

## УТВЕРЖДАЮ



Проректор по научной работе  
Университета ИТМО  
д.т.н., профессор

В.О. Никифоров

\_\_\_\_\_ 2024 г.

## ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет ИТМО» на диссертационную работу КРАЕВОЙ Яны Александровны «Масштабируемые методы и алгоритмы поиска аномалий во временных рядах», представленную на соискание ученой степени кандидата физико-математических наук по специальности 2.3.5 – математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

### Актуальность темы диссертации

Исследование, выполненное Я.А. Краевой в рамках диссертационной работы, направлено на решение задачи, связанной с поиском аномалий во временном ряде. Соискателем обоснованно указаны следующие факторы, подтверждающие актуальность темы диссертации. Временные ряды являются одним из наиболее важных видов данных, подвергаемых интеллектуальному анализу, и встречаются в широком спектре научных и практических задач: цифровая индустрия, финансовое прогнозирование, персональное здравоохранение и др. Поиск аномалий является одной из основных задач интеллектуального анализа временных рядов, востребованной в различных областях. Важным и актуальным аспектом проблемы поиска аномалий во временных рядах является случай потокового временного ряда, элементы которого поступают в систему в режиме реального времени. Как правило, в таких системах за короткое время продуцируются больше временные ряды,

которые не могут быть размещены в оперативной памяти одного процессора. В настоящее время одной из наиболее распространенных многоядерных архитектур являются графические процессоры, значительно опережающие центральные процессоры по производительности. Кроме того, одной из современных тенденций организации высокопроизводительных вычислений является объединение нескольких вычислительных узлов посредством высокоскоростной соединительной сети в кластер. В соответствии с этим актуальной является задача разработки методов и алгоритмов поиска аномалий во временных рядах на платформе современных высокопроизводительных вычислительных систем.

### **Цель, задачи и основные результаты диссертации**

Цель исследования Я.А. Краевой состояла в разработке масштабируемых методов и алгоритмов поиска аномалий во временных рядах. Соискателем были решены следующие задачи для достижения этой цели:

1. Разработан параллельный алгоритм поиска диссонансов временного ряда, имеющих фиксированную длину, для графического процессора.
2. Разработан параллельный алгоритм поиска диссонансов временного ряда, имеющих произвольную длину, для графического процессора.
3. Разработан параллельный алгоритм поиска диссонансов временного ряда, имеющих произвольную длину, для высокопроизводительного вычислительного кластера с графическими процессорами.
4. Разработан метод поиска аномалий в потоковом временном ряде на основе нейросетевых технологий и интеллектуального анализа данных.
5. Проведены вычислительные эксперименты, показавшие, что предложенные решения эффективны и не уступают известным аналогам.

### **Содержание текста диссертации**

*Первая глава* посвящена обзору современных методов и алгоритмов поиска аномалий во временных рядах. Даны определения различных видов аномалий, которые могут содержать временные ряды. Дается классификация подходов к поиску аномалий. Вводятся формальные определения и нотация

для обозначения используемых в последующих главах понятий. Дается детальный обзор последовательных и параллельных алгоритмов поиска аномалий на основе концепции диссонанса временного ряда, наиболее близко относящихся к теме диссертации.

*Во второй главе* представлен новый параллельный алгоритм PD3, разработанный для поиска диссонансов фиксированной длины во временном ряде, для графического процессора. Рассмотрены основные принципы распараллеливания алгоритма PD3: параллелизм по данным, векторизация вычислений и выравнивание данных в памяти. Представлены введенные соискателем векторные структуры данных для обеспечения возможности векторизации вычислений. Приведена схема сегментации временного ряда, в которой применяется предлагаемый способ разбиения с перекрытием. Дано подробное описание CUDA-ядер, реализующих основные фазы алгоритма PD3, на уровне блоков сетки и на уровне работы нитей в блоке. Приведены описание и анализ экспериментов. Результаты экспериментов с реальными и синтетическими данными показывают, что разработанный алгоритм опережает известных аналогов по быстродействию.

*В третьей главе* представлен новый параллельный алгоритм PALMAD, решающий задачу поиска диссонансов произвольной длины из заданного диапазона на графических процессорах. Представлены принципы распараллеливания алгоритма PALMAD. Доказано утверждение о рекуррентных формулах, связывающих вычисление среднего арифметического и среднеквадратичного отклонения подпоследовательностей соседних длин временного ряда, что позволяет сократить объем вычислений при поиске диссонансов. Представлены метод для визуализации диссонансов различных длин на основе построения тепловой карты и алгоритм ранжирования диссонансов, имеющих длину из заданного диапазона. Описаны результаты тематических исследований по поиску диссонансов в сенсорных данных на примере временных рядов из следующих областей: машиностроение, металлургия и энергетика. Результаты

показали, что найденные диссонансы адекватно отражают необычное поведение систем из исследуемых областей. Представлены результаты вычислительных экспериментов, подтверждающие высокую эффективность алгоритма PALMAD.

*Четвертая глава* представляет новый параллельный алгоритм PADDi, осуществляющий поиск диссонансов временного ряда, имеющих длину в заданном диапазоне, для высокопроизводительного вычислительного кластера с графическими процессорами. Данный алгоритм разработан для случая, когда временной ряд не может быть целиком размещен в оперативной памяти одного графического процессора. Описано предложенное соискателем двухуровневое распараллеливание вычислений: на уровне узлов кластерной системы используется технология MPI, в рамках одного графического процессора узла кластера используется технология CUDA. Приведена общая схема вычислений алгоритма PADDi. Также представлено описание распараллеливания новой введенной фазы очистки диссонансов в алгоритме PADDi. Представлены результаты вычислительных экспериментов над реальными и синтетическими временными рядами, подтверждающие высокую эффективность алгоритма PADDi.

*Пятая глава* посвящена новому методу поиска аномалий в потоковом временном ряде, получивший название DiSSiD. Данный метод основан на концепции диссонансов и снippetов. Подробно описаны компоненты метода DiSSiD: нейросетевая модель и алгоритм формирования обучающей выборки для этой модели. Представлена архитектура модели, обучение и применение нейросетевой модели DiSSiD, а также пошаговое выполнение алгоритма формирования обучающей выборки, основанного на параллельных алгоритмах поиска снippetов и диссонансов во временном ряде. Описана модификация алгоритма поиска снippetов, позволяющая более точно по сравнению с оригинальным алгоритмом находить указанные подпоследовательности заданного временного ряда. Представлены результаты вычислительных экспериментов, подтверждающие высокую

точность DiSSiD при поиске аномалий и быстроедействие, достаточное для ее применения в задачах потоковой обработки временных рядов.

В *заключении* резюмируются итоги исследования, представляются отличия данной работы от ранее выполненных родственных работ других авторов, рассматриваются направления дальнейших исследований в данной области.

### **Обоснованность и достоверность полученных результатов**

Обоснованность и достоверность полученных результатов подтверждаются вычислительными экспериментами, которые соискатель выполнил в строгом соответствии с общепринятыми стандартами, а также объективным сравнением полученных результатов с известными аналогами. Кроме того, для репродуцируемости результатов экспериментов соискатель разместил исходные тексты своих разработок и наборы данных в свободном доступе в сети Интернет.

### **Научная новизна работы**

Научная новизна диссертационной работы состоит в разработке следующих новых алгоритмов и методов:

1. Параллельный алгоритм PD3 для поиска диссонансов фиксированной длины на графическом процессоре, который в отличие от известных аналогов позволяет находить все диссонансы временного ряда, имеющие заданную длину.
2. Параллельные алгоритмы PALMAD и PADDi для поиска диссонансов произвольной длины на графическом процессоре и вычислительном кластере с графическими процессорами соответственно. Данные алгоритмы позволяют находить во временном ряде все диссонансы, которые имеют длину в заданном диапазоне.
3. Метод DiSSiD для поиска аномалий в потоковом временном ряде. В отличие от известных аналогов, метод позволяет выявлять аномальные подпоследовательности ряда, отражающие нетипичную и редко встречающуюся активности исследуемого субъекта.

## **Теоретическая и практическая ценность работы**

Теоретическая значимость исследования Я.А. Краевой состоит в следующем. Предложенные соискателем параллельные алгоритмы достигают высокой эффективности благодаря предложенным оригинальным схемам распараллеливания вычислений и организации данных в виде векторно-матричных структур, которые обеспечивают сокращение избыточных вычислений. Практическая значимость исследования Я.А. Краевой состоит в том, что разработанные параллельные алгоритмы поиска диссонансов временных рядов и метод обнаружения аномальных подпоследовательностей в потоковом временном ряде базируются на математически строго определяемой концепции диссонанса и потому не зависят от предметной области. Разработки соискателя минимизируют участие аналитика-эксперта (от которого требуется задать длину искомых аномалий) и тем самым обеспечивают применение результатов диссертационного исследования в широком спектре приложений интеллектуального анализа временных рядов. Кроме того, проведено экспериментальное исследование, подтвердившее применимость полученных для поиска аномалий в сенсорных данных цифровой индустрии.

## **Публикации и апробации**

Основные результаты, полученные Я.А. Краевой в рамках диссертационного исследования, достаточно полно опубликованы в авторитетных рецензируемых научных изданиях: три статьи в российских научных журналах Перечня ВАК (К-1) и две статьи в зарубежных журналах, приравненных к категории К-1 Перечня ВАК (квартили Q1 и Q2 Scopus и Web of Science соответственно). По результатам исследования соискателем сделано девять докладов на международных и всероссийских научных конференциях. Получено два свидетельства о государственной регистрации программ для ЭВМ.

## **Автореферат и текст диссертации**

Текст диссертации Я.А. Краевой оформлен в соответствии с требованиями Минобрнауки и написан в научном стиле, для которого

характерно четко структурированное, ясное, логичное и точное изложение материала. Автореферат диссертации в полной мере отражает содержание диссертации.

### **Соответствие диссертации паспорту научной специальности**

Результаты, полученные в диссертации, относятся к агностическим (не зависящим от предметной области) методам выявления аномалий во временных рядах и потому могут считаться основой для новых эффективных проблемно-ориентированных программно-аппаратных комплексов обработки больших временных рядов в различных предметных областях. Содержание и результаты диссертации соответствуют паспорту специальности 2.3.5 по следующим областям:

4. Интеллектуальные системы машинного обучения, управления базами данных и знаний, инструментальные средства разработки цифровых продуктов;
8. Модели и методы создания программ и программных систем для параллельной и распределенной обработки данных, языки и инструментальные средства параллельного программирования.

### **Замечания по диссертации**

В качестве *замечаний* к диссертации, которые не снижают общей значимости и высокого научного уровня полученных результатов, необходимо отметить следующее.

1. В формуле 1.7 на странице 27 указано, что в качестве функции расстояния между подпоследовательностями используется квадрат нормированного евклидова расстояния. Однако выбор данной функции автор ничем не аргументирует, кроме как скоростью вычисления. Так же не приведены аналоги функций расстояния и критерии при которых они могут использоваться в предложенном алгоритме. Отсутствия аналогов данной функции ведет к последующему снижению масштабируемости алгоритма.

2. В формуле 1.9 на странице 28 и далее по тексту раздела 1.4.3 вводится понятие длины подпоследовательности снippets. Однако выбор значения длины подпоследовательности, а так же оценку влияния этого значения на результаты построения профиля снippets автор не приводит и единственным критерием выбора является большая кратность отношения всей длины ряда к длине снippets. Таким образом предложенный алгоритм реализует поведение семейства переборных алгоритмов, что так же не является оптимальным подходом и может приводить к проблемам с масштабированием и устойчивостью работы алгоритма.
3. На рис. 2.4 и рис. 2.5 (стр. 54) приведены результаты сравнения предложенного алгоритма с существующими аналогами. Однако сравнения проведено только по 1 критерию (времы выполнения или среднее время на поиск 1 диссонанса), что не полностью отражает общую картину эффективности разработанного подхода. Приведенный анализ возможно дополнить результатам сравнения алгоритмов по точности и заблаговременности детектирования аномалий. Аналогичные замечания к рисункам 3.2 и 3.3 на страницах 69-70.
4. В формулах 5.2-5.4 (стр. 104) предложена модификация функции потерь архитектура нейросетевой модели DiSSiD для поиска аномалий в потоквых данных. Одним из предложенных автором гиперпараметров модели является минимальное (пороговое) расстояние функции MPdist. Однако при этом сама функция расстояние не рассматривается как гиперпараметр модели, что может приводить к эффекту «переобучения» предложенной сети, в следствии ограниченности пространства поиска гиперпараметров модели.
5. На рис. 5.7 (стр. 123) приведены результаты сравнения метода DiSSid с существующими аналогами. Выбор и показанные результаты определенных алгоритмов (в частности AE/IE-CAE или LSTM-AD) демонстрируют либо неактуальность предложенного метода, либо

дублирование подхода для решаемой задачи. Возможным решением является сокращение и актуализация списка предложенных методов (например добавления упомянутого в 1 главе метода MERLIN).

## **ЗАКЛЮЧЕНИЕ**

Диссертационная работа Я.А. Краевой представляет собой самостоятельную законченную научно-квалификационную работу, которая выполнена на высоком научно-технологическом уровне и в полной мере соответствует требованиям Положения о порядке присуждения ученых степеней, в том числе п. 9. В рамках исследования Я.А. Краевой разработаны новые масштабируемые методы и алгоритмы поиска аномалий во временных рядах, которые имеют существенное значение для решения задач интеллектуального анализа временных рядов на платформе современных высокопроизводительных вычислительных систем. Соискатель, Яна Александровна Краева, заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 2.3.5 – математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Диссертация и отзыв обсуждены и одобрены на заседании научного семинара Научно-исследовательского института наукоемких компьютерных технологий Университета ИТМО, протокол № 4 от 11.04. 2024.

### **Отзыв составили:**

Директор научно-исследовательского  
института наукоемких компьютерных  
технологий Университета ИТМО,  
д.т.н., профессор



А.В. Бухановский

**Наименование организации, предоставившей отзыв:**

Федеральное государственное автономное образовательное учреждение  
высшего образования «Национальный исследовательский университет  
ИТМО»

Адрес: 197101, Санкт-Петербург, Кронверкский пр., дом 49, литер А

Телефон: +7 (812) 480-00-00

Email: [od@itmo.ru](mailto:od@itmo.ru)

WWW: <https://www.itmo.ru/>