

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

кандидата физико-математических наук Никитенко Дмитрия Александровича **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 диссертационного исследования Я.А. Краевой, направленного на поиск аномалий во временных рядах, обусловлена следующими основными факторами. Во-первых, временные ряды встречаются в широком спектре практических задач, связанных с анализом данных: мониторинг производительности суперкомпьютерных систем, интеллектуальное управление зданиями, умные города и др. Во-вторых, поиск аномалий является одной из задач интеллектуального анализа временных рядов, решение которой востребовано во многих приложениях обработки временных рядов, в частности, в случае, когда элементы ряда поступают в режиме реального времени. В-третьих, наличие больших временных рядов, которые не могут быть целиком размещены в оперативной памяти, приводит к необходимости разработки параллельных алгоритмов для эффективного поиска аномалий на платформе массово распространенных в настоящее время графических процессоров и кластеров с узлами на их основе.

Цель диссертационного исследования Я.А. Краевой заключается в разработке новых методов и параллельных алгоритмов поиска аномалий во временных рядах на платформе современных высокопроизводительных вычислительных систем. Для выполнения данной цели Я.А. Краевой были решены следующие **задачи**:

1. Разработан комплекс параллельных алгоритмов поиска диссонансов фиксированной и произвольной длин для графического процессора и вычислительного кластера с графическими процессорами.

2. Разработаны нейросетевая модель для поиска аномалий в потоковом временном ряде в режиме реального времени.
3. Проведены вычислительные эксперименты, результаты которых показали эффективность предложенных методов и алгоритмов.

Первая глава диссертации посвящена обзору методов и алгоритмов поиска аномалий во временных рядах. Рассматриваются различные типы аномалий временного ряда, приводятся их определения. Дана таксономия существующих методов поиска аномалий, описаны их основные идеи и представители. Приведены формальные определения и нотация. Дан обзор публикаций, посвященных последовательным и параллельным алгоритмам поиска аномалий, которые формализуют понятие аномалии с помощью концепции диссонанса.

Во **второй главе** представлен новый параллельный алгоритм PD3 для поиска диссонансов фиксированной длины для графических процессоров. Алгоритм PD3 использует матрично-векторные структуры данных, позволяющие эффективно распараллелить вычисления на графическом процессоре за счет векторизации вычислений. Представлена схема сегментирования ряда, с помощью которой реализуется параллелизм по данным. Приведено описание параллельной реализации каждой из фаз, составляющих алгоритм PD3: предварительная обработка данных, отбор и очистка кандидатов. Представлены результаты вычислительных экспериментов, показывающих преимущество алгоритма PD3 в быстродействии по сравнению с известными аналогами.

В **третьей главе** представлен новый параллельный алгоритм PALMAD, предназначенный для поиска диссонансов временного ряда, имеющих длину в заданном диапазоне. Описана вычислительная схема параллельного алгоритма PALMAD, который в своей основе использует представленный в предыдущей главе алгоритм PD3. Доказаны рекуррентные формулы, связывающие вычисления расстояний подпоследовательностей ряда соседних длин, что позволило увеличить эффективность алгоритма благодаря

сокращению объема вычислений. Представлены матрично-векторные структуры данных для эффективной параллельной обработки данных. Предложены метод построения тепловой карты диссонансов для их визуализации и алгоритм нахождения наиболее значимых диссонансов независимо от их длин. Представлены результаты вычислительных экспериментов, показывающих высокую эффективность разработанного алгоритма PALMAD. Выполнены тематические исследования по поиску диссонансов во временных рядах, взятых из металлургии, энергетики и машиностроения.

Четвертая глава содержит описание нового параллельного алгоритма PADDi для поиска диссонансов, имеющих длину в заданном диапазоне, во временном ряде, который не может быть целиком размещен в оперативной памяти одного графического процессора. Описаны два уровня распараллеливания по данным: фрагментация распределяет вычислительную нагрузку и данные между узлами кластерной системы, сегментация – между графическими процессорами каждого узла. Дана общая схема вычислений алгоритма PADDi. Представлены результаты вычислительных экспериментов, показывающих высокую производительность и масштабируемость разработанного алгоритма PADDi на различных аппаратных платформах.

В пятой главе представлен новый метод DiSSiD, предназначенный для поиска аномальных подпоследовательностей в потоковом временном ряде в режиме реального времени. Дано описание алгоритма формирования обучающей выборки, в котором задействованы модифицированный алгоритм поиска снippetов и параллельные алгоритмы поиска диссонансов, представленные в предыдущих главах диссертационной работы. Описана архитектура нейронной сети, для которой предложена модифицированная функция контрастных потерь. Представлены результаты вычислительных экспериментов над реальными временными рядами из различных предметных областей и выполнено сравнение с известными аналогами, показавшие превосходство метода DiSSiD над ними.

Текст диссертации хорошо структурирован и изложен грамотным научным языком. **Текст автореферата** достаточно полно отражает содержание диссертации. Содержание соответствует паспорту специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Достоверность и обоснованность полученных результатов подтверждены вычислительными экспериментами, которые выполнены согласно общепринятым стандартам как с реальными, так и с синтезированными временными рядами. Произведено сравнение алгоритмов автора с алгоритмами–передовыми аналогами. Обеспечена возможность повторного проведения экспериментов со стороны научного сообщества, поскольку исходные тексты алгоритмов и наборы данных размещены соискателем в свободных Интернет-репозиториях.

Научная новизна диссертационного исследования заключается в следующем:

1. Разработан комплекс параллельных алгоритмов поиска аномалий во временных рядах, основанных на концепции диссонанса:
 - поиск диссонансов фиксированной длины на графическом процессоре;
 - поиск диссонансов произвольной длины на графическом процессоре;
 - поиск диссонансов произвольной длины на высокопроизводительном вычислительном кластере с графическими процессорами.
2. Разработана нейросетевая модель для поиска аномалий в потоковом временной ряде, использующая предложенные в исследовании параллельные алгоритмы для подготовки обучающей выборки.
3. Проведены вычислительные эксперименты, подтвердившие более высокую эффективность предложенных подходов по сравнению с известными аналогами.

Теоретическая ценность диссертационного исследования состоит в том, что предложены параллельные алгоритмы поиска аномалий, основанные

на оригинальных схемах распараллеливания и организации данных, благодаря которым обеспечивается сокращение избыточных вычислений и, в целом, увеличение быстродействия поиска.

Практическая ценность диссертационной работы заключается в следующем. Предложенные подходы к поиску аномалий могут быть применены в широком круге практических задач и приложений, связанных с обработкой временных рядов, например: медицина (анализ и мониторинг ЭКГ пациентов), моделирование климата и погоды (предсказание природных катаклизмов), администрирование суперкомпьютерных комплексов (выявление нештатных ситуаций при мониторинге вычислительных систем), умное производство (поиск и предсказание сбоев и отказов технологически сложного оборудования) и др. Разработки соискателя по поиску аномалий во временных рядах предназначены для массово распространенных в настоящее время графических процессоров и кластеров с узлами на их основе.

Публикации и апробации. Соискателем опубликовано 5 статей: 3 статьи в российских научных журналах Перечня ВАК (категории К-1) и 2 статьи в зарубежных журналах квартилей Q1 и Q2 по Scopus и Web of Science соответственно (приравнены к категории К-1 Перечня ВАК). Публикации адекватно отражают результаты диссертации. Сделано 9 докладов на международных и всероссийских научных конференциях. Зарегистрированы две программы для ЭВМ.

В качестве **замечаний** к диссертационной работе можно отметить следующие моменты.

1. Формальное определение диссонанса, на котором базируется исследование, задействует симметричную неотрицательную функцию расстояния между двумя подпоследовательностями (см. раздел 1.4.2). В исследовании в качестве такой функции применяется квадрат z -нормализованного евклидова расстояния. Вместе с тем хотелось бы прояснить, почему для вычисления расстояния не применяется иная

- функция, например, евклидово расстояние или динамическая трансформация времени (DTW, Dynamic Time Warping)?
2. В тексте диссертации определения некоторых терминов даны дважды: термин "матричный профиль" определен в гл. 1 на стр. 30 и в гл. 3 на стр. 62, термин "профиль снippets" определен в гл. 1 на стр. 28 и в гл. 5 на стр. 109. Более логичным выглядели бы однократное определение указанных терминов в гл. 1, где дается теоретический базис исследования, и перекрестная ссылка на соответствующие формулы в гл. 3 и 5.
 3. Апробация разработанных методов проводилась на общеизвестных наборах данных, упомянутых в табл. 2.1 в гл. 2 на стр. 52, в табл. 3.1 в гл. 3 на стр. 67, табл. 4.1 в гл. 4 на стр. 92, табл. 5.1 в гл. 5 на стр. 119. Вместе с тем хотелось бы видеть более явные требования к входным данным, на случай применения разработанных методов в реальной жизни.
 4. Приводя данные апробации методов на GPU (разделы 2.4, 3.3, 4.3, 5.4), хотелось бы в выводах увидеть оценку, какой из ресурсов GPU является приоритетным – число CUDA-ядер, объем памяти на GPU или их соотношение?

Указанные замечания не снижают общей значимости результатов диссертационного исследования, выполненного Я.А. Краевой.

ЗАКЛЮЧЕНИЕ

Считаю, что диссертационное исследование и полученные в нем результаты имеют существенное значение для решения задач интеллектуального анализа временных рядов, возникающих в широком спектре предметных областях. Диссертационное исследование по своим актуальности, научной новизне, практической значимости, личному вкладу автора, публикациям в рецензируемых научных изданиях и апробациям на научных конференциях в полной мере отвечает требованиям п. 9 Положения о порядке присуждения ученых степеней и соответствует специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей», а автор диссертации, Краева Яна

Александровна, заслуживает присуждения ученой степени кандидата физико-математических наук по указанной специальности.

Официальный оппонент:



Никитенко Дмитрий Александрович
«19» апрель 2024 г.

Кандидат физ.-мат. наук
ведущий научный сотрудник Лаборатории параллельных информационных технологий Научно-исследовательского вычислительного центра Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В. Ломоносова», г. Москва
Адрес организации: 119991, Москва, Ленинские горы, 1 стр. 4
Телефон: +7 (495) 939-54-24
Email: dan@parallel.ru

