УДК 004.272.25, 004.421, 004.032.24 DOI: 10.14529/cmse220203

# 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 разработанного ранее авторами параллельного алгоритма для аннотирования сенсорных данных с помощью графического процессора на основе концепции сниппетов. Сниппет 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, сниппет, 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 подпосле-

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

В нашей предыдущей работе [12] предложен параллельный алгоритм поиска сниппетов 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:

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

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 $i$:

$$T_{i,m} = (t_i, \ldots, t_{i+m-1}),\ 1 \leqslant m \ll n, 1 \leqslant i \leqslant n - m + 1. \tag{2}$$

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 $S_1, \ldots, S_{n/m}$:

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

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 $C_1, \ldots, C_K$:

$$C_T^m = (C_1, \ldots, C_K),\ C_i \in S_T^m. \tag{4}$$

Число $K$ $(1 \leqslant K \leqslant n/m)$ 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 $\ell$ $(3 \leqslant \ell \leqslant m)$, 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].

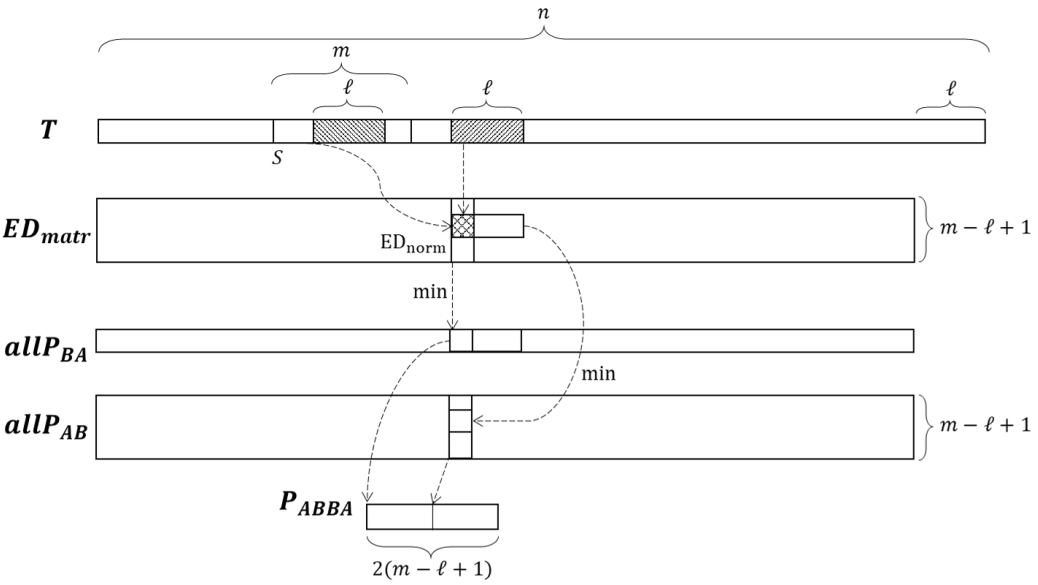



**Рис. 1.** Структуры данных алгоритма PSF

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:

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

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 $allP_{BA}$:

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

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

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

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):

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

где $1 \leqslant j \leqslant n-\ell+1$. 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:

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

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 сниппетов соответствует общему числу активностей: $K = 4$. Длина сегмента соответствует пяти секундам: $m = 4000$, длина подпоследовательности равна $0.5m$: $\ell = 2000$.

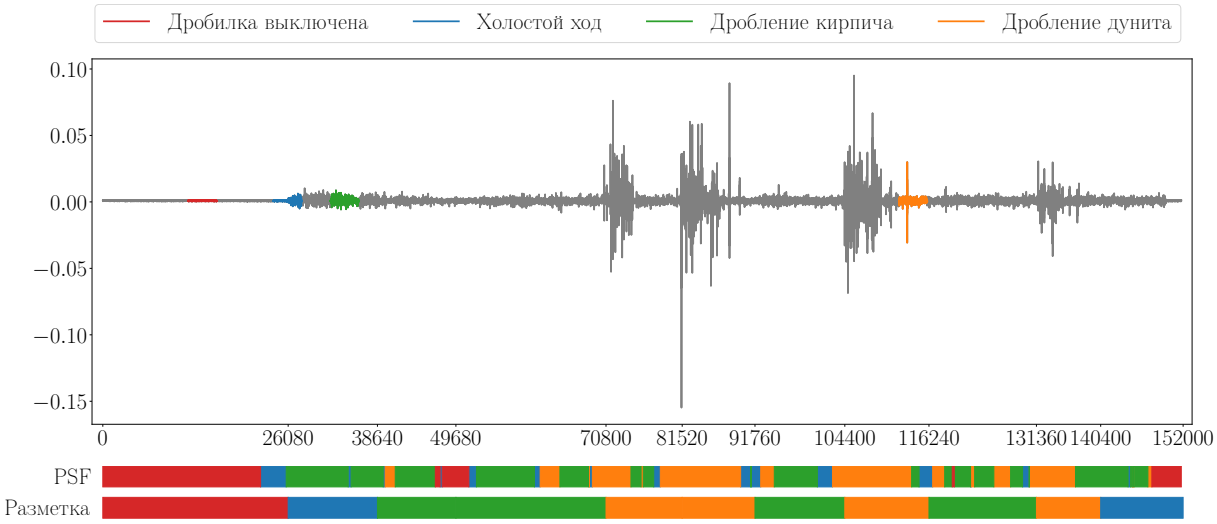
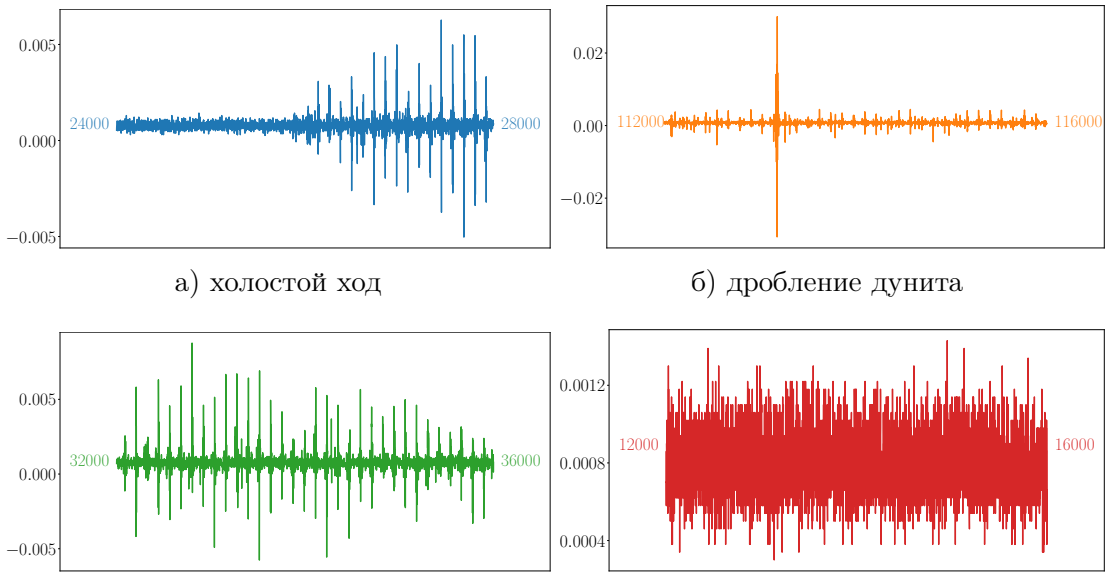



**Рис. 2.** Аннотирование показаний сенсора, установленного на дробильной установке (при $K = 4$)




а) холостой ход

б) дробление дунита

в) дробление кирпича

г) дробилка выключена

**Рис. 3.** Сниппеты 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 искомых 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.

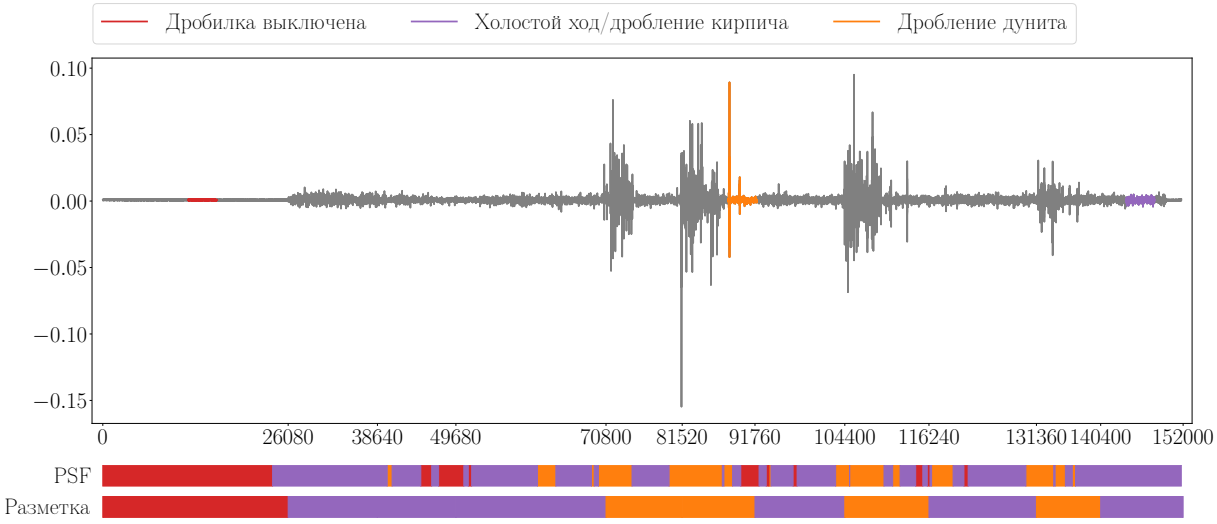
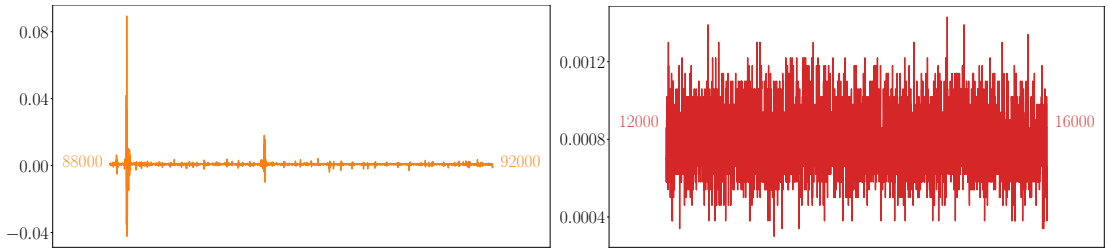
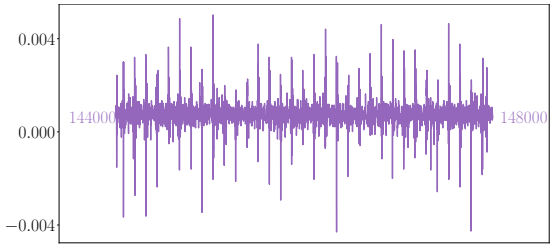



**Рис. 4.** Аннотирование показаний виброакселерометра, установленного на дробильной установке (при $K = 3$)




а) дробление дунита б) дробилка выключена




в) холостой ход/дробление кирпича

**Рис. 5.** Сниппеты 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 PAMAP [14], представляющего собой показания закрепленного на человеке виброакселерометра. Данный ряд содержит показан ия при трех видах физической активности: глажка белья, подъем по лестнице, спуск по лестнице. Количество сниппетов 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 сниппеты, соответствующие активностям на временном ряду, с индексами начала и конца сниппетов.

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

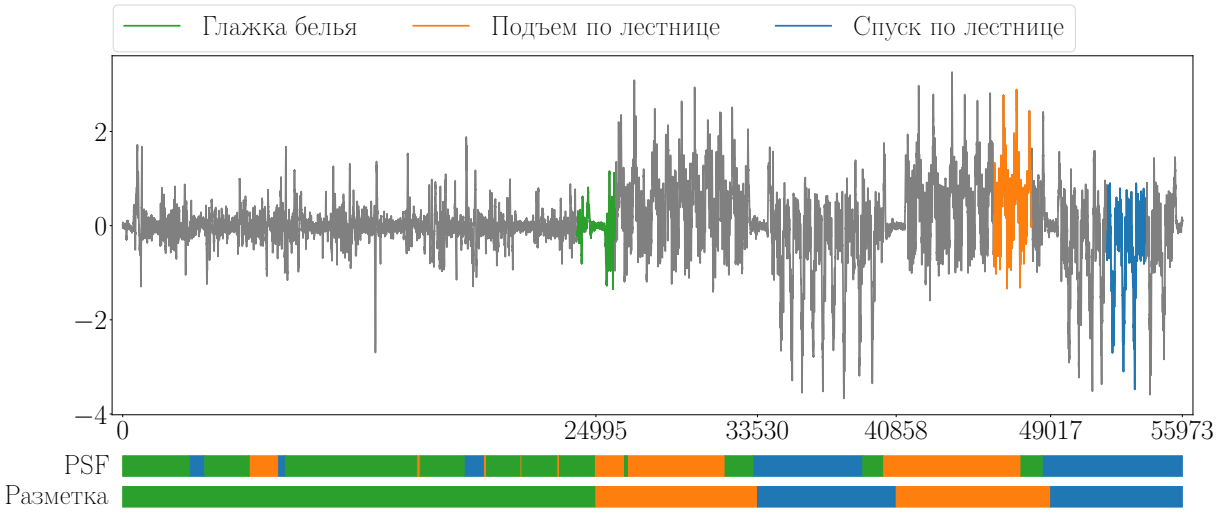
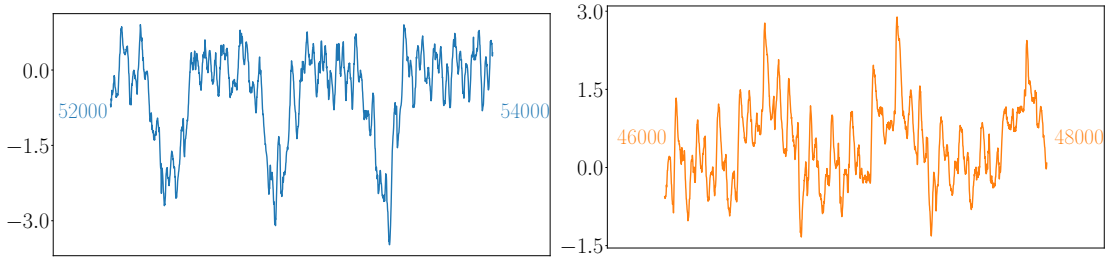
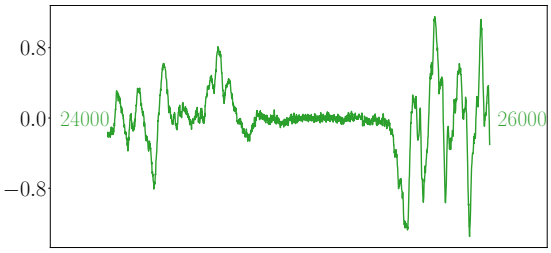



**Рис. 6.** Аннотирование показаний акселерометра, закрепленного на человеке




а) спуск по лестнице б) подъем по лестнице




в) глажка белья

**Рис. 7.** Сниппеты 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 $0.3m \leqslant \ell \leqslant 0.8m$. 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 PAMAP. 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 PAMAP 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 искомых 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., Миниахметов 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 трасляции 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. Нояб. № 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)

DOI: 10.14529/cmse220203

# 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.



## 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 Programmirovanie). 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.