

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Южно-Уральский государственный университет (национальный исследовательский университет)»
Высшая школа электроники и компьютерных наук
Кафедра системного программирования

Разработка параллельного алгоритма поиска типичных подпоследовательностей временного ряда для графического процессора

Рецензент:
Заведующий кафедрой ТУиО ЧелГУ,
д.ф.-м.н., профессор
В.И. Ухоботов

Автор работы:
студент группы КЭ-220
А.И. Гоглачев

Научный руководитель:
профессор кафедры СП,
д.ф.-м.н., доцент
М.Л. Цымблер

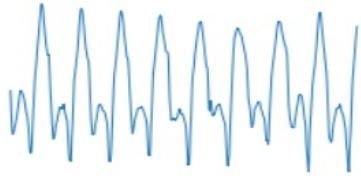
Челябинск, 2022 г.

Поиск типичных подпоследовательностей

Показания носимого акселерометра во время различных физических активностей:



Бег:



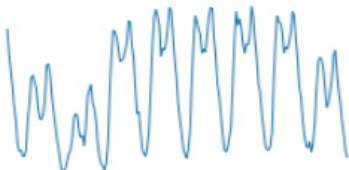
Шаг:



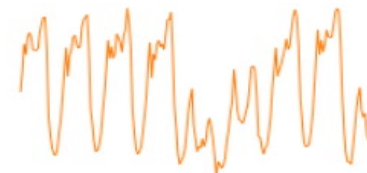
Энергопотребление в холодные и теплые сезоны:



Холодный
сезон:



Теплый
сезон:



Цель и задачи исследования

Цель: разработать параллельный алгоритм поиска типичных подпоследовательностей временного ряда для графического процессора.

Задачи:

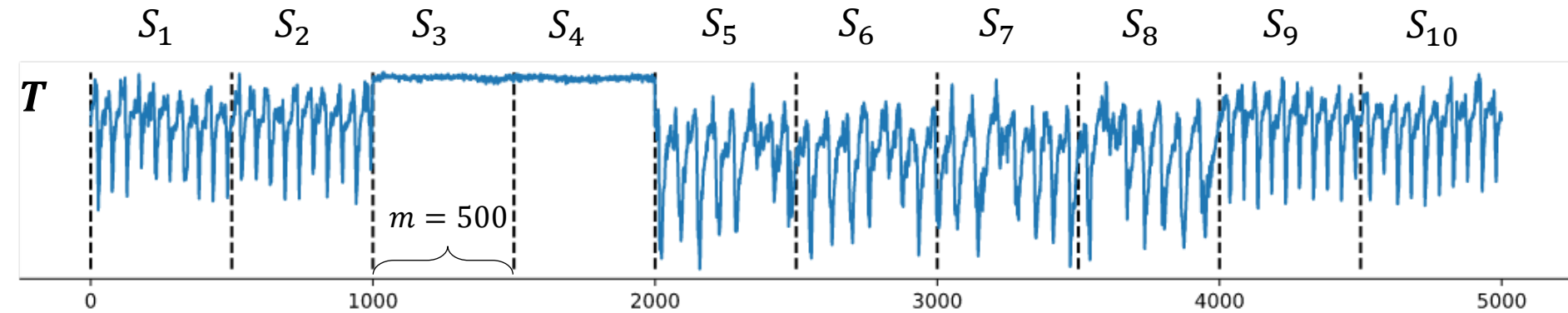
1. Провести обзор по тематике исследования
2. Изучить технологию программирования NVIDIA CUDA
3. Реализовать параллельный алгоритм поиска типичных подпоследовательностей временного ряда для графического процессора
4. Провести вычислительные эксперименты, исследующие эффективность разработанного алгоритма

Методы поиска типичных подпоследовательностей

Метод	Разметка	Без обучения
Простая случайная выборка	нет	да
Кластеризация ¹	нет	да
Шейплеты ²	да	нет
Средняя тенденция ³	да	нет
Лейтмотивы ⁴	нет	да
Нейросетевые методы (энкодеры ⁵ и др.)	да	нет
Сниппеты ⁶	да	да

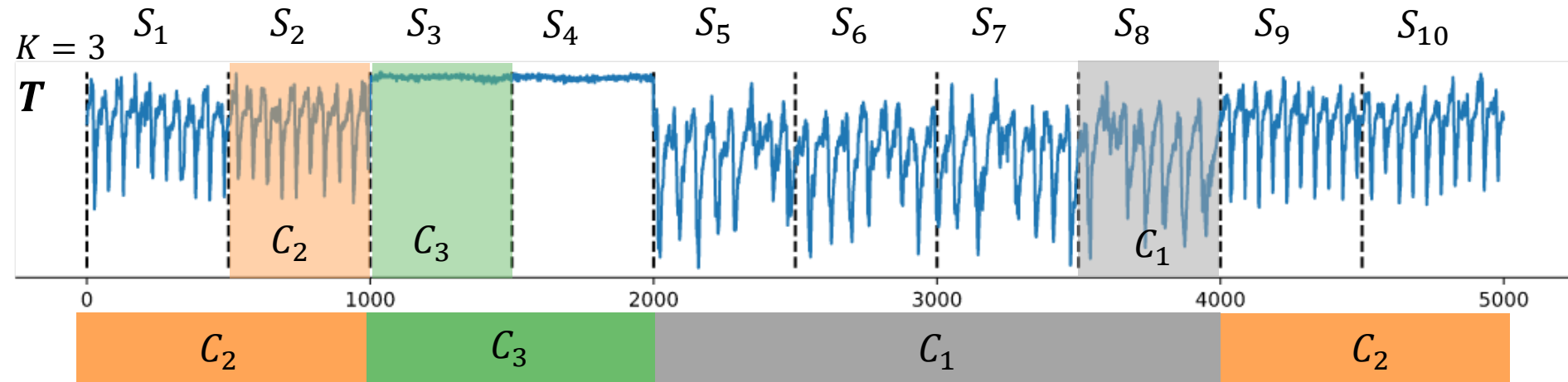
1. Keogh, E.J., et. al.: Clustering of time-series subsequences is meaningless: implications for previous and future research. Knowl. Inf. Syst. 8(2), 154–177 (2005)
2. Yeh, C.M., et. al.: Matrix profile I: all pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. IEEE 16th International Conference on Data Mining, pp. 1317–1322.
3. Indyk, P., et. al.: Identifying representative trends in massive time series data sets using sketches. Proceedings of 26th International Conference on Very Large Data Bases. pp. 363–372
4. Mueen, A., et. al. Exact discovery of time series motifs. Proceedings of the SIAM International Conference on Data Mining. pp. 473–484.
5. Noering, F.K., et. Al.: Pattern discovery in time series using autoencoder in comparison to nonlearning approaches. Integr. Comput. Aided Eng. 2021, 28, 237–256.
6. Imani, S., et.al.: Introducing time series snippets: a new primitive for summarizing long time series. Data Min. Knowl. Discov. 34(6), 1713–1743

Поиск типичных подпоследовательностей при помощи снippetов



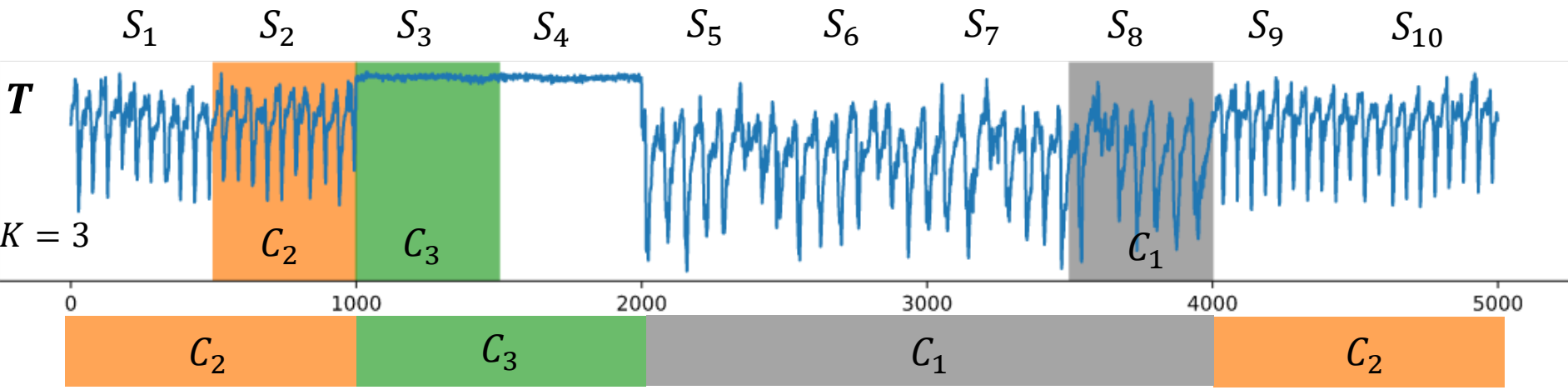
1. Представим ряд как набор непересекающихся сегментов

Поиск типичных подпоследовательностей при помощи снippetов

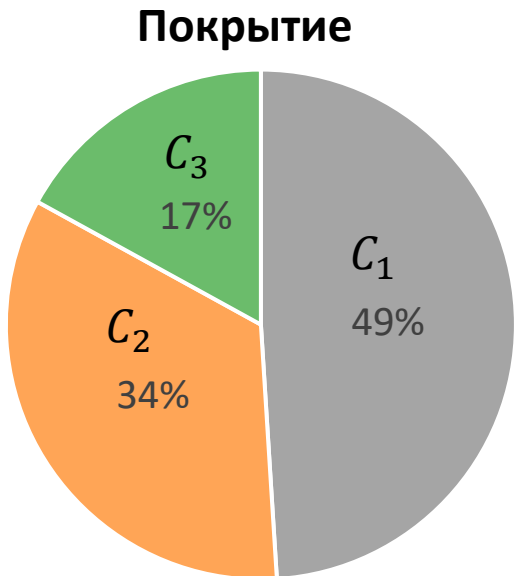


1. Представим ряд как набор непересекающихся сегментов
2. Для каждого сегмента найдем его ближайших соседей

Поиск типичных подпоследовательностей при помощи снippetов



1. Представим ряд как набор непересекающихся сегментов
2. Для каждого сегмента найдем его ближайших соседей
3. Возьмем **top-K** сегментов по убыванию **покрытия** (доли ближайших соседей)

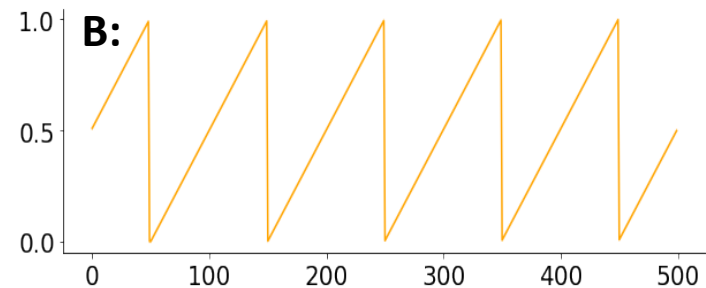
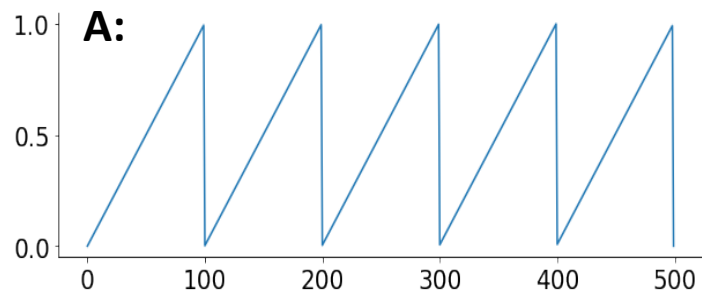


Мера MPdist*

Временные ряды длины m тем более похожи друг на друга по мере **MPdist**,

чем больше в них имеется **нормализованных подпоследовательностей** длины ℓ ($3 \leq \ell \leq m$),

близких друг к другу по **евклидовой метрике**



$ED(A, B)$	MPdist (A, B)
11.2	0

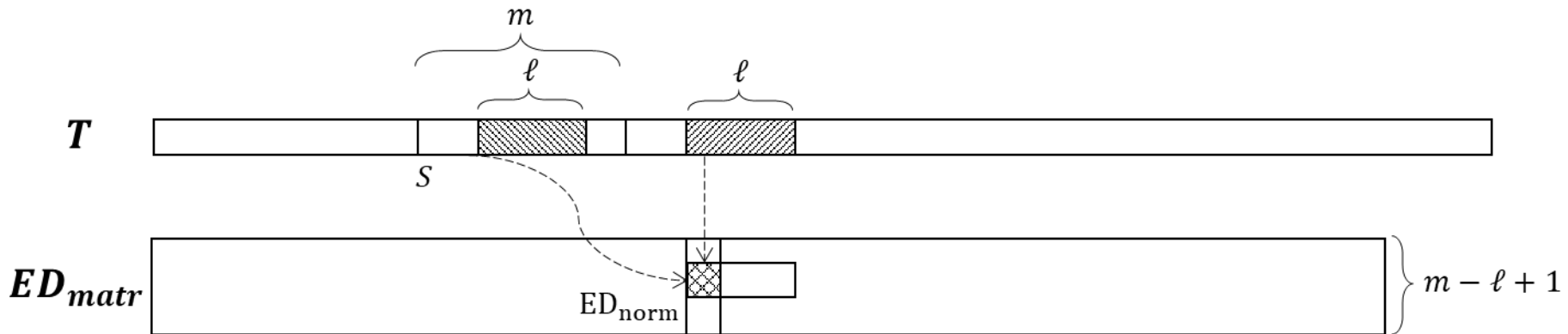
* Gharghabi S., Imani S., Bagnall A.J. et al. An ultra-fast time series distance measure to allow data mining in more complex real-world deployments // Data Mining and Knowledge Discovery. 2020. Vol. 34. P. 1104–1135

PSF: параллельный алгоритм поиска типичных подпоследовательностей для GPU

Временная сложность последовательного алгоритма: $O(n^2 \cdot (n - m)/m)$

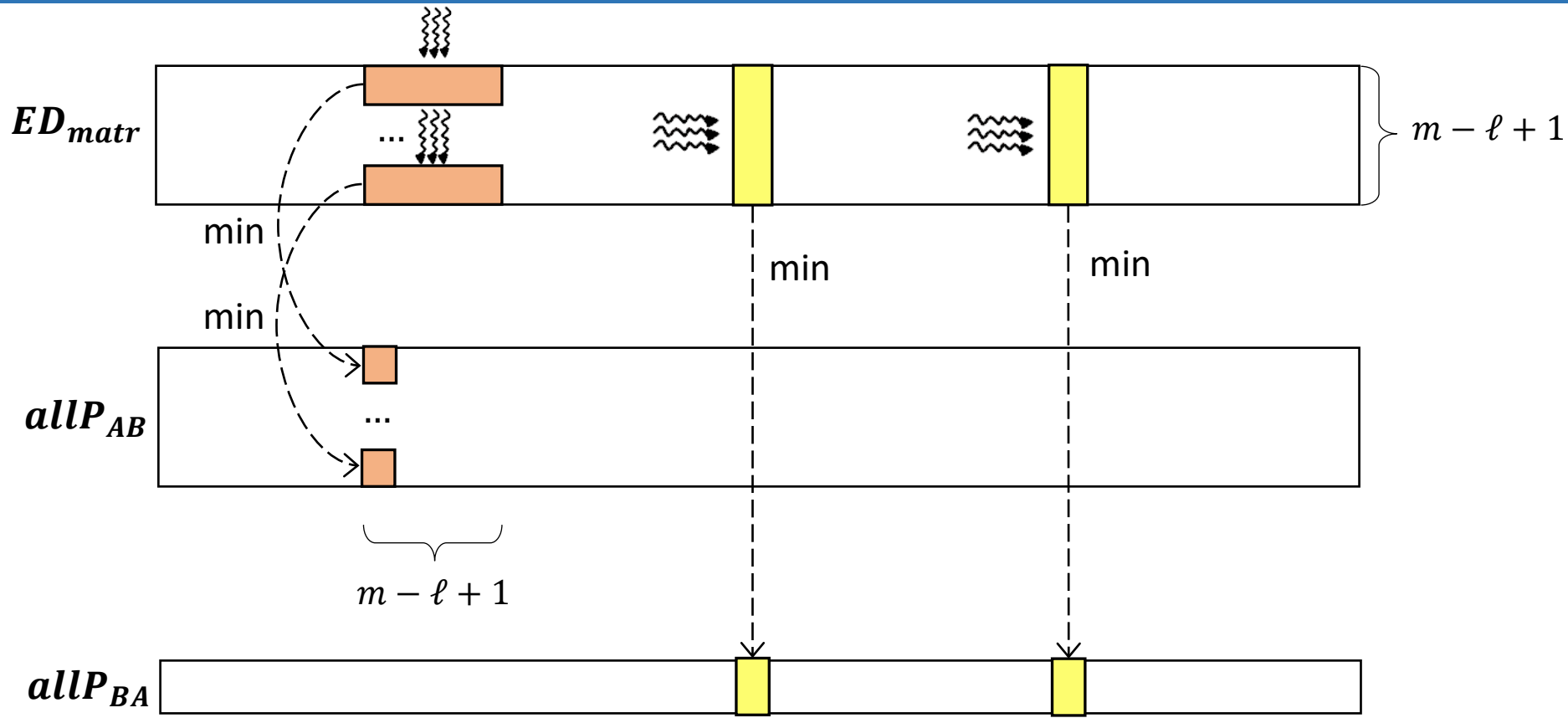
Шаг	Snippet-Finder	PSF
1. Вычисление матричного профиля P_{AB}	$\{P_{AB}(i) = \text{ED}_{\text{norm}}(A_{i,\ell}, B_{j,\ell})\}_{i=1}^{m-\ell+1},$ $B_{j,\ell} = \arg \min_{1 \leq q \leq m-\ell+1} \text{ED}_{\text{norm}}(A_{i,\ell}, B_{q,\ell})$	<p>Вычисление матрицы нормализованных евклидовых расстояний ED_{matr}</p> $allP_{AB}(i, j) = \min_{j \leq c \leq j+m-\ell+1} \text{ED}_{\text{matr}}(i, c)$
2. Вычисление матричного профиля P_{BA}	$\{P_{BA}(i) = \text{ED}_{\text{norm}}(B_{i,\ell}, A_{j,\ell})\}_{i=1}^{m-\ell+1},$ $A_{j,\ell} = \arg \min_{1 \leq q \leq m-\ell+1} \text{ED}_{\text{norm}}(B_{i,\ell}, A_{q,\ell})$	$allP_{BA}(j) = \min_{1 \leq i \leq m-\ell+1} \text{ED}_{\text{matr}}(i, j)$
3. Вычисление матричного профиля конкатенации P_{ABBA}	$P_{ABBA} = P_{AB} \odot P_{BA}$	$P_{ABBA} = allP_{AB}(i, m - \ell) \odot allP_{BA}(i)$
4. Вычисление MPdist профиля	$\text{MPdist}(A, B, \ell) = \begin{cases} \text{Sorted}P_{ABBA}(k), & P_{ABBA} > k \\ \text{Sorted}P_{ABBA}(2(m - \ell + 1)), & P_{ABBA} \leq k \end{cases}$	

Параллельный поиск сниппетов: вычисление ED_{matr}



$$\begin{aligned} \overline{QT}_{i,j} &= \overline{QT}_{i-1,j-1} + df_i \cdot dg_j + df_j \cdot dg_i, \\ df_0 &= 0; df_i = \frac{t_{i+m-1} - t_{i-1}}{2}, \\ dg_0 &= 0; dg_i = (t_{i+m-1} - \mu_i) + (t_{i-1} - \mu_{i-1}), \\ \mu_i &= \frac{1}{m} \sum_{j=i}^{i+m} t_j, \\ T_{i,m} - \mu_i &= (t_i - \mu_i, \dots, t_{i+m-1} - \mu_i), \\ P_{i,j} &= \overline{QT}_{i,j} \cdot \frac{1}{\|T_{i,m} - \mu_i\|} \cdot \frac{1}{\|T_{j,m} - \mu_j\|}, \\ ED_{norm}(T_{i,m}, T_{j,m}) &= \sqrt{2m(1 - P_{i,j})} \end{aligned}$$

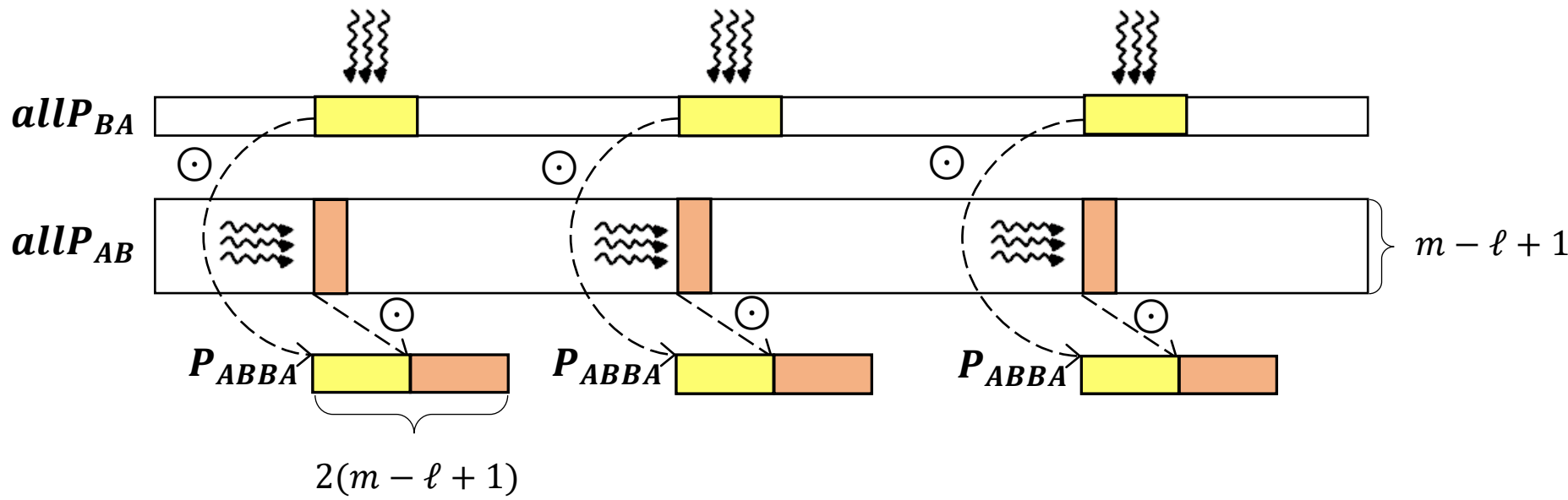
Параллельный поиск сниппетов: вычисление $allP_{AB}$ и $allP_{BA}$



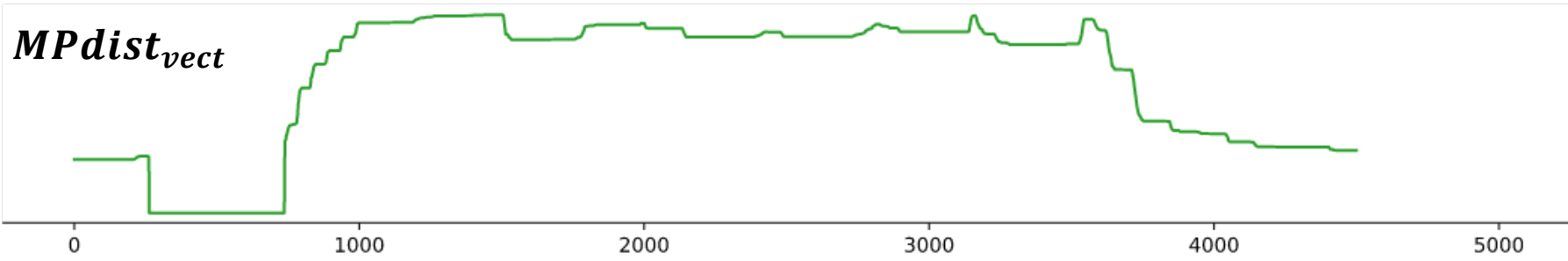
$$allP_{AB}(i, j) = \min_{j \leq c \leq j+m-\ell+1} ED_{matr}(i, c)$$

$$allP_{BA}(j) = \min_{1 \leq i \leq m-\ell+1} ED_{matr}(i, j)$$

Параллельный поиск сниппетов: вычисление P_{ABBA}



$$MPdist(A, B, \ell) = \begin{cases} SortedP_{ABBA}(k), & |P_{ABBA}| > k \\ SortedP_{ABBA}(2(m - \ell + 1)), & |P_{ABBA}| \leq k \end{cases}$$
 где $k = \lceil 0.05 \cdot 2m \rceil = \lceil 0.1m \rceil$.



Эксперимент: временной ряд РАМАР

- Физические активности:

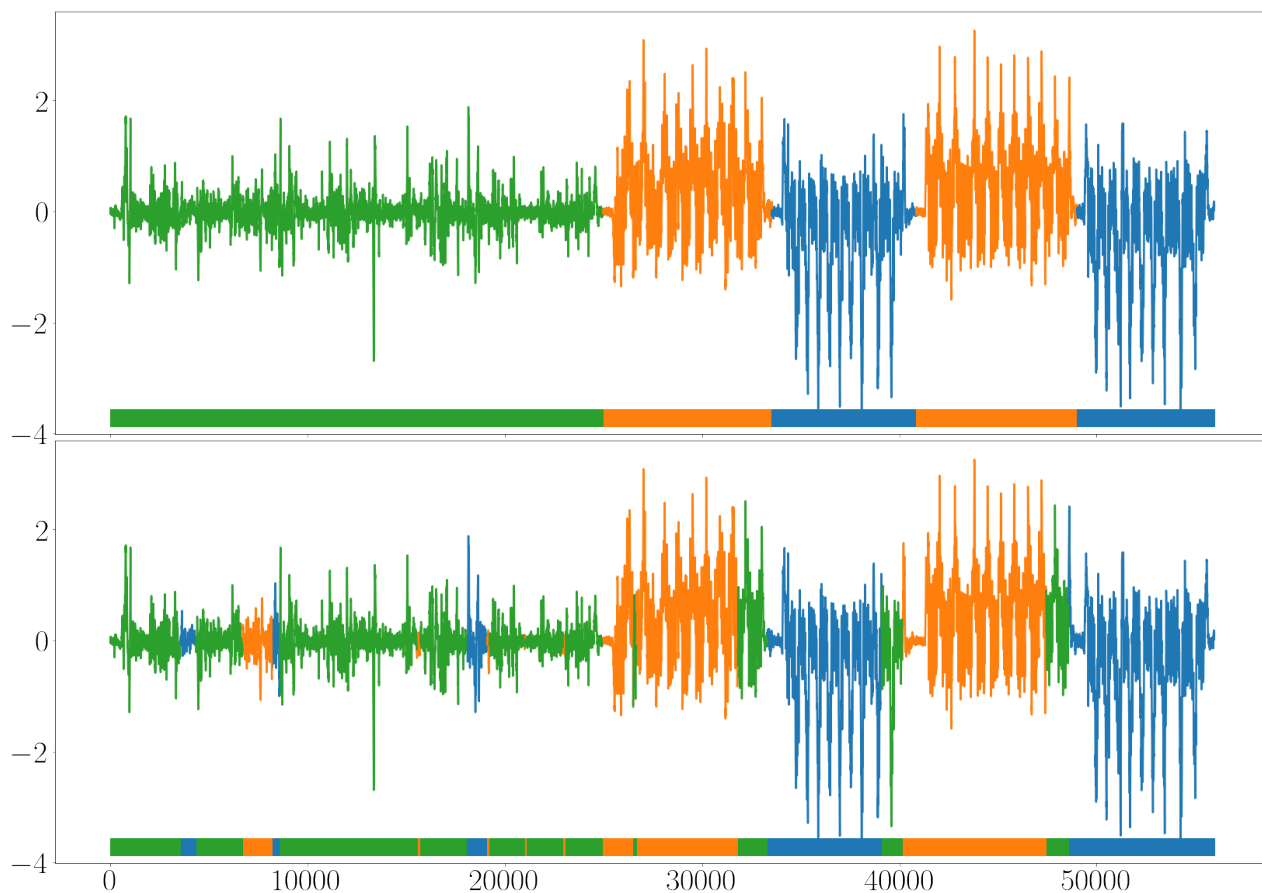
- спуск по лестнице
- подъем по лестнице
- глажка белья

Алгоритм	Время, с
Последовательный Snippet-Finder	817
Параллельный Snippet-Finder	52

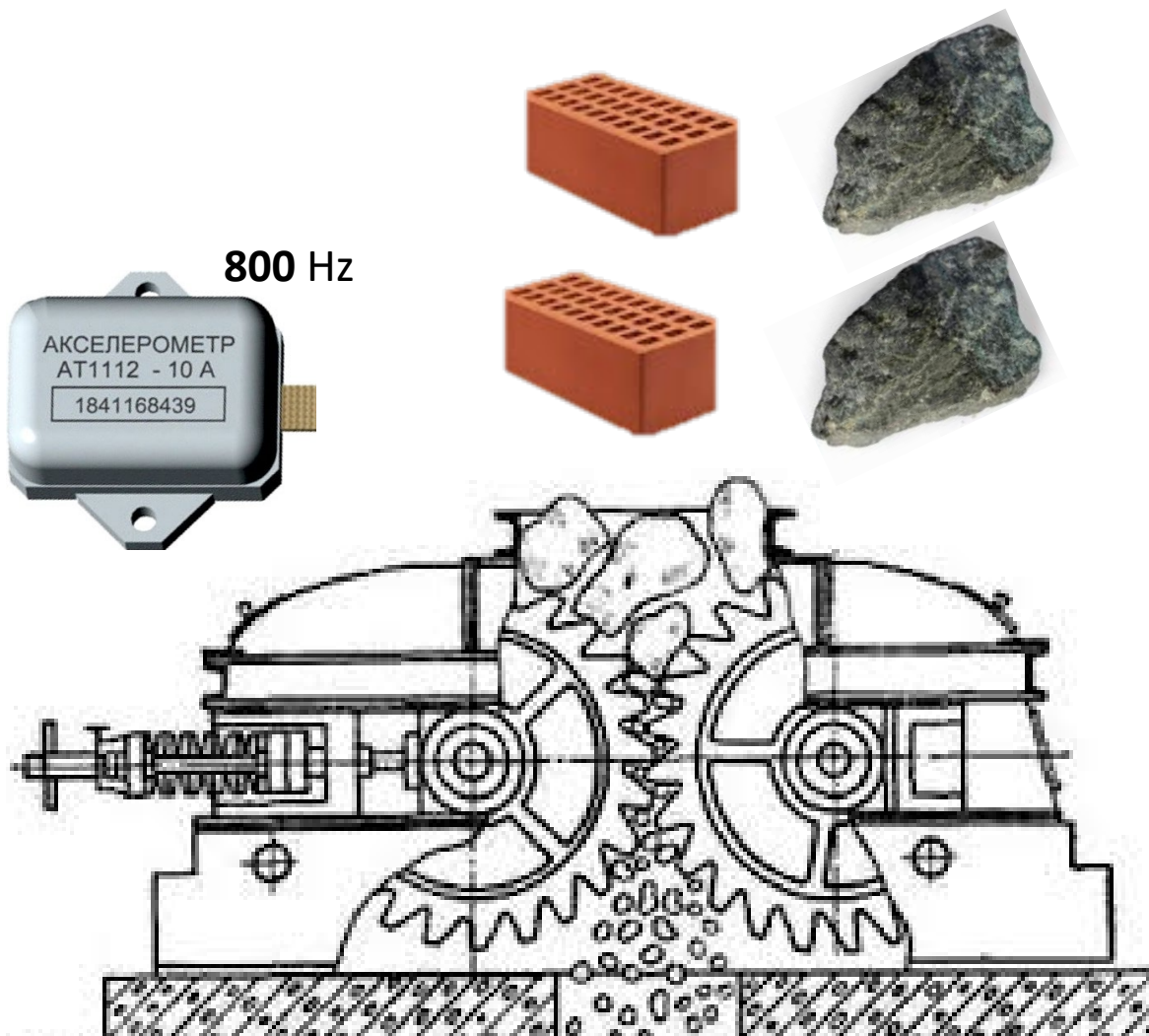
- Исходная разметка:

Длина ряда, n : 55 974
Длина сегмента, m : 2 000
Число снippetов, K : 3

- Полученная разметка:
(точность 83%)



Эксперимент: временной ряд Crusher



Протокол

Время	Действие
00:00:00	Дробилка выключена
00:32.60	Включение, холостой ход
00:48.34	Заброс кирпича
01:02.09	Заброс кирпича
01:28.05	Заброс дунита
01:41.86	Заброс дунита
01:54.68	Заброс кирпича
02:10.49	Заброс дунита
02:25.32	Заброс кирпича
02:44.20	Заброс дунита
02:55.49	Завершение дробления, холостой ход
03:07.06	Останов

Эксперимент: временной ряд Crusher

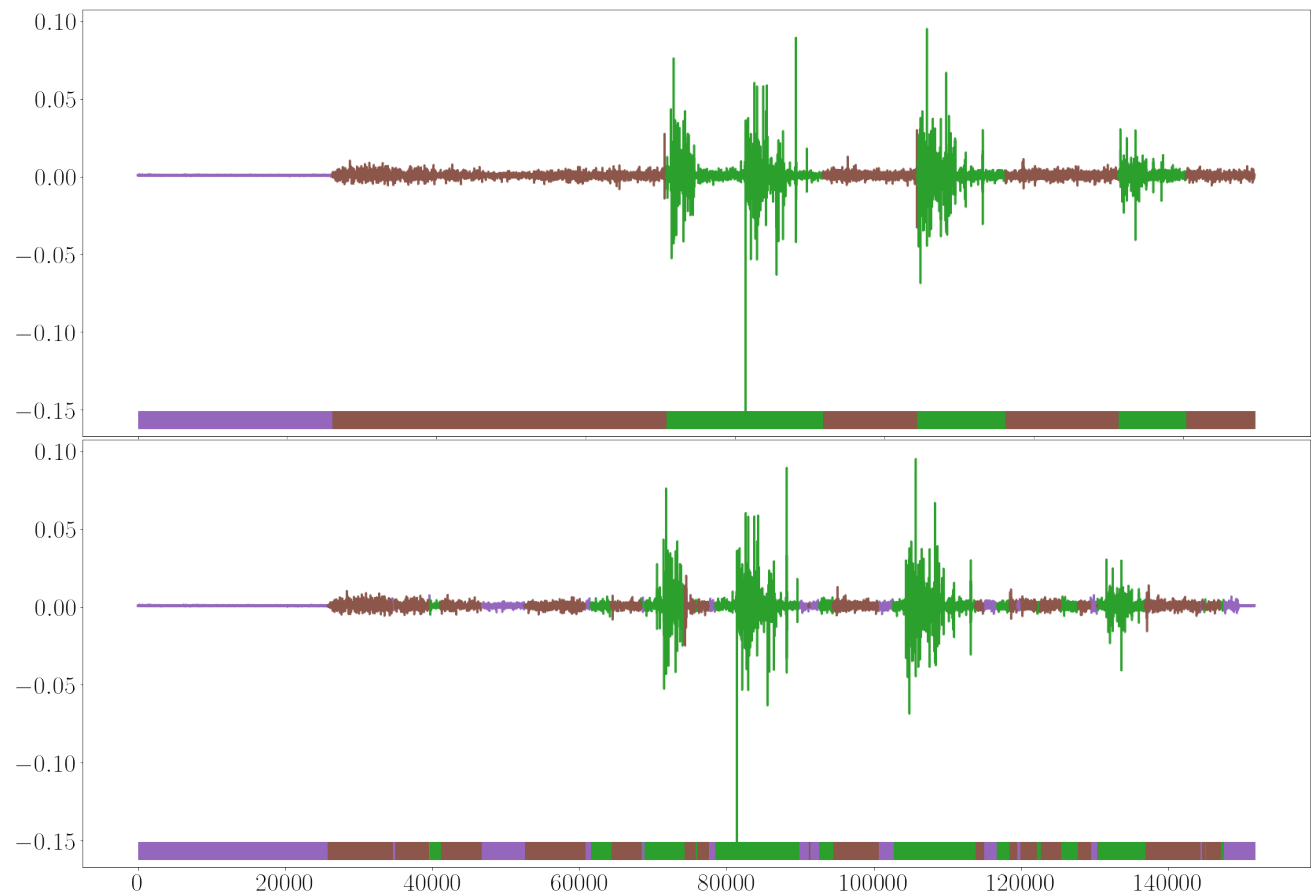
- Активности:
 - дробление дунита
 - дробилка выключена
 - холостой ход *или* дробление кирпича

- Исходная разметка:

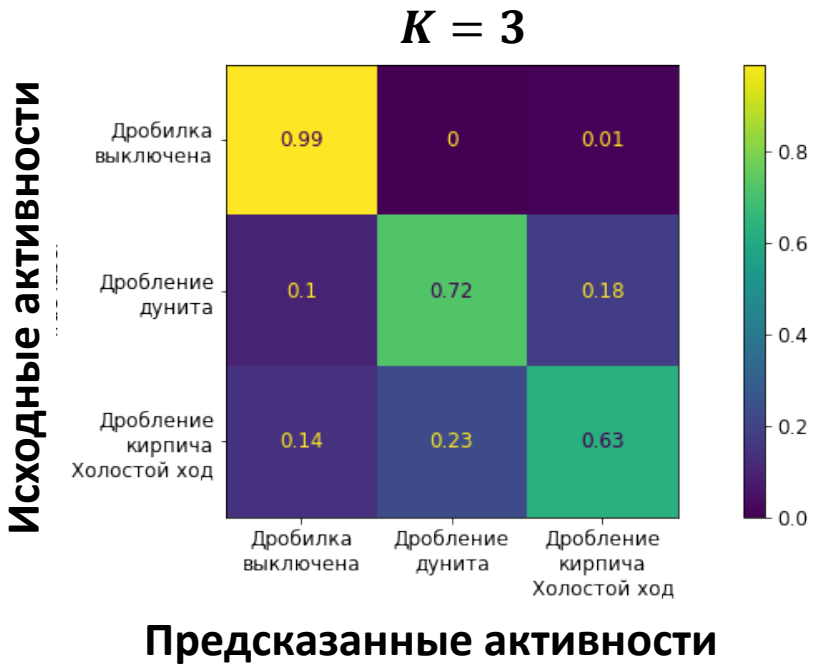
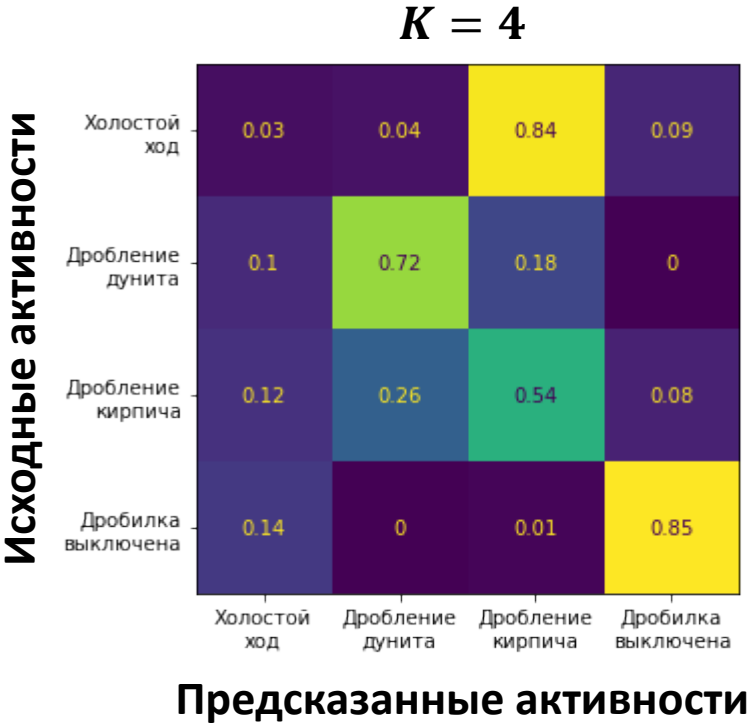
Длина ряда, n : 151 744
Длина сегмента, m : 4 000
Число снippetов, K : 3

- Полученная разметка (точность 72%):

Алгоритм	Время, с
Последовательный Snippet-Finder	12 796
Параллельный Snippet-Finder	726

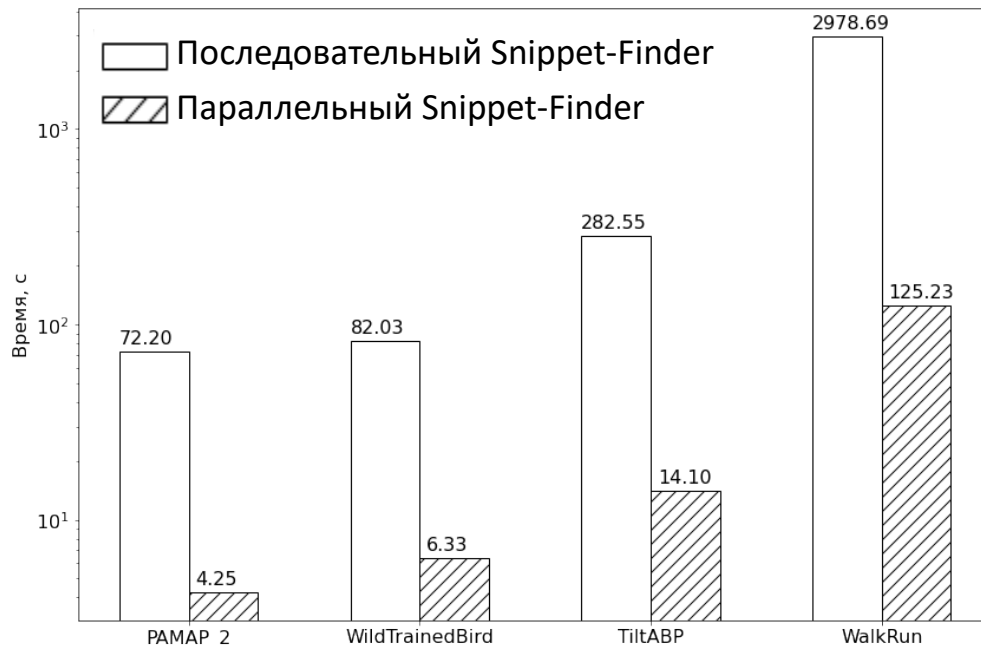


Выбор значения K для временного ряда Crusher



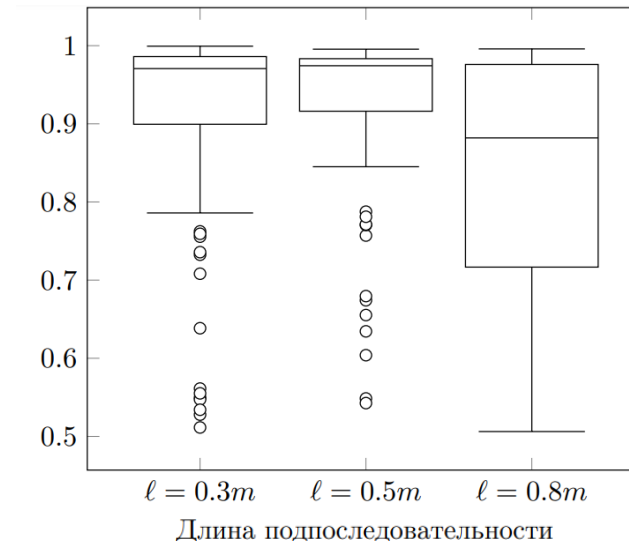
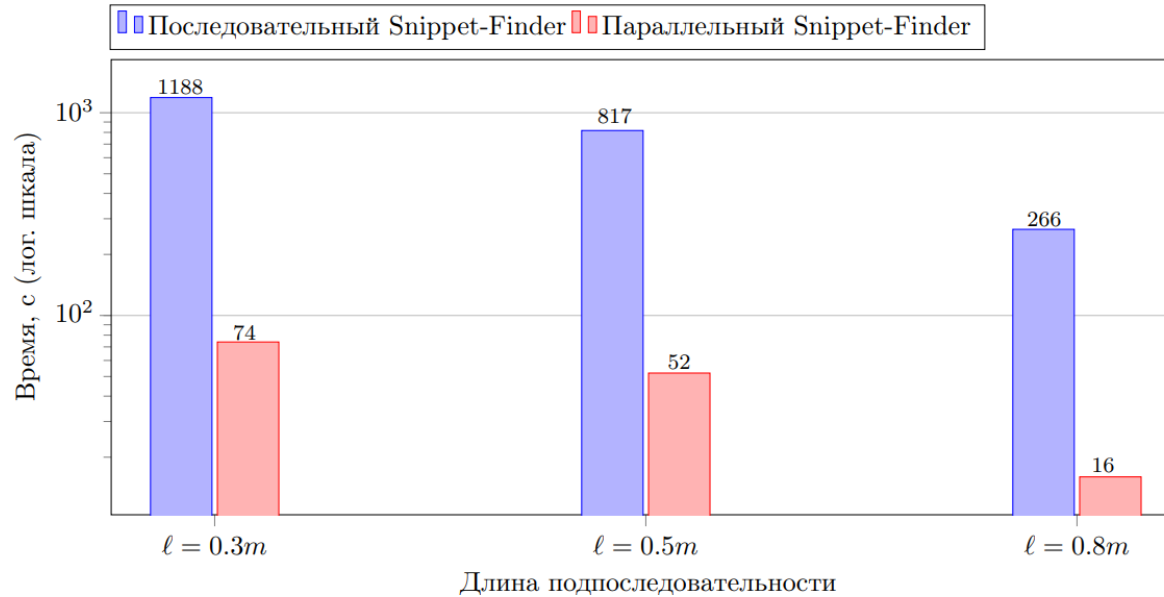
Эксперименты: производительность

Название	Длина ряда n	Длина сегмента m	Описание
PAMAP_2	20 002	600	Показания акселерометра во время различных видов физической активности
WalkRun	100 000	240	
WildVTrainedBird	20 002	900	Физиологические показатели жизнедеятельности птиц
TiltABP	40 000	630	Показания кровяного давления человека во время быстрых наклонов



Параллельный алгоритм дает прирост производительности в 18—23 раза по сравнению с последовательной реализацией

Эксперименты: влияние значения параметра ℓ



Публикации

- **Опубликованы научные статьи:**
 1. Цымблер М.Л., Гоглачев А.И. Поиск типичных подпоследовательностей временного ряда на графическом процессоре // Вычислительные методы и программирование, 2021. – Т. 22. – № 4. – С. 344–359 (индексируется в **РИНЦ**).
 2. Цымблер М.Л., Гоглачев А.И. Применение параллельных вычислений для аннотирования сенсорных данных. // Параллельные вычислительные технологии – XVI международная конференция (ПаВТ'2022): Короткие статьи и описания плакатов. – С. 121–128 (индексируется в **РИНЦ**).
 3. Zymbler M., Goglachev A. Fast Summarization of Long Time Series with Graphics Processor. // Mathematics, 2022. – Vol. 10. – No. 10. – Article 1781 (индексируется в **Scopus, Q1**).
- **Сделан доклад на международной научной конференции** «Параллельные вычислительные технологии (ПаВТ'2022)» (29 марта – 31 марта 2022 г., Дубна).

Заключение

1. Проведен обзор по тематике исследования
2. Изучена технология программирования NVIDIA CUDA
3. Реализован параллельный алгоритм поиска типичных подпоследовательностей временного ряда для графического процессора
4. Проведены вычислительные эксперименты, исследующие эффективности разработанного алгоритма