

Международная научная конференция
Параллельные вычислительные технологии (ПавТ'2022)
Дубна, 29–31 марта 2022

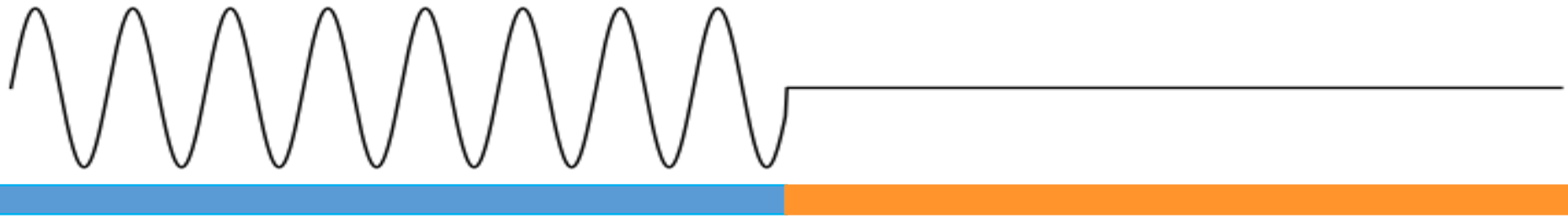
Применение параллельных вычислений для аннотирования сенсорных данных

А.И. Гоглачев, М.Л. Цымблер

Южно-Уральский государственный университет (Челябинск)

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант №20-07-00140) и Министерства образования и науки РФ (гос. задание FENU-2020-0022).

Аннотирование сенсорных данных



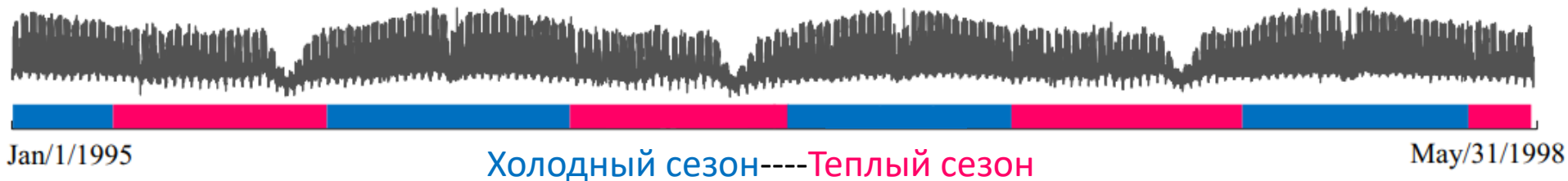
Активность 1 ---- Активность 2

Двигательная активность пациента по показаниям нагрудного акселерометра



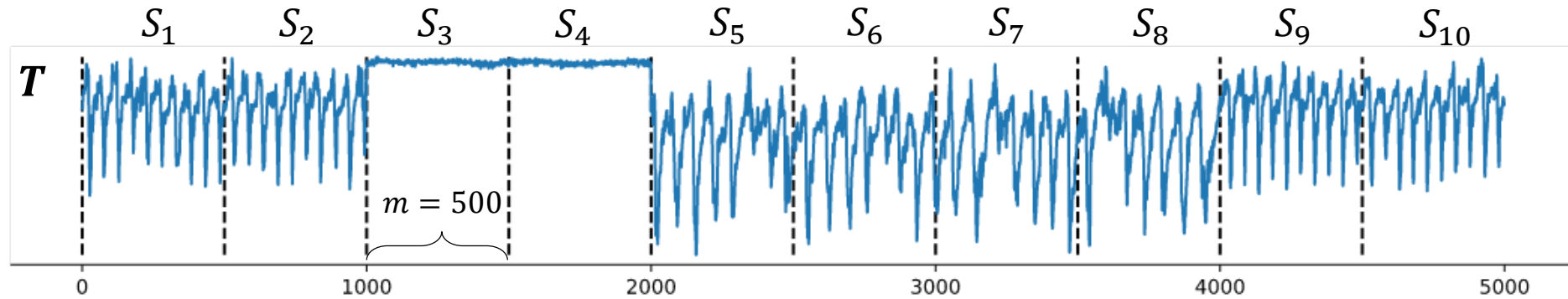
Ходьба ---- Скандинавская ходьба ---- Бег ---- Прыжки на скакалке ---- Скандинавская ходьба

Почасовое энергопотребления в Италии в течение 3 лет



Холодный сезон ---- Теплый сезон

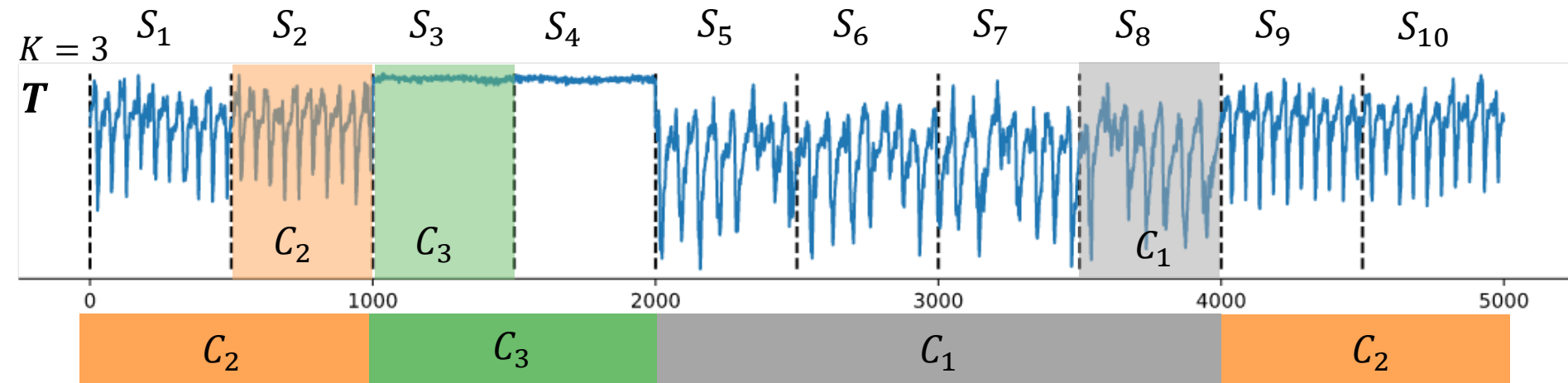
Аннотирование при помощи сниппетов*



1. Представим ряд как набор непересекающихся сегментов (сниппетов)

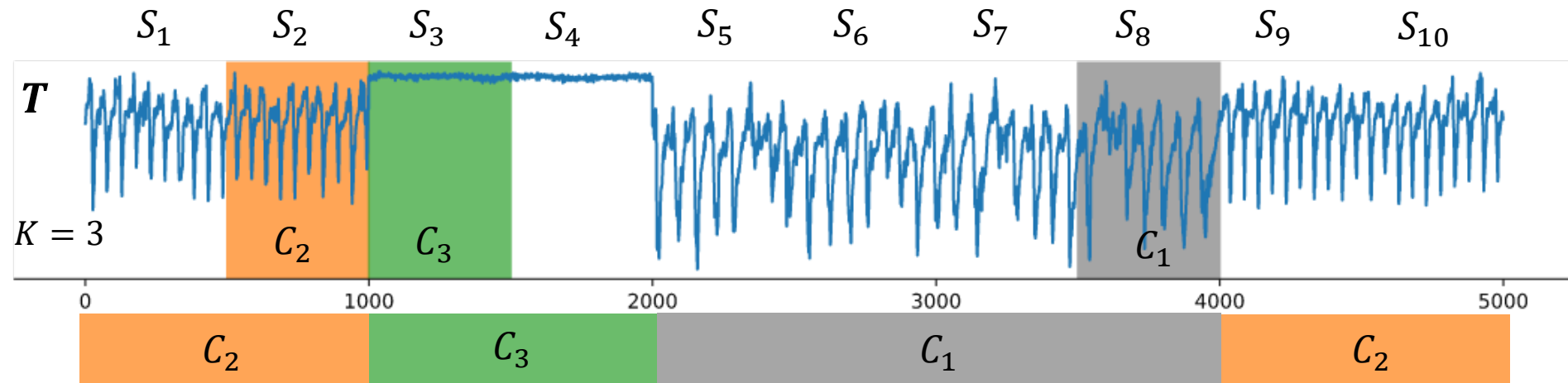
* Imani S., Madrid F., Ding W., Crouter S.E., Keogh E.J. Introducing time series snippets: a new primitive for summarizing long time series. Data Min. Knowl. Discov. 34(6): 1713-1743 (2020). doi: [10.1007/s10618-020-00702-y](https://doi.org/10.1007/s10618-020-00702-y)

Аннотирование при помощи сниппетов

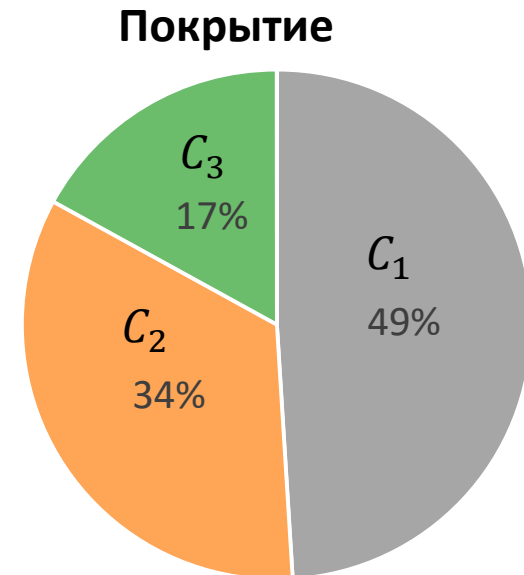


1. Представим ряд как набор непересекающихся сегментов
2. Для каждого сегмента найдем его **ближайших соседей** (наиболее похожих подпоследовательностей)

Аннотирование при помощи сниппетов



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



Мера MPdist*

Два ряда длины m тем более похожи друг на друга в смысле меры MPdist, чем больше в них имеется подпоследовательностей меньшей длины ℓ , близких друг к другу в смысле нормализованного евклидова расстояния

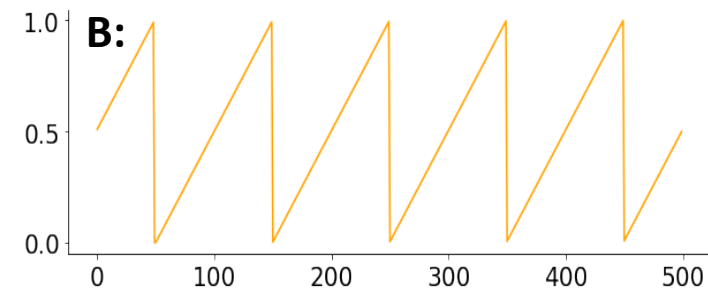
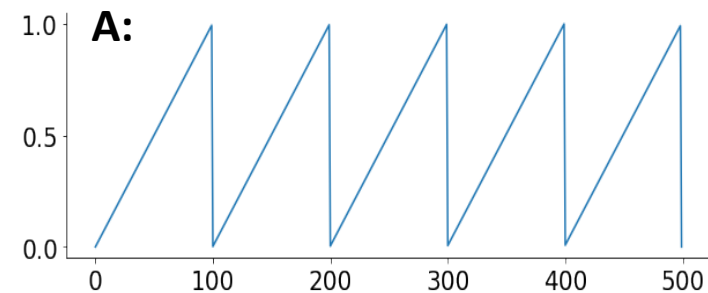
$$\text{MPdist}(A, B, \ell) = \begin{cases} \text{Sorted}P_{ABBA}(k), & |P_{ABBA}| > k \\ \text{Sorted}P_{ABBA}(2(m - \ell + 1)), & |P_{ABBA}| \leq k \end{cases}$$

где $k = \lceil 0.05 \cdot 2m \rceil = \lceil 0.1m \rceil$.

$$P_{ABBA} = P_{AB} \odot P_{BA}$$

$$\{P_{BA}(i) = \text{ED}_{\text{norm}}(B_{i,\ell}, A_{j,\ell})\}_{i=1}^{m-\ell+1},$$
$$A_{j,\ell} = \arg \min_{1 \leq q \leq m-\ell+1} \text{ED}_{\text{norm}}(B_{i,\ell}, A_{q,\ell})$$

$$\{P_{AB}(i) = \text{ED}_{\text{norm}}(A_{i,\ell}, B_{j,\ell})\}_{i=1}^{m-\ell+1},$$
$$B_{j,\ell} = \arg \min_{1 \leq q \leq m-\ell+1} \text{ED}_{\text{norm}}(A_{i,\ell}, B_{q,\ell})$$



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. DOI: [10.1007/s10618-020-00695-8](https://doi.org/10.1007/s10618-020-00695-8)

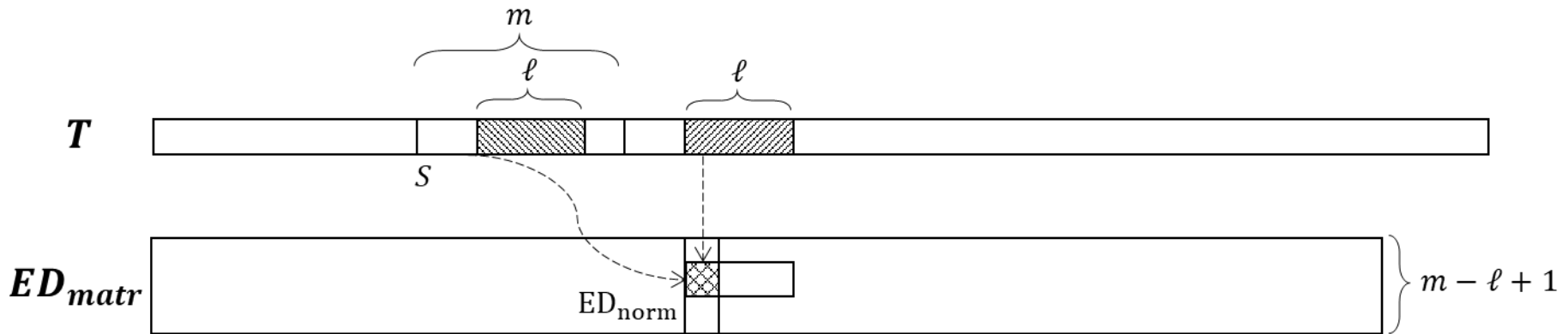
PSF: параллельный алгоритм аннотирования ряда для GPU

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	$\text{MPdist}(A, B, \ell) = \begin{cases} \text{Sorted}P_{ABBA}(k), & P_{ABBA} > k \\ \text{Sorted}P_{ABBA}(2(m - \ell + 1)), & P_{ABBA} \leq k \end{cases}$	

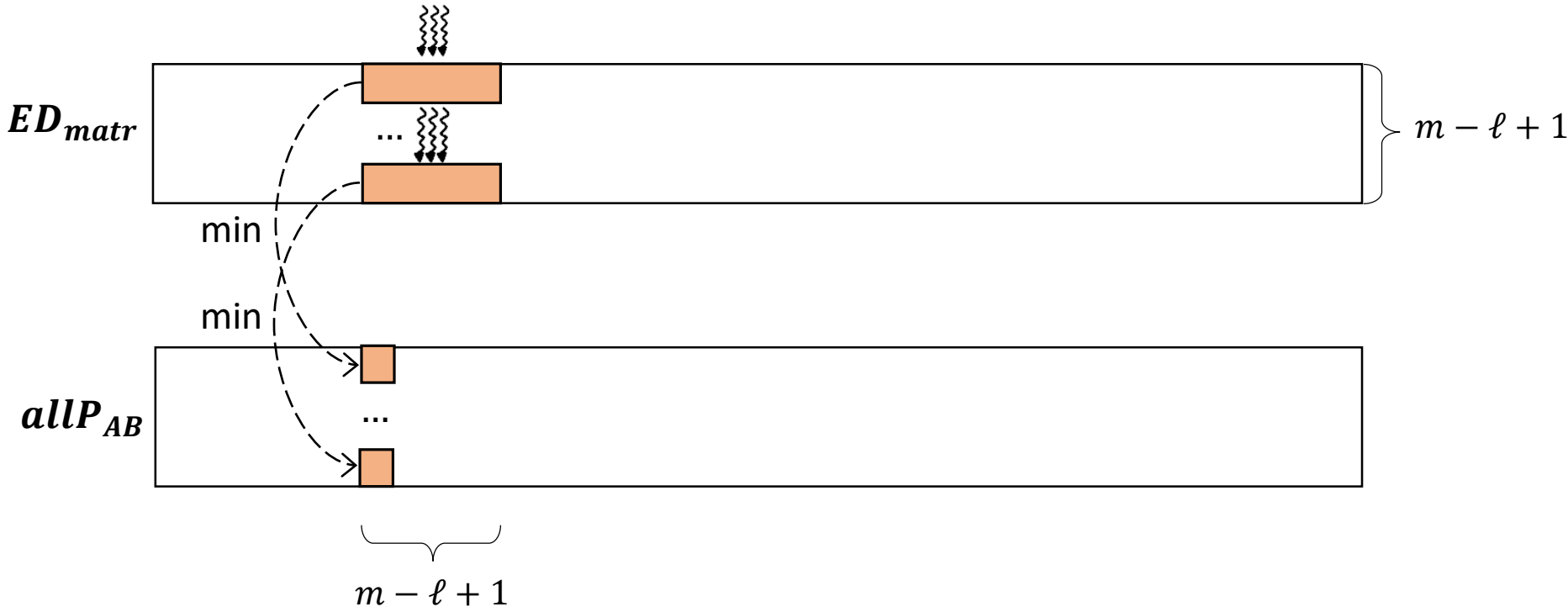
* Imani S., Madrid F., Ding W., Crouter S.E., Keogh E.J. Introducing time series snippets: a new primitive for summarizing long time series. Data Min. Knowl. Discov. 34(6): 1713-1743 (2020). doi: [10.1007/s10618-020-00702-y](https://doi.org/10.1007/s10618-020-00702-y)

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 $allP_{AB}$

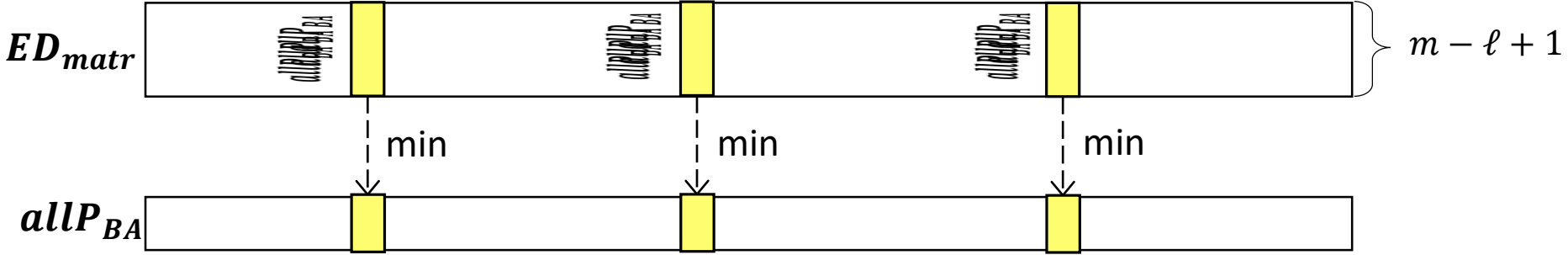


$$allP_{AB}(i, j) = \min_{j \leq c \leq j+m-\ell+1} ED_{matr}(i, c)$$

$$P_{BA}(i) = \{ED_{norm}(B_{i,\ell}, A_{j,\ell})\}_{j=1}^{m-\ell+1},$$

$$A_{j,\ell} = \arg \min_{1 \leq q \leq m-\ell+1} ED_{norm}(B_{i,\ell}, A_{q,\ell})$$

Параллельный поиск сниппетов: вычисление $allP_{BA}$

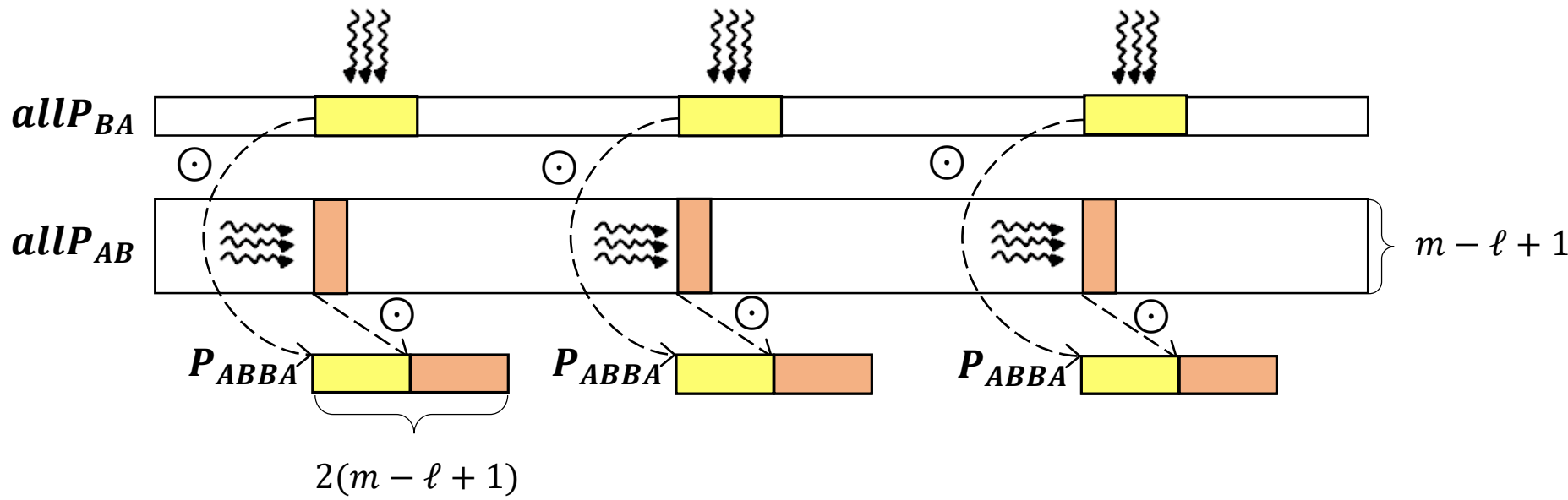


$$allP_{BA}(j) = \min_{1 \leq i \leq m-l+1} ED_{matr}(i, j)$$

$$\{P_{BA}(i) = ED_{norm}(B_{i,\ell}, A_{j,\ell})\}_{i=1}^{m-l+1},$$

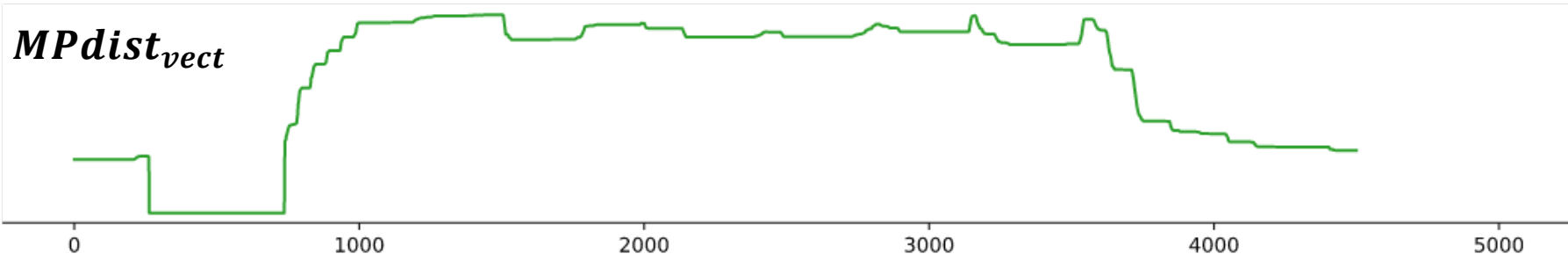
$$A_{j,\ell} = \arg \min_{1 \leq q \leq m-l+1} ED_{norm}(B_{i,\ell}, A_{q,\ell})$$

Параллельный поиск сниппетов: вычисление P_{ABBA}



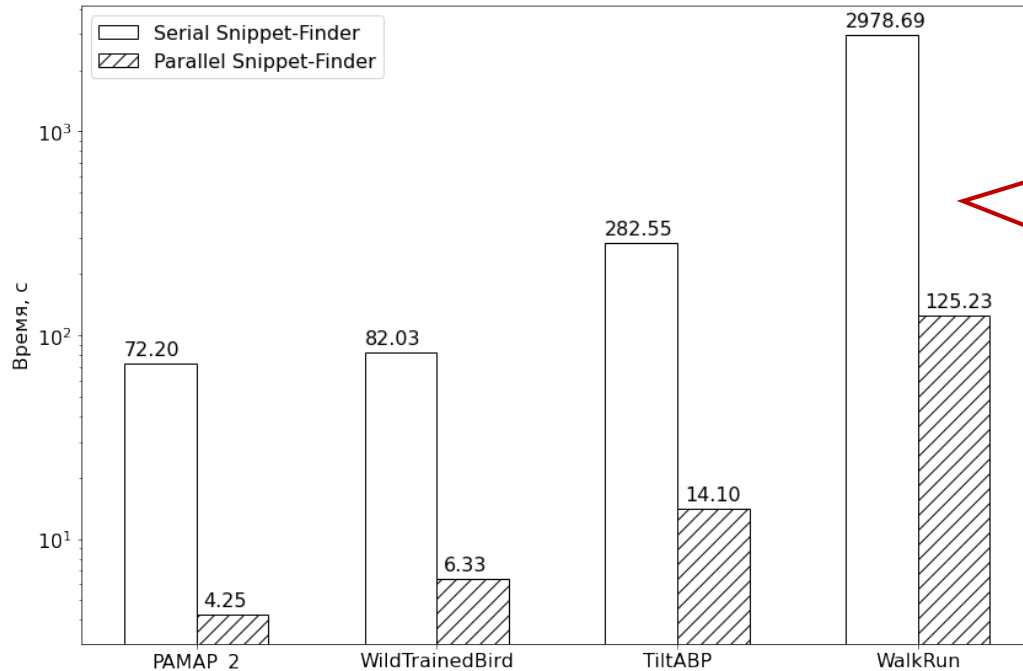
$$MPdist(A, B, \ell) = \begin{cases} SortedP_{ABBA}(k), & |P_{ABBA}| > k \\ SortedP_{ABBA}(2(m - \ell + 1)), & |P_{ABBA}| \leq k \end{cases}$$

где $k = \lceil 0.05 \cdot 2m \rceil = \lceil 0.1m \rceil$.



Эксперименты: производительность

Ряд	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: аннотирование

- Ряды для аннотирования

Название	Длина ряда, n	Длина сегмента, m	Описание
РАМАР*	55 974	800	Показания нагрудного акселерометра во время физической активности человека
Crusher	151 744	4 000	Данные акселерометра во время работы дробильной установки

- Аппаратная платформа

Характеристика	Центральный процессор	Графический процессор
Модель	Intel Xeon Gold 6254	NVIDIA Tesla V100 SXM2
Количество ядер	18	5120
Тактовая частота, GHz	4.0	1.3
Оперативная память, GB	64	32
Пиковая производительность, TFLOPS	1.2	15.7

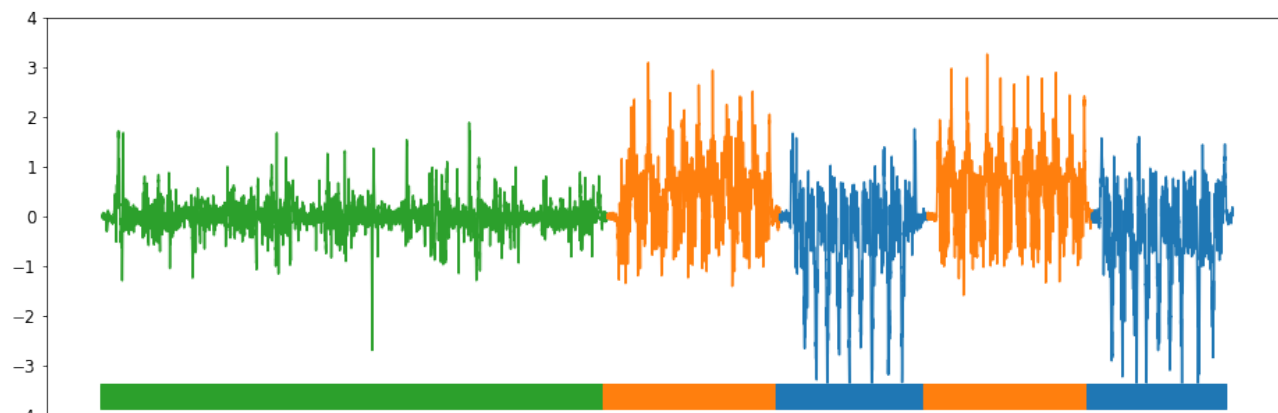
* Reiss A., Stricker D. Introducing a New Benchmarked Dataset for Activity Monitoring // Proc. of the 16th International Symposium on Wearable Computers (ISWC). 2012. doi: 10.1109/ISWC.2012.13.

Аннотирование РАМАР

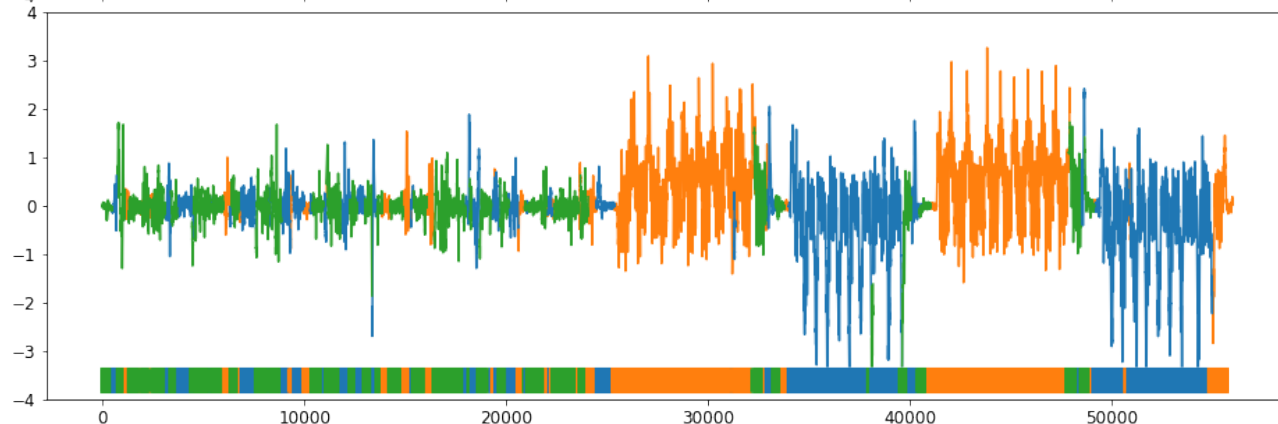
- Физические активности:
 - спуск по лестнице
 - подъем по лестнице
 - глажка белья

Алгоритм	Время, с
Serial Snippet-Finder	817
Parallel Snippet-Finder	51

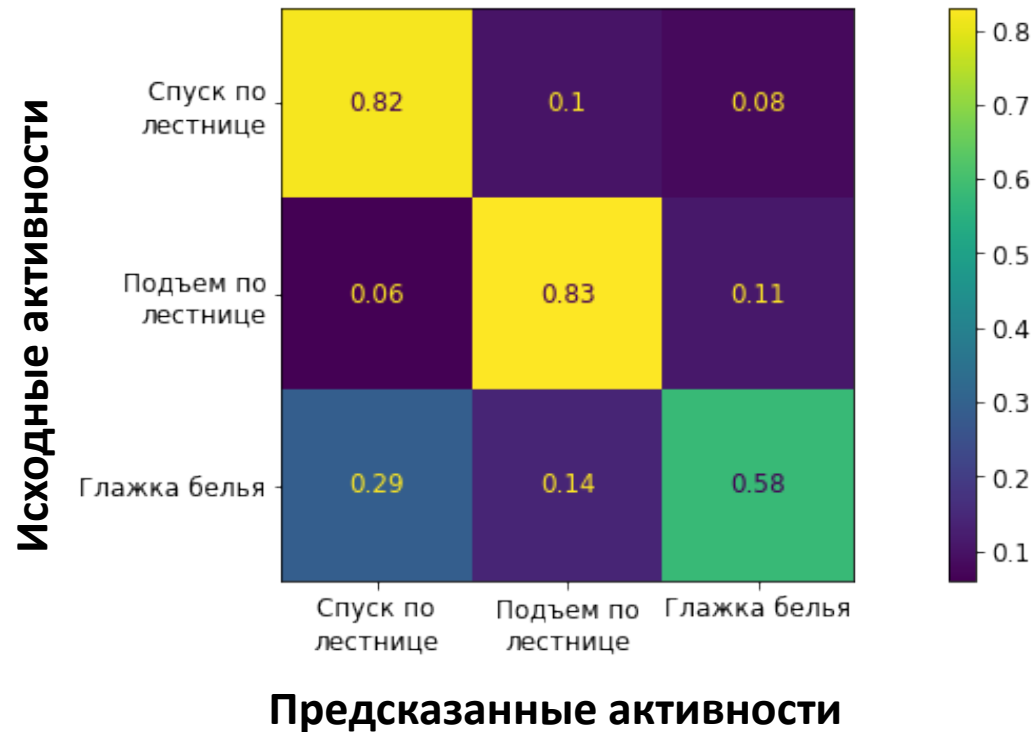
- Исходное аннотирование:



- Полученное аннотирование:

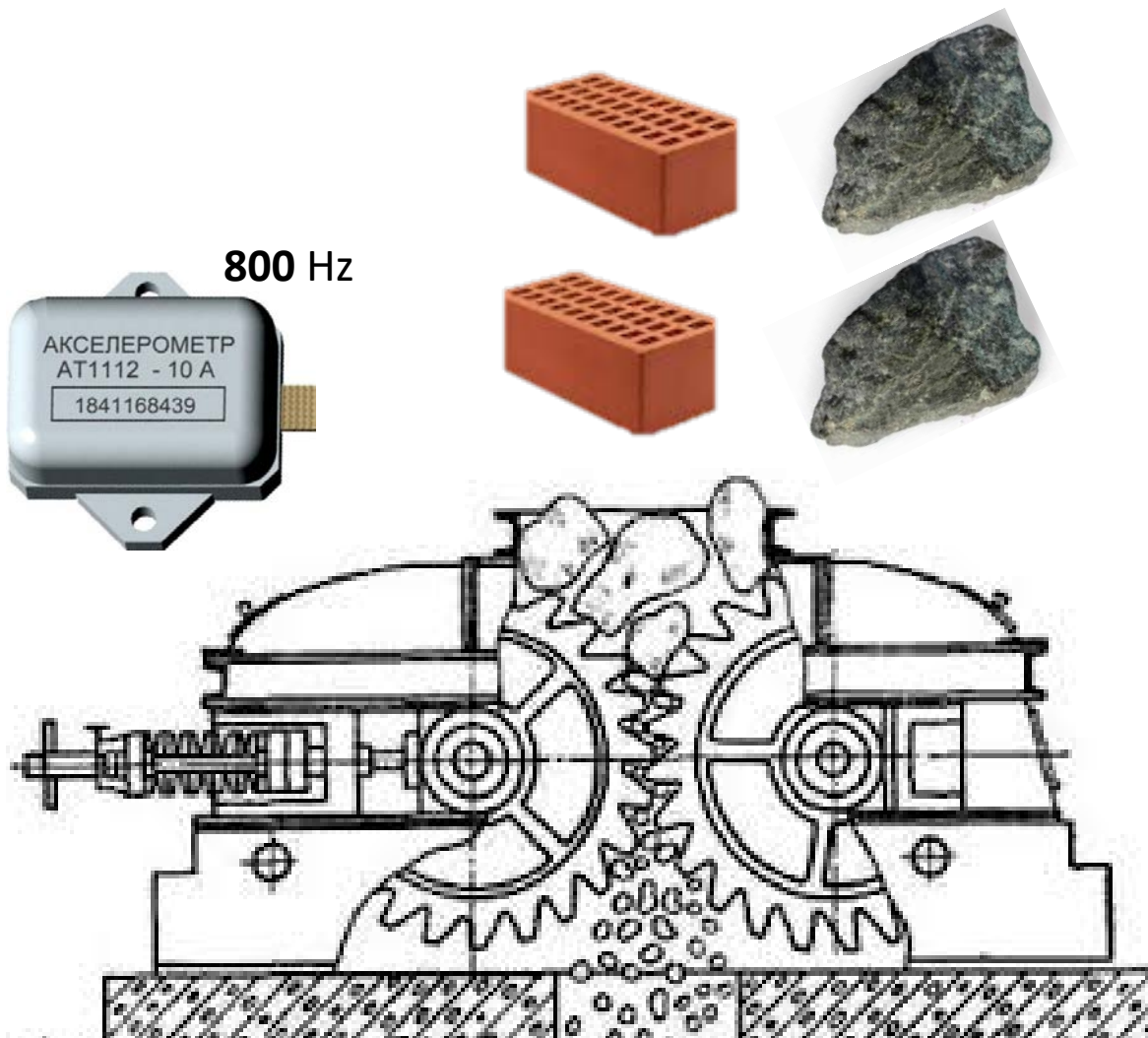


Аннотирование RA-MAP: оценка точности



Активность	Точность	Полнота	F1-мера
Спуск по лестнице	0.59	0.82	0.68
Подъем по лестнице	0.74	0.83	0.78
Глажка белья	0.83	0.58	0.68

Аннотирование Crusher

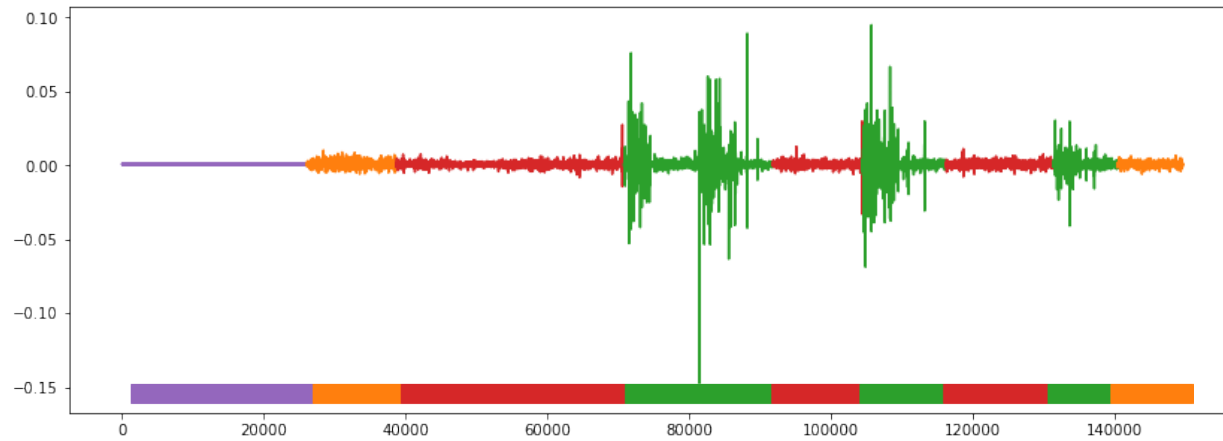


Протокол

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

Аннотирование Crusher

- Исходное аннотирование:

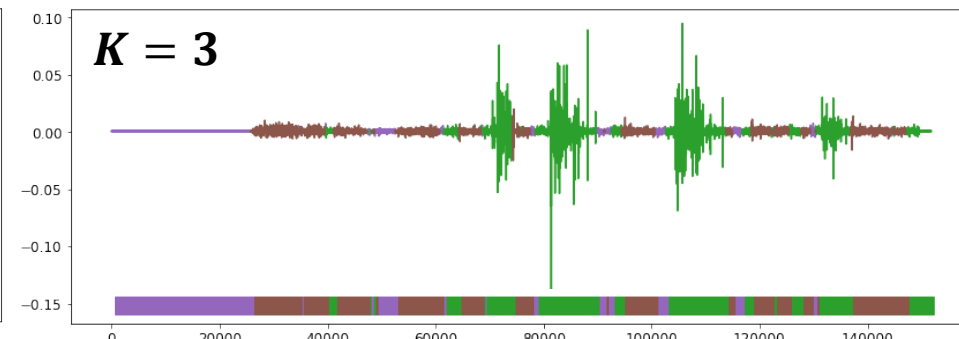
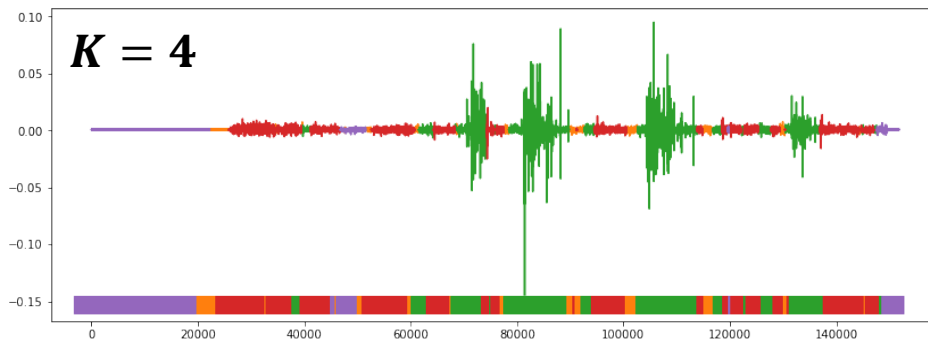


Активности:

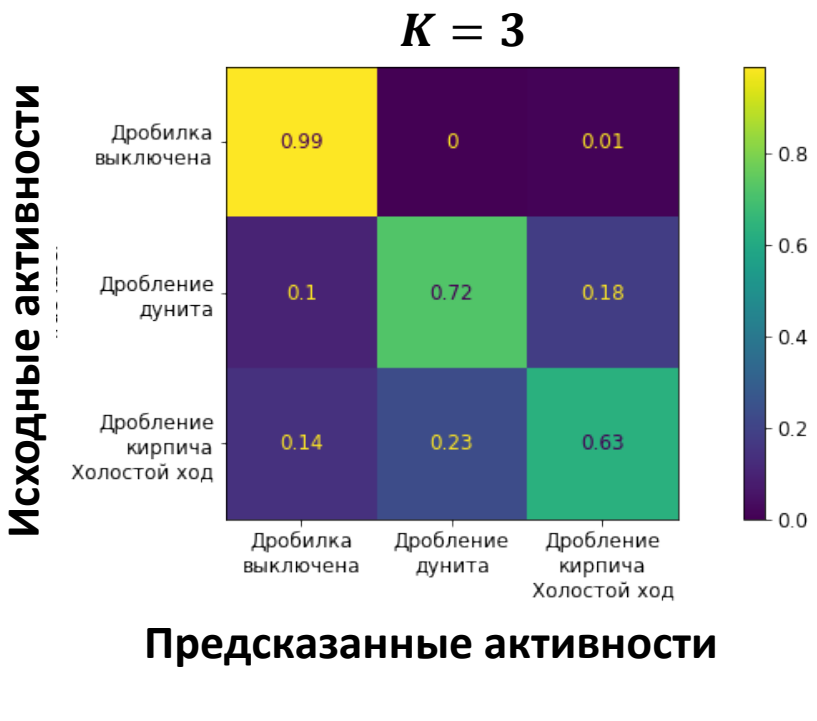
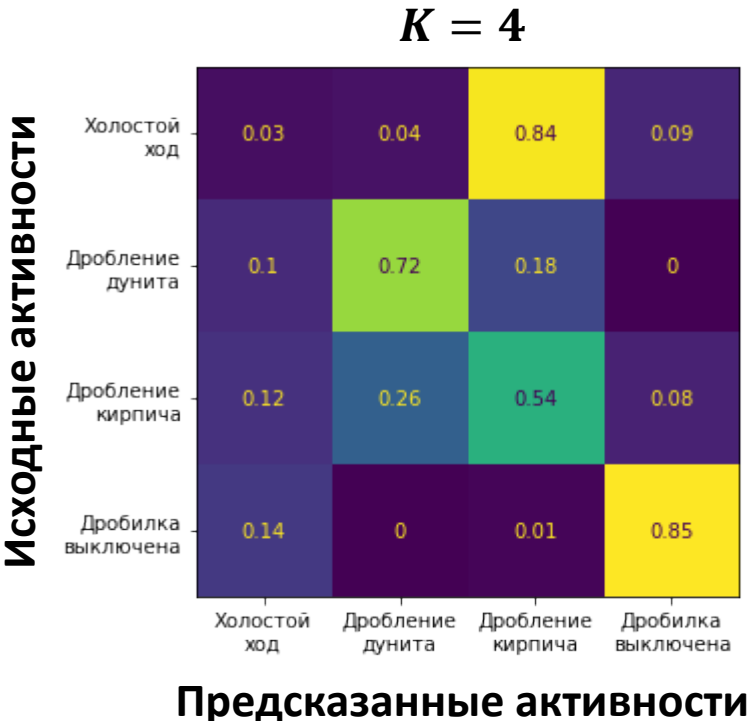
- холостой ход
- дробление кирпича
- дробление дунита
- дробилка выключена
- при $K = 3$:
холостой ход *или*
дробление кирпича

Алгоритм	Время, с
Snippet-Finder	12 796
PSF	726

- Полученное аннотирование:



Аннотирование Crusher: оценка точности



Активность	K = 4			K = 3		
	Точность	Полнота	F1-мера	Точность	Полнота	F1-мера
Холостой ход	0.05	0.03	0.04	0.87	0.63	0.73
Дробление кирпича	0.55	0.54	0.55			
Дробление дунита	0.65	0.72	0.68	0.62	0.72	0.66
Дробилка выключена	0.77	0.85	0.81	0.63	0.99	0.77

Заключение

- PSF (Parallel Snippet Finder) параллельный алгоритм поиска типичных подпоследовательностей временного ряда на GPU
- Высокая производительность PSF в вычислительных экспериментах
- Будущие исследования: распределенный алгоритм

- Спасибо за внимание! Вопросы?
 - Андрей Гоглачев, goglachevai@susu.ru