

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

доктора технических наук, доцента Баркалова Константина 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 темы диссертации

Диссертационная работа Я.А. Краевой посвящена разработке масштабируемых методов и параллельных алгоритмов для интеллектуального анализа временных рядов для суперкомпьютерных систем с вычислительными узлами на базе графических процессоров. Тема исследования, несомненно, является актуальной в силу всепроникающего характера временных рядов и высокой востребованности эффективных решений задачи поиска в них аномалий в широком спектре предметных областей: медицина, моделирование климата, цифровая индустрия и др.

Цель, задачи и результаты диссертационного исследования

Целью диссертационного исследования Я.А. Краевой является разработка и исследование новых методов и алгоритмов поиска аномальных подпоследовательностей временных рядов для современных высокопроизводительных вычислительных систем на основе концепции диссонанса. Для выполнения данной цели Я.А. Краевой были решены следующие задачи.

1. Разработаны следующие параллельные алгоритмы поиска диссонансов временного ряда:
 - 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 временном ряде на основе концепции диссонанса.
3. Проведены вычислительные эксперименты, результаты которых показали эффективность предложенных методов и алгоритмов.

Соответствие специальности

Диссертационное исследование и полученные результаты соответствуют специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Содержание текста диссертации

Диссертация состоит из введения, пяти глав, заключения, списка литературы, списков рисунков, таблиц и алгоритмов, и приложения. Объем диссертации – 158 страниц, объем библиографии – 151 наименование.

Во **введении** обоснованы актуальность темы диссертационной работы и степень ее разработанности, сформулированы цель и задачи исследования, раскрыты новизна, теоретическая и практическая значимость полученных результатов, сформулирована методологическая основа диссертационного исследования, сделан обзор содержания диссертации.

Первая глава диссертации посвящена обзору методов и алгоритмов поиска аномалий во временных рядах. Дана классификация существующих подходов к поиску аномалий и представлены основные из них. Приведены формальные определения и нотация для обозначения основных понятий, используемых в диссертационной работе. Приведен обзор работ, близких к теме диссертации.

Во **второй главе** представлен новый параллельный алгоритм поиска диссонансов фиксированной длины PD3 для графических процессоров. Рассмотрены основные принципы распараллеливания алгоритма PD3 и приведены матрично-векторные структуры данных, позволившие эффективно распараллелить вычисления на графическом процессоре. Описана параллельная реализация фаз, составляющих алгоритм PD3. Представлены результаты вычислительных экспериментов на реальных и синтетических

временных рядах, в которых исследуется производительность алгоритма PD3 в сравнении с известными аналогами.

В третьей главе представлен новый параллельный алгоритм PALMAD, предназначенный для поиска диссонансов временного ряда, имеющих длину в заданном диапазоне. Описаны принципы распараллеливания алгоритма PALMAD, который в своей основе использует представленный в предыдущей главе алгоритм PD3. Доказано утверждение о рекуррентных формулах, позволяющих сократить объем вычислений и представлены матрично-векторные структуры данных для эффективной параллельной обработки данных. Предложен метод построения тепловой карты диссонансов для их визуализации и алгоритм ранжирования найденных диссонансов независимо от их длин. Представлены результаты вычислительных экспериментов, демонстрирующие эффективность разработанного алгоритма PALMAD, и результаты тематических исследований для поиска диссонансов в сенсорных данных из различных областей цифровой индустрии.

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. Описаны принципы применения разработанных ранее алгоритмов PD3 и PALMAD для распараллеливания указанной задачи. Представлены результаты вычислительных экспериментов, в которых исследуется производительность и масштабируемость разработанного алгоритма PADDi на различных аппаратных платформах.

В пятой главе представлен новый метод поиска аномальных подпоследовательностей в потоковом временном ряде, получивший название DiSSiD. Описаны компоненты метода DiSSiD. Для обучения нейросетевой модели предложена модифицированная функция контрастных потерь. Описана модификация алгоритма поиска снippetов, позволяющая более точно находить указанные подпоследовательности временного ряда. Предложена эвристика по подбору гиперпараметров для модифицированного

алгоритма поиска сниппетов. Представлены результаты вычислительных экспериментов над временными рядами из различных предметных областей.

Автореферат диссертации

Текст автореферата полно и адекватно отражает содержание диссертации.

Обоснованность и достоверность результатов диссертации

Достоверность и обоснованность полученных результатов подтверждаются результатами вычислительных экспериментов над реальными и синтетическими данными, проведенными в соответствии с общепринятыми стандартами, и выполнено сравнение полученных результатов с известными аналогами. Исходные тексты разработанных алгоритмов вместе с наборами данных из проведенных экспериментов размещены в свободно доступных Интернет-репозиториях.

Научная новизна результатов диссертации

Научная новизна диссертационного исследования заключается в следующем. Разработано семейство новых параллельных алгоритмов, которые, в отличие от известных аналогов, позволяют находить все диссонансы во временном ряде, имеющие заданную и произвольную длины, на графическом процессоре, и все диссонансы, имеющие произвольную длину, на высокопроизводительном вычислительном кластере с графическими процессорами во временном ряде, который не может быть целиком размещен в оперативной памяти вычислительного узла. Разработан новый метод поиска аномалий в потоковом временном ряде, который, в отличие от известных аналогов, позволяет выявлять аномальные подпоследовательности ряда, отражающие нетипичную и редко встречающуюся активности исследуемого субъекта.

Теоретическая и практическая ценность диссертации

Теоретическая ценность диссертационного исследования состоит в том, что в разработанных алгоритмах поиска диссонансов временного ряда предложены оригинальные схемы распараллеливания вычислений и

организации данных, которые обеспечивают сокращение избыточных вычислительных операций и существенное увеличение быстродействия поиска.

Практическая ценность диссертационного исследования заключается в том, что предложенные разработки могут применяться для выявления аномалий временных рядов, возникающих в широком спектре предметных областей: приложения цифровой индустрии и Интернета вещей, медицина, сейсмология и др. В указанных приложениях датчики, как правило, снимают показания с высокой частотой, генерируя тем самым большие временные ряды, которые требуют эффективной обработки, в т.ч. в режиме реального времени.

Публикации и апробация

По теме диссертационного исследования автором опубликованы 5 статей, в том числе 3 статьи в российских научных журналах категории К-1 Перечня ВАК и 2 статьи в зарубежных журналах, приравненных к журналам категории К-1 Перечня ВАК, индексируемых в квартилях Q1 и Q2 библиографических баз Scopus и Web of Science соответственно. Получено два свидетельства Роспатента о государственной регистрации программ для ЭВМ. Публикации с достаточной полнотой отражают основные результаты диссертации. Результаты исследования прошли апробацию на 9 международных и всероссийских научных конференциях.

Замечания по диссертации

По диссертационной работе Я.А. Краевой имеются следующие замечания, которые, однако, не снижают общего высокого научного уровня работы и значимости полученных результатов:

1. В работе предложен алгоритм PD3 для поиска всех диссонансов заданной длины, при этом эффективность его работы сравнивается с алгоритмами, обеспечивающими поиск одного (наиболее важного) диссонанса. Способ, основанный на усреднении времени работы алгоритма

PD3, дает возможность сравнить эти алгоритмы, но здесь корректно было бы сравнивать именно алгоритмы для поиска всех диссонансов.

2. Предложенный алгоритм PALMAD для поиска диссонансов произвольной длины предполагает задание интервала поиска для длин диссонансов, однако методика подбора указанного интервала не приведена.

3. В тексте отражены результаты экспериментов, показывающие влияние длины интервала для длин диссонансов на производительность алгоритма. В частности, рис. 3.5 и 3.6. показывают масштабируемость алгоритма PALMAD в зависимости от длины временного ряда и ширины диапазона длин диссонансов. Однако из приведенных результатов не очевидно, как длина интервала влияет на качество найденного решения и производительность в расчете на один найденный диссонанс.

4. Для нейронных сетей, задействованных в алгоритме DiSSiD, были использованы размеры ядер свертки 8×1 , 5×1 и 3×1 , а также использовалось значение $k = 95$ в качестве порогового значения. Выбор именно таких значений параметров не пояснен.

5. Эксперименты с методом DiSSiD проводились на временных рядах из различных предметных областей, в частности, сигналов ЭКГ. Из текста работы неясно, встречаются ли сигналы ЭКГ, снятые с одного человека, одновременно в обучающей и тестовой выборке. Если встречаются, то нейронная сеть могла «натренироваться» на этого человека, что искусственно повысит исследуемые метрики качества сети.

6. При описании параллельных алгоритмов в тексте диссертации используются специальные термины, связанные с программированием на графических процессорах, (например, варп, тайл), значение которых было бы лучше предварительно пояснить.

ЗАКЛЮЧЕНИЕ

Диссертация Я.А. Краевой представляет собой самостоятельно выполненную и законченную научно-квалификационную работу, в которой автором разработаны новые масштабируемые методы поиска аномалий во

временных рядах для современных высокопроизводительных вычислительных платформ, имеющие существенное значение для решения задач, связанных с интеллектуальным анализом временных рядов в широком спектре предметных областей. Диссертационная работа Я.А. Краевой в полной мере соответствует требованиям Положения о порядке присуждения ученых степеней, включая п. 9. Я.А. Краева заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Официальный оппонент:



Константин Александрович Баркалов
«11» апреля 2024 г.

Доктор техн. наук, доцент
Заведующий кафедрой математического обеспечения и суперкомпьютерных технологий Института информационных технологий, математики и механики Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского», г. Нижний Новгород
Адрес организации: 603022, г. Нижний Новгород, пр. Гагарина, 23
Телефон: +7 (831) 462-33-56
Email: barkalov@vmk.unn.ru

