обработка) разбиваются на подмножества типа **conp** (параллельно-зависимая обработка), размерность которых определяется латентностями операционных вершин  $l_\alpha$  (1a) и  $l_\beta$  (1b), а затем преобразование **FG**, описанное на рис. 6–9, выполняется над каждым подмножеством (**sub**):

```

(1)  attribute (vcut (a1, a2) | a1=beta (c, f, e)  &&  a2=beta (c, f, ee) ) :
(2)    par (a1, a2)=[beta (c, f, k) | k=par (e, ee)  &&  e=ee];
(3)  end (vcut);
(4)  Rec [scut (B) | set (B) ] :
(5)    break [cap (B)=1];
(6)    main [scut (B)=seq{vcut (Head (B) , Head (Tail (B) ) ) , scut (Tail (B) ) }]];
(7)  end (scut);
(9)  Rec [Gcut (G) | set (G) ] :
(10)   B=(a, b, ... | a, b in G && a=beta (c, f, e)  &&  b=beta (c, f, ee) );
(11)   break [cap (B)=1];
(12)   main [Gcut (G)=seq (Gcut (G) , scut (B) ) ];
(13)  end (Gcut);

```

Рис. 9. Фрагмента кода программы на языке Set@l, реализующий удаление повторяющихся вершин  $\beta$  (см. рис. 4)

```

XD=pipe(union[conc(x(a) ...x(b))|(x(a),x(b)) in X &&
              a in set(1,1a+1b+1, ...) &&
              b in set(1a+1b,2*(1a+1b), ...)]];
BD=pipe(union[conc(b(c) ...b(d))|(b(c),b(d)) in B &&
              c in set(1,1a+1b+1, ...) &&
              d in set(1a+1b,2*(1a+1b), ...)]]);
CS=FG(x,b,r| forall sub(x) in XD && forall sub(b) in BD);

```

В итоге будет получена вычислительная структура, изображенная на рис. 5, а. Чтобы перейти к конечной вычислительной структуре, изображенной на рис. 5, б, необходимо выделить множество входов  $A$  пирамиды из вершин  $\alpha$  и модифицировать его следующим образом:

```

A=(a|beta(c,d,a) && alpha(a,b,e));
A*=(N/T(DIV2(a1...ak-1), ak) | k=cap(A));

```

Представленный в данном разделе код на языке программирования Set@l учитывает все случаи, выделенные в формуле (9), и позволяет описать реализацию информационных графов с дистрибутивными операциями на реконфигурируемых ВС в компактной ресурснезависимой форме.

## Заключение

Таким образом, в данной статье показан метод, который за счет перестановки вершин в информационном графе с дистрибутивными операциями и дальнейшей оптимизации вычислительной структуры позволяет ускорить выполнение всех операций графа, причем степень этого ускорения зависит от того, сколько вычислительных блоков возможно реализовать на имеющемся аппаратном ресурсе. Язык архитектурно-независимого программирования Set@l обеспечивает описание преобразований в компактной ресурснезависимой форме. В отличие от традиционных языков параллельного программирования, в которых изменение топологии информационного графа требует модификации исходного кода программы, Set@l задает множество вариантов реализации в одной программе, а синтез конкретной вычислительной структуры осуществляется автоматически в соответствии с заданными пользователем параметрами конфигурации — объемом доступного вычислительного ресурса и латентностью базовых операций.

*Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-07-00545.*

## Литература

1. Levin I.I., Dordopulo A.I., Pisarenko I., *et al.* Resource-Independent Description of Information Graphs with Associative Operations in Set@l Programming Language // Parallel Computing Technologies, 16th International Conference, PaCT 2021, Kaliningrad, Russia, September 13–18, 2021. Proceedings. Vol. 12942 / ed. by V. Malyshkin. Springer, 2021. Lecture Notes in Computer Science. P. 74–87. DOI: 10.1007/978-3-030-86359-3\_6.
2. Tan L., Jiang J. Digital Signal Processing: Fundamentals and Applications. 2nd ed. Oxford: Elsevier Science, 2013. 896 p.
3. Winser A., Cranos M.W. Digital Signal Processing: Principles, Algorithms and System Design. London: Elsevier, 2017. 634 p.
4. Levin I.I., Dordopulo A.I., Pisarenko I.V., Melnikov A.K. Aspect-Oriented Set@l Language for Architecture-Independent Programming of High-Performance Computer Systems // Supercomputing. RuSCDays 2019. Vol. 1129 / ed. by L. Sokolinsky, M. Zymbler. Cham: Springer, 2019. P. 517–528. Communications in Computer and Information Science. DOI: 10.1007/978-3-030-36592-9\_42.
5. Levin I.I., Dordopulo A.I., Pisarenko I.V., Melnikov A.K. Objects of Alternative Set Theory in Set@l Programming Language // Parallel Computing Technologies, 15th International Conference, PaCT 2019, Almaty, Kazakhstan, August 19–23, 2019. Proceedings. Vol. 11657 / ed. by V. Malyshkin. Springer, 2019. P. 18–31. Lecture Notes in Computer Science. DOI: 10.1007/978-3-030-25636-4\_3.
6. Левин И.И., Дордопуло А.И., Писаренко И.В., Мельников А.К. Язык архитектурно-независимого программирования вычислительных систем Set@l // Вестник компьютерных и информационных технологий. 2019. № 3. С. 48–56. DOI: 10.14489/vkit.2019.03.pp.048-056.
7. Кареева Е.Д. Основы многопоточного и параллельного программирования. Красноярск: Сиб. федер. ун-т, 2016. 356 с.
8. Levin I.I., Dudko S.A. Equivalent Transformations of Some Kinds of Computing Structures of Non-linear Recurrent Expressions for Reconfigurable Computing Systems // Parallel Computational Technologies. Vol. 1437 / ed. by L. Sokolinsky, M. Zymbler. Cham: Springer, 2021. P. 3–17. Communications in Computer and Information Science. DOI: 10.1007/978-3-030-81691-9\_1.

Левин Илья Израилевич, д.т.н., профессор, кафедры интеллектуальных и многопроцессорных систем, Южный федеральный университет (Таганрог, Российская Федерация)

Писаренко Иван Вадимович, «НИЦ супер-ЭВМ и нейрокомпьютеров» (Таганрог, Российская Федерация)

Михайлов Денис Васильевич, «НИЦ супер-ЭВМ и нейрокомпьютеров» (Таганрог, Российская Федерация)

Мельников Андрей Кимович, к.т.н., доцент, АО «Вычислительные решения» (Москва, Российская Федерация)

Дордопуло Алексей Игоревич, к.т.н., «НИЦ супер-ЭВМ и нейрокомпьютеров» (Таганрог, Российская Федерация)

# RESOURCE-INDEPENDENT DESCRIPTION OF INFORMATION GRAPHS WITH DISTRIBUTIVE OPERATIONS IN THE SET@L PROGRAMMING LANGUAGE

© 2022 I.I. Levin<sup>1</sup>, I.V. Pisarenko<sup>2</sup>, D.V. Mikhailov<sup>2</sup>,  
A.K. Melnikov<sup>3</sup>, A.I. Dordopulo<sup>2</sup>

<sup>1</sup>*Southern Federal University (Nekrasovsky lane 44, Taganrog, 347928 Russia),*

<sup>2</sup>*Supercomputers and Neurocomputers Research Center  
(Italyansky lane 106, Taganrog, 347922 Russia),*

<sup>3</sup>*“Computational Solutions” JSC (Varshavskoe highway 125c17, Moscow, 119991 Russia)*

*E-mail: iilevin@sfnedu.ru, pisarenko@superevm.ru, mixailow.den@gmail.com,  
ak@comp-sol.ru, dordopulo@superevm.ru*

Received: 05.05.2022

In the paper, we suggest to transform a standard sequential topology of an information graph with distributive operations into its hybrid version with sequential and parallel fragments. Such transformation allows to provide efficient description of calculations in the resource-independent form. The result topology depends on available hardware resource of a reconfigurable system and provides an increase in the special performance in comparison to the initial topology. We have developed an algorithm of a linear structure transformation into various hybrid topologies according to the configuration of the computing system. The algorithm is described in the Set@l programming language.

*Keywords: distributive operations, resource-independent programming, reconfigurable computing systems, performance reduction, Set@l, “half-splitting” and “head/tail” attributes.*

## FOR CITATION

Levin I.I., Pisarenko I.V., Mikhailov D.V., Melnikov A.K., Dordopulo A.I. Resource-independent Description of Information Graphs with Distributive Operations in the Set@l Programming Language. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2022. Vol. 11, no. 2. P. 5–17. (in Russian) DOI: 10.14529/cmse220201.

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Levin I.I., Dordopulo A.I., Pisarenko I., *et al.* Resource-Independent Description of Information Graphs with Associative Operations in Set@l Programming Language. Parallel Computing Technologies, 16th International Conference, PaCT 2021, Kaliningrad, Russia, September 13–18, 2021. Proceedings. Vol. 12942 / ed. by V. Malyskin. Springer, 2021. Lecture Notes in Computer Science. P. 74–87. DOI: 10.1007/978-3-030-86359-3\_6.
2. Tan L., Jiang J. Digital Signal Processing: Fundamentals and Applications. 2nd ed. Oxford: Elsevier Science, 2013. 896 p.
3. Winser A., Cranos M.W. Digital Signal Processing: Principles, Algorithms and System

Design. London: Elsevier, 2017. 634 p.

4. Levin I.I., Dordopulo A.I., Pisarenko I.V., Melnikov A.K. Aspect-Oriented Set@l Language for Architecture-Independent Programming of High-Performance Computer Systems. Supercomputing. RuSCDays 2019. Vol. 1129 / ed. by L. Sokolinsky, M. Zymbler. Cham: Springer, 2019. P. 517–528. Communications in Computer and Information Science. DOI: 10.1007/978-3-030-36592-9\_42.
5. Levin I.I., Dordopulo A.I., Pisarenko I.V., Melnikov A.K. Objects of Alternative Set Theory in Set@l Programming Language. Parallel Computing Technologies, 15th International Conference, PaCT 2019, Almaty, Kazakhstan, August 19–23, 2019. Proceedings. Vol. 11657 / ed. by V. Malyshkin. Springer, 2019. P. 18–31. Lecture Notes in Computer Science. DOI: 10.1007/978-3-030-25636-4\_3.
6. Levin I.I., Dordopulo A.I., Pisarenko I.V., Melnikov A.K. Architecture-independent Set@l programming language for computer systems. Vestnik komp'yuternykh i informatsionnykh tekhnologii. 2019. No. 3. P. 48–56. DOI: 10.14489/vkit.2019.03.pp.048-056. (in Russian)
7. Karepova E.D. Fundamentals of Multithreaded and Parallel Programming. Krasnoyarsk: Publishing of the Siberian Federal University, 2016. 356 p. (in Russian)
8. Levin I.I., Dudko S.A. Equivalent Transformations of Some Kinds of Computing Structures of Non-linear Recurrent Expressions for Reconfigurable Computing Systems. Parallel Computational Technologies. Vol. 1437 / ed. by L. Sokolinsky, M. Zymbler. Cham: Springer, 2021. P. 3–17. Communications in Computer and Information Science. DOI: 10.1007/978-3-030-81691-9\_1.

## СРАВНЕНИЕ ЦЕЛЕВЫХ ФУНКЦИЙ В ЗАДАЧЕ ПРОНИ ДЛЯ АППРОКСИМАЦИИ ДАННЫХ ИЗМЕРЕНИЙ

© 2022 А.А. Ломов<sup>1,2</sup>, Е.А. Русинова<sup>2</sup>

<sup>1</sup>Институт математики им. С. Л. Соболева СО РАН  
(630090 Новосибирск, пр. Академика Коптюга, д. 4),

<sup>2</sup>Новосибирский государственный университет  
(630090 Новосибирск, ул. Пирогова, д. 1)

E-mail: lomov@math.nsc.ru, a.lomov@g.nsu.ru

Поступила в редакцию: 13.01.2022

В работе проводится сравнение двух целевых функций в задаче Прони аппроксимации данных измерений решениями линейного дифференциального уравнения заданного порядка с постоянными коэффициентами. Целевые функции различаются типом зависимости градиента от коэффициентов уравнения (линейная или со сложной нелинейностью) и являются 1) нормой невязки уравнения (линейный метод наименьших квадратов) или 2) нормой ошибки аппроксимации по А. Хаусхолдеру (вариационный метод идентификации). В последнем случае производится совместная оптимизация коэффициентов дифференциального уравнения и начальных условий решения. Для рассмотренных целевых функций вычислены константы локальной устойчивости решения задачи Прони с использованием локальных разложений зависимостей оптимальных коэффициентов уравнения как неявных функций от данных из условия равенства градиента целевой функции нулю. На этой основе предложен способ определения допустимой погрешности в данных задачи для обеспечения заданного уровня отклонения решения от истинного значения. На примере К. Ланцоша вычисления показателей экспонент по наблюдениям суммы трех экспонент с ошибками округления показано существенное преимущество (с точки зрения допустимой погрешности в данных) использования вариационной целевой функции. Адекватность используемых локальных показателей устойчивости для немалых возмущений проверяется численным экспериментом.

*Ключевые слова:* аппроксимация данных измерений, задача Прони, пример К. Ланцоша выделения показательных функций, локальная устойчивость, метод наименьших квадратов, вариационный метод.

### ОБРАЗЕЦ ЦИТИРОВАНИЯ

Ломов А.А., Русинова Е.А. Сравнение целевых функций в задаче Прони для аппроксимации данных измерений // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2022. Т. 11, № 2. С. 18–29. DOI: 10.14529/cmse220202.

### Введение

При аппроксимации данных измерений широко используется принцип минимизации невязки подбираемого уравнения. В задачах аппроксимации типа Прони этот подход приводит к переопределенной системе уравнений с решением в виде матричной формулы линейного метода наименьших квадратов [1, 2]. В отличие от целевой функции по невязке уравнения, целевые функции по ошибкам в переменных (невязке сигналов) не имеют явных формул для точек минимума, вследствие чего возникают вопросы сходимости вычислительных алгоритмов [1, 3, 4].

Интуитивные предположения (подтверждаемые асимптотической теорией [5]) состоят в том, что при случайных аддитивных возмущениях в аппроксимируемых данных предпочтительны целевые функции по ошибкам в переменных. В статье предлагается обоснование этого факта в неасимптотическом случае произвольных конечных выборок путем сравнения целевых функций по количественным показателям локальной устойчивости решений [6, 7].

В пределе больших объемов выборок предлагаемые показатели совпадают с известными асимптотическими критериями, основанными на информационной матрице [6]. В качестве примера применения методики рассматривается задача К. Ланцоша восстановления экспоненциальных слагаемых из возмущенных наблюдений суммы трех экспонент [8]. Теоретические расчеты показывают, что в задаче К. Ланцоша более сложная целевая функция (по ошибкам в переменных) приводит к решениям, которые на порядок более точны, чем решения, полученные по невязке уравнения. Результаты подтверждаются вычислительным экспериментом.

Заметим, что в настоящее время, учитывая мощность доступных вычислительных устройств, вопросы трудоемкости решения отходят на второй план, и поэтому в статье они не рассматриваются. Главным становится вопрос о гарантиях сходимости алгоритма. Если имеет место сходимость «почти наверное» за конечное число шагов [9], то вычислительный алгоритм, если пренебречь трудоемкостью, приближается по эффективности к явной формуле.

Статья организована следующим образом. В разделе 1 кратко описаны известные постановки задачи Прони, приведены соответствующие целевые функции, охарактеризованы методы решения. В разделе 2 предложена методика сравнения целевых функций через вычисление констант локальной устойчивости решений задачи Прони. В разделе 3 описан вычислительный эксперимент с приложением методики сравнения к задаче К. Ланцоша восстановления показателей экспонент по наблюдениям их суммы с возмущениями. В заключении приводится краткая сводка результатов, полученных в работе, и указаны направления дальнейших исследований.

## 1. Целевые функции в задаче Прони

Задача Прони [10] является исторически первой задачей аппроксимации наблюдений решениями линейного дифференциального или разностного уравнения заданного порядка с неопределенными коэффициентами. Задачи такого типа возникают во многих областях: в геологии [11], радиоэлектронике, спектральном анализе, медицине и биоинформатике [12] и др. Гаспар Риш де Прони<sup>1</sup> предложил метод построения разностного уравнения, описывающего наблюдения процесса испарения смеси жидкостей [10]. Он использовал предположение о малости возмущений в наблюдениях; тем не менее, в предложенном им решении были заложены идеи, позволившие впоследствии решить задачу в общем случае немалых возмущений с использованием метода наименьших квадратов, опубликованного немного позже в работах К. Гаусса и П. Лежандра [13].

В литературе нет устоявшейся традиции употребления словосочетаний «задача Прони», «метод Прони». Мы предпочитаем говорить о «задачах типа Прони». Использование различных целевых функций для нахождения коэффициентов аппроксимирующего уравнения формально приводит к разным оптимизационным задачам, или, в другой терминологии, к разным методам решения «задачи Прони».

*Классическая задача и метод Прони для малых возмущений.* Введем вектор наблюдений  $f = (f_1 \dots f_N)^\top \in \mathbb{R}^N$ , где  $N$  — число моментов времени (длина процесса), и  $f_i$  — значения некоторой функции на равномерной временной сетке (неравномерная сетка сводится к равномерной аффинными преобразованиями [2]). Классическая задача Прони

<sup>1</sup>Сам Г. Р. де Прони подписал свою работу как «Par R. Prony» («Автор Р. Прони») [10]

состоит в аппроксимации  $f$  линейной комбинацией  $z = (z_1 \dots z_N)^\top$  синусоид или экспонент:

$$z_k \doteq \sum_{i=1}^l (C_i e^{\lambda_i k} + \overline{C_i} e^{\overline{\lambda_i} k}) \approx f_k, \quad (1)$$

где  $k = \overline{1, N}$ ,  $l = \lceil \frac{N}{2} \rceil$ ,  $\lambda_i, C_i \in \mathbb{C}$ ,  $\forall i = \overline{1, l}$ , черта в  $\overline{\lambda_i}, \overline{C_i}$  означает комплексное сопряжение. Коэффициенты  $\lambda_1, \dots, \lambda_l, C_1, \dots, C_l$  подлежат определению, их найденные значения вместе с  $z_1, \dots, z_N$  считаются решением задачи Прони.

Решение достигается за два шага. Сначала определяются показатели  $\lambda_1, \dots, \lambda_l$ . Ключевая идея Прони состоит в том, чтобы сумму (1) рассматривать как решение некоторого разностного уравнения:

$$\alpha_n z_{k+n} + \alpha_{n-1} z_{k+n-1} + \dots + \alpha_0 z_k = 0, \quad (2)$$

где  $n = 2l$ ,  $k = \overline{1, N-n}$ ,  $\alpha_n = 1$ . Коэффициенты  $\alpha_0, \dots, \alpha_{n-1}$  находятся из предположения, что измерения  $f_k$  близки к точному процессу  $z_k$ . Составляется система из  $n$  уравнений для  $n$  неизвестных:

$$\begin{cases} f_{n+1} + \alpha_{n-1} f_n + \dots + \alpha_0 f_1 \approx 0, \\ f_{n+2} + \alpha_{n-1} f_{n+1} + \dots + \alpha_0 f_2 \approx 0, \\ \vdots \\ f_{2n} + \alpha_{n-1} f_{2n-1} + \dots + \alpha_0 f_n \approx 0. \end{cases} \quad (3)$$

Определив коэффициенты  $\alpha_0, \dots, \alpha_{n-1}$  из последней системы уравнений, можно найти значения  $\lambda_1 = \ln \zeta_1, \dots, \lambda_l = \ln \zeta_l$  через корни  $\zeta_1, \dots, \zeta_l, \overline{\zeta_1}, \dots, \overline{\zeta_l}$  характеристического многочлена  $\alpha(\zeta) = \zeta^n + \alpha_{n-1} \zeta^{n-1} + \dots + \alpha_0$ .

На втором шаге вычисляются коэффициенты  $C_1, \dots, C_l$ , исходя из полученных на первом шаге значений  $\lambda_1, \dots, \lambda_l$ . Составляется представление, устанавливающее взаимно-однозначное соответствие между комплексными  $C_1, \dots, C_l$  и вещественными коэффициентами  $A_{11}, \dots, A_{1l}, A_{21}, \dots, A_{2l}$ :

$$C_i e^{\lambda_i k} + \overline{C_i} e^{\overline{\lambda_i} k} = (A_{1i} \sin(\omega_i k) + A_{2i} \cos(\omega_i k)) e^{\tau_i k}, \\ \omega = \arg(\lambda), \quad \tau = \text{mod}(\lambda).$$

Для нахождения  $A_{1i}, A_{2i}$ ,  $i = \overline{1, l}$ , используется система из  $n = 2l$  уравнений:

$$\begin{cases} f_1 \approx \sum_{i=1}^l e^{\tau_i} (A_{1i} \sin(\omega_i) + A_{2i} \cos(\omega_i)), \\ f_2 \approx \sum_{i=1}^l e^{2\tau_i} (A_{1i} \sin(2\omega_i) + A_{2i} \cos(2\omega_i)), \\ \vdots \\ f_n \approx \sum_{i=1}^l e^{n\tau_i} (A_{1i} \sin(n\omega_i) + A_{2i} \cos(n\omega_i)). \end{cases} \quad (4)$$

Найдя решение  $A_{1i}, A_{2i}$ , восстанавливаем коэффициенты  $C_i$ ,  $i = \overline{1, l}$ .

«Обобщенный» метод Прони [1] (раздел 11.5), [3]. Для нахождения коэффициентов  $\alpha_0, \dots, \alpha_{n-1}$  вместо (3) используется переопределенная система уравнений

$$\begin{pmatrix} f_1 & \dots & f_n & | & f_{n+1} \\ f_2 & \dots & f_{n+1} & | & f_{n+2} \\ \vdots & \ddots & \vdots & | & \vdots \\ f_{N-n} & \dots & f_{N-1} & | & f_N \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_{n-1} \\ 1 \end{pmatrix} \doteq \begin{pmatrix} V_1 & | & V_2 \end{pmatrix} \begin{pmatrix} \alpha \\ 1 \end{pmatrix} \doteq e. \quad (5)$$

Из условия минимума целевой функции метода наименьших квадратов

$$J = \|e\|^2 \rightarrow \min_{\alpha} \quad (6)$$

следует решение

$$\hat{\alpha} = - \left( V_1^T V_1 \right)^{-1} \left( V_1^T V_2 \right). \quad (7)$$

Для нахождения  $A_{1i}, A_{2i}, i = \overline{1, l}$ , вместо (4) используется система уравнений

$$\begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix} - \begin{pmatrix} e^{\tau_1} \sin(\omega_1) & \dots & e^{\tau_1} \cos(\omega_l) \\ \vdots & & \vdots \\ e^{N\tau_1} \sin(N\omega_1) & \dots & e^{N\tau_1} \cos(N\omega_l) \end{pmatrix} \begin{pmatrix} A_{11} \\ \vdots \\ A_{2l} \end{pmatrix} \doteq \begin{pmatrix} f \\ \vdots \\ f \end{pmatrix} - H\theta \doteq \varepsilon,$$

По методу наименьших квадратов из условия минимума  $\|\varepsilon\|^2 \rightarrow \min_{\theta}$  получаем

$$\hat{\theta} = \left( H^T H \right)^{-1} H^T f, \quad z = H\hat{\theta}.$$

Если нужно найти только аппроксимирующую функцию  $z$ , можно обойтись без поиска корней многочлена  $\alpha(\zeta)$  и вычисления матрицы  $H$ :

$$z = \left[ I - G^T \left( GG^T \right)^{-1} G \right] f, \quad (8)$$

$$G \doteq \begin{pmatrix} \alpha_0 & \dots & \alpha_{n-1} & 1 & & 0 \\ & \alpha_0 & \dots & \alpha_{n-1} & 1 & \\ & & \ddots & & \ddots & \ddots \\ 0 & & & \alpha_0 & \dots & \alpha_{n-1} & 1 \end{pmatrix}.$$

Полная задача и «модифицированный» метод Прони. Из постановки задачи аппроксимации (1) вытекает целевая функция [3]

$$J_{\Pi} \doteq \|f - z\|^2 \rightarrow \min_{Gz=0}, \quad (9)$$

которую можно использовать для поиска коэффициентов  $\alpha$  вместо целевой функции МНК (5), (6). Соответствующая задача минимизации уже не имеет решения в виде явных формул [1] (раздел 11.3), [3]. А. Хаусхолдер [3] использовал алгоритм типа Ньютона в пространстве переменных  $z, \alpha$  и множителей Лагранжа. Задачу (9) будем называть полной задачей

Прони (ПП) («вариационная задача идентификации», «вариационный метод» в [4], «модифицированный метод Прони» в [14]), а целевую функцию  $J_{\Pi}$  — вариационной. Задача (9) была исследована М. Осборном [15] и А. О. Егоршиным [16], предложившими независимо друг от друга<sup>2</sup> эффективные вычислительные алгоритмы решения ПП с гарантированной при малых возмущениях сходимостью из «почти любого» начального приближения [9]. Для алгоритмов, основанных на локальных оценках производных, в частности, алгоритма А. Хаусхолдера [3], насколько известно, такого рода гарантий не существует.

## 2. Устойчивость решений и вычисление допустимой погрешности

Рассмотрим оценку параметров  $\hat{\alpha}(f) = \arg \min_{\alpha} J(\alpha, f)$  как функцию наблюдений  $f$ . Целевая функция  $J(\alpha, f)$  может иметь вид МНК (6), ПП (9). Предположим, что  $f = z + \Delta z$ , где  $\Delta z$  — малое возмущение, и выполнено условие единственности<sup>3</sup>  $\hat{\alpha}(z) = \alpha$ . Тогда  $\hat{\alpha}(f) = \alpha + \Delta\alpha$ , где  $\Delta\alpha$  — отклонение оценки параметров. Определим матрицу производных (см. [17, формула (1.2.5)])

$$S^{\top} = \frac{d\hat{\alpha}}{df} = - (J''_{\alpha\alpha})^{-1} J''_{\alpha f} \Big|_{f=z} \in \mathbb{R}^{n \times N}.$$

Пусть  $B_{\sigma} \doteq \{\|\Delta z\| \leq \sigma\}$  — шар возмущений. Для малых  $\sigma$  множеству наблюдений  $f = z + B_{\sigma}$  при отображении  $f \mapsto \hat{\alpha}(f)$  соответствует эллипсоид  $E_{\sigma}$  значений оценки  $\hat{\alpha}(f) = \alpha + E_{\sigma}$ , который определяется с помощью матрицы  $S^{\top}$ :

$$E_{\sigma} = S^{\top} B_{\sigma} = \left\{ \Delta\alpha : \Delta\alpha^{\top} (S^{\top} S)^{-1} \Delta\alpha \leq \sigma^2 \right\}.$$

Матрицу  $S^{\top} S$  будем называть *матрицей эллипсоида рассеяния*. Наибольшее отклонение оценки  $\|\Delta\alpha\|_{\max}$  определяется вдоль наибольшей полуоси эллипсоида  $E_{\sigma}$ , откуда следует соотношение

$$\|\Delta z\|_{\max} = \sigma = \frac{\|\Delta\alpha\|_{\max}}{\sqrt{\lambda_{\max}(S^{\top} S)}}. \quad (10)$$

Как следствие, для нахождения допустимого значения погрешности наблюдений  $\|\Delta z\| \leq \|\Delta z\|_{\max}$  необходимо задать желаемый относительный уровень  $p = \frac{\|\Delta\alpha\|_{\max}}{\|\alpha\|}$  погрешности вычисления коэффициентов и найти наибольшее собственное значение матрицы эллипсоида рассеяния  $S^{\top} S$ , которое и определяет искомую константу устойчивости решения.

**Утверждение 1.** [7] Для целевых функций МНК (6) и ПП (9) верны следующие формулы нахождения матриц эллипсоидов рассеяния.

$$S^{\top} S \Big|_{\text{МНК}} = (V_1^{\top} V_1)^{-1} V_1^{\top} G G^{\top} V_1 (V_1^{\top} V_1)^{-1}, \quad (11)$$

$$S^{\top} S \Big|_{\text{ПП}} = \left[ V_1^{\top} (G G^{\top})^{-1} V_1 \right]^{-1}. \quad (12)$$

*Доказательство.* Из формул (19)–(21) статьи [7] прямо следуют выражения

$$\Delta\hat{\alpha} \Big|_{\text{МНК}} = - (V_1^{\top} V_1)^{-1} V_1^{\top} \begin{pmatrix} \Delta V_1 \\ \Delta V_2 \end{pmatrix} \begin{pmatrix} \alpha \\ 1 \end{pmatrix} = - (V_1^{\top} V_1)^{-1} V_1^{\top} G \Delta z,$$

<sup>2</sup>Работы [3, 15] были практически недоступны в СССР в 1970-е гг.

<sup>3</sup>Это условие равносильно линейной независимости столбцов матрицы наблюдений  $V_1$ .

$$\Delta\hat{\alpha}|_{\text{III}} = - \left( V_1^\top C V_1 \right)^{-1} V_1^\top C G \Delta z, \quad C = \left( G G^\top \right)^{-1},$$

из которых получаем

$$S^\top|_{\text{MHK}} = - \left( V_1^\top V_1 \right)^{-1} V_1^\top G, \quad S^\top|_{\text{III}} = - \left( V_1^\top C V_1 \right)^{-1} V_1^\top C G.$$

Отсюда сразу следуют формулы (11), (12). □

**Замечание 1.** Пусть вектор коэффициентов  $\tilde{\alpha} \doteq (\alpha_0 \ \dots \ \alpha_{n-1} \ \alpha_n)^\top$  зависит аффинным образом от некоторого векторного параметра

$$a \doteq (a_0 \ \dots \ a_{n-1})^\top, \quad \tilde{\alpha} = D a + d, \tag{13}$$

где матрица  $D \in \mathbb{R}^{(n+1) \times n}$  и столбец  $d \in \mathbb{R}^{(n+1) \times 1}$  заданы,  $\text{rank} \begin{pmatrix} D & d \end{pmatrix} = n + 1$ . Тогда в формулах (7), (11), (12) следует заменить  $\alpha$  на  $a$ , матрицу  $V_1$  и столбец  $V_2$  соответственно на матрицу  $W_1 \doteq \begin{pmatrix} V_1 & | & V_2 \end{pmatrix} D$  и столбец  $W_2 \doteq \begin{pmatrix} V_1 & | & V_2 \end{pmatrix} d$  [7].

Линейное соотношение (10) применимо в предельном случае малых возмущений  $\sigma \rightarrow 0$ . Остается открытым вопрос, для каких величин  $\sigma$  линейное приближение  $\Delta\hat{\alpha} = S^\top \Delta z$  зависимости  $\Delta\hat{\alpha}(\Delta z)$  и соответствующая оценка (10) сохраняют актуальность. Известные теоретические оценки допустимых  $\sigma$  крайне малы [7]. В рассматриваемом ниже примере адекватность соотношения (10) проверяется вычислительным экспериментом.

### 3. Пример К. Ланцоша

Рассмотрим пример К. Ланцоша ([8], гл. IV, раздел 23) неустойчивости задачи восстановления трех экспонент при погрешности в третьем разряде наблюдений  $f$ . Дан точный процесс

$$z(t) = 0.0951e^{-t} + 0.8607e^{-3t} + 1.5576e^{-5t}, \tag{14}$$

описываемый дифференциальным уравнением

$$z''' + a_2 z'' + a_1 z' + a_0 z = 0 \tag{15}$$

с «истинными» значениями коэффициентов

$$a_2 = 9, \quad a_1 = 23, \quad a_0 = 15. \tag{16}$$

Для восстановления показателей и амплитуд экспонент используются наблюдения  $f_k = z_k + \Delta z_k$ ,  $k = 1, \dots, 24$  процесса  $z(t)$  в равноотстоящие с шагом  $h = 0.05$  моменты времени с ошибками округления  $\Delta z_k$  (табл. 1).

Решая задачу, К. Ланцош получил аппроксимацию  $\hat{z}$ :

$$\hat{z}_k = 2.202 e^{-4.45 kh} + 0.305 e^{-1.58 kh}, \quad k = \overline{1, 24}. \tag{17}$$

Простая проверка показывает, что расхождения между найденным решением  $\hat{z}$  и наблюдениями  $f$  нигде не превышают 0.006, а среднее отклонение равно 0.0026, что не превышает пределов ошибок данных, при этом в полученной функции (17) не просто утеряна третья экспонента с показателем  $-5t$ , но ее отсутствие оказало влияние на показатели двух других экспонент. Отношение амплитуд экспонент также искажено.

Таблица 1. Данные наблюдений

$k$	$f_k$	$k$	$f_k$	$k$	$f_k$	$k$	$f_k$
1	2.51	7	0.77	13	0.27	19	0.11
2	2.04	8	0.64	14	0.23	20	0.10
3	1.67	9	0.53	15	0.20	21	0.09
4	1.37	10	0.45	16	0.17	22	0.08
5	1.12	11	0.38	17	0.15	23	0.07
6	0.93	12	0.32	18	0.13	24	0.06

Рассмотрение случая, когда заранее известно число экспонент в искомой функции, также не дает решения, близкого к истинному; полученная К.Ланцошем функция из трех экспонент имеет вид

$$\hat{z}_k = 0.041 e^{-0.50 kh} + 0.79 e^{-2.73 kh} + 1.68 e^{-4.96 kh},$$

также с отличием от наблюдений в пределах ошибок данных.

Анализируя полученные решения, К.Ланцош делает вывод, что никакие математические методы не смогут привести к лучшим результатам, т.к. проблема сокрыта именно в округлениях наблюдений.

Далее обоснуем этот вывод К.Ланцоша на основе соотношения (10) и вычислим погрешности наблюдений, при которых можно получить приемлемые по точности значения коэффициентов дифференциального уравнения (15) и показателей экспонент в выражении (14).

Будем использовать аффинное приближение вида (13) для связи коэффициентов разностного (2) и дифференциального (15) уравнений (следствие формул (21), (23), (25) из [18]):

$$\tilde{\alpha} = T^{-\top} \tilde{a}, \quad \tilde{a} = \begin{pmatrix} I \\ 0 \end{pmatrix} a + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \doteq D_a a + d_a, \quad (18)$$

$$T \doteq \|t_{ij}\|_{i,j \in \overline{1, n+1}}, \quad t_{ij} = \frac{\left(-\frac{n}{2}h + (i-1)h\right)^{j-1}}{(j-1)!},$$

$$D = T^{-\top} D_a, \quad d = T^{-\top} d_a.$$

Заметим, что ввиду приближенности дискретизации (18) никакое решение разностного уравнения (2) с коэффициентами  $\tilde{\alpha}$  (18) точно не совпадает с функцией  $z(t)$  (14) в узлах сетки  $t_k = kh$ . Поэтому выберем решение  $\hat{z}_*$ , ближайшее к точному  $z_*$  (14), и назовем  $\hat{z}_*$  эталоном, прибавляя в вычислительных экспериментах возмущения к эталону. Этим будет обеспечено, что при уменьшении возмущений до нуля получаемое в результате решения обратной задачи разностное уравнение будет иметь решение, совпадающее с эталоном, и коэффициенты  $a$  из формул (18) будут совпадать при нулевых возмущениях с «истинными» значениями (16).

Эталонное решение  $\hat{z}_*$  вычисляем из точного  $z_*$ , взятого в узлах сетки  $t_k = kh$  (14), путем проецирования  $z_*$  на множество решений разностного уравнения (2) с коэффициентами  $\tilde{\alpha}_* = T^{-\top} \begin{pmatrix} 15 & 23 & 9 & 1 \end{pmatrix}^\top$ . Эта операция соответствует второму шагу решения

задачи Прони, см. (8):

$$\hat{z}_* = \left[ I - G^\top (GG^\top)^{-1} G \right] z_*, \quad G = G(\tilde{\alpha}_*).$$

Для расчетов допустимой погрешности в наблюдениях  $f$  при идентификации коэффициентов  $a$  (18) дифференциального уравнения (15) с заданной точностью  $p = \frac{\|\Delta a\|_{\max}}{\|a\|} = 0.1$  использовались формулы (10)–(12) с учетом замечания 1. В результате получено, что допустимая погрешность наблюдений при решении полной задачи Прони ограничена величиной

$$\|\Delta z\|_{\max}^{\text{ПП}} \simeq 3.2 \cdot 10^{-5}. \quad (19)$$

Если использовать метод наименьших квадратов, то получится на порядок менее выгодное значение

$$\|\Delta z\|_{\max}^{\text{МНК}} \simeq 3.6 \cdot 10^{-6}. \quad (20)$$

Теоретические расчеты подтверждаются результатами моделирования. Для минимизации ЦФ ПП использовался вычислительный алгоритм E73-HP [9]. В табл. 2 приведены средние значения для норм относительных отклонений вектора параметров  $a$  (13) и показателей экспонент от истинных значений (16),  $\lambda_1 = -1$ ,  $\lambda_2 = -3$ ,  $\lambda_3 = -5$ . Усреднение проводилось по  $M = 100$  экспериментам с независимыми случайными наблюдениями  $f(i) = \hat{z}_* + \Delta z(i)$ ,  $i \in \overline{1, M}$ . Возмущения  $\Delta z(i)$  были распределены равномерно на интервале  $[-\sigma, \sigma]$ . Относительные отклонения  $|\Delta_{\text{отн}} x|$  оценок  $x \in \{\hat{a}, \hat{\lambda}_{1 \div 3}\}$  вычислялись по формуле  $|\Delta_{\text{отн}} x| \doteq \frac{\|x - x_*\|}{\|x_*\|}$ , где  $x_*$  — истинное значение, а черта означает усреднение по  $M$  экспериментам. В теле таблицы в числителях приведены результаты расчетов с ЦФ ПП, в знаменателях — с ЦФ МНК. Если относительное смещение  $\frac{\|\Delta x\|}{\|x\|}$  превышало единицу, то данные заменялись звездочками \* \* \*. Для уровней возмущений вблизи теоретических границ (19), (20) отличия относительных смещений оценок с ЦФ ПП, МНК от расчетной величины  $p = 0.1$  оказываются несущественными.

В аналогичных расчетах для разностного уравнения (когда в (18) заменяется  $T$  на  $I$  и, соответственно,  $a$  на  $\alpha$ ) для разбросов оценок корней  $\lambda_{1 \div 3}$  в показателях экспонент (14) получаются качественно те же результаты, что и в табл. 2. Разбросы оценок коэффициентов  $\alpha$  разностного уравнения получаются на порядок меньше, чем разбросы оценок коэффициентов  $a$  дифференциального уравнения. Допустимые погрешности наблюдений (19) и (20) для оценок коэффициентов разностного уравнения получаются на порядок более выгодными, чем для коэффициентов дифференциального уравнения ( $4.7 \cdot 10^{-4}$  в (19) и  $5.5 \cdot 10^{-5}$  в (20)).

## Заключение

Количественные показатели локальной устойчивости решения выступают как удобный инструмент для сравнения целевых функций в задаче аппроксимации данных измерений. В статье на основе линейных приближений получены выражения для констант локальной устойчивости решений задачи Прони при использовании целевой функции МНК по норме невязки уравнения и вариационной целевой функции по норме ошибки аппроксимации. Исходя из полученных выражений на примере задачи К. Ланцоша показано, что для восстановления коэффициентов и корней характеристического многочлена дифференциального уравнения, решения которого наилучшим образом аппроксимируют данные возмущенных

**Таблица 2.** Результаты оценивания коэффициентов и корней характеристического многочлена ДУ для целевых функций ПП (в числителе) и МНК (в знаменателе); \*\*\* — относительное смещение  $> 1$

$\sigma$	$ \Delta_{\text{отн}} \hat{a} $ $\frac{\text{ПП}}{\text{МНК}}$	$ \Delta_{\text{отн}} \hat{\lambda}_3 $ $\frac{\text{ПП}}{\text{МНК}}$	$ \Delta_{\text{отн}} \hat{\lambda}_2 $ $\frac{\text{ПП}}{\text{МНК}}$	$ \Delta_{\text{отн}} \hat{\lambda}_1 $ $\frac{\text{ПП}}{\text{МНК}}$
$3 \cdot 10^{-7}$	$\frac{4.1 \cdot 10^{-4}}{3.9 \cdot 10^{-3}}$	$\frac{5.8 \cdot 10^{-6}}{4.7 \cdot 10^{-5}}$	$\frac{1.3 \cdot 10^{-4}}{1.2 \cdot 10^{-3}}$	$\frac{5.3 \cdot 10^{-4}}{5.1 \cdot 10^{-3}}$
$3 \cdot 10^{-6}$	$\frac{4.6 \cdot 10^{-3}}{3.6 \cdot 10^{-2}}$	$\frac{7.2 \cdot 10^{-5}}{4.9 \cdot 10^{-4}}$	$\frac{1.5 \cdot 10^{-3}}{1.1 \cdot 10^{-2}}$	$\frac{5.9 \cdot 10^{-3}}{4.7 \cdot 10^{-2}}$
$3 \cdot 10^{-5}$	$\frac{0.043}{0.47}$	$\frac{5.8 \cdot 10^{-4}}{1.5 \cdot 10^{-2}}$	$\frac{0.013}{0.20}$	$\frac{0.055}{0.45}$
$3 \cdot 10^{-4}$	$\frac{0.37}{***}$	$\frac{0.0091}{***}$	$\frac{0.12}{0.71}$	$\frac{0.44}{***}$

наблюдений, предпочтительней использовать вариационную целевую функцию. Дальнейшие исследования направлены на изучение допустимых уровней возмущений в данных, при которых полученные по линейным приближениям оценки констант устойчивости сохраняют актуальность.

*Работа выполнена в рамках государственного задания Института математики им. С.Л. Соболева СО РАН (проект № FWNF-2022-0008).*

## Литература

1. Марпл С.Л. Цифровой спектральный анализ и его приложения. М.: Мир, 1990. 265 с.
2. Бердышев В.И., Петрак Л.В. Аппроксимация функций, сжатие численной информации, приложения. Екатеринбург: Уро РАН, 1999. 296 с.
3. Householder A.S. On Prony's method of fitting exponential decay curves and multiple-hit survival curves. Oak Ridge National Lab. Report ORNL-455. 1950. Oak Ridge, Tennessee. URL: <http://www.technicalreports.org/trail/detail/11105> (дата обращения: 15.12.2021).
4. Егоршин А.О. Метод наименьших квадратов и «быстрые» алгоритмы в вариационных задачах идентификации и фильтрации (метод ВИ) // Автометрия. 1988. № 1. С. 30–42. URL: [http://www.iae.nsk.su/images/stories/5\\_Autometria/5\\_Archives/1988/1/30-42.pdf](http://www.iae.nsk.su/images/stories/5_Autometria/5_Archives/1988/1/30-42.pdf) (дата обращения: 15.12.2021).
5. Fuller W.A. Measurement Error Models. New York: Wiley, 1987. 440 p. DOI: 10.1002/9780470316665.
6. Ломов А.А. О количественных априорных показателях идентифицируемости коэффициентов линейных динамических систем // Известия РАН ТСУ. 2011. № 1. С. 3–15. DOI: 10.1134/S106423071101014X.

7. Ломов А.А., Федосеев А.В. Сравнение методов параметрической идентификации линейных динамических систем в условиях смешанных возмущений // Сибирский журнал чистой и прикладной математики. 2018. Т. 18, № 3. С. 45–59. DOI: 10.17377/PAM.2018.18.6.
8. Ланцош К. Практические методы прикладного анализа. Справочное руководство. М.: Физматгиз, 1961. 524 с.
9. Ломов А.А. О сходимости вычислительных алгоритмов в вариационной задаче идентификации коэффициентов разностных уравнений // Сибирский журнал индустриальной математики. 2020. Т. 83, № 3. С. 77–90. DOI: 10.33048/SIBJIM.2020.23.307.
10. de Prony G. Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures // Journal de l'école Polytechnique. 1795. Vol. 1, no. 22. P. 24–76. URL: <http://users.polytech.unice.fr/~leroux/PRONY.pdf> (дата обращения: 15.12.2021).
11. Mitrofanov G., Priimenko V. Prony Filtering of Seismic Data // Acta Geophysica. 2015. Vol. 63, no. 3. P. 652–678. DOI: 10.1515/acgeo-2015-0012.
12. Коломейцева А.В., Мишугова Г.В., Мул А.П., Рябых Г.Ю. Применение вейвлет-преобразования и метода Прони для идентификации биогенных сигналов // Вестник ДГТУ. 2010. Т. 10, № 4. С. 455–465.
13. Bjorck A. Numerical methods for least squares problems. USA: SIAM, 1996. 425 p. DOI: 10.1137/1.9781611971484.ch1.
14. Keller I., Plonka G. Modifications of Prony's Method for the Recovery and Sparse Approximation with Generalized Exponential Sums // Approximation Theory XVI / ed. by G.E. Fasshauer, M. Neamtu, L.L. Schumaker. Cham: Springer International Publishing, 2021. P. 123–152. DOI: 10.1007/978-3-030-57464-2\_7.
15. Osborne M.R. A class of nonlinear regression problems // Data representation / ed. by R.S. Anderssen, M.R. Osborne. St. Lucia: University of Queensland Press, 1970. P. 94–101.
16. Егоршин А.О., Будянов В.П. Сглаживание сигналов и оценивание динамических параметров в автоматических системах с помощью ЦВМ // Автометрия. 1973. № 1. С. 78–82. URL: [https://www.iae.nsk.su/images/stories/5\\_Autometria/5\\_Archives/1973/1/78-82.pdf](https://www.iae.nsk.su/images/stories/5_Autometria/5_Archives/1973/1/78-82.pdf) (дата обращения: 15.12.2021).
17. Брайсон А., Хо Ю-Ши. Прикладная теория оптимального управления: оптимизация, оценка и управление. М.: Мир, 1972. 544 с.
18. Егоршин А.О. О дискретизации линейных дифференциальных уравнений // Вестник ЮУрГУ. Серия: Матем. моделирование и программирование. 2012. Т. 40(299), № 14. С. 59–72. URL: <https://mmp.susu.ru/article/en/179> (дата обращения: 15.12.2021).

Ломов Андрей Александрович, д.ф.-м.н., доцент, с.н.с. Института математики им. С.Л. Соболева СО РАН, профессор кафедры теоретической кибернетики механико-математического факультета и кафедры компьютерных систем факультета информационных технологий Новосибирского государственного университета (национальный исследовательский университет) (Новосибирск, Российская Федерация)

Русинова Елизавета Александровна, магистрант кафедры теоретической кибернетики механико-математического факультета Новосибирского государственного университета (национальный исследовательский университет) (Новосибирск, Российская Федерация)

## COMPARISON OF THE TARGET FUNCTIONS IN THE PRONY'S PROBLEM OF MEASUREMENT DATA APPROXIMATION

© 2022 A.A. Lomov<sup>1,2</sup>, E.A. Rusinova<sup>2</sup>

<sup>1</sup>*Sobolev Institute of Mathematics (pr. Acad. Koptyuga 4, Novosibirsk, 630090 Russia),*

<sup>2</sup>*Novosibirsk State University (Pirogova str. 1, Novosibirsk, 630090 Russia)*

*E-mail: lomov@math.nsc.ru, a.lomov@g.nsu.ru*

Received: 13.01.2022

In the article, we compare two objective functions in the Prony's problem of approximation of measurement data by solutions of a linear differential equation of a given order with constant coefficients. The target functions differ in the type of dependence of the gradient on the coefficients of the equation (linear or with complex nonlinearity) and are 1) the norm of the residual of the equation (linear least squares method) or 2) the norm of the approximation error according to A. Householder (variational identification method). In the latter case, the coefficients of the differential equation and the initial conditions of the solution are jointly optimized. For the considered objective functions, the local stability constants of the solution to the Prony's problem are calculated using local expansions of the dependencies of the optimal coefficients of the equation as implicit functions of the data with the condition that the gradient of the objective function is identically equal to zero. On this basis, a method is proposed for determining the permissible error in the data to ensure a given level of deviation of the solution from the true value. We use the example of K. Lanczos of calculating the exponents given observations of the sum of three exponents with rounding errors to confirm a significant advantage (in terms of the allowable errors in the data) of using the variational objective function. The adequacy of the used local stability indices for considerable perturbations is verified by numerical experiment.

*Keywords: measurement data approximation, Prony's problem, C. Lanczos example of separation of exponentials, local stability, least squares method, variational method.*

### FOR CITATION

Lomov A.A., Rusinova E.A. Comparison of the Target Functions in the Prony's Problem of Measurement Data Approximation. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2022. Vol. 11, no. 2. P. 18–29. (in Russian) DOI: 10.14529/cmse220202.

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

### References

1. Marple S.L. Digital spectral analysis: with applications. USA: Prentice-Hall, 1986. 492 p.
2. Berdyshev V.I., Petrak L.V. Function approximation, compression of numerical information, applications. Ekaterinburg: UB RAN, 1999. 296 p. (in Russian)
3. Householder A.S. On Prony's method of fitting exponential decay curves and multiple-hit survival curves. Oak Ridge National Lab. Report ORNL-455. 1950. Oak Ridge, Tennessee. URL: <http://www.technicalreports.org/trail/detail/11105> (accessed: 15.12.2021).
4. Egorshin A.O. Least squares method and the "fast" algorithms in variational problems of identification and filtration (VI method). Avtometriya. 1988. No. 1. P. 30–42.

- (in Russian). URL: [http://www.iae.nsk.su/images/stories/5\\_Autometria/5\\_Archives/1988/1/30-42.pdf](http://www.iae.nsk.su/images/stories/5_Autometria/5_Archives/1988/1/30-42.pdf) (accessed: 15.12.2021).
5. Fuller W.A. Measurement Error Models. New York: Wiley, 1987. 440 p.
  6. Lomov A.A. On Quantitative A Priori Measures of Identifiability of Coefficients of Linear Dynamic Systems. *Journal of Computer and System Sciences International*. 2011. Vol. 50, no. 1. P. 1–13. DOI: 10.1134/S106423071101014X.
  7. Lomov A.A., Fedoseev A.V. Comparison of Parameter Identification Methods for Linear Dynamic Systems Under Mixed Noise. *Springer Journal of Mathematical Sciences*. 2021. Vol. 253, no. 3. P. 407–418. DOI: 10.1007/s10958-021-05238-0.
  8. Lanczos C. Applied analysis. USA: Prentice Hall, 1956. 539 p.
  9. Lomov A.A. On Convergence of Computational Algorithms for a Variational Problem of Identifying the Coefficients of Difference Equations. *Journal of Applied and Industrial Mathematics*. 2020. Vol. 14, no. 3. P. 541–554. DOI: 10.1134/S1990478920030138.
  10. de Prony G. Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures. *Journal de l'école Polytechnique*. 1795. Vol. 1, no. 22. P. 24–76. URL: <http://users.polytech.unice.fr/~leroux/PRONY.pdf> (accessed: 15.12.2021).
  11. Mitrofanov G.M., Priimenko V.I. Basics and Applications of the Prony filtering. *Tehnologii seismorazvedki*. 2011. No. 3. P. 93–108. (in Russian)
  12. Kolomeytseva A.V., Mishugova G.V., Mool A.P., Ryabykh G.Yu. Application of the wavelet transform and the Prony method for the identification of biogenic signals. *Vestnik DGTU*. 2010. Vol. 10, no. 4. P. 455–465. (in Russian)
  13. Bjorck A. Numerical methods for least squares problems. USA: SIAM, 1996. 425 p. DOI: 10.1137/1.9781611971484.ch1.
  14. Keller I., Plonka G. Modifications of Prony's Method for the Recovery and Sparse Approximation with Generalized Exponential Sums. *Approximation Theory XVI* / ed. by G.E. Fasshauer, M. Neamtu, L.L. Schumaker. Cham: Springer International Publishing, 2021. P. 123–152. DOI: 10.1007/978-3-030-57464-2\_7.
  15. Osborne M.R. A class of nonlinear regression problems. *Data representation* / ed. by R.S. Anderssen, M.R. Osborne. St. Lucia: University of Queensland Press, 1970. P. 94–101.
  16. Egorshin A.O., Budyanov V.P. Smoothing of signals and estimation of dynamic parameters in automatic systems using a digital computer. *Avtometriya*. 1973. No. 1. P. 78–82. (in Russian) URL: [https://www.iae.nsk.su/images/stories/5\\_Autometria/5\\_Archives/1973/1/78-82.pdf](https://www.iae.nsk.su/images/stories/5_Autometria/5_Archives/1973/1/78-82.pdf) (accessed: 15.12.2021).
  17. Bryson A.E., Ho Y.-C. Applied Optimal Control: Optimization, Estimation and Control. Waltham, MA: Blaisdell. 1969. 481 p.
  18. Egorshin A.O. On discretization of linear differential equations. *Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming & Computer Software*. 2012. Vol. 40(299), no. 14. P. 59–72. (in Russian) URL: <https://mmp.susu.ru/article/en/179> (accessed: 15.12.2021).

## ПРИМЕНЕНИЕ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ ДЛЯ АННОТИРОВАНИЯ СЕНСОРНЫХ ДАННЫХ\*

© 2022 А.И. Гоглачев, М.Л. Цымблер

*Южно-Уральский государственный университет*

*(454080 Челябинск, пр. им. В.И. Ленина, д. 76)*

*E-mail: goglachevai@susu.ru, mzym@susu.ru*

Поступила в редакцию: 04.04.2022

Аннотирование сенсорных данных предполагает автоматизированную разметку временного ряда показаний, снятых с сенсора, которая выделяет различные активности, заданные указанным рядом. Разметка активностей имеет широкий спектр практического применения: предиктивное техническое обслуживание оборудования в приложениях цифровой индустрии, интеллектуальное управление зданиями в приложениях Интернета вещей, мониторинг состояния человека и упреждающая диагностика заболеваний в приложениях персональной медицины и др. В данной статье описаны два тематических исследования по аннотированию временных рядов: показания носимого виброакселерометра, закрепленного на человеке, и стационарного виброакселерометра, установленного на малогабаритной дробильной установке. Данные исследования выполнены разработанным ранее авторами параллельного алгоритма для аннотирования сенсорных данных с помощью графического процессора на основе концепции снippetов. Снippet представляет собой подпоследовательность, на которую похожи многие другие подпоследовательности данного ряда в смысле специализированной меры схожести, основанной на евклидовом расстоянии. Представлены также результаты вычислительных экспериментов по исследованию производительности и качества разметки разработанного алгоритма.

*Ключевые слова: временной ряд, аннотирование, снippet, параллельный алгоритм, графический процессор.*

### ОБРАЗЕЦ ЦИТИРОВАНИЯ

Гоглачев А.И., Цымблер М.Л. Применение параллельных вычислений для аннотирования сенсорных данных // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2022. Т. 11, № 2. С. 30–42. DOI: 10.14529/cmse220203.

### Введение

Аннотирование сенсорных данных предполагает автоматизированную разметку временного ряда показаний, снятых с сенсора, которая выделяет различные активности, заданные указанным рядом. Разметка активностей позволяет кратко описать и визуализировать сенсорные данные и поэтому имеет широкий спектр практического применения: предиктивное техническое обслуживание в цифровой индустрии, умное управление системами жизнеобеспечения [1, 2], мониторинг показателей функциональной диагностики организма [3] человека, моделирование климата [4] и др.

Для решения задачи аннотирования предложены различные подходы: лейтмотивы [5], шейпелеты [6], ослабленные периоды и средние тенденции [7] и др. Указанные подходы, однако, не являются независимыми от предметной области либо не обеспечивают количественную оценку покрытия выделенных активностей. Например, лейтмотив представляет собой пару наиболее похожих друг на друга подпоследовательностей временного ряда, но доля ряда, покрываемая таким шаблоном, неизвестна. Шейплет определяется как подпоследо-

\*Статья рекомендована к публикации программным комитетом Международной научной конференции «Параллельные вычислительные технологии (ПаВТ) 2022».

довательность, одновременно наиболее похожая на большинство подпоследовательностей данного класса и наиболее отличающаяся от подпоследовательностей из других классов. Шейплеты допускают количественную оценку покрытия, однако требуют знаний о предметной области. Мы также упоминаем исследования [8, 9], направленные на обнаружение типичных шаблонов временных рядов с помощью сверточных автоэнкодеров (SAE). SAE используются для восстановления входного временного ряда с помощью фильтров сверточного кодирования и декодирования, в то время как фильтры содержат интерпретируемые признаки (шаблоны) входного временного ряда. Такой подход не зависит от предметной области, но для получения хороших результатов необходимо тщательно настроить более десяти параметров нейронной сети [9].

В недавней работе [10] предложена концепция снippets (snippet), которая свободна от указанных выше недостатков. Снippet представляет собой подпоследовательность заданной длины, на которую похожи многие другие подпоследовательности данного ряда в смысле специальной меры схожести [11]. Эксперименты показывают, что снippets позволяют адекватно аннотировать сенсорные данные из широкого спектра предметных областей [10]. Однако оригинальный алгоритм поиска снippets имеет высокую вычислительную сложность, что критично при обработке временных рядов, насчитывающих от сотни тысяч элементов.

В нашей предыдущей работе [12] предложен параллельный алгоритм поиска снippets для графического процессора, названный PSF. В данной статье продолжается начатая работа и описаны два тематических исследования по аннотированию сенсорных данных, выполненные с помощью алгоритма PSF: анализ показаний носимого виброакселерометра, закрепленного на человеке, и стационарного виброакселерометра, установленного на малогабаритной дробильной установке. Остаток статьи имеет следующую структуру. Раздел 1 содержит формальные определения и описание алгоритма PSF. В разделе 2 описаны проведенные исследования. Заключение резюмирует полученные результаты.

## 1. Параллельный алгоритм поиска активностей

### 1.1. Формальные определения и обозначения

*Временной ряд (time series)  $T$*  представляет собой последовательность хронологически упорядоченных вещественных значений:

$$T = (t_1, \dots, t_n), t_i \in \mathbb{R}. \quad (1)$$

Число  $n$  обозначается как  $|T|$  и называется длиной ряда.

*Подпоследовательность (subsequence)  $T_{i,m}$*  временного ряда  $T$  представляет собой непрерывное подмножество  $T$  из  $m$  элементов, начиная с позиции  $i$ :

$$T_{i,m} = (t_i, \dots, t_{i+m-1}), 1 \leq m \leq n, 1 \leq i \leq n - m + 1. \quad (2)$$

Временной ряд  $T$  может быть логически разбит на сегменты — непересекающиеся подпоследовательности заданной длины  $m$ . Здесь и далее без существенного ограничения общности мы можем считать, что  $n$  кратно  $m$ , поскольку  $m \ll n$ . Множество сегментов ряда, имеющих длину  $m \ll n$ , обозначим как  $S_T^m$ , элементы этого множества как  $S_1, \dots, S_{n/m}$ :

$$S_T^m = (S_1, \dots, S_{n/m}), S_i = T_{m \cdot (i-1) + 1, m}. \quad (3)$$

Концепция *сниппетов* (*snippet*) предложена Кеогом и др. в работе [10] и уточняет понятие типичных подпоследовательностей временного ряда следующим образом. Каждый сниппет представляет собой один из сегментов временного ряда. Со сниппетом ассоциируются его ближайшие соседи — подпоследовательности ряда, имеющие ту же длину, что и сниппет, которые более похожи на данный сниппет, чем на другие сегменты. Для вычисления схожести подпоследовательностей используется специализированная мера схожести MPdist, основанная на евклидовом расстоянии. Сниппеты упорядочиваются по убыванию мощности множества своих ближайших соседей. Множество сниппетов ряда  $T$ , имеющих длину  $m$  обозначается, как  $C_T^m$ , а элементы этого множества — как  $C_1, \dots, C_K$ :

$$C_T^m = (C_1, \dots, C_K), C_i \in S_T^m. \quad (4)$$

Число  $K$  ( $1 \leq K \leq n/m$ ) представляет собой параметр, задаваемый прикладным программистом, и отражает соответствующее количество наиболее типичных сниппетов. С каждым сниппетом ассоциированы следующие атрибуты: индекс сниппета, ближайшие соседи и значимость данного сниппета. Сниппеты упорядочиваются по убыванию их значимости.

Мера MPdist [11], используемая для вычисления схожести подпоследовательностей при нахождении сниппетов, неформально определяется следующим образом. Два временных ряда равной длины  $m$  тем более похожи друг на друга в смысле меры MPdist, чем больше в каждом из них имеется подпоследовательностей заданной длины  $\ell$  ( $3 \leq \ell \leq m$ ), близких друг к другу в смысле нормализованного евклидова расстояния.

## 1.2. Реализация

В данном разделе приводится краткое описание параллельного алгоритма PSF поиска сниппетов на графическом процессоре, предложенного в нашей предыдущей работе [12].

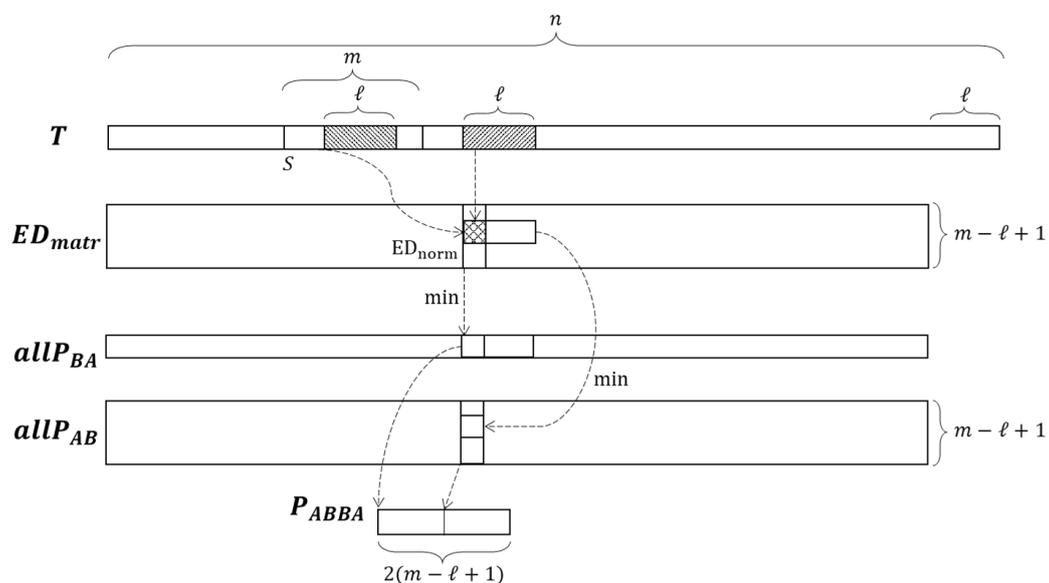


Рис. 1. Структуры данных алгоритма PSF

Структуры данных алгоритма PSF представлены на рис. 1. Ключевой для распараллеливания структурой данных является матрица  $ED_{matr}$ , в которой хранятся нормализованные евклидовы расстояния между каждой подпоследовательностью длины  $\ell$  сегмента  $S$  и

каждой подпоследовательностью длины  $\ell$  исходного ряда:

$$ED_{matr} \in \mathbb{R}^{(m-\ell+1) \times (n-\ell+1)} : ED_{matr}(i, j) = ED_{\text{norm}}(S_{i, \ell}, T_{j, \ell}). \quad (5)$$

Параллелизм вычислений матрицы расстояний  $ED_{matr}$  реализован на основе следующей техники, предложенной в работе [13]. Сначала вычисляется матрица центрированных сумм произведений значений ряда, которая используется для вычисления корреляции по Пирсону между подпоследовательностями ряда. Далее значения корреляции по Пирсону между двумя подпоследовательностями ряда преобразуются в z-нормализованное евклидово расстояние.

На втором шаге в каждом столбце матрицы  $ED_{matr}$ , полученной на первом шаге, находится минимум. Обозначим вектор таких минимумов за  $allP_{BA}$ :

$$allP_{BA} \in \mathbb{R}^{n-\ell+1} : allP_{BA}(j) = \min_{1 \leq i \leq m-\ell+1} ED_{matr}(i, j). \quad (6)$$

На третьем шаге в каждой строке  $ED_{matr}$  выполняется поиск минимумов в скользящем окне длины  $\ell$ . Обозначим матрицу таких минимумов за  $allP_{AB}$ :

$$allP_{AB} \in \mathbb{R}^{(m-\ell+1) \times (n-\ell+1)} : allP_{AB}(i, j) = \min_{j \leq c \leq j+m-\ell+1} ED_{matr}(i, c). \quad (7)$$

На четвертом шаге для каждой подпоследовательности ряда, имеющей длину  $\ell$ , и сегмента  $S$  выполняется построение матричного профиля. Для построения одного матричного профиля выполняется сцепление соответствующих данной подпоследовательности столбца матрицы  $allP_{AB}$  и подпоследовательности длины  $m - \ell + 1$ , входящей в вектор  $allP_{BA}$ . Результат сцепления обозначим как вектор  $P_{ABBA}$  (здесь символ  $\odot$  означает операцию конкатенации):

$$P_{ABBA} \in \mathbb{R}^{2(m-\ell+1)} : P_{ABBA}(T_{j, \ell}) = allP_{AB}(1, j) \odot \dots \odot allP_{AB}(m - \ell + 1, j) \odot \odot allP_{BA}(j) \odot \dots \odot allP_{BA}(m - \ell + 1), \quad (8)$$

где  $1 \leq j \leq n - \ell + 1$ . Для финального вычисления меры схожести MPdist между сегментом и подпоследовательностью необходимо выполнить сортировку вектора  $P_{ABBA}$  по возрастанию и взять его  $k$ -е значение.

## 2. Вычислительные эксперименты

В данном разделе описаны тематические исследования по применению параллельного алгоритма PSF для аннотирования сенсорных данных. Вычислительные эксперименты были проведены на графическом процессоре NVIDIA Tesla V100 SXM2 (5120 ядер, тактовая частота 1.3 GHz, пиковая производительность 15.7 TFLOPS). Для оценки эффективности аннотирования нами использованы стандартные меры качества классификации, определяемые следующим образом:

$$\text{Precision} = \frac{TP}{TP + FP}, \text{ Recall} = \frac{TP}{TP + FN}, \text{ F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (9)$$

где  $TP$ ,  $FP$ ,  $TN$  и  $FN$  — количество истинно-положительных, ложно-положительных, истинно-отрицательных и ложно-отрицательных элементов ряда соответственно при сравнении истинной и полученной при помощи алгоритма разметок ряда.

## 2.1. Аннотирование показаний сенсора, установленного на промышленном оборудовании

Для первого тематического исследования нами используются данные виброакселерометра, установленного на малогабаритной дробильной установке. Показания записаны во время заброса двух материалов различной твердости: дунита и кирпича. Помимо дробления указанных материалов, записаны два других вида активности: установка выключена и холостой ход. Количество снипшетов соответствует общему числу активностей:  $K = 4$ . Длина сегмента соответствует пяти секундам:  $m = 4000$ , длина подпоследовательности равна  $0.5m$ :  $\ell = 2000$ .

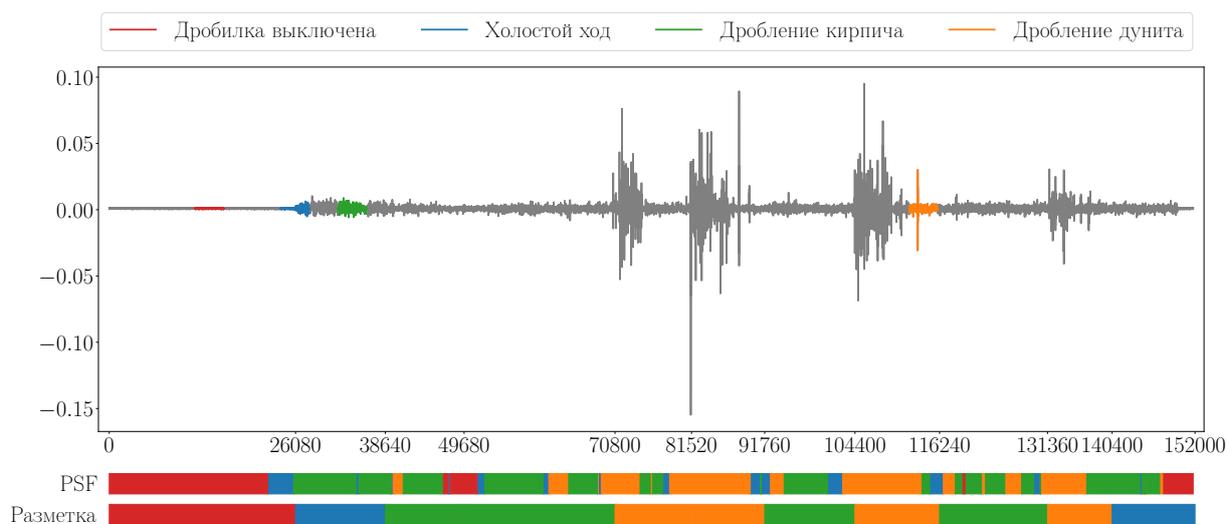


Рис. 2. Аннотирование показаний сенсора, установленного на дробильной установке (при  $K = 4$ )

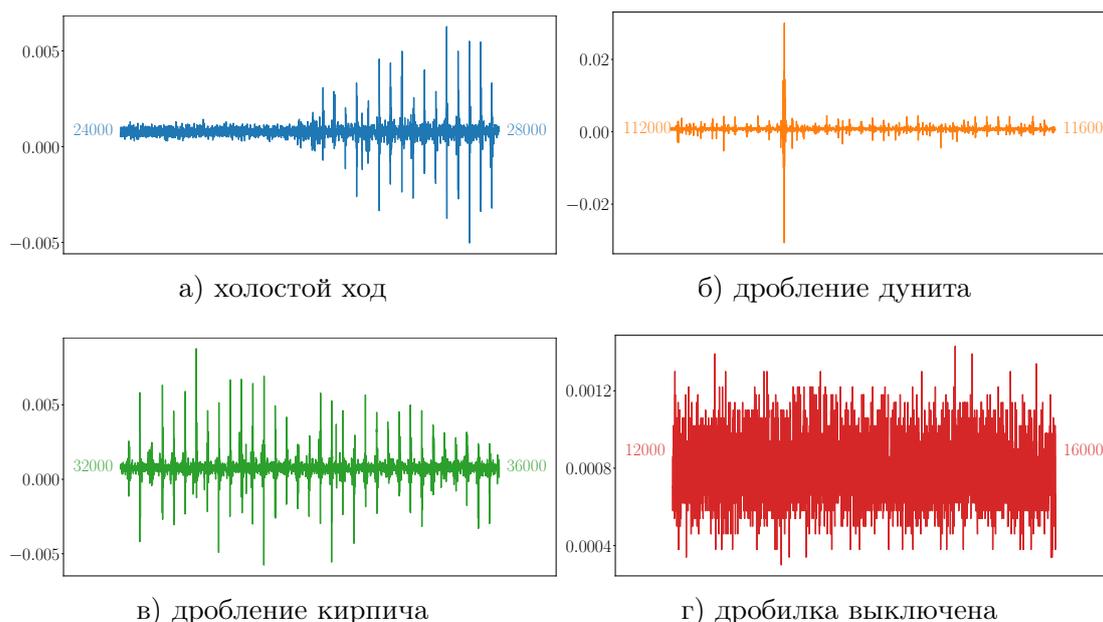


Рис. 3. Снипшеты показаний виброакселерометра, установленного на дробильной установке (при  $K = 4$ )

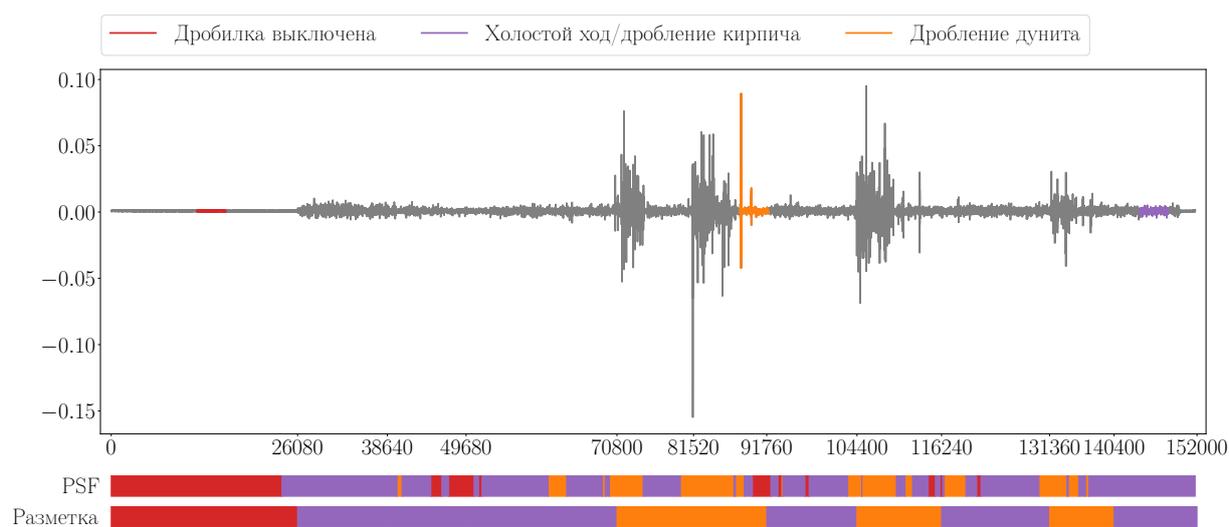
На рис. 2 представлены исходная разметка ряда и результат аннотирования ряда при помощи алгоритма PSF. На рис. 3 представлены найденные сниппеты. Числа, данные цветом, показывают индексы начала и конца сниппетов.

**Таблица 1.** Показатели качества аннотирования при  $K = 4$

Активность	Точность	Полнота	F1-мера
Холостой ход	0.05	0.03	0.04
Дробление дунита	0.65	0.72	0.68
Дробление кирпича	0.55	0.54	0.55
Установка выключена	0.77	0.85	0.81

В табл. 1 представлена оценка эффективности аннотирования по мерам, указанным в формуле (9). Можно видеть, что дробление дунита и выключенное состояние дробильной установки имеют наиболее высокие точность и полноту распознавания. Интегрально наилучшим образом распознается дробление дунита, наихудшим — холостой ход дробильной установки. Можно также заметить, что алгоритм PSF плохо распознает активности, связанные с холостым ходом и дроблением кирпича. Поэтому с целью повышения качества аннотирования разметка с помощью алгоритма PSF была проведена повторно для меньшего количества искомым сниппетов,  $K = 3$ .

По полученным результатам аннотирования (см. разметку ряда и найденные сниппеты на рис. 4 и рис. 5 соответственно) видно, что алгоритм обобщает показания, соответствующие холостому ходу и дроблению кирпича. Предположительно, это обусловлено низкой чувствительностью установленного датчика, и тем, что кирпич является более мягким материалом по сравнению с дунитом.



**Рис. 4.** Аннотирование показаний виброакселерометра, установленного на дробильной установке (при  $K = 3$ )

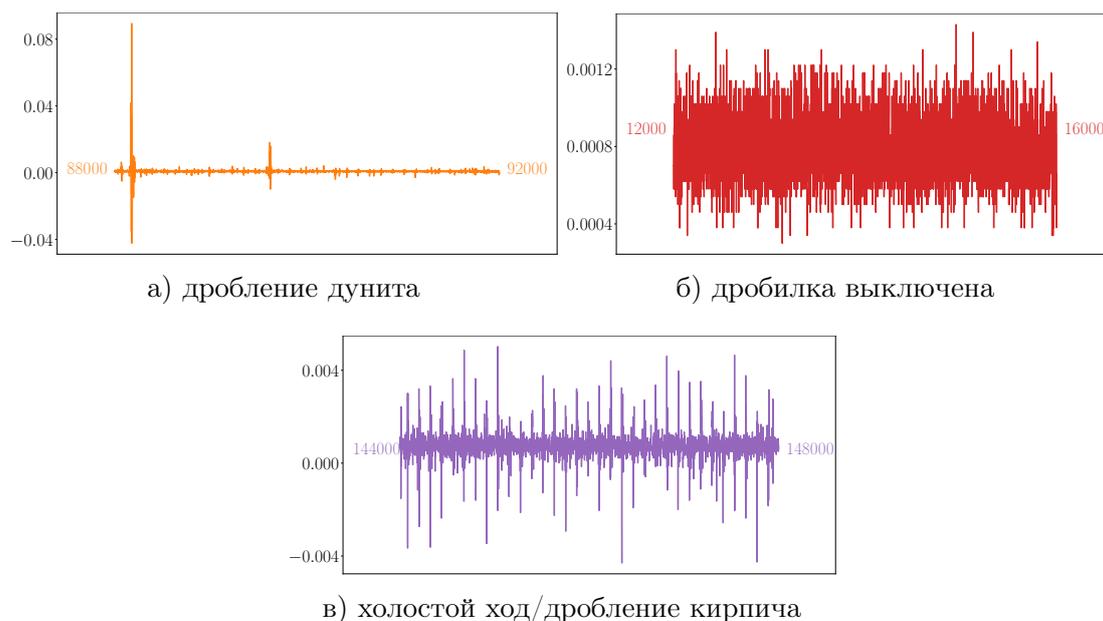


Рис. 5. Сниметы показаний виброакселерометра, установленного на дробильной установке (при  $K = 3$ )

Таблица 2. Показатели качества аннотирования при  $K = 3$

Активность	Точность	Полнота	F1-мера
Дробление дунита	0.68	0.56	0.61
Установка выключена	0.68	0.92	0.78
Холостой ход/дробление кирпича	0.78	0.77	0.78

В табл. 2 представлены полученные показатели качества аннотирования для случая  $K = 3$ . Можно видеть, что показатели качества существенно улучшены по сравнению с предыдущим случаем  $K = 4$  (см. табл. 1) и наилучшим образом распознается активность «холостой ход или дробление кирпича».

## 2.2. Аннотирование сенсорных данных носимого акселерометра

Для второго тематического исследования нами взят отрезок временного ряда РАМАР [14], представляющего собой показания закрепленного на человеке виброакселерометра. Данный ряд содержит показания при трех видах физической активности: глажка белья, подъем по лестнице, спуск по лестнице. Количество сниметов соответствовало числу активностей, отражаемых временным рядом, т.е.  $K = 3$ . Длина сегмента  $m = 2000$ , длина подпоследовательности равна  $0.5m$ :  $\ell = 1000$ .

На рис. 6 представлены исходная разметка ряда и результат аннотирования ряда при помощи алгоритма PSF. На рис. 7 представлены найденные сниметы, соответствующие активностям на временном ряду, с индексами начала и конца сниметов.

В табл. 3 представлена оценка точности аннотирования по мерам, указанным в формуле (9). Можно видеть, что наилучшим образом распознается подъем по лестнице, наименее — глажка белья. Можно видеть высокую полноту при распознавании подъема и спуска по лестнице и меньшую — при глажке белья.

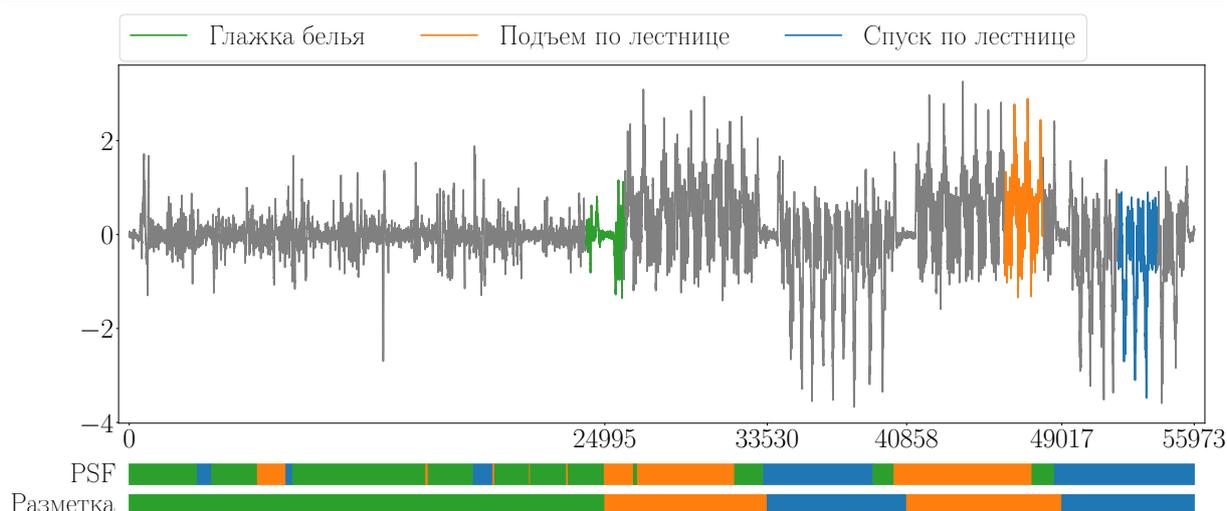


Рис. 6. Аннотирование показаний акселерометра, закрепленного на человеке

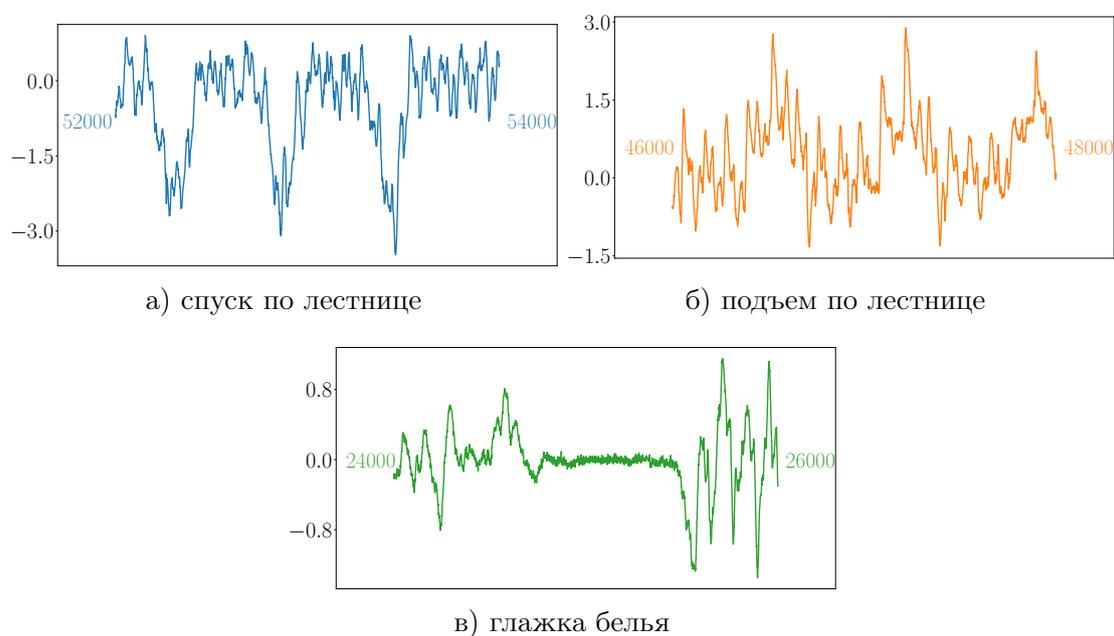


Рис. 7. Сниметы показаний акселерометра, закрепленного на человеке

Таблица 3. Показатели качества аннотирования

Активность	Точность	Полнота	F1-мера
Подъем по лестнице	0.74	0.83	0.78
Спуск по лестнице	0.59	0.82	0.68
Глажка белья	0.83	0.58	0.68

В работе [11] указывается, что мера MPdist дает адекватные результаты аннотирования ряда при значении длины подпоследовательности на интервале  $0.3m \leq \ell \leq 0.8m$ . Поэтому нами были проведены эксперименты по определению зависимости производительности и точности алгоритма PSF в зависимости от длины подпоследовательности  $\ell$  с использованием значений  $0.3m$ ,  $0.5m$  и  $0.8m$ .

**Таблица 4.** Производительность алгоритма PSF

Алгоритм	Время выполнения, с		
	$\ell = 0.3m$	$\ell = 0.5m$	$\ell = 0.8m$
Snippet-Finder	1188	817	266
PSF	74	52	16

В табл. 4 показана зависимость производительности параллельного алгоритма от длины подпоследовательности  $\ell$  для временного ряда РАМАР. По полученным результатам можно сделать вывод, что большее значение длины подпоследовательности повышает производительность алгоритма.

**Таблица 5.** Матрица ошибок алгоритма PSF при распознавании активностей

		Предсказанные активности			
		Длина подп-ти, $\ell$	Спуск по лестнице	Подъем по лестнице	Глажка белья
Исходные активности	Спуск по лестнице	$0.3m$	0.78	0.01	0.21
		$0.5m$	0.82	0.1	0.08
		$0.8m$	0.46	0.09	0.45
	Подъем по лестнице	$0.3m$	0.06	0.84	0.1
		$0.5m$	0.06	0.83	0.11
		$0.8m$	0.02	0.54	0.44
	Глажка белья	$0.3m$	0.2	0.07	0.73
		$0.5m$	0.29	0.14	0.58
		$0.8m$	0.21	0.41	0.38

В табл. 5 приведена матрица ошибок для различных значений длины подпоследовательности. По полученным результатам можно видеть, что для временного ряда РАМАР алгоритм дает более высокую точность аннотирования при меньших значениях длины подпоследовательности.

## Заключение

Статья посвящена проблеме применения параллельных вычислений на графическом процессоре для повышения производительности аннотирования сенсорных данных. Аннотирование сенсорных данных предполагает автоматизированную разметку временного ряда показаний, снятых с сенсора, которая выделяет различные активности, заданные указанным рядом. Разметка активностей имеет широкий спектр практического применения: предиктивное техническое обслуживание оборудования в приложениях цифровой индустрии, интеллектуальное управление зданиями в приложениях Интернета вещей, мониторинг состояния человека и упреждающая диагностика заболеваний в приложениях персональной медицины и др.

В статье представлены результаты двух тематических исследований, посвященных применению разработанного ранее авторами параллельного алгоритма PSF [12] для аннотирования сенсорных данных на графическом процессоре. Приведено краткое описание методов реализации разработанного алгоритма. Первое исследование связано с аннотированием дан-

ных виброакселерометра, установленного на малогабаритной дробильной установке. Записанные данные включают дробление дунита и кирпича (твердого и мягкого материалов соответственно). Наибольшая точность аннотирования была достигнута при количестве искомым сниппетов  $K = 3$ . Второе исследование связано с аннотированием показаний носимого виброакселерометра, закрепленного на человеке. В среднем точность классификации не ниже 74%. Были проведены эксперименты по определению зависимости производительности и точности разработанного алгоритма от входного значения длины подпоследовательности. В результате наибольшая производительность была достигнута при  $\ell = 0.8m$ , а точность аннотирования при  $\ell = 0.3m$ . Также во всех исследованных случаях параллельный алгоритм [12] показал большую эффективность по сравнению с оригинальной последовательной версией [10].

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 20-07-00140) и Министерства науки и высшего образования РФ (государственное задание FENU-2020-0022).*

## Литература

1. Цымблер М.Л., Краева Я.А., Латыпова Е.А. и др. Очистка сенсорных данных в интеллектуальных системах управления отоплением зданий // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2021. Т. 10, № 3. С. 16–36. DOI: 10.14529/cmse210302.
2. Иванов С.А., Никольская К.Ю., Радченко Г.И. и др. Концепция построения цифрового двойника города // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2020. Т. 9, № 4. С. 5–23. DOI: 10.14529/cmse200401.
3. Епишев В.В., Исаев А.П., Минахметов Р.М. и др. Система интеллектуального анализа данных физиологических исследований в спорте высших достижений // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2013. Т. 2, № 1. С. 44–54. DOI: 10.14529/cmse130105.
4. Абдуллаев С.М., Ленская О.Ю., Гаязова А.О. и др. Алгоритмы краткосрочного прогноза с использованием радиолокационных данных: оценка траектории и композиционный дисплей жизненного цикла // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2014. Т. 3, № 1. С. 17–32. DOI: 10.14529/cmse140102.
5. Mueen A., Keogh E.J., Zhu Q., *et al.* Exact Discovery of Time Series Motifs // Proceedings of the SIAM International Conference on Data Mining, SDM 2009, April 30 - May 2, 2009, Sparks, Nevada, USA. SIAM, 2009. P. 473–484. DOI: 10.1137/1.9781611972795.41.
6. Ye L., Keogh E.J. Time series shapelets: a new primitive for data mining // Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009 / ed. by J.F. Elder IV, F. Fogelman-Soulié, P.A. Flach, M.J. Zaki. ACM, 2009. P. 947–956. DOI: 10.1145/1557019.1557122.
7. Indyk P., Koudas N., Muthukrishnan S. Identifying Representative Trends in Massive Time Series Data Sets Using Sketches // VLDB 2000, Proceedings of 26th International Conference on Very Large Data Bases, September 10-14, 2000, Cairo, Egypt / ed. by A.E.

- Abbadi, M.L. Brodie, S. Chakravarthy, *et al.* Morgan Kaufmann, 2000. P. 363–372. URL: <http://www.vldb.org/conf/2000/P363.pdf>.
8. Bascol K., Emonet R., Fromont É., Odobez J. Unsupervised Interpretable Pattern Discovery in Time Series Using Autoencoders // Structural, Syntactic, and Statistical Pattern Recognition - Joint IAPR International Workshop, S+SSPR 2016, Mérida, Mexico, November 29 - December 2, 2016, Proceedings. Vol. 10029 / ed. by A. Robles-Kelly, M. Loog, B. Biggio, *et al.* 2016. P. 427–438. Lecture Notes in Computer Science. DOI: 10.1007/978-3-319-49055-7\\_38.
  9. Noering F.K., Schröder Y., Jonas K., Klawonn F. Pattern discovery in time series using autoencoder in comparison to nonlearning approaches // Integr. Comput. Aided Eng. 2021. Vol. 28, no. 3. P. 237–256. DOI: 10.3233/ICA-210650.
  10. Imani S., Madrid F., Ding W., *et al.* Matrix Profile XIII: Time Series Snippets: A New Primitive for Time Series Data Mining // 2018 IEEE International Conference on Big Knowledge, ICBK 2018, Singapore, November 17-18, 2018 / ed. by X. Wu, Y. Ong, C.C. Aggarwal, H. Chen. IEEE Computer Society, 2018. P. 382–389. DOI: 10.1109/ICBK.2018.00058.
  11. Gharghabi S., Imani S., Bagnall A.J., *et al.* An ultra-fast time series distance measure to allow data mining in more complex real-world deployments // Data Min. Knowl. Discov. 2020. Vol. 34, no. 4. P. 1104–1135. DOI: 10.1007/s10618-020-00695-8.
  12. Цымблер М.Л., Гоглачев А.И. Поиск типичных подпоследовательностей временного ряда на графическом процессоре // Вычислительные методы и программирование. 2021. Ноябрь. № 4. С. 344–359. DOI: 10.26089/NumMet.v22r423.
  13. Yeh C.M., Zhu Y., Ulanova L., *et al.* Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets // IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain / ed. by F. Bonchi, J. Domingo-Ferrer, R. Baeza-Yates, *et al.* IEEE Computer Society, 2016. P. 1317–1322. DOI: 10.1109/ICDM.2016.0179.
  14. Reiss A., Stricker D. Introducing a New Benchmarked Dataset for Activity Monitoring // 16th International Symposium on Wearable Computers, ISWC 2012, Newcastle, United Kingdom, June 18-22, 2012. IEEE Computer Society, 2012. P. 108–109. DOI: 10.1109/ISWC.2012.13.

Гоглачев Андрей Игоревич, программист отдела интеллектуального анализа данных и виртуализации Лаборатории суперкомпьютерного моделирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Цымблер Михаил Леонидович, д.ф.-м.н., доцент, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

## APPLICATION OF PARALLEL COMPUTING FOR SENSOR DATA ANNOTATION

© 2022 A.I. Goglachev, M.L. Zymbler

*South Ural State University (pr. Lenina 76, Chelyabinsk, 454080 Russia)*

*E-mail: goglachevai@susu.ru, mzym@susu.ru*

Received: 04.04.2022

Sensor data annotation involves automated marking of a time series of readings taken from the sensor, which highlights various activities specified by the specified series. Activity marking has a wide range of practical applications: predictive maintenance, intelligent management of life support systems, climate modeling, etc. Previously, we developed a parallel PSF algorithm for annotating sensor data using a GPU based on the concept of snippets. Snippet is a subsequence that many other subsequences of a given series resemble in the sense of a specialized similarity measure based on Euclidean distance. This article describes two case studies performed using the PSF algorithm: annotation of the readings of a wearable vibration accelerometer mounted on a person and a stationary vibration accelerometer mounted on a small crusher. As part of the research, computational experiments were conducted to evaluate the speed and accuracy of the developed algorithm. Also there was the research on the dependence of the efficiency of the algorithm on the values of the input parameters: the number of the desired snippets and the length of the subsequence.

*Keywords: time series, annotation, snippet, parallel algorithm, GPU.*

### FOR CITATION

Goglachev A.I., Zymbler M.L. Application of Parallel Computing for Sensor Data Annotation. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2022. Vol. 11, no. 2. P. 30–42. (in Russian) DOI: 10.14529/cmse220203.

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

### References

1. Zymbler M.L., Kraeva Y.A., Latypova E.A., *et al.* Cleaning Sensor Data in Intelligent Heating Control System. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2021. Vol. 10, no. 3. P. 16–36. (in Russian) DOI: 10.14529/cmse210302.
2. Ivanov S.A., Nikolskaya K.Y., Radchenko G.I., *et al.* Digital Twin of a City: Concept Overview. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2020. Vol. 9, no. 4. P. 5–23. (in Russian) DOI: 10.14529/cmse200401.
3. Epishev V.V., Isaev A.P., Miniakhmetov R.M., *et al.* Physiological Data Mining System For Elite Sports. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2013. Vol. 2, no. 1. P. 44–54. (in Russian) DOI: 10.14529/cmse130105.
4. Abdoulaev S.M., Lenskaia O.U., Gayazova A.O., *et al.* Short-Range Forecasting Algorithms Using Radar Data: Translation Estimate And Life-Cycle Composite Display. Bulletin of the

- South Ural State University. Series: Computational Mathematics and Software Engineering. 2014. Vol. 3, no. 1. P. 17–32. (in Russian) DOI: 10.14529/cmse140102.
5. Mueen A., Keogh E.J., Zhu Q., *et al.* Exact Discovery of Time Series Motifs. Proceedings of the SIAM International Conference on Data Mining, SDM 2009, April 30 - May 2, 2009, Sparks, Nevada, USA. SIAM, 2009. P. 473–484. DOI: 10.1137/1.9781611972795.41.
  6. Ye L., Keogh E.J. Time series shapelets: a new primitive for data mining. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009 / ed. by J.F. Elder IV, F. Fogelman-Soulié, P.A. Flach, M.J. Zaki. ACM, 2009. P. 947–956. DOI: 10.1145/1557019.1557122.
  7. Indyk P., Koudas N., Muthukrishnan S. Identifying Representative Trends in Massive Time Series Data Sets Using Sketches. VLDB 2000, Proceedings of 26th International Conference on Very Large Data Bases, September 10-14, 2000, Cairo, Egypt / ed. by A.E. Abbadi, M.L. Brodie, S. Chakravarthy, *et al.* Morgan Kaufmann, 2000. P. 363–372. URL: <http://www.vldb.org/conf/2000/P363.pdf>.
  8. Bascol K., Emonet R., Fromont É., Odobez J. Unsupervised Interpretable Pattern Discovery in Time Series Using Autoencoders. Structural, Syntactic, and Statistical Pattern Recognition - Joint IAPR International Workshop, S+SSPR 2016, Mérida, Mexico, November 29 - December 2, 2016, Proceedings. Vol. 10029 / ed. by A. Robles-Kelly, M. Loog, B. Biggio, *et al.* 2016. P. 427–438. Lecture Notes in Computer Science. DOI: 10.1007/978-3-319-49055-7\_38.
  9. Noering F.K., Schröder Y., Jonas K., Klawonn F. Pattern discovery in time series using autoencoder in comparison to nonlearning approaches. Integr. Comput. Aided Eng. 2021. Vol. 28, no. 3. P. 237–256. DOI: 10.3233/ICA-210650.
  10. Imani S., Madrid F., Ding W., *et al.* Matrix Profile XIII: Time Series Snippets: A New Primitive for Time Series Data Mining. 2018 IEEE International Conference on Big Knowledge, ICBK 2018, Singapore, November 17-18, 2018 / ed. by X. Wu, Y. Ong, C.C. Aggarwal, H. Chen. IEEE Computer Society, 2018. P. 382–389. DOI: 10.1109/ICBK.2018.00058.
  11. Gharghabi S., Imani S., Bagnall A.J., *et al.* An ultra-fast time series distance measure to allow data mining in more complex real-world deployments. Data Min. Knowl. Discov. 2020. Vol. 34, no. 4. P. 1104–1135. DOI: 10.1007/s10618-020-00695-8.
  12. Zymbler M.L., Goglachev A.I. Discovery of typical subsequences of time series on graphical processor. Numerical Methods and Programming (Vychislitel'nye Metody i Programirovanie). 2021. Nov. No. 4. P. 344–359. (in Russian).
  13. Yeh C.M., Zhu Y., Ulanova L., *et al.* Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets. IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain / ed. by F. Bonchi, J. Domingo-Ferrer, R. Baeza-Yates, *et al.* IEEE Computer Society, 2016. P. 1317–1322. DOI: 10.1109/ICDM.2016.0179.
  14. Reiss A., Stricker D. Introducing a New Benchmarked Dataset for Activity Monitoring. 16th International Symposium on Wearable Computers, ISWC 2012, Newcastle, United Kingdom, June 18-22, 2012. IEEE Computer Society, 2012. P. 108–109. DOI: 10.1109/ISWC.2012.13.

## МЕТОД ПОДАВЛЕНИЯ АКУСТИЧЕСКОГО ЭХА НА ОСНОВЕ РЕКУРРЕНТНОЙ НЕЙРОННОЙ СЕТИ И АЛГОРИТМА КЛАСТЕРИЗАЦИИ

© 2022 Д.М. Шаход, О.Л. Ибряева

*Южно-Уральский государственный университет*

*(454080 Челябинск, пр. им. В.И. Ленина, д. 76)*

*E-mail: ghiathlovealaa@gmail.com, ibriaevaol@susu.ru*

Поступила в редакцию: 01.04.2022

В статье решается задача подавления акустического эха на основе нейронной сети оценивающей идеальную двоичную маску IBM из признаков, извлеченных из смеси сигналов ближнего и дальнего конца. Новизна предложенного метода заключается в использовании алгоритма кластеризации дополнительно с двунаправленной рекуррентной нейронной сетью BLSTM. Для оценки использования алгоритмов кластеризации EM, Mean-Shift, k-Means, модели были обучены и протестированы на базе данных TIMIT. Для каждой модели были вычислены метрики ERLE, PESQ, STOI, характеризующие ее качество. Использование алгоритмов кластеризации EM, Mean-Shift оказалось неэффективным по сравнению с алгоритмом BLSTM при соотношении сигнал/эхо 10 дБ. При соотношении сигнал/эхо 6 дБ BLSTM+Mean-Shift привел к незначительному улучшению метрики PESQ по сравнению с алгоритмом BLSTM. Результаты экспериментов показали эффективность предложенной модели BLSTM при использовании сети с алгоритмом K-Means, по сравнению с использованием чистой BLSTM для подавления эха в сценариях с двойным разговором. При соотношении сигнал/эхо 10 дБ метрика STOI, характеризующая разборчивость речи, улучшилась на 7%, а метрика PESQ, характеризующая качество восстановления речи, на 18.8%.

*Ключевые слова: идеальная двоичная маска, сигнал ближнего конца, сигнал дальнего конца, двунаправленная рекуррентная нейронная сеть, кластеризация, двойной разговор.*

### ОБРАЗЕЦ ЦИТИРОВАНИЯ

Шаход Д.М., Ибряева О.Л. Метод подавления акустического эха на основе рекуррентной нейронной сети и алгоритма кластеризации // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2022. Т. 11, № 2. С. 43–58. DOI: 10.14529/cmse220204.

### Введение

Алгоритмы восстановления речевого сигнала, искаженного аддитивным некоррелированным шумом, в случае, когда доступен только зашумленный сигнал, широко применяются в различных областях цифровой обработки речевых сигналов, таких как распознавание речи, распознавание говорящего, детектирование речевой активности, улучшение качества и разборчивости речевых сигналов и др. [1]. С развитием эффективных методов машинного обучения широкое распространение стали получать алгоритмы подавления шума на основе глубоких нейронных сетей [2–4]. Одними из наиболее используемых методов шумоподавления являются методы, основанные на оценке частотно-временных масок [5]. Например, в работах [6, 7] в роли целевого выхода нейросетевой модели выступает идеальная двоичная маска (ideal binary mask, IBM).

В настоящей статье разработан алгоритм на основе двунаправленной рекуррентной сети (Bidirectional Long Short-Term Memory, BLSTM) выходом которой является маска IBM. Ключевой особенностью нашего алгоритма является использование кластеризации на выходе нейронной сети. В работе рассмотрены три метода кластеризации (EM, Mean-Shift, k-Means) и проведено сравнение алгоритмов на сигналах базы данных TIMIT на основе об-

щепринятых метрик в обработке речи: ERLE, STOI, PESQ. Показано, что дополнительное использование кластеризации k-Means улучшает работу модели BLSTM.

Статья организована следующим образом. Раздел 1 содержит краткий обзор работ по тематике исследования. В разделе 2 представлен метод BLSTM с кластеризацией. В разделе 3 приведены результаты вычислительных экспериментов по оценке эффективности предложенных методов. Заключение резюмирует полученные результаты и описывает направления будущих исследований.

## 1. Обзор связанных работ

Традиционно задача акустического эхоподавления решается за счет адаптации акустической импульсной характеристики между громкоговорителем и микрофоном с использованием фильтра с конечной импульсной характеристикой [8]. Одним из наиболее широко используемых адаптивных алгоритмов является NLMS (Normalized Least Mean Square) [9], имеющий хорошую надежность при низкой сложности. Однако он, как и все адаптивные алгоритмы [8], имеет недостаток возможной расходимости из-за корреляции между полезным сигналом и подавляемым эхом. Эта корреляция имеет место, например, во время так называемого «двойного разговора» и большинство алгоритмов строится на попытке сторонними средствами распознать двойную речь в конкретный момент времени и приостановить обучение адаптивного фильтра [10]. С помощью такого метода можно предотвратить расходимость фильтра, однако обучение значительно замедляется.

Сигнал, поступающий на микрофон, содержит не только эхо и речь на ближнем конце, но и фоновый шум, с которым система акустического эхоподавления сама по себе не способна справиться. Для подавления фонового шума и остаточных эхо-сигналов, которые существуют на выходе системы обычно используют пост-фильтры. Например, в работе [11] авторы объединили адаптивный алгоритм с методом подавления шума на основе кратковременного спектрального затухания и получили высокую степень удаления эха при наличии фонового шума.

Другим вариантом устранения остаточного эха является использование мощных моделей глубокого обучения [12], которые способны моделировать сложные нелинейные искажения, вносимые усилителями мощности и громкоговорителями. Такие модели являются все более популярной альтернативой нелинейному моделированию систем акустического эхоподавления. Так, авторы работы [13] смоделировали нелинейную систему как модель Хаммерштейна и использовали полносвязную нейронную сеть с двумя слоями, за которой следует адаптивный фильтр для идентификации параметров модели. В недавней работе [2] использована глубокая нейронная сеть для оценки усиления остаточного шума как на дальнем конце сигнала, так и на выходе системы [14], чтобы удалить нелинейные компоненты эхо-сигнала.

Глубокое обучение показало большой потенциал для разделения речи [15, 16]. Способность рекуррентных нейронных сетей RNN моделировать изменяющиеся во времени функции может играть важную роль в решении проблем акустического эхоподавления. LSTM (Long short-Term memory) [17] — вариант RNN, разработанный для решения проблемы исчезновения градиента, присущей традиционным RNN, уже показал хорошую производительность в задачах распознавания и улучшения речи в шумных условиях [18, 19]. Работа [20] также отмечает лучшее качество модели, использующей LSTM по сравнению с глубокой полносвязной нейронной сетью.

## 2. Предложенная модель

### 2.1. Постановка задачи

Рассматриваемая модель приведена на рис. 1. Сигнал микрофона  $y(n)$  состоит из сигнала на ближнем конце  $s(n)$ , эха  $d(n)$  и фонового шума  $v(n)$ :

$$y(n) = d(n) + s(n) + v(n). \quad (1)$$

Для простоты в этой работе будем считать  $v(n) = 0$ .

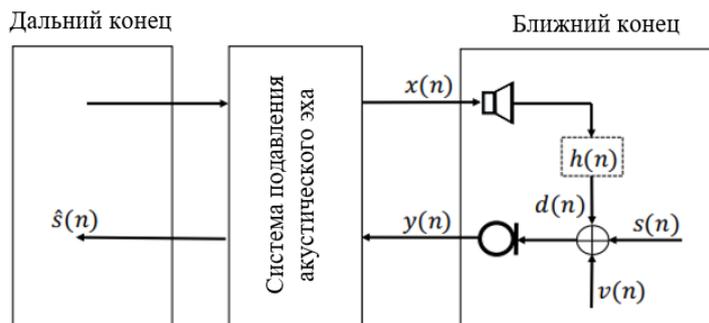


Рис. 1. Аддитивная модель акустического эха

Эхо-сигнал  $d(n)$  формируется за счет отражения сигнала с дальнего конца  $x(n)$  от стенок комнаты и моделируется путем свертки  $x(n)$  с импульсной характеристикой  $h(n)$  помещения RIR (Room Impulse Responses).

Наша задача — выделить из смеси  $y(n)$  полезный сигнал  $s(n)$ , убрав нежелательную помеху  $d(n)$ .

На рис. 2 представлена схема предлагаемой нами модели для решения данной задачи. Из сигнала микрофона  $y(n)$  с помощью кратковременного преобразования Фурье STFT (Short Time Fourier Transform) извлекаются признаки, которые служат входными данными для двунаправленной рекуррентной нейронной сети BLSTM. Выходом нейронной сети является бинарная маска Ideal Binary Mask (IBM), которая часто используется в качестве цели в задаче разделения речи от помех. Используя IBM маску, можно оценить спектр сигнала ближнего конца и, с помощью обратного преобразования Фурье ISTFT, восстановить  $s(n)$ .

### 2.2. Входные данные для модели

Исходные данные представляют собой аудиофайлы из базы данных TIMIT (описанной в параграфе 3.1). Из этих данных мы (случайным образом) выбирали сигналы ближнего конца  $s(n)$ , дальнего  $x(n)$  и формировали соответствующие сигналы микрофона  $y(n)$ .

К этим сигналам, передискретизированным с частоты 16 кГц до 8 кГц (с целью уменьшения времени обработки данных) было применено кратковременное преобразование Фурье с окном Ханнинга шириной 256 точек, что соответствует временной длине в 32 миллисекунды и 129 элементам разрешения по частоте. Для увеличения обучающей выборки была проведена аугментация данных, заключающаяся в перекрытии временных сигналов на 50%.

Полученные спектрограммы  $Y$  сигналов микрофона  $y(n)$  были разбиты на блоки  $100 \times 129$  для того, чтобы у нейронной сети всегда был вход фиксированного размера.

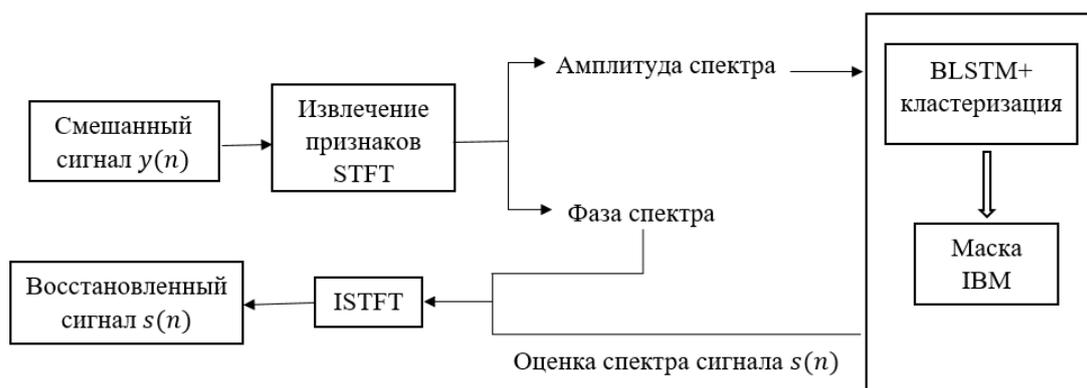


Рис. 2. Схема предложенной модели выделения сигнала  $s(n)$

Итоговый объем обучающей выборки составил 13606 образцов, объем тестовых данных — 3093 образца, объем валидационного набора данных — 1855 образцов. Таким образом, набор данных разбит на тренировочный, тестовый, валидационный датасеты в соотношении примерно 75% — 15% — 10%.

Также были найдены спектрограммы цели  $s(n)$  и помехи  $d(n)$ , которые использовались, чтобы найти маску IBM, являющуюся выходом модели для данного входа  $Y$ .

### 2.3. Выход модели

Выходом нейронной сети (целью обучения, target) является идеальная двоичная маска IBM — одна из наиболее часто используемых масок в задачах распознавания речи. IBM определяется как [21]:

$$IBM = \begin{cases} 1, & \text{if } \frac{S_T(t,f)}{S_I(t,f)} > 1 \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Здесь  $S_T = S_T(t, f)$  и  $S_I = S_I(t, f)$  — спектрограммы цели  $s(n)$  и помехи  $d(n)$ , соответственно. Некоторые исследователи, например [22], считают, что при использовании IBM восстановленная речь звучит не естественно, однако разборчивость речи при этом очень хорошая.

Если мы обозначим за  $Y$  спектрограмму сигнала микрофона, то

$$Y = S_T + S_I \quad (3)$$

и, с помощью маски IBM, можно восстановить спектрограмму полезного сигнала  $s(n)$  [21]:

$$S_T = IBM \odot Y. \quad (4)$$

Здесь оператор  $\odot$  представляет собой поэлементное умножение.

По известной спектрограмме для  $s(n)$  можно далее восстановить сам сигнал с помощью обратного преобразования Фурье.

## 2.4. Описание модели BLSTM+clustering

Модель двунаправленной рекуррентной нейронной сети BLSTM имеет два слоя, Каждый слой содержит две однонаправленные рекуррентные сети LSTM с 300 нейронами, одна из которых обрабатывает сигнал в прямом направлении, а другая — в обратном. Выходной полносвязный слой имеет сигмоидную функцию активации, и диапазон значений в  $[0, 1]$ , который легко (с установкой порогового значения 0.5) трансформируется в дискретный выход размером  $100 \times 129$  из нулей и единиц, соответствующий IBM маске.

Для обучения сети был выбран оптимизатор adam, в качестве функции потерь использовалась среднеквадратичная ошибка MSE (Mean Square Error). Скорость обучения была равна 0.01. Количество эпох обучения было равно 100.

Результаты описанной модели чистой BLSTM оказались неудовлетворительными (они, как и результаты для других моделей, представлены в параграфе 3.4) и нами было решено использовать дополнительно глубокую кластеризацию.

На рис. 3 приведена структура нейронной сети BLSTM+clustering. В этом случае мы увеличили размеры матрицы весов и смещения в три раза на последнем слое нейронной сети и теперь ее выход представляет собой матрицу размера  $(12900, 3)$  (в отличие от матрицы  $(12900, 1)$  для модели BLSTM). Далее применяется алгоритм кластеризации к 12900 точкам в трехмерном пространстве. Разделяя данные на два класса, получаем вектор из нулей и единиц, который затем преобразуем в матрицу, соответствующую маске IBM. В качестве алгоритмов кластеризации в работе использовались известные алгоритмы K-Means, Mean-shift, EM (Expectation-maximization).

## 2.5. Метрики качества

Для оценки качества модели в работе используются три метрики.

### 1. Уровень ослабления эхо-сигнала ERLE (Echo Return Loss Enhancement).

ERLE измеряет, насколько хорошо акустическое эхо удаляется из смешанного сигнала и определяется по формуле [23]:

$$ERLE = \lim_{n \rightarrow \infty} 10 \log_{10} \left( \frac{E[y^2(n)]}{E[e^2(n)]} \right), \quad (5)$$

где  $E$  — среднее значение. Чем выше этот показатель, тем лучше, он говорит о меньшем остаточном эхе  $e$ .

### 2. Перцепционная оценка качества речи PESQ (Perceptual Evaluation of Speech Quality).

PESQ — общепринятая метрика для оценки качества речи, которая сравнивает исходный сигнал  $s(n)$  с полученной для него оценкой  $\hat{s}(n)$ . Процедура вычисления PESQ является довольно сложной, ее описание приводится, в частности, в работе [24].

В большинстве практических случаев величина PESQ принимает значения в диапазоне от 0.5 до 4.5. Более высокий балл указывает на лучшее качество.

### 3. Кратковременная объективная разборчивость речи STOI (Short-Time Objective Intelligibility).

STOI практически всегда идет рядом с PESQ и отвечает за «понимаемость» речи [25]. Обычно значения STOI находятся в диапазоне  $[0, 1]$  и, можно считать, означают процент правильности работы модели.

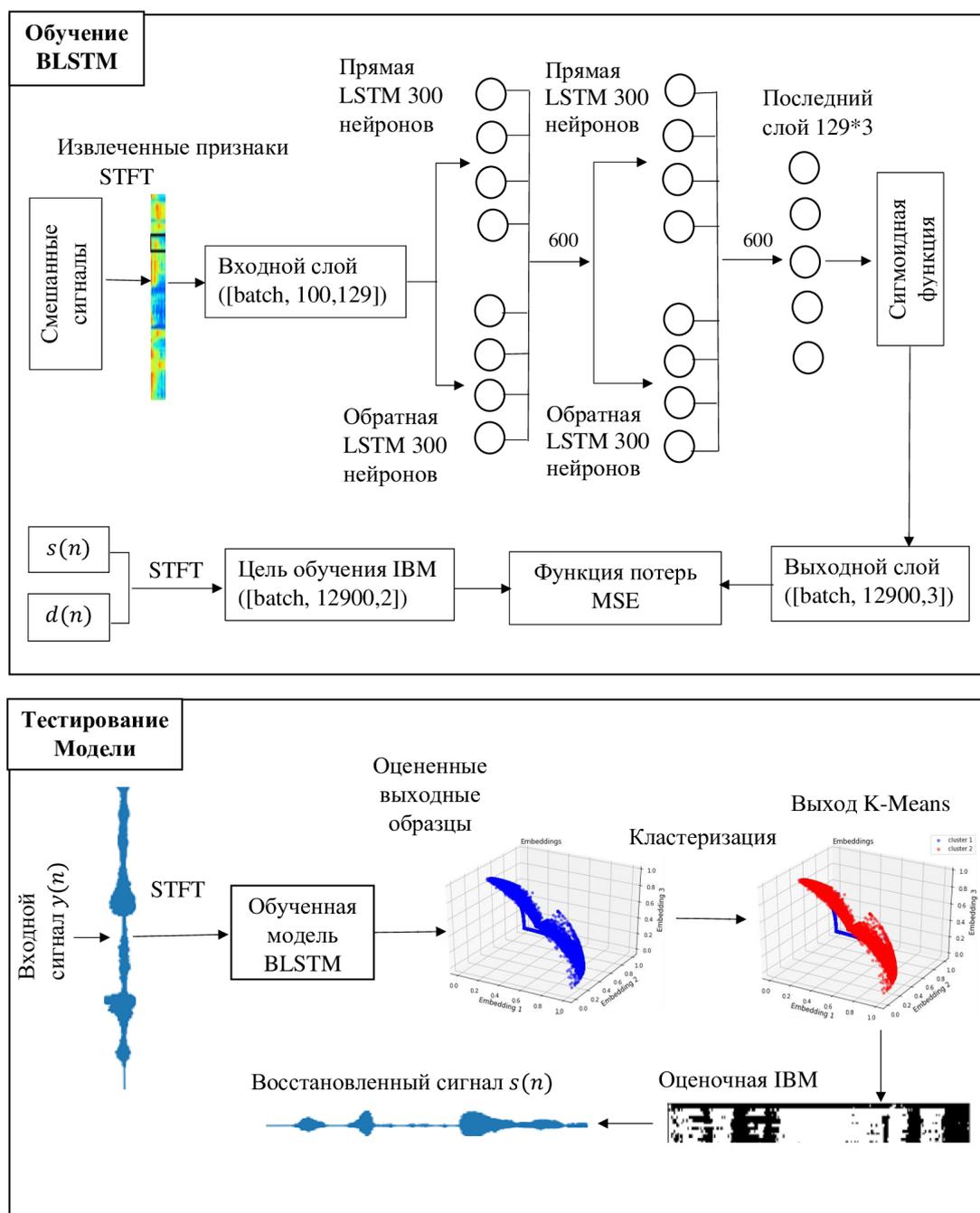


Рис. 3. Структура предлагаемой модели

### 3. Результаты экспериментов

#### 3.1. Набор данных

Набор данных ТИМІТ — один из известных наборов данных для акустико-фонетических исследований, а также для разработки и оценки систем автоматического распознавания речи. ТИМІТ содержит записи 630 носителей восьми основных диалектов американского английского языка, каждый из которых читает десять предложений с фонетически богатым звучанием. Набор данных записан с частотой 16 КНз в Texas Instruments, Inc, Массачусетским технологическим институтом [21].

### 3.2. Оценка эффективности модели BLSTM

Для обучения, валидации и тестирования из исходного набора данных с 630 носителями были случайным образом выбраны записи 462, 68 и 100 человек, соответственно. Поскольку каждый человек читает 10 предложений, у нас имеется 4620 аудиофайлов для обучения, 680 для валидации и 1000 для тестирования. Случайно выбранная пара из этого набора представляет собой, как правило, аудиофайлы с речью разных людей, которые мы берем в качестве сигналов ближнего  $s(n)$  и дальнего конца  $x(n)$ . Из сигнала  $x(n)$  путем свертки с импульсной характеристикой  $h(n)$  помещения RIR формируется эхо-сигнал  $d(n)$ . Смешивая  $d(n)$  с  $s(n)$ , получаем сигнал микрофона. Таким образом, у нас имеется 2310, 340 и 500 пар аудиофайлов для обучения, валидации и тестирования. Поскольку длительность аудиофайлов различна, то из каждого из них мы получаем различное число спектрограмм (в среднем, около 3, но с учетом аугментации и перекрытия в 50%, около 6). Окончательно, объем обучающей, валидационной и тестовой выборок составил в наших экспериментах 13606, 1855 и 3093, соответственно.

RIR генерируется при времени реверберации  $T60 = 0.5$ с (время, необходимое для уменьшения RIR на 60 dB) с использованием метода источника изображения ISM (Image Source Method) [26]. Размер комнаты для моделирования составляет (9, 7.5, 3.5) м, микрофон находится в позиции (6.3, 4.87, 1.2) м внутри комнаты, источник помехи в (2.5, 3.73, 1.76) м. Источник помехи озвучивает содержимое wav-файла (из TIMIT), начиная с 1.3 секунды.

Отношение сигнал/эхо  $SER$  вычисляется по формуле:

$$SER = 10 \lg \left( \frac{E[s^2(n)]}{E[d^2(n)]} \right). \quad (6)$$

Далее в работе будут приведены результаты работы моделей для случая  $SER = 6$  дБ.

На рис. 4 представлены графики изменения метрики Ассигасу (доли правильных ответов) для тренировочной и валидационной выборок в процессе 100 эпох обучения. Видно, что метрика достигла почти постоянного порога, который однако составляет немногим более 50% для валидационной выборки, что совершенно неудовлетворительно.

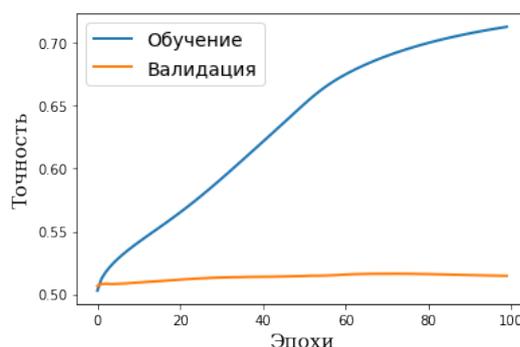


Рис. 4. Оценка эффективности модели BLSTM

На рис. 5 показан пример спектрограммы сигнала микрофона, оценочной маски IBM, являющейся целевым выходом модели BLSTM, а также настоящей маски (Real IBM).

На рис. 6 показаны сигнал входа для модели BLSTM  $y(n)$  (Mixture signal), сигнал ближнего конца  $s(n)$  (Near-end signal), ресинтезированный целевой сигнал  $\hat{s}(n)$  (Target signal),

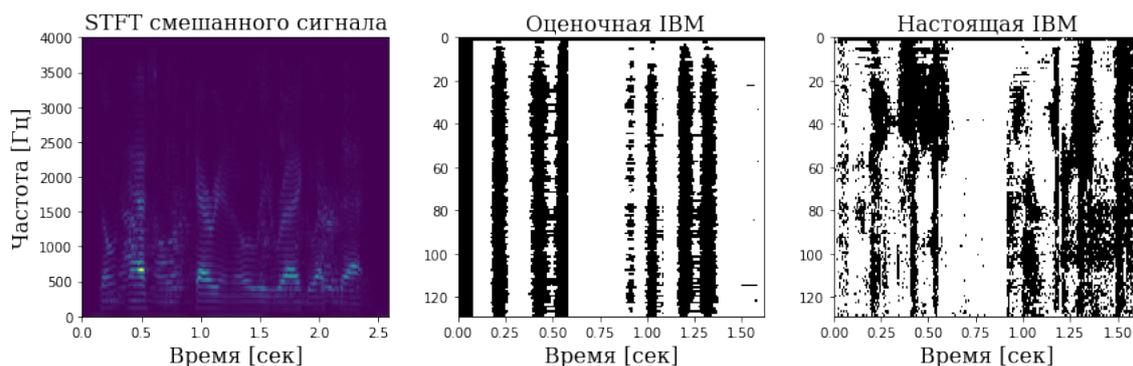


Рис. 5. Спектрограмма сигнала микрофона, оценочная маска IBM, полученная методом BLSTM и настоящая маска IBM

и эхо (Echo)  $d(n)$ . Можно видеть, что сигналы  $s(n)$  и  $\hat{s}(n)$  очень похожи. Значение метрик PESQ и STOI составило 1.03 и 0.846, соответственно.

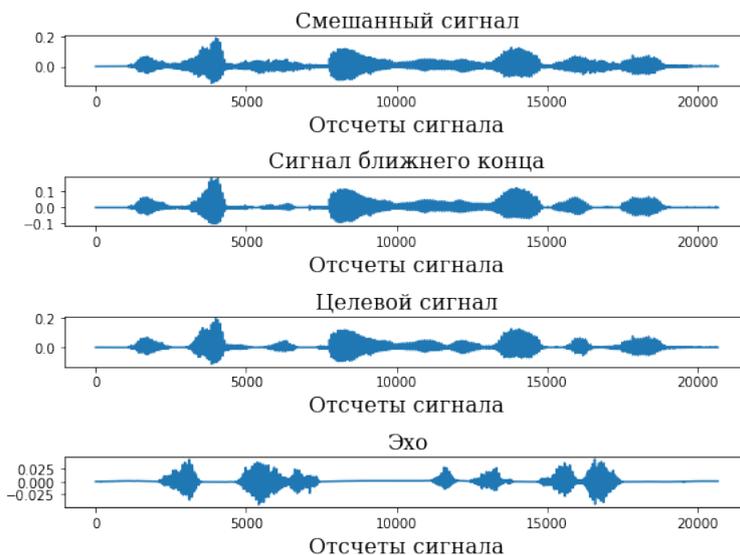


Рис. 6. Входной и выходной сигналы модели BLSTM

### 3.3. Оценка эффективности модели BLSTM+K-Means

Перейдем к построению предложенной модели после добавления кластеризации на выходе BLSTM. На рис. 7 представлены графики изменения метрики Ассигасу для тренировочной и валидационной выборок в процессе 100 эпох обучения. Отметим, что в отличие от аналогичного рис. 4 для простой модели BLSTM, сейчас достигнута точность около 78% на валидационной выборке.

На рис. 8 показаны точки в трехмерном пространстве, являющиеся выходом обучающей модели, а также показано разбиение этих точек на два класса, полученное с помощью алгоритма кластеризации K-Means.

На рис. 9 показан пример спектрограммы сигнала микрофона, оценочной маски IBM, являющейся целевым выходом модели BLSTM+k-Means, а также настоящей маски (Real IBM).

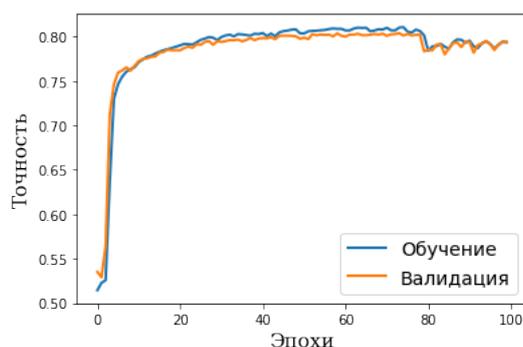


Рис. 7. Оценка эффективности модели BLSTM+K-Means

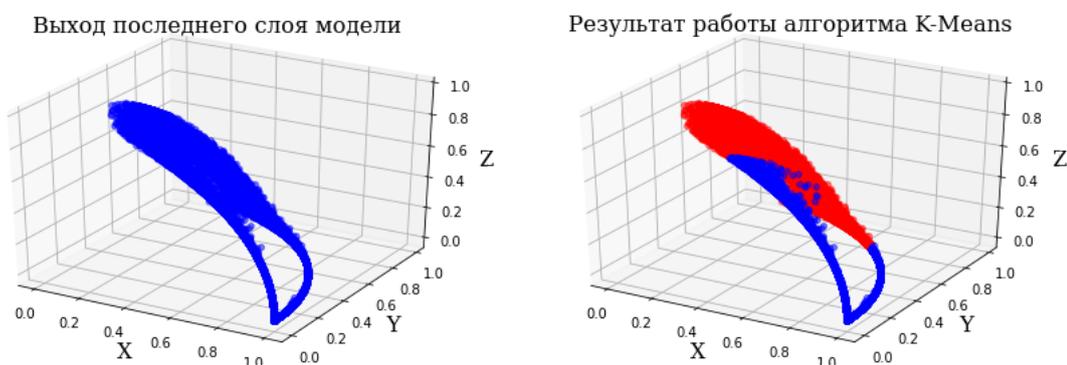


Рис. 8. Выход модели и результат k-Means кластеризации

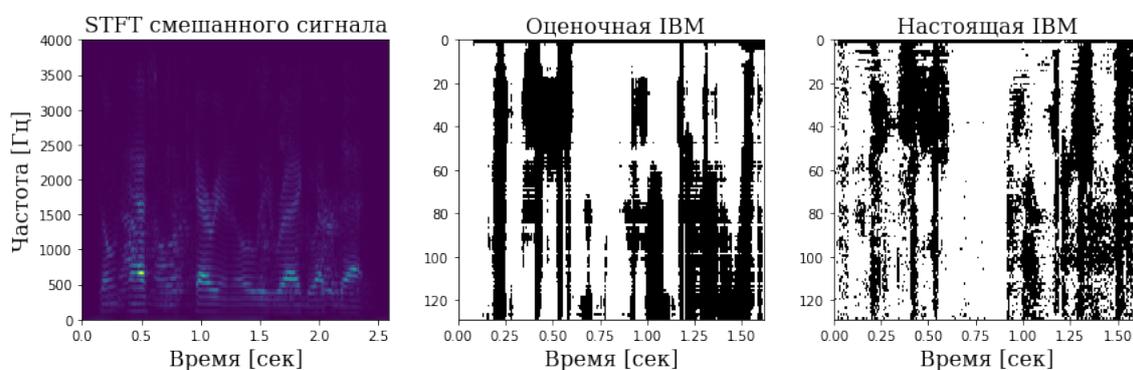


Рис. 9. Спектрограмма сигнала микрофона, оценочная маска ИВМ, полученная методом BLSTM+K-Means и настоящая маска ИВМ

На рис. 10 показаны сигнал входа BLSTM+K-Means, сигнал ближнего конца, ресинтезированный целевой сигнал и эхо. Значение метрик PESQ и STOI, характеризующих качество восстановления сигнала, составило 2.1 и 0.911, соответственно, и превысило их значения в случае простой модели BLSTM.

Аналогично были проанализированы алгоритмы BLSTM+EM, BLSTM+MeanShift, основанные на использовании методов кластеризации EM и MeanShift. Полученные значения метрик приведены в табл. 1.

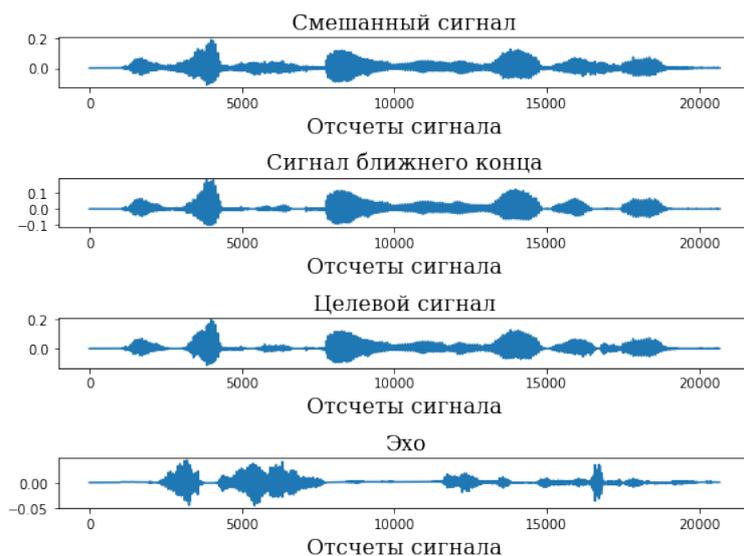


Рис. 10. Входной и выходной сигналы модели BLSTM+K-Means

### 3.4. Сравнение эффективностей моделей

Таблица 1 содержит значения метрик ERLE, PESQ и STOI, полученных четырьмя рассматриваемыми методами при  $SER = 6$  дБ и  $SER = 10$  дБ.

Таблица 1. Сравнение эффективности моделей

Метод	ERLE	PESQ	STOI
SER = 6 дБ			
BLSTM	6.8	1.03	0.846
BLSTM+EM	3.5	0.91	0.714
BLSTM+Mean-Shift	-2.6	1.17	0.808
BLSTM+K-Means	<b>8.1</b>	<b>2.1</b>	<b>0.911</b>
SER = 10 дБ			
BLSTM	8.7	2.23	0.865
BLSTM+EM	5.3	1.59	0.770
BLSTM+Mean-Shift	-1.8	2.14	0.846
BLSTM+K-Means	<b>11.2</b>	<b>2.65</b>	<b>0.924</b>

Отметим, что для вычисления метрик использовались данные, которые не участвовали в процессе обучения. Как можно видеть, использование алгоритма k-Means улучшило все показатели, в то время как другие алгоритмы кластеризации почти всегда ухудшают работы модели BLSTM.

Можно видеть, что и в случае  $SER = 10$  дБ добавление k-Means к BLSTM показывает улучшение значений ERLE примерно на 2.5 дБ, PESQ на 0.42 и STOI на 0.059. Таким образом, метрика STOI, характеризующая разборчивость речи, улучшилась на 7%, а метрика PESQ, характеризующая качество восстановления речи, на 18.8%. Использование алгоритмов Mean-Shift и EM не улучшило производительность модели BLSTM.

## Заключение

В статье предложена модель восстановления зашумленного сигнала на основе двунаправленной рекуррентной нейронной сети BLSTM с IBM маской на выходе. Сеть обучалась и тестировалась на наборе данных TIMIT и показала недостаточную эффективность.

Далее модель была модифицирована добавлением дополнительного этапа кластеризации данных. Были рассмотрены три метода кластеризации: k-Means, Mean-Shift, EM. Использование метода k-Means привело к существенному улучшению показателей ERLE, PESQ, STOI, в отличие от методов Mean-Shift, EM.

В дальнейшем предложенная модель BLSTM+k-Means будет использована для задачи подавления акустического эха при наличии шума и нелинейных искажений.

## Литература

1. Benesty J., Jensen J., Christensen M., Chen J. *Speech Enhancement: A Signal Subspace Perspective*. Elsevier Academic Press, 2014. 129 p. DOI: 10.1016/C2013-0-16082-5.
2. Lee C.M., Shin J.W., Kim N.S. DNN-based residual echo suppression // *Interspeech 2015*, Dresden, Germany, September 6–10, 2015. ISCA, 2015. P. 1775–1779. DOI: 10.21437/Interspeech.2015-412.
3. Zhang H., Wang D. Deep learning for acoustic echo cancellation in noisy and double-talk scenarios // *Interspeech 2018*, Hyderabad, India, September 2–6, 2018. ISCA, 2018. P. 3239–3243. DOI: 10.21437/Interspeech.2018-1484.
4. Zhang H., Tan K., Wang D. Deep learning for joint acoustic echo and noise cancellation with nonlinear distortions // *Interspeech 2019*, Graz, Austria, September 15–19, 2019. ISCA, 2019. P. 4255–4259. DOI: 10.21437/Interspeech.2019-2651.
5. Wang D. On Ideal Binary Mask As the Computational Goal of Auditory Scene Analysis // *Speech Separation by Humans and Machines* / ed. by P. Divenyi. Springer, Boston, MA, 2005. P. 181–197. DOI: 10.1007/0-387-22794-6\_12.
6. Li N., Loizou P.C. Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction // *J. Acoust. Soc. Am.* 2008. Vol. 123, no. 3. P. 1673–1682. DOI: 10.1121/1.2832617.
7. Brungart D.S., Chang P.S., Simpson B.D., Wang D. Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation // *J. Acoust. Soc. Am.* 2006. Vol. 120, no. 6. P. 4007–4018. DOI: 10.1121/1.2363929.
8. Benesty J., Gänslér T., Morgan D.R., *et al.* *Advances in network and acoustic echo cancellation*. Springer, Berlin, Heidelberg, 2001. 222 p. DOI: 10.1007/978-3-662-04437-7.
9. Enzner G., Buchner H., Favrot A., Kuech F. Chapter 30 - Acoustic Echo Control // *Academic Press Library in Signal Processing: Volume 4* / ed. by J. Trussell, A. Srivastava, A.K. Roy-Chowdhury, *et al.* Elsevier, 2014. P. 807–877. DOI: 10.1016/B978-0-12-396501-1.00030-3.
10. Hamidia M., Amrouche A. A new robust double-talk detector based on the Stockwell transform for acoustic echo cancellation // *Digital Signal Processing*. 2017. Vol. 60. P. 99–112. DOI: 10.1016/j.dsp.2016.09.001.

11. Ykhlef F., Ykhlef H. A post-filter for acoustic echo cancellation in frequency domain // 2014 Second World Conference on Complex Systems (WCCS), Agadir, Morocco, Nov. 10–12, 2014. IEEE, 2014. P. 446–450. DOI: 10.1109/ICoCS.2014.7060938.
12. Kuech F., Kellermann W. Nonlinear residual echo suppression using a power filter model of the acoustic echo path // 2007 International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, Honolulu, HI, USA, April 15–20, 2007. IEEE, 2007. P. I-73–I-76. DOI: 10.1109/ICASSP.2007.366619.
13. Malek J., Koldovský Z. Hammerstein model-based nonlinear echo cancelation using a cascade of neural network and adaptive linear filter // 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), Xi'an, China, Sept. 13–16, 2016. IEEE, 2016. P. 1–5. DOI: 10.1109/IWAENC.2016.7602906.
14. Yang F., Wu M., Yang J. Stereophonic acoustic echo suppression based on wiener filter in the short-time fourier transform domain // IEEE Signal Processing Letters. 2012. Vol. 19, no. 4. P. 227–230. DOI: 10.1109/LSP.2012.2187446.
15. Wang D., Chen J. Supervised speech separation based on deep learning: an overview // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2018. Vol. 26, no. 10. P. 1702–1726. DOI: 10.1109/TASLP.2018.2842159.
16. Wang Y., Narayanan A., Wang D. On training targets for supervised speech separation // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2014. Vol. 22, no. 12. P. 1849–1858. DOI: 10.1109/TASLP.2014.2352935.
17. Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. 1997. Vol. 9, no. 8. P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
18. Erdogan H., Hershey J.R., Watanabe S., Roux J.L. Phase-sensitive and recognition-boosted speech separation using deep recurrent neural networks // 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, April 19–24, 2015. IEEE, 2015. P. 708–712. DOI: 10.1109/ICASSP.2015.7178061.
19. Weninger F., Erdogan H., Watanabe S., *et al.* Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR // Latent Variable Analysis and Signal Separation. Vol. 9237 / ed. by E. Vincent, A. Yeredor, Z. Koldovský, P. Tichavský. Cham: Springer International Publishing, 2015. P. 91–99. Lecture Notes in Computer Science. DOI: 10.1007/978-3-319-22482-4\_11.
20. Chen J., Wang D. Long short-term memory for speaker generalization in supervised speech separation // The Journal of the Acoustical Society of America. 2017. Vol. 141, no. 6. P. 4705–4714. DOI: 10.1121/1.4986931.
21. Zermini A. Deep Learning for Speech Separation: PhD thesis / Zermini Alfredo. University of Surrey, faculty of engineering, physical sciences, Centre for Vision, Speech, Signal Processing (CVSSP), South East of England, UK, 2020. URL: <https://openresearch.surrey.ac.uk/esploro/outputs/doctoral/99512310402346#file-0>.
22. Xia S., Li H., Zhang X. Using Optimal Ratio Mask as Training Target for Supervised Speech Separation // 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, Malaysia, Dec. 12–15, 2017. IEEE, 2017. P. 163–166. DOI: 10.1109/APSIPA.2017.8282021.

23. Palmqvist M. Methods and algorithms for quality and performance evaluation of audio conferencing systems: PhD thesis / Palmqvist Maria. Umeå University, Faculty of Science, Technology, Department of Physics, Sweden, 2013. URL: <http://umu.diva-portal.org/smash/get/diva2:630382/FULLTEXT01.pdf>.
24. ITU-T Recommendation P. 862, Perceptual Evaluation of Speech Quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. 2001. URL: <https://www.itu.int/rec/T-REC-P.862-200102-I/en>.
25. Fu S.-W., Liao C.-F., Tsao Y. Learning with Learned Loss Function: Speech Enhancement with Quality-Net to Improve Perceptual Evaluation of Speech Quality // IEEE Signal Processing Letters. 2020. Vol. 27. P. 26–30. DOI: 10.1109/LSP.2019.2953810.
26. Allen J.B., Berkley D.A. Image method for efficiently simulating small-room acoustics // The Journal of the Acoustical Society of America. 1998. Vol. 65, no. 4. P. 943–950. DOI: 10.1121/1.382599.

Шаход Джаих Михаил, магистрант, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Ибряева Ольга Леонидовна, к.ф.-м.н., доцент, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

## METHOD OF AN ACOUSTIC ECHO SUPPRESSION BASED ON RECURRENT NEURAL NETWORK AND CLUSTERING

© 2022 Gh.M. Shahoud, O.L. Ibryaeva

*South Ural State University*

*(pr. Lenina 76, Chelyabinsk, 454080 Russia)*

*E-mail: ghiathlovealaa@gmail.com, ibriaevaol@susu.ru*

Received: 01.04.2022

The article solves the problem of acoustic echo suppression based on a neural network that evaluates an ideal binary mask IBM using features extracted from a mixture of near-end and far-end signals. The novelty of the proposed method lies in the use of the clustering algorithm in addition to the bidirectional recurrent neural network BLSTM. To evaluate the use of the EM, Mean-Shift, k-Means clustering algorithms, the models have been trained and tested on the TIMIT database. For each model, the ERLE, PESQ, STOI metrics have been calculated to characterize its quality. The use of the EM and Mean-Shift clustering algorithms appeared to be inefficient compared to the BLSTM algorithm at a signal-to-echo ratio of 10 dB. With a signal-to-echo ratio of 6 dB, BLSTM+Mean-Shift resulted in a marginal improvement in the PESQ metric compared to the BLSTM algorithm. The results of the experiments show the effectiveness of the proposed BLSTM model when using a network with the K-Means algorithm, compared to using a pure BLSTM for echo cancellation in double-talk scenarios. With a signal-to-echo ratio of 10 dB, the STOI metric, which characterizes speech intelligibility, has improved by 7%, and the PESQ metric, which characterizes the quality of speech restoration, by 18.8%.

*Keywords: ideal binary mask, near-end signal, far-end signal, bidirectional recurrent neural network, clustering, double-talk.*

### FOR CITATION

Shahoud Gh.M., Ibryaeva O.L. Method of an Acoustic Echo Suppression Based on Recurrent Neural Network and Clustering. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2022. Vol. 11, no. 2. P. 43–58. (in Russian) DOI: 10.14529/cmse220204.

### References

1. Benesty J., Jensen J., Christensen M., Chen J. Speech Enhancement: A Signal Subspace Perspective. Cambridge: Elsevier Academic Press, 2014. 129 p. DOI: 10.1016/C2013-0-16082-5.
2. Lee C.M., Shin J.W., Kim N.S. DNN-based residual echo suppression. Interspeech 2015, Dresden, Germany, September 6–10, 2015. ISCA, 2015. P. 1775–1779. DOI: 10.21437/Interspeech.2015-412.
3. Zhang H., Wang D. Deep learning for acoustic echo cancellation in noisy and double-talk scenarios. Interspeech 2018, Hyderabad, India, September 2–6, 2018. ISCA, 2018. P. 3239–3243. DOI: 10.21437/Interspeech.2018-1484.
4. Zhang H., Tan K., Wang D. Deep learning for joint acoustic echo and noise cancellation with nonlinear distortions. Interspeech 2019, Graz, Austria, September 15–19, 2019. ISCA, 2019. P. 4255–4259. DOI: 10.21437/Interspeech.2019-2651.

5. Wang D. On Ideal Binary Mask As the Computational Goal of Auditory Scene Analysis. *Speech Separation by Humans and Machines* / ed. by P. Divenyi. Springer, Boston, MA, 2005. P. 181–197. DOI: 10.1007/0-387-22794-6\_12.
6. Li N., Loizou P.C. Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. *J. Acoust. Soc. Am.* 2008. Vol. 123, no. 3. P. 1673–1682. DOI: 10.1121/1.2832617.
7. Brungart D.S., Chang P.S., Simpson B.D., Wang D. Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *J. Acoust. Soc. Am.* 2006. Vol. 120, no. 6. P. 4007–4018. DOI: 10.1121/1.2363929.
8. Benesty J., Gänslér T., Morgan D.R., *et al.* *Advances in network and acoustic echo cancellation*. Springer, Berlin, Heidelberg, 2001. 222 p. DOI: 10.1007/978-3-662-04437-7.
9. Enzner G., Buchner H., Favrot A., Kuech F. Chapter 30 - Acoustic Echo Control. *Academic Press Library in Signal Processing: Volume 4* / ed. by J. Trussell, A. Srivastava, A.K. Roy-Chowdhury, *et al.* Elsevier, 2014. P. 807–877. DOI: 10.1016/B978-0-12-396501-1.00030-3.
10. Hamidia M., Amrouche A. A new robust double-talk detector based on the Stockwell transform for acoustic echo cancellation. *Digital Signal Processing*. 2017. Vol. 60. P. 99–112. DOI: 10.1016/j.dsp.2016.09.001.
11. Ykhlef F., Ykhlef H. A post-filter for acoustic echo cancellation in frequency domain. 2014 Second World Conference on Complex Systems (WCCS), Agadir, Morocco, Nov. 10–12, 2014. IEEE, 2014. P. 446–450. DOI: 10.1109/ICoCS.2014.7060938.
12. Kuech F., Kellermann W. Nonlinear residual echo suppression using a power filter model of the acoustic echo path. 2007 International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, Honolulu, HI, USA, April 15–20, 2007. IEEE, 2007. P. I-73–I-76. DOI: 10.1109/ICASSP.2007.366619.
13. Malek J., Koldovský Z. Hammerstein model-based nonlinear echo cancellation using a cascade of neural network and adaptive linear filter. 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), Xi'an, China, Sept. 13–16, 2016. IEEE, 2016. P. 1–5. DOI: 10.1109/IWAENC.2016.7602906.
14. Yang F., Wu M., Yang J. Stereophonic acoustic echo suppression based on wiener filter in the short-time fourier transform domain. *IEEE Signal Processing Letters*. 2012. Vol. 19, no. 4. P. 227–230. DOI: 10.1109/LSP.2012.2187446.
15. Wang D., Chen J. Supervised speech separation based on deep learning: an overview. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2018. Vol. 26, no. 10. P. 1702–1726. DOI: 10.1109/TASLP.2018.2842159.
16. Wang Y., Narayanan A., Wang D. On training targets for supervised speech separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2014. Vol. 22, no. 12. P. 1849–1858. DOI: 10.1109/TASLP.2014.2352935.
17. Hochreiter S., Schmidhuber J. Long Short-Term Memory. *Neural Computation*. 1997. Vol. 9, no. 8. P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

18. Erdogan H., Hershey J.R., Watanabe S., Roux J.L. Phase-sensitive and recognition-boosted speech separation using deep recurrent neural networks. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, April 19–24, 2015. IEEE, 2015. P. 708–712. DOI: 10.1109/ICASSP.2015.7178061.
19. Weninger F., Erdogan H., Watanabe S., *et al.* Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR. Latent Variable Analysis and Signal Separation. Vol. 9237 / ed. by E. Vincent, A. Yeredor, Z. Koldovský, P. Tichavský. Cham: Springer, 2015. P. 91–99. Lecture Notes in Computer Science. DOI: 10.1007/978-3-319-22482-4\_11.
20. Chen J., Wang D. Long short-term memory for speaker generalization in supervised speech separation. The Journal of the Acoustical Society of America. 2017. Vol. 141, no. 6. P. 4705–4714. DOI: 10.1121/1.4986931.
21. Zermini A. Deep Learning for Speech Separation: PhD thesis / Zermini Alfredo. University of Surrey, faculty of engineering, physical sciences, Centre for Vision, Speech, Signal Processing (CVSSP), South East of England, UK, 2020. URL: <https://openresearch.surrey.ac.uk/esploro/outputs/doctoral/99512310402346#file-0>.
22. Xia S., Li H., Zhang X. Using Optimal Ratio Mask as Training Target for Supervised Speech Separation. 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, Malaysia, Dec. 12–15, 2017. IEEE, 2017. P. 163–166. DOI: 10.1109/APSIPA.2017.8282021.
23. Palmqvist M. Methods and algorithms for quality and performance evaluation of audio conferencing systems: PhD thesis / Palmqvist Maria. Umeå University, Faculty of Science, Technology, Department of Physics, Sweden, 2013. URL: <http://umu.diva-portal.org/smash/get/diva2:630382/FULLTEXT01.pdf>.
24. ITU-T Recommendation P. 862, Perceptual Evaluation of Speech Quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. 2001. URL: <https://www.itu.int/rec/T-REC-P.862-200102-I/en>.
25. Fu S.-W., Liao C.-F., Tsao Y. Learning with Learned Loss Function: Speech Enhancement with Quality-Net to Improve Perceptual Evaluation of Speech Quality. IEEE Signal Processing Letters. 2020. Vol. 27. P. 26–30. DOI: 10.1109/LSP.2019.2953810.
26. Allen J.B., Berkley D.A. Image method for efficiently simulating small-room acoustics. The Journal of the Acoustical Society of America. 1998. Vol. 65, no. 4. P. 943–950. DOI: 10.1121/1.382599.

# ДИАГНОСТИКА НЕИСПРАВНОСТЕЙ ПОДШИПНИКОВ КАЧЕНИЯ С ИСПОЛЬЗОВАНИЕМ ПИКОВ СПЕКТРА И НЕЙРОННЫХ СЕТЕЙ

© 2022 О.Л. Ибряева, М.Н. Мохаммад

*Южно-Уральский государственный университет*

*(454080 Челябинск, пр. им. В.И. Ленина, д. 76)*

*E-mail: ibriaevaol@susu.ru, mohnoufl@gmail.com*

Поступила в редакцию: 20.04.2022

Важнейшими составляющими деталей машин являются подшипники качения, контроль за состоянием которых необходим, так как возможные дефекты в их конструкции могут привести к неправильной работе или общему выходу машин из строя. Современные решения по диагностике неисправностей подшипников обычно используют сложные процессы извлечения признаков, например, построение их изображений спектра Гильберта и дальнейшую мощную нейронную сеть для их классификации. В этой статье мы предлагаем простой, но, тем не менее, эффективный алгоритм решения данной задачи. Для выделения признаков из сигнала мы делим спектр сигнала на равные подинтервалы и находим максимум амплитуды и соответствующее значение частоты в каждом из них. В статье, на основе метода t-SNE, показано, что выделенные таким образом признаки, несмотря на свой небольшой размер, хорошо представляют разного типа сигналы. На втором этапе выделенные признаки поступают на вход простой нейронной сети классификатора. Предложенный метод обладает простотой в вычислительном отношении, как на этапе выделения признаков, так и на этапе обучения нейронной сети. Несмотря на это, метод дает 100% точность для всех типов сигналов на коротких данных из набора данных IMS.

*Ключевые слова: диагностика неисправностей, подшипник качения, спектр сигнала, нейронные сети, t-SNE.*

## ОБРАЗЕЦ ЦИТИРОВАНИЯ

Ибряева О.Л., Мохаммад М.Н. Диагностика неисправностей подшипников качения с использованием пиков спектра и нейронных сетей // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2022. Т. 11, № 2. С. 59–71. DOI: 10.14529/cmse220205.

## Введение

Подшипники качения играют ключевую роль практически в любой системе с механизмом вращения, такой как корабли, станы холодной прокатки и многие другие [1]. Ввиду чрезвычайной важности обнаружения неисправностей подшипников в области машиностроения к настоящему времени существует множество подходов к их диагностике. В погоне за большей точностью современные подходы часто используют вычислительно затратные методы, например, получают признаки с помощью построения спектральных изображений, на которых затем обучают мощную сверточную нейронную сеть [2, 3].

Целью настоящей статьи явилась разработка вычислительно простого, но эффективного метода диагностики подшипников качения, который основан на применении быстрого преобразования Фурье (БПФ) к вибрационным сигналам подшипника. Извлеченные из спектра сигнала признаки используются далее в качестве входных данных нейронной сети для определения типа неисправности. Репрезентативность выделенных признаков показана с помощью метода t-Distributed Stochastic Neighbor (t-SNE) [4]. Предложенный метод протестирован на эталонном наборе данных IMS [5] для диагностики подшипников и по-

казал высокую эффективность. Дополнительные преимущества нашего нового метода заключаются в простой процедуре выделения признаков, их небольшом количестве и простой архитектуре нейронной сети.

Основные ключевые моменты данной статьи:

- Разработан новый подход к выделению признаков из вибрационных сигналов подшипника.
- Отличительной особенностью нового метода является простота процесса извлечения признаков.
- Новый метод использует небольшое количество признаков и простую нейронную сеть по сравнению с другими современными подходами.
- Эксперименты, проведенные на наборе данных IMS, доказывают высокую эффективность метода.

Структура данной работы такова. Раздел 1 содержит обзор аналогичных работ. В разделе 2 описана экспериментальная установка и набор данных IMS. Раздел 3 содержит описание процесса извлечения признаков из вибрационных сигналов подшипника и структуры нейронной сети. Экспериментальные исследования представлены в разделе 4. Последний раздел содержит заключение и выводы исследования.

## 1. Обзор аналогов

По сути, диагностику неисправностей подшипников качения можно рассматривать как проблему распознавания образов, связанную с состоянием вращающегося оборудования. Обычно метод диагностики неисправностей состоит из двух ключевых этапов: извлечение признаков и классификация неисправностей.

Обработка сигналов на основе вибрации является одним из наиболее часто используемых методов на первом этапе, обычно с использованием анализа во временной области, частотной области или частотно-временной области. Анализ в частотной области обычно производится с помощью быстрого преобразования Фурье (БПФ), а также методов на основе БПФ, методов спектрального анализа [6] и т. д. Анализ во временной области рассчитывает статистические параметры, такие как среднеквадратичное значение виброускорения, эксцесс, ударные импульсы [7] и т. д. Частотно-временной анализ используется для расширения возможностей анализа в частотной области на нестационарные вибрационные сигналы. и включает в себя такие методы, как краткосрочное преобразование Фурье [8], вейвлет-преобразование [9], разложение по эмпирическим модам преобразования Гильберта—Хуанга [10] и т. д.

На втором этапе извлеченные признаки используются в качестве входных данных для методов машинного обучения с целью распознавания ошибок. Среди этих методов есть метод k-NN, метод опорных векторов и методы на основе дерева решений, но, конечно, преобладают нейронные сети, особенно сверточные и рекуррентные нейронные сети [11–13]. Например, Лю в [14] предложил систему диагностики неисправностей подшипников качения с рекуррентной нейронной сетью (RNN) и автоматическим кодировщиком. В работе [15] представлен метод, использующий разложение по вариационным модам на английском языке — Variational Mode Decomposition (VMD) и глубокие сверточные нейронные сети для выполнения классификации неисправностей подшипников качения в ветряных турбинах.

В целом, в диагностике неисправностей подшипников наблюдается тенденция к более сложным и глубоким моделям машинного обучения. Но глубокое обучение обнажает про-

блемы большого размера выборки и длительного времени обучения, а в реальных условиях работы количество доступных данных с помеченными ошибками относительно невелико, поэтому модель сложной архитектуры с хорошей обобщающей способностью и высокой точностью трудно обучить. Некоторые исследователи уже говорят о необходимости других, более простых и прикладных методов [16–18].

## 2. Описание набора данных

Для проверки работоспособности предлагаемого метода в данной статье используются экспериментальные данные IMS (Intelligent Maintenance Systems) [5], представленные Центром интеллектуальных сервисных систем Университета Цинциннати. Экспериментальная установка представлена на рис. 1. На валу были установлены четыре двухрядных подшипника Rexnord ZA-2115 с двумя акселерометрами PCB 353B33 на каждом. Скорость вращения вала поддерживалась постоянной на уровне 2000 оборотов в минуту. Он приводился в движение двигателем переменного тока, который был соединен с валом фрикционными ремнями. К подшипникам и валу с помощью пружинного механизма добавлялась нагрузка в 6000 фунтов. Частота дискретизации была установлена равной 20480 Гц. После вращения более чем на 100 миллионов оборотов в системе проявились такие неисправности подшипников, как дефект внутреннего кольца, внешнего кольца и шарика.

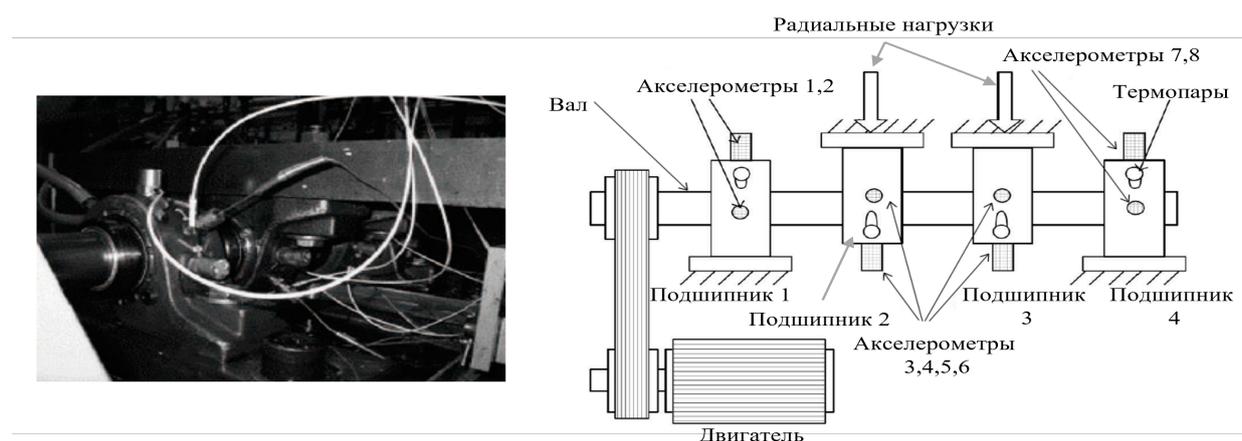
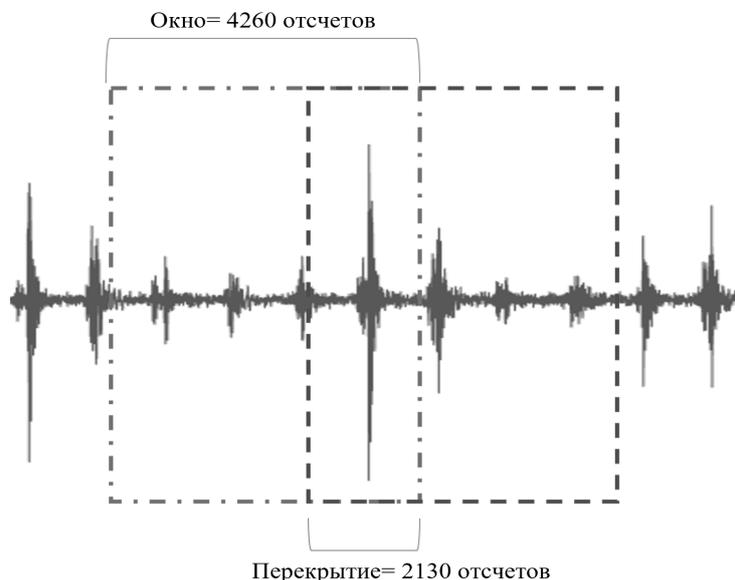


Рис. 1. Тестовая платформа IMS

IMS содержит наборы вибрационных сигналов подшипника с появившимися дефектами. Каждый набор содержит все записи от состояния исправного подшипника до его поломки. Мы рассмотрели сигналы каждого типа продолжительностью 20 секунд, как это было сделано в работе [19]. Для сигналов, отвечающих нормальному состоянию подшипника, мы взяли первые 20 секунд из набора данных для подшипника 1, когда он (как и все подшипники) был еще исправен. Последние 20 секунд записи для того же подшипника уже соответствуют его дефекту внутреннего кольца. Для сигнала с дефектом шарика были выбраны последние 20 секунд записи для подшипника 4. Наконец, запись с 2004-04-17 21:52:55 до 2004-04-18 01:02:55 с подшипника 3 содержит данные с дефектом внешнего кольца.

Использование всего 20 секунд данных каждого типа сигналов усложняет задачу их классификации, но является важным для практики, т.к. в реальных условиях, как правило, у нас очень мало образцов дефектных сигналов.

Для расширения данных сигналы вибрации были сегментированы на перекрывающиеся кадры по 4260 отсчетов, что эквивалентно  $(4260)/20480 = 0.208$  секунды, с процентом перекрытия 50%, т. е. 2130 отсчетов, как показано на рис. 2. В табл. 1 представлены классы сигналов набора данных IMS и количество образцов в каждом.



**Рис. 2.** Сегментация сигнала

**Таблица 1.** Классы набора данных IMS

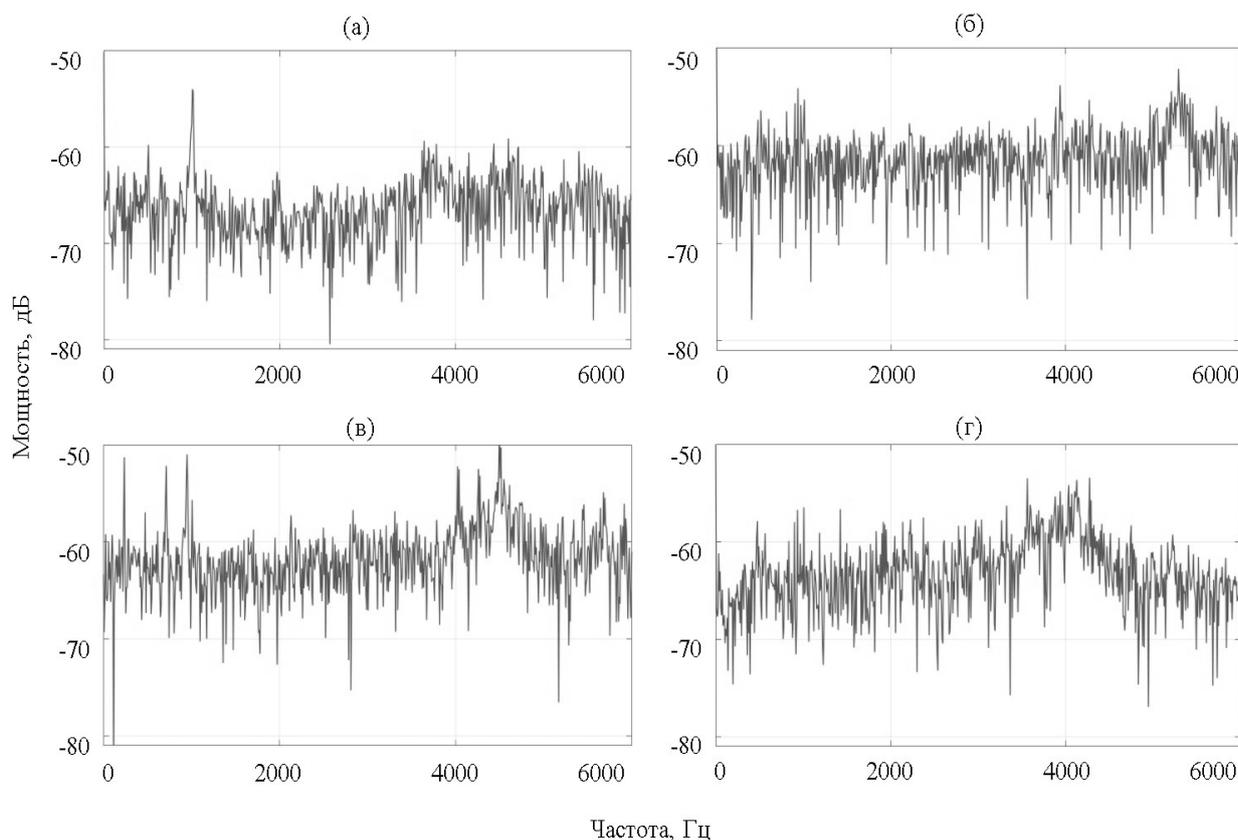
Класс	Тип дефекта	Число образцов в обучающей выборке	Число образцов в тестовой выборке
1	нет	134	56
2	дефект внутреннего кольца	134	56
3	дефект внешнего кольца	134	56
4	дефект шарика	134	56

### 3. Предлагаемый метод

Предлагаемое решение для классификации типов дефектов подшипников качения основано на нахождении пиков спектра сигнала и дальнейшем использовании выделенных признаков на входе нейронной сети. В этом разделе мы опишем оба этих этапа.

### 3.1. Извлечение признаков

На рис. 3 приведены примеры спектров сигналов разного типа. Можно видеть, что сигналы разного типа имеют существенно разные спектральные формы. Преобладающие спектральные компоненты каждого типа сигнала расположены в разных частотных диапазонах.



**Рис. 3.** Спектр различных сигналов; а) нормальный; б) дефект внутреннего кольца; в) дефект внешнего кольца; г) дефект шарика

Наш метод выделения признаков основан на этой особенности сигналов. Мы разделяем ось частот на равные интервалы и внутри каждого ищем частоту с максимальным значением амплитуды. Спектр сигнала при этом предварительно сглажен.

Извлекаемые признаки представляют собой два вектора: первый содержит значения частот с максимальной амплитудой в каждом частичном интервале, второй вектор содержит значения этих амплитуд, другими словами, признаки представляют собой пары  $(x, y) \in R^{N \times 2}$ , где  $N$  — количество интервалов.

Мы поделили весь интервал частот в спектре от 0 до 10240 Гц (половина частоты дискретизации) на  $N = 106$  равных интервалов, каждый из которых имеет длину примерно 100 Гц (рис. 4). Таким образом, каждому образцу сигнала из табл. 1 мы поставили в соответствие два 106 мерных вектора (значений частот и амплитуд), вычисленные по его спектру.

Чтобы проверить, что извлеченные признаки адекватно представляют объекты различных классов, мы использовали метод t-SNE [4], который позволяет нам отобразить точки из многомерного пространства в трехмерное и посмотреть на их расположение в пространстве.

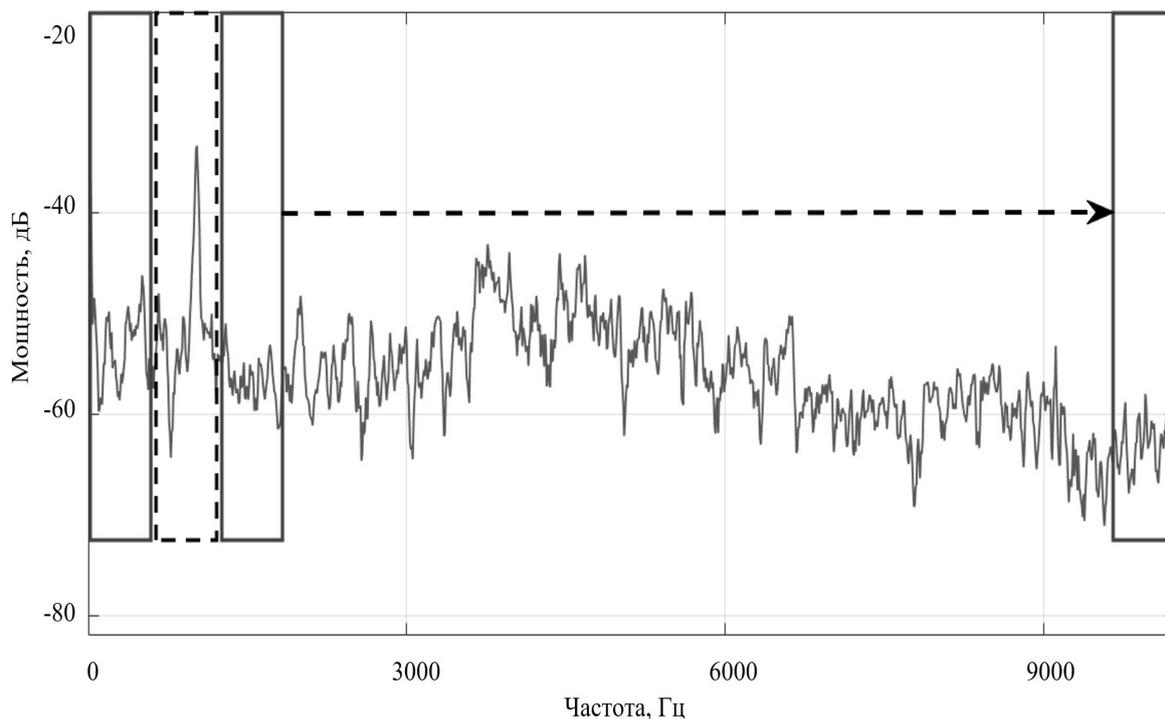


Рис. 4. Разбиение спектра на интервалы и нахождение максимальных значений каждого интервала

На рис. 5 построено такое представление для значений  $x$  пар  $(x, y)$ , полученное с помощью официального сайта TensorFlow (<https://projector.tensorflow.org/>). Можно видеть, что 4 типа сигналов сгруппированы в 4 отдельных кластера, что указывает на хорошее качество выделенных признаков и на то, что их можно успешно использовать для решения задачи классификации.

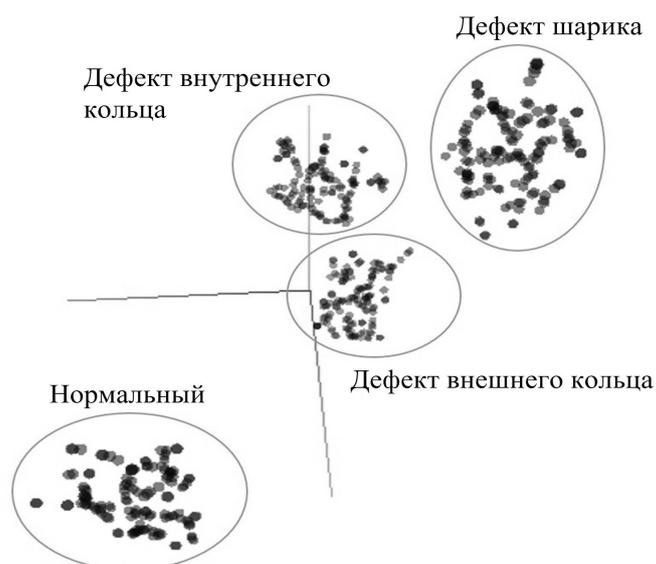


Рис. 5. Трехмерная визуализация извлеченных признаков с использованием t-SNE всех типов сигналов IMS

### 3.2. Модель нейронной сети

В качестве классификатора, который принимает на вход признаки, извлеченные методом описанным в прошлом разделе, мы использовали нейронную сеть простой архитектуры. Сеть состоит из двух отдельных модулей, обрабатывающих значения максимальных частот и их амплитуд, соответственно. За входным слоем каждого модуля следуют два скрытых слоя, после чего модули объединяются. Далее следуют еще два скрытых слоя и, наконец, выходной слой из 4 нейронов (по количеству классов), как показано на рис. 6.

При обучении мы традиционно использовали категориальную кросс-энтропию в качестве функции потерь и оптимизатор Адам. Первоначально количество эпох обучения было установлено на 2000, но мы использовали коллбэки Keras: EarlyStopping и ModelCheckpoint, которые на практике уменьшили это число эпох. Эти коллбэки сохраняют лучшую нейронную сеть (с наименьшим значением функции потерь на валидационном наборе). При отсутствии улучшения (уменьшения функции потерь) в течение patience эпох, они останавливают обучение и берут ранее сохраненную лучшую модель. Параметр patience («терпение») был установлен равным 100 эпох.

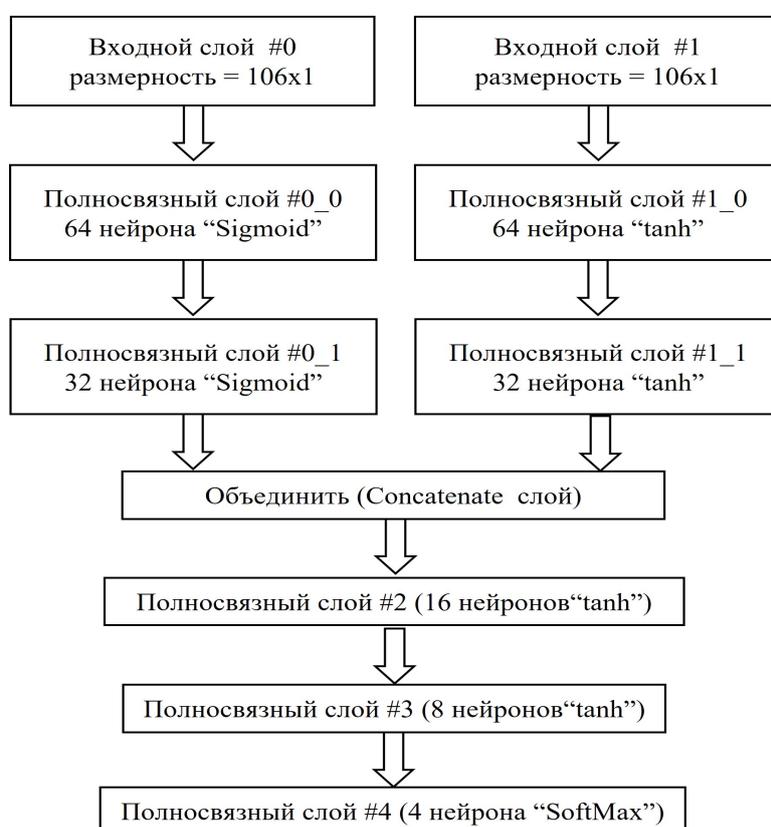


Рис. 6. Схема используемой нейронной сети

## 4. Результаты экспериментов

Используем новый предложенный метод для классификации сигналов из набора данных IMS и, чтобы иметь возможность сравнить с уже существующим результатом в [19], проведем эксперимент в похожих условиях. Мы также, как и авторы статьи [19], взяли небольшие отрезки данных (20 секунд) и разделили выборку на обучающие выборки (70%) и тестовые выборки (30%) набора.

Для оценки нашей модели используется метрика Accuracy. Accuracy вычисляется по следующей формуле:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} ,$$

где TP — истинно положительные образцы (true positive), TN — истинно отрицательные образцы (true negative), FP — ложно положительные образцы (false positive), и FN — ложно отрицательные образцы (false negative).

Авторы статьи [19] решали данную задачу классификации на основе глубокой нейронной сети и, как показал их эксперимент (табл. 2), наилучшие результаты показала сеть с 5 слоями со средней точностью 94.4% на тестовых данных. В табл. 2 приведен также результат, полученный нашим методом, он дает точность 100% для всех типов сигналов.

Итак, предложенный метод обеспечивает высокую точность классификации неисправностей, несмотря на его простоту.

**Таблица 2.** Сравнение результатов предложенного решения с [19]

Класс	Глубокие нейронные сети, учитывающие временную когерентность								Наш метод	
	3-х слойная модель		4-х слойная модель		5-х слойная модель		6-х слойная модель		проверка, %	тест, %
	проверка, %	тест, %	проверка, %	тест, %	проверка, %	тест, %	проверка, %	тест, %		
1	98.9	98.2	99.3	98.2	98.2	98.9	97.8	98.9	100	100
2	92.3	92.6	91.9	91.5	93.8	91.5	90.1	90.8	100	100
3	100	100	100	100	100	100	100	100	100	100
4	93.0	84.9	90.8	85.7	91.9	87.1	87.9	83.5	100	100
среднее	<b>96.6</b>	93.9	95.5	93.8	<b>96.0</b>	<b>94.4</b>	93.9	93.3	<b>100</b>	<b>100</b>

## Заключение

В данной работе был разработан новый метод для обработки вибрационных сигналов подшипников качения с целью классификации их неисправностей.

Основная идея предложенного решения заключается в определении новых признаков, которые могут эффективно характеризовать особенности спектров сигналов, отвечающих различным состояниям подшипников качения. Для извлечения признаков мы разделили ось частот на равные интервалы и внутри каждого нашли частоту с максимальным значением амплитуды. Извлеченные признаки представляют собой два вектора: первый содержит значения частот с максимальной амплитудой в каждом частичном интервале, второй вектор содержит значения этих амплитуд. Как показано в работе, данные признаки являются информативными и позволяют хорошо различить сигналы с подшипников разного типа.

В качестве классификатора, который принимает на вход извлеченные признаки, мы использовали нейронную сеть простой архитектуры. Сеть состоит из двух отдельных модулей: первый обрабатывает значения максимальных амплитуд спектра, а второй — значения соответствующих им частот. За входным слоем каждого модуля следуют два скрытых слоя, после чего модули объединяются. Далее следуют еще два скрытых слоя и, наконец, выходной слой из 4 нейронов.

Достоинствами предложенного подхода являются вычислительная простота, небольшое количество признаков, легкость в возможной реализации на микропроцессоре. Несмотря на простоту, метод показывает высокую эффективность на коротких наборах данных IMS, превосходящую современные аналоги.

В дальнейшем мы планируем применить предложенный метод к другому набору данных, который является более сложным и содержит данные с подшипников с разным уровнем нагрузки и степенью дефекта.

## Литература

1. Yu X., Dong F., Ding E., *et al.* Rolling Bearing Fault Diagnosis Using Modified LFDA and EMD With Sensitive Feature Selection // IEEE Access. 2018. Vol. 6. P. 3715–3730. DOI: 10.1109/ACCESS.2017.2773460.
2. Xu Y., Li Z., Wang S., *et al.* A hybrid deep-learning model for fault diagnosis of rolling bearings // Measurement. 2021. Vol. 169. P. 108502. DOI: 10.1016/j.measurement.2020.108502.
3. Han H., Wang H., Liu Z., Wang J. Intelligent vibration signal denoising method based on non-local fully convolutional neural network for rolling bearings // ISA Transactions. 2022. Vol. 122. P. 13–23. DOI: 10.1016/j.isatra.2021.04.022.
4. Hamad R., Jarpe E., Lundstrom J. Stability Analysis of the t-SNE Algorithm for Human Activity Pattern Data // Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018. 2019. P. 1839–1845. DOI: 10.1109/SMC.2018.00318.
5. Prognostics Center of Excellence - Data Repository. URL: <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#bearing> (accessed: 21.01.2022).
6. Wang W., Ismail F., *et al.* An enhanced bispectrum technique with auxiliary frequency injection for induction motor health condition monitoring // IEEE Transactions on Instrumentation and Measurement. 2015. Vol. 64, no. 10. P. 2679–2687. DOI: 10.1109/TIM.2015.2419031.
7. Zhen L., Zhengjia H., Yanyang Z., Xuefeng C. Bearing condition monitoring based on shock pulse method and improved redundant lifting scheme // Mathematics and Computers in Simulation. 2008. Vol. 79, no. 3. P. 318–338. DOI: 10.1016/j.matcom.2007.12.004.
8. Wang W., McFadden P. Early detection of gear failure by vibration analysis i. calculation of the time-frequency distribution // Mechanical Systems and Signal Processing. 1993. Vol. 7, no. 3. P. 193–203. DOI: 10.1006/mssp.1993.1008.
9. Luo G.Y., Osypiw D., Irle M. On-line vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing // Journal of vibration and control. 2003. Vol. 9, no. 8. P. 931–947.
10. Lei Y., Lin J., He Z., Zuo M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery // Mechanical Systems and Signal Processing. 2013. Vol. 35, no. 1-2. P. 108–126. DOI: 10.1016/j.ymsp.2012.09.015.
11. Yuan Z., Zhang L., Duan L., Li T. Intelligent fault diagnosis of rolling element bearings based on HHT and CNN // 2018 Prognostics and System Health Management Conference (PHM-Chongqing). IEEE, 2018. P. 292–296.

12. Liang P., Deng C., Wu J., *et al.* Intelligent Fault Diagnosis of Rolling Element Bearing Based on Convolutional Neural Network and Frequency Spectrograms // 2019 IEEE International Conference on Prognostics and Health Management (ICPHM). IEEE, 2019. P. 1–5. DOI: 10.1109/ICPHM.2019.8819444.
13. Shao H., Jiang H., Zhao H., Wang F. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis // Mechanical Systems and Signal Processing. 2017. Vol. C, no. 95. P. 187–204. DOI: 10.1016/J.YMSSP.2017.03.034.
14. Liu H., Zhou J., Zheng Y., *et al.* Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders // ISA transactions. 2018. Vol. 77. P. 167–178. DOI: 10.1016/J.ISATRA.2018.04.005.
15. Xu Z., Li C., Yang Y. Fault diagnosis of rolling bearing of wind turbines based on the Variational Mode Decomposition and Deep Convolutional Neural Networks // Applied Soft Computing. 2020. Vol. 95. P. 106515. DOI: 10.1016/J.ASOC.2020.106515.
16. Chen Z., Cen J., Xiong J. Rolling Bearing Fault Diagnosis Using Time-Frequency Analysis and Deep Transfer Convolutional Neural Network // IEEE Access. 2020. Vol. 8. P. 150248–150261. DOI: 10.1109/ACCESS.2020.3016888.
17. Zhang W., Peng G., Li C., *et al.* A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals // Sensors. 2017. Vol. 17, no. 2. P. 425. DOI: 10.3390/s17020425.
18. Li X., Wang J., Zhang B. Fault diagnosis of rolling element bearing weak fault based on sparse decomposition and broad learning network // Transactions of the Institute of Measurement and Control. 2020. Vol. 42, no. 2. P. 169–179. DOI: 10.1177/0142331219864820.
19. Zhang R., Peng Z., Wu L., *et al.* Fault Diagnosis from Raw Sensor Data Using Deep Neural Networks Considering Temporal Coherence // Sensors. 2017. Vol. 17, no. 3. P. 549. DOI: 10.3390/S17030549.

Ибряева Ольга Леонидовна, к.ф.-м.н., доцент, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Мохаммад Мохаммад Науфол, студент, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

# FAULT DIAGNOSIS OF ROLLING BEARINGS USING SPECTRAL PEAKS AND NEURAL NETWORKS

© 2022 O.L. Ibryaeva, M.N. Mohammad

*South Ural State University (pr. Lenina 76, Chelyabinsk, 454080 Russia)*

*E-mail: ibriaevaol@susu.ru, mohnoufl@gmail.com*

Received: 20.04.2022

The most important components of machine parts are rolling bearings, the condition of which is necessary to control, since possible defects in their design can lead to incorrect operation or general failure of machines. Modern solutions on fault diagnosis of bearings typically use complex feature extraction processes, such as their Hilbert spectrum imaging and a further powerful neural network to classify them. In this article, we propose a simple, but, nevertheless, an effective algorithm for solving this problem. To extract features from a signal, we divide the signal spectrum into equal subintervals and find the amplitude maximum and the corresponding frequency value in each of them. In the article, based on the t-SNE method, it is shown that the features selected in this way, despite their small size, represent different types of signals well. At the second stage, the selected features are fed to the input of a simple classifier neural network. The proposed method is computationally simple, both at the stage of feature extraction and at the stage of neural network training. Despite this, the method gives 100% accuracy for all types of signals on short data from the IMS dataset.

*Keywords: fault diagnosis, rolling bearing, signal spectrum, neural networks, t-SNE.*

## FOR CITATION

Ibryaeva O.L., Mohammad M.N. Fault Diagnosis of Rolling Bearings Using Spectral Peaks and Neural Networks. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2022. Vol. 11, no. 2. P. 59–71. (in Russian) DOI: 10.14529/cmse220205.

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Yu X., Dong F., Ding E., *et al.* Rolling Bearing Fault Diagnosis Using Modified LFDA and EMD With Sensitive Feature Selection. IEEE Access. 2018. Vol. 6. P. 3715–3730. DOI: 10.1109/ACCESS.2017.2773460.
2. Xu Y., Li Z., Wang S., *et al.* A hybrid deep-learning model for fault diagnosis of rolling bearings. Measurement. 2021. Vol. 169. P. 108502. DOI: 10.1016/j.measurement.2020.108502.
3. Han H., Wang H., Liu Z., Wang J. Intelligent vibration signal denoising method based on non-local fully convolutional neural network for rolling bearings. ISA Transactions. 2022. Vol. 122. P. 13–23. DOI: 10.1016/j.isatra.2021.04.022.
4. Hamad R., Jarpe E., Lundstrom J. Stability Analysis of the t-SNE Algorithm for Human Activity Pattern Data. Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018. 2019. P. 1839–1845. DOI: 10.1109/SMC.2018.00318.

5. Prognostics Center of Excellence - Data Repository. URL: <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#bearing> (accessed: 21.01.2022).
6. Wang W., Ismail F., *et al.* An enhanced bispectrum technique with auxiliary frequency injection for induction motor health condition monitoring. *IEEE Transactions on Instrumentation and Measurement*. 2015. Vol. 64, no. 10. P. 2679–2687. DOI: 10.1109/TIM.2015.2419031.
7. Zhen L., Zhengjia H., Yanyang Z., Xuefeng C. Bearing condition monitoring based on shock pulse method and improved redundant lifting scheme. *Mathematics and Computers in Simulation*. 2008. Vol. 79, no. 3. P. 318–338. DOI: 10.1016/j.matcom.2007.12.004.
8. Wang W., McFadden P. Early detection of gear failure by vibration analysis i. calculation of the time-frequency distribution. *Mechanical Systems and Signal Processing*. 1993. Vol. 7, no. 3. P. 193–203. DOI: 10.1006/mssp.1993.1008.
9. Luo G.Y., Osypiw D., Irle M. On-line vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing. *Journal of vibration and control*. 2003. Vol. 9, no. 8. P. 931–947.
10. Lei Y., Lin J., He Z., Zuo M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*. 2013. Vol. 35, no. 1-2. P. 108–126. DOI: 10.1016/j.ymssp.2012.09.015.
11. Yuan Z., Zhang L., Duan L., Li T. Intelligent fault diagnosis of rolling element bearings based on HHT and CNN. 2018 Prognostics and System Health Management Conference (PHM-Chongqing). IEEE, 2018. P. 292–296.
12. Liang P., Deng C., Wu J., *et al.* Intelligent Fault Diagnosis of Rolling Element Bearing Based on Convolutional Neural Network and Frequency Spectrograms. 2019 IEEE International Conference on Prognostics and Health Management (ICPHM). IEEE, 2019. P. 1–5. DOI: 10.1109/ICPHM.2019.8819444.
13. Shao H., Jiang H., Zhao H., Wang F. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*. 2017. Vol. C, no. 95. P. 187–204. DOI: 10.1016/J.YMSSP.2017.03.034.
14. Liu H., Zhou J., Zheng Y., *et al.* Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA transactions*. 2018. Vol. 77. P. 167–178. DOI: 10.1016/J.ISATRA.2018.04.005.
15. Xu Z., Li C., Yang Y. Fault diagnosis of rolling bearing of wind turbines based on the Variational Mode Decomposition and Deep Convolutional Neural Networks. *Applied Soft Computing*. 2020. Vol. 95. P. 106515. DOI: 10.1016/J.ASOC.2020.106515.
16. Chen Z., Cen J., Xiong J. Rolling Bearing Fault Diagnosis Using Time-Frequency Analysis and Deep Transfer Convolutional Neural Network. *IEEE Access*. 2020. Vol. 8. P. 150248–150261. DOI: 10.1109/ACCESS.2020.3016888.
17. Zhang W., Peng G., Li C., *et al.* A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals. *Sensors*. 2017. Vol. 17, no. 2. P. 425. DOI: 10.3390/s17020425.

18. Li X., Wang J., Zhang B. Fault diagnosis of rolling element bearing weak fault based on sparse decomposition and broad learning network. Transactions of the Institute of Measurement and Control. 2020. Vol. 42, no. 2. P. 169–179. DOI: 10.1177/0142331219864820.
19. Zhang R., Peng Z., Wu L., *et al.* Fault Diagnosis from Raw Sensor Data Using Deep Neural Networks Considering Temporal Coherence. Sensors. 2017. Vol. 17, no. 3. P. 549. DOI: 10.3390/S17030549.

## СВЕДЕНИЯ ОБ ИЗДАНИИ

Научный журнал «Вестник ЮУрГУ. Серия «Вычислительная математика и информатика» основан в 2012 году.

Учредитель — Федеральное государственное автономное образовательное учреждение высшего образования «Южно-Уральский государственный университет» (национальный исследовательский университет).

Главный редактор — Л.Б. Соколинский.

Свидетельство о регистрации ПИ ФС77-57377 выдано 24 марта 2014 г. Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций.

Журнал включен в Реферативный журнал и Базы данных ВИНИТИ; индексируется в библиографической базе данных РИНЦ. Журнал размещен в открытом доступе на Всероссийском математическом портале MathNet. Сведения о журнале ежегодно публикуются в международной справочной системе по периодическим и продолжающимся изданиям «Ulrich's Periodicals Directory».

Решением Президиума Высшей аттестационной комиссии Министерства образования и науки Российской Федерации журнал включен в «Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук» по научным специальностям и соответствующим им отраслям науки: 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей (физико-математические науки), 05.13.17 – Теоретические основы информатики (физико-математические науки).

Подписной индекс научного журнала «Вестник ЮУрГУ», серия «Вычислительная математика и информатика»: 10244, каталог «Пресса России». Периодичность выхода — 4 выпуска в год.

Адрес редакции, издателя: 454080, г. Челябинск, проспект Ленина, 76, Издательский центр ЮУрГУ, каб. 32.

## ПРАВИЛА ДЛЯ АВТОРОВ

1. Правила подготовки рукописей и пример оформления статей можно загрузить с сайта серии <http://vestnikvmi.susu.ru>. Статьи, оформленные без соблюдения правил, к рассмотрению не принимаются.
2. Адрес редакционной коллегии научного журнала «Вестник ЮУрГУ», серия «Вычислительная математика и информатика»:  
Россия 454080, г. Челябинск, пр. им. В.И. Ленина, 76, ЮУрГУ, кафедра СП,  
ответственному секретарю Цымблеру М.Л.
3. Адрес электронной почты редакции: [vestnikvmi@susu.ru](mailto:vestnikvmi@susu.ru)
4. Плата с авторов за публикацию рукописей не взимается, и гонорары авторам не выплачиваются.

ВЕСТНИК  
ЮЖНО-УРАЛЬСКОГО  
ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА  
Серия  
«ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА И ИНФОРМАТИКА»  
Том 11, № 2  
2022

16+

Техн. редактор *А.В. Миних*

Издательский центр Южно-Уральского государственного университета

Подписано в печать 31.05.2022. Дата выхода в свет 17.06.2022. Формат 60×84 1/8. Печать цифровая.  
Усл. печ. л. 8,83. Тираж 500 экз. Заказ 151/170. Цена свободная.

Отпечатано в типографии Издательского центра ЮУрГУ.  
454080, г. Челябинск, проспект Ленина, 76.