

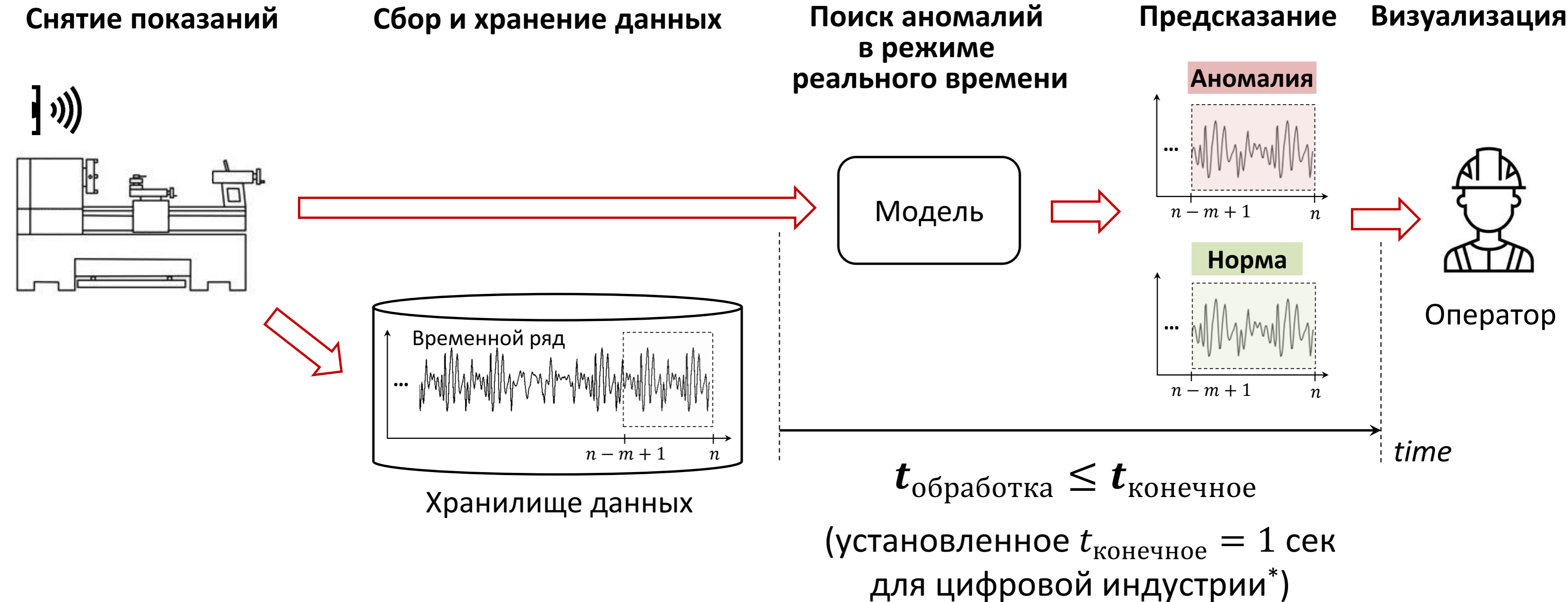
Всероссийская научная конференция с международным участием
Параллельные вычислительные технологии (ПаВТ'2024)
Челябинск, 2–4 апреля 2024 г.

Метод поиска аномалий временного ряда в режиме реального времени

Я.А. Краева

Южно-Уральский государственный университет (Челябинск)

Поиск аномалий в потоковых временных рядах



* Rosemount 248 wireless temperature transmitter.

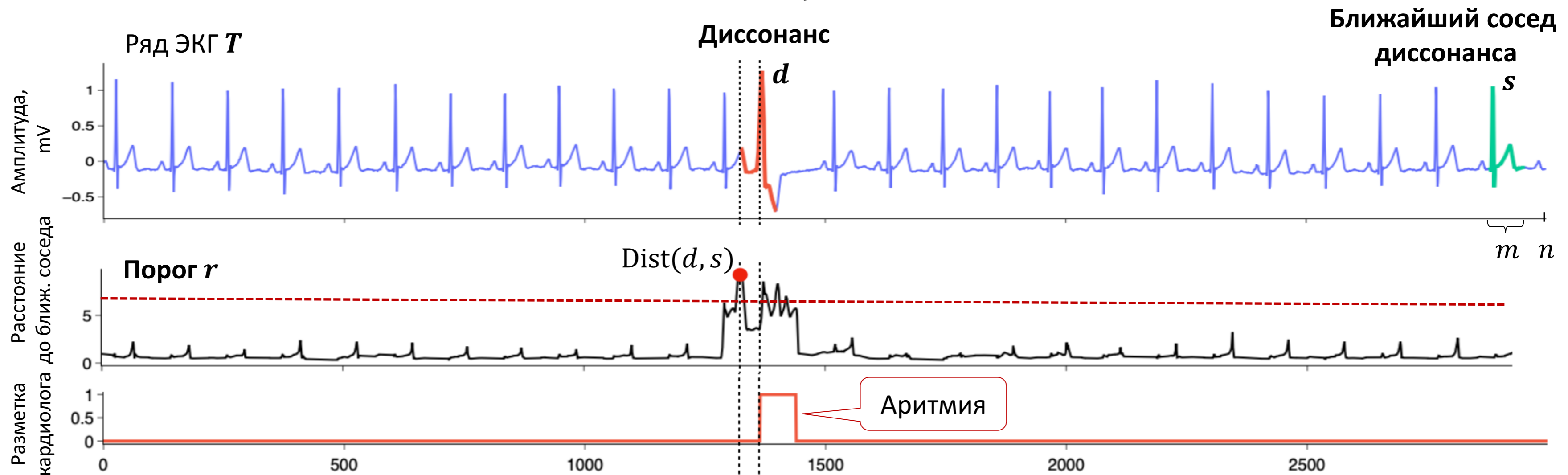
URL: <https://www.emerson.com/documents/automation/product-data-sheet/rosemount-248-wireless-temperature-transmitter-en-87848.pdf>

Аномальная подпоследовательность: диссонанс

Диссонанс d ряда T – подпоследовательность длины m , расстояние от которой до *ближайшего соседа* не ниже заданного порога r ;

ближайший сосед s – подпоследовательность, расстояние от которой до d минимально, причем d и s не пересекаются:

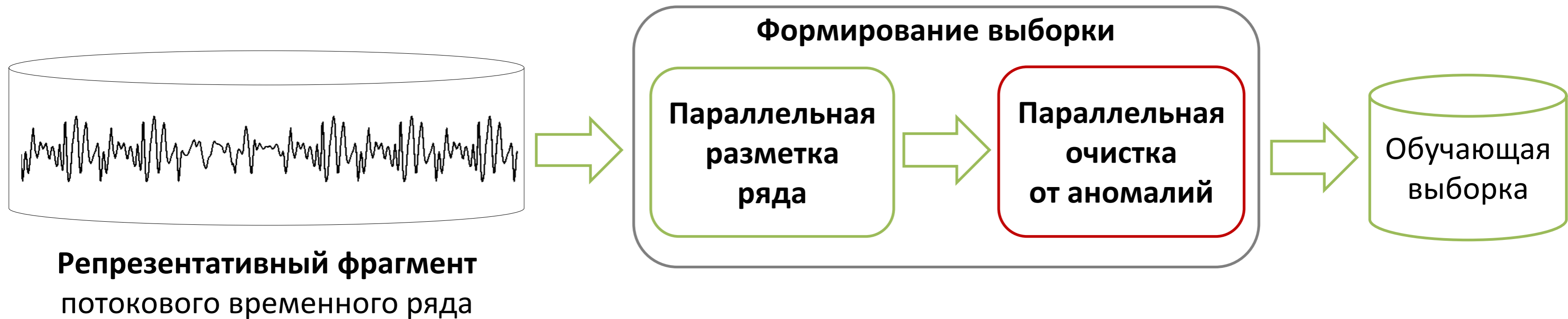
$$\forall s \in T \quad \min_{\{s \in T \mid s \cap d = \emptyset\}} \text{Dist}(d, s) \geq r$$



Работы по теме исследования

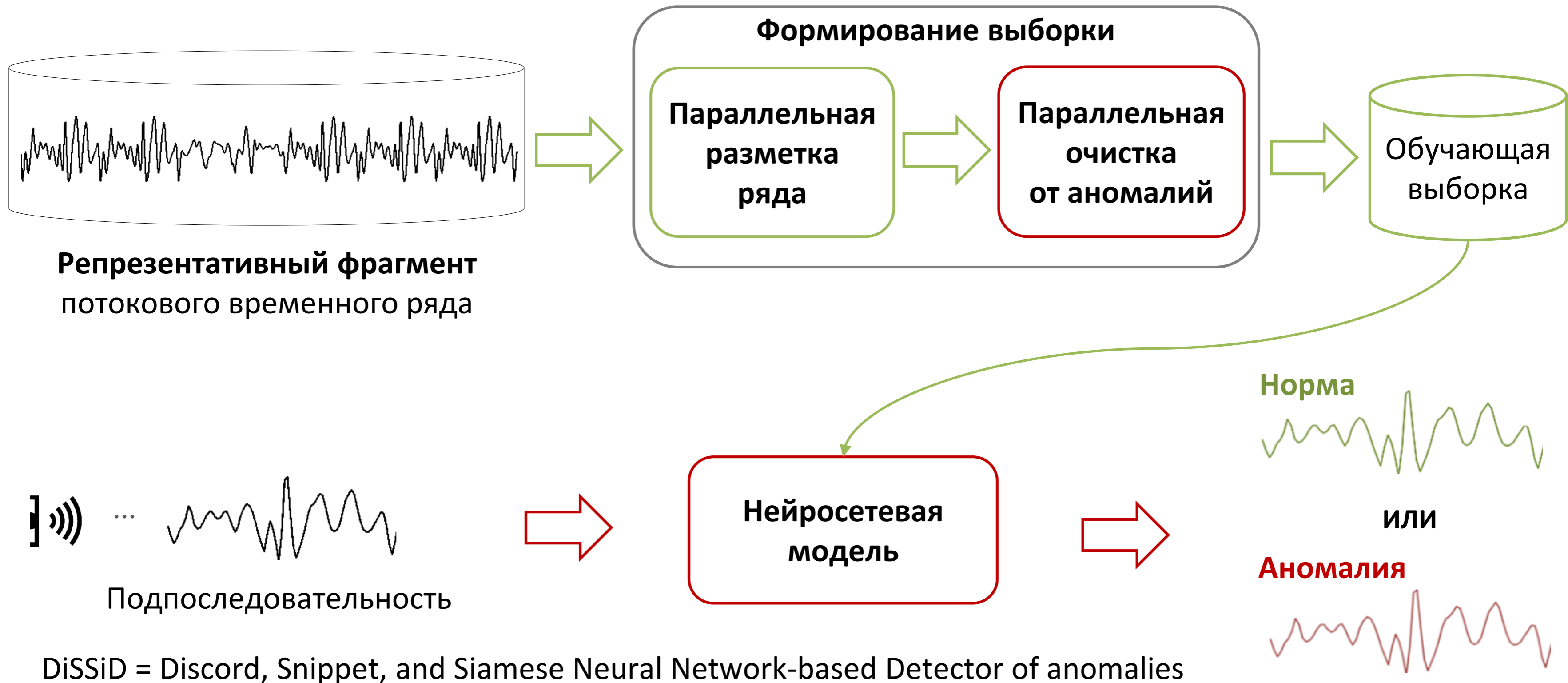
	Алгоритм	Особенности	Критика
Параллельные алгоритмы	PhiDD: Zymbler M. et al. A Parallel Approach to Discords Discovery in Massive Time Series Data. Computers, Materials & Continua 66(2): 1867-1876. 2021.	<ul style="list-style-type: none"> • Матрично-векторные структуры данных • Аппаратная платформа: кластер Intel MIC 	Производительность не достаточна для режима реального времени
	Жу и др.: Zhu B. et al. A GPU Acceleration framework for motif and discord based pattern mining. IEEE Trans. on Paral. and Distr. Systems 32(8): 1987-2004. 2021.	<ul style="list-style-type: none"> • Вычислительные шаблоны • Поиск одного диссонанса ряда • Аппаратная платформа: GPU 	
	KBF_GPU: Thuy T.T.H. et al. A new discord definition and an efficient time series discord detection method using GPUs. ICSED 2021. pp. 63-70.	<ul style="list-style-type: none"> • Полный перебор подпоследовательностей ряда • Аппаратная платформа: GPU 	
Нейросетевые методы	AE: Sakurada M. et al. Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction. MLSDA 2014. pp. 4.	<ul style="list-style-type: none"> • Основан на восстановлении временного ряда • Архитектура сети: автокодировщик, где энкодер и декодер – многослойные перцептроны 	Требуется предварительная очистка временного ряда от аномалий (возможна с участием эксперта)
	DeepAnT: Munir M. et al. DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series. IEEE Access 7: 1991-2005. 2019.	<ul style="list-style-type: none"> • Основан на прогнозировании временного ряда • Архитектура сети: сверточная нейросеть 	
	LSTM-AD: Malhotra P., et al. Long Short Term Memory Networks for Anomaly Detection in Time Series. ESANN 2015.	<ul style="list-style-type: none"> • Основан на прогнозировании временного ряда • Архитектура сети: рекуррентная нейросеть LSTM 	
	TAnoGAN: Bashar M.A., et al. TAnoGAN: Time Series Anomaly Detection with Generative Adversarial Networks. SSCI 2020. IEEE, 2020. pp. 1778-1785.	<ul style="list-style-type: none"> • Основан на восстановлении временного ряда • Архитектура сети: нейросеть GAN, где генератор и дискриминатор – рекуррентные нейросети LSTM 	

Общая схема метода DiSSiD



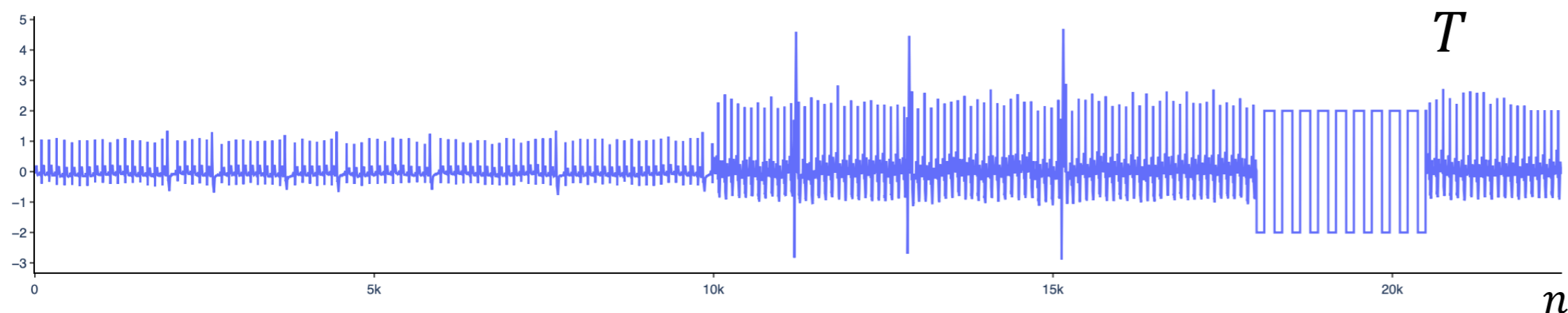
DiSSiD = Discord, Snippet, and Siamese Neural Network-based Detector of anomalies

Общая схема метода DiSSiD

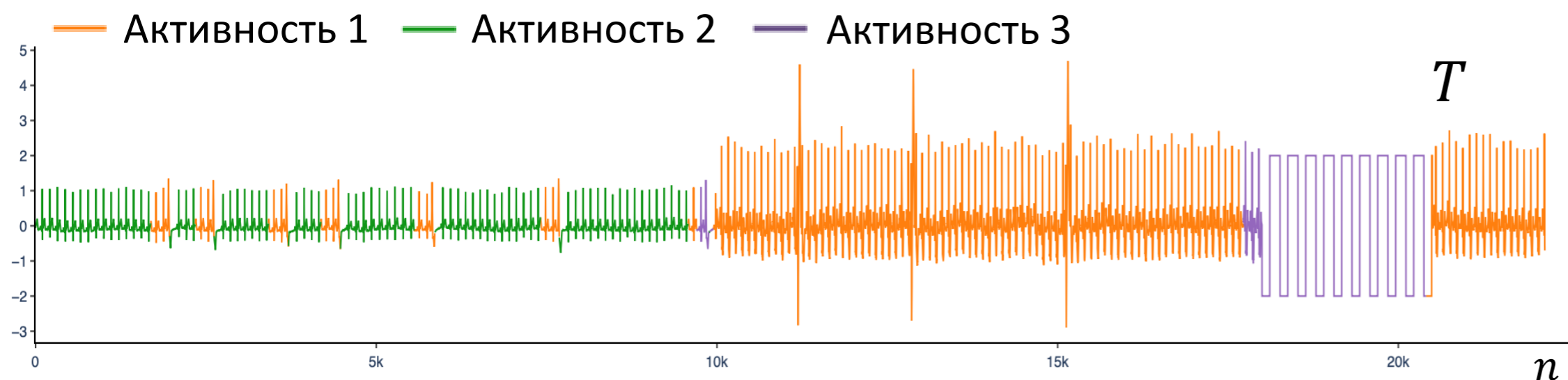


Параллельная разметка ряда

- Разметка ряда на основе поведенческих шаблонов (сниппетов)¹⁾



PSF (Parallel Snippet Finder)²⁾



Мера MPdist

$$\text{MPdist}_\ell^k(A, B) = \text{AscSort}(P_{ABBA})(k),$$
$$P_{ABBA} = P_{AB} \cdot P_{BA}, \quad P_{AB} = \{\text{ED}_{\text{norm}}^2(A_{i,\ell}, B_{j,\ell})\}_{i=1}^{m-\ell+1},$$
$$B_{j,\ell} = \arg \min_{1 \leq q \leq m-\ell+1} \text{ED}_{\text{norm}}^2(A_{i,\ell}, B_{q,\ell}),$$

где $0 < k < m$, $k_{\text{default}} = \lceil 0.1m \rceil$

Множество сниппетов длины m :

$$C_T^m = \{C_i \mid C_i \in T, 1 \leq i \leq K\}$$

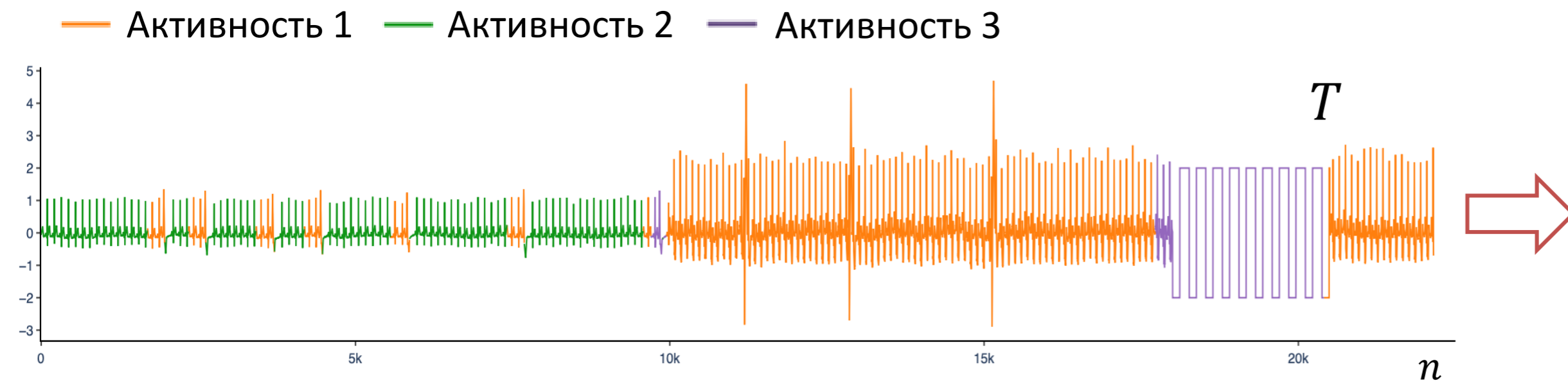
Атрибуты сниппета C_i :

- $C_i.index$ – индекс
- $C_i.NN$ – ближайшие соседи
- $C_i.frac$ – мощность сниппета

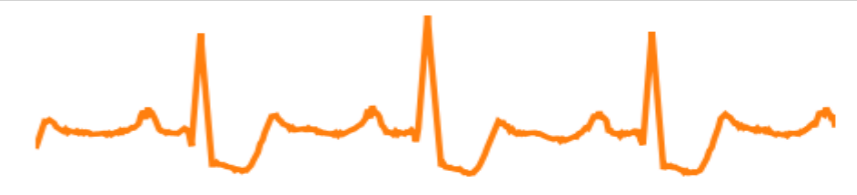
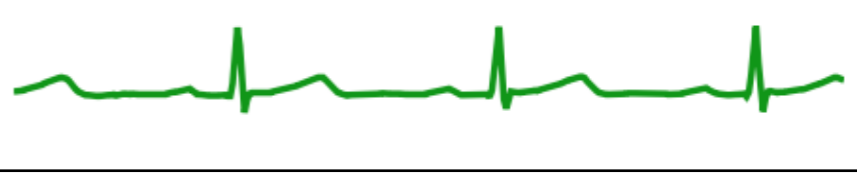
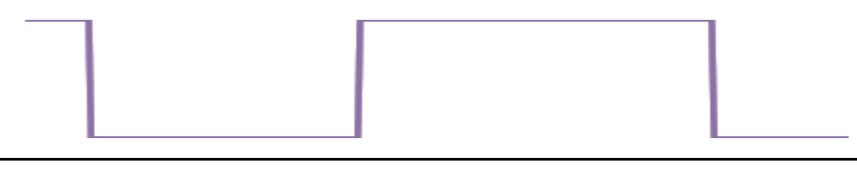
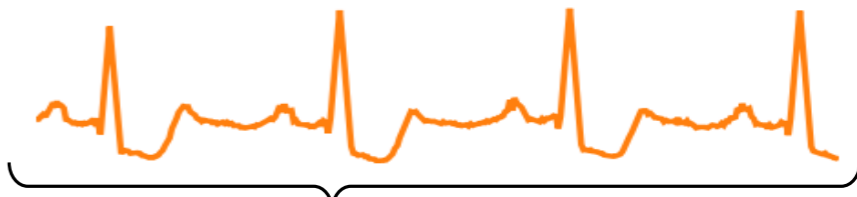
¹⁾ Imani S., et al. Introducing time series snippets: a new primitive for summarizing long time series. Data Min. Knowl. Discov. 34(6): 1713-1743 (2020). DOI: [10.1007/s10618-020-00702-y](https://doi.org/10.1007/s10618-020-00702-y)

²⁾ Zymbler M., Gogachev A. Fast Summarization of Long Time Series with Graphics Processor. Mathematics 10(10). Article 1781. 2022. DOI: [10.3390/math10101781](https://doi.org/10.3390/math10101781)

Параллельная разметка ряда



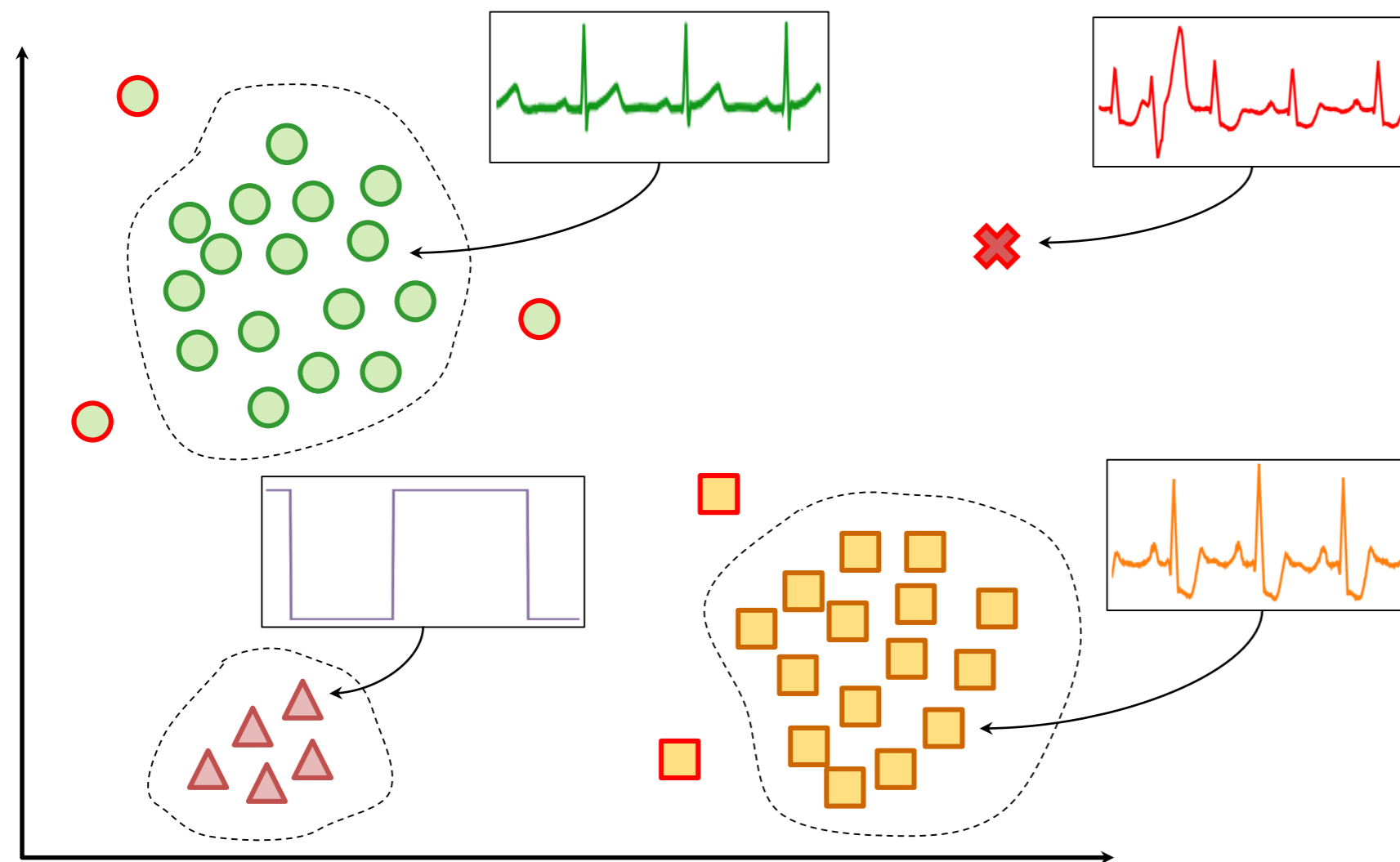
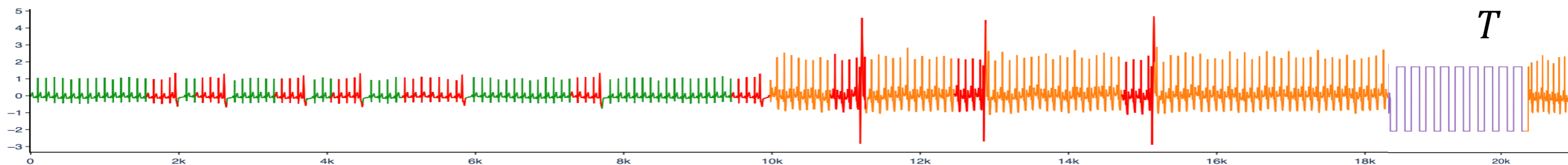
Множество
подпоследовательностей S

Подпоследовательность	Номер класса
	1
	2
	3
...	...
	1

Сниппеты и их
ближайшие соседи

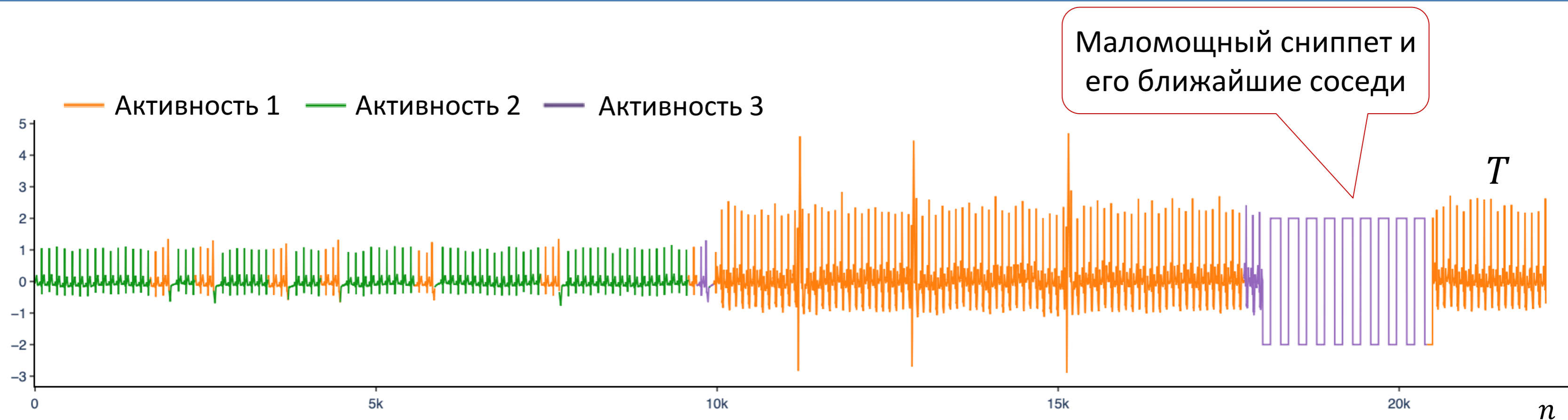
Номер
сниппета

Таксономия аномалий



- □ – типичные активности
- △ – нетипичные активности
- □ – шумы в каждой активности
- × – аномальные активности (диссонансы)

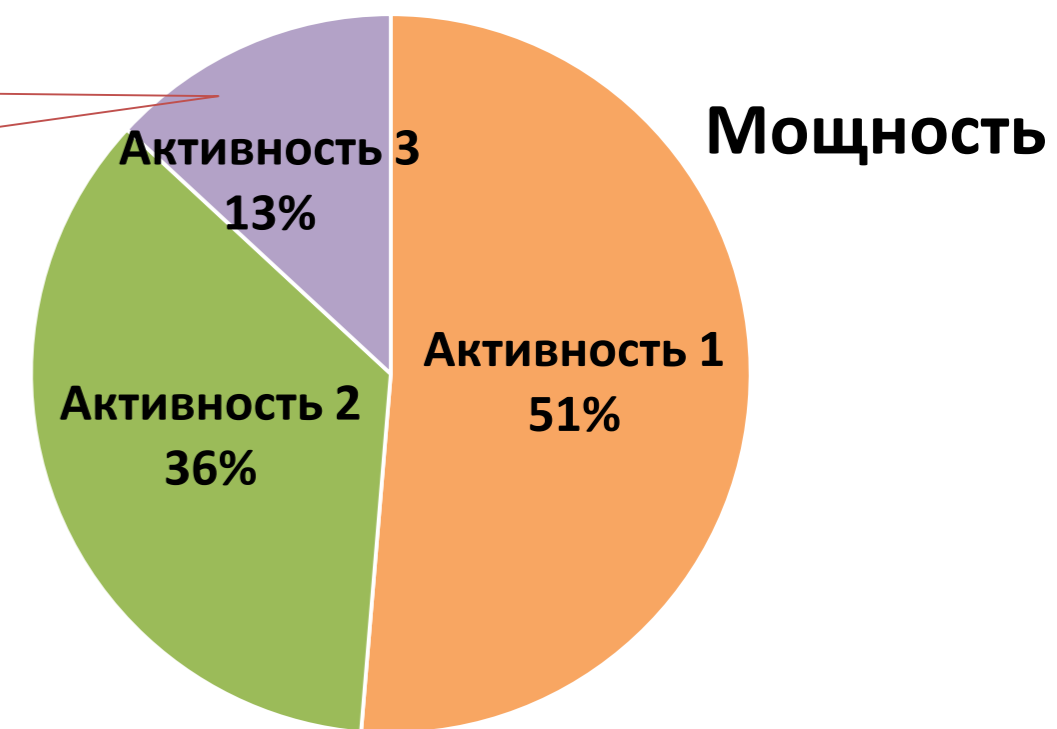
Поиск нетипичных активностей



$$C_3.frac < \varepsilon$$

$$C_{weak} \leftarrow \{C_i \in C_T^m \mid C_i.frac \leq \varepsilon, 1 \leq i \leq K\}$$

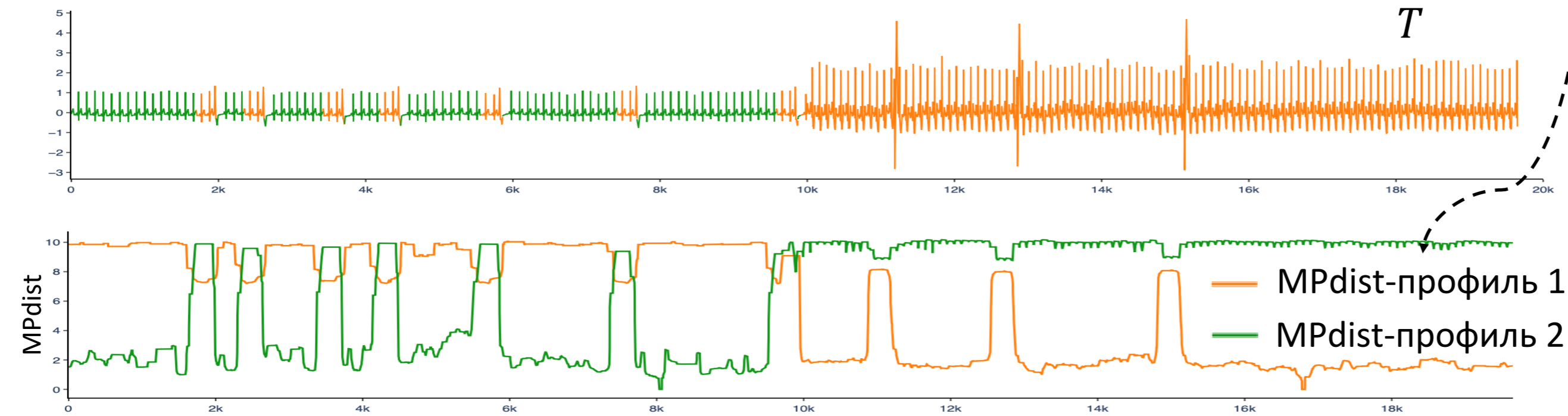
ε – пороговая мощность снимета, $0 < \varepsilon \leq \frac{1}{K}$



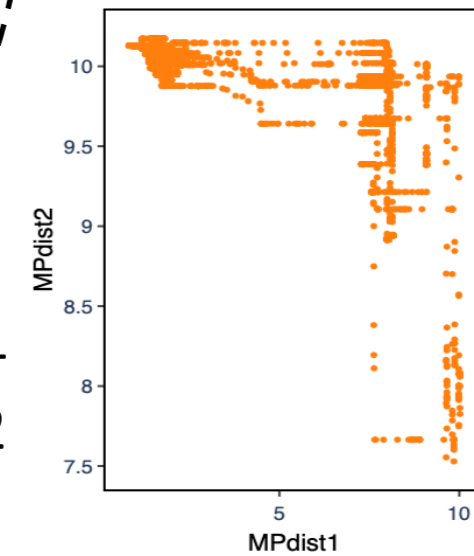
Поиск шумов в активностях

MPdist-профиль

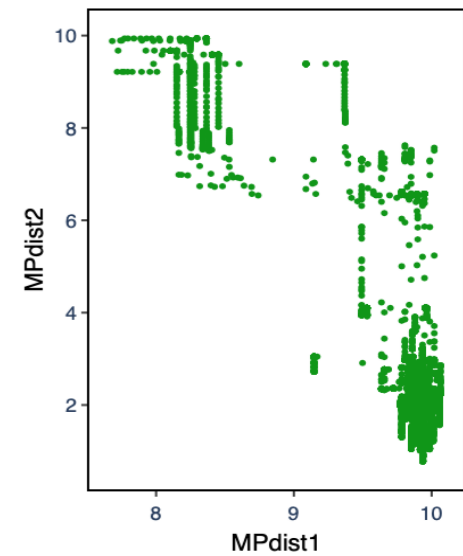
$$MPD_i = \left\{ MPdist(C_i, T_{j,m}) \right\}_{j=1}^{n-m+1}, 1 \leq i \leq K$$



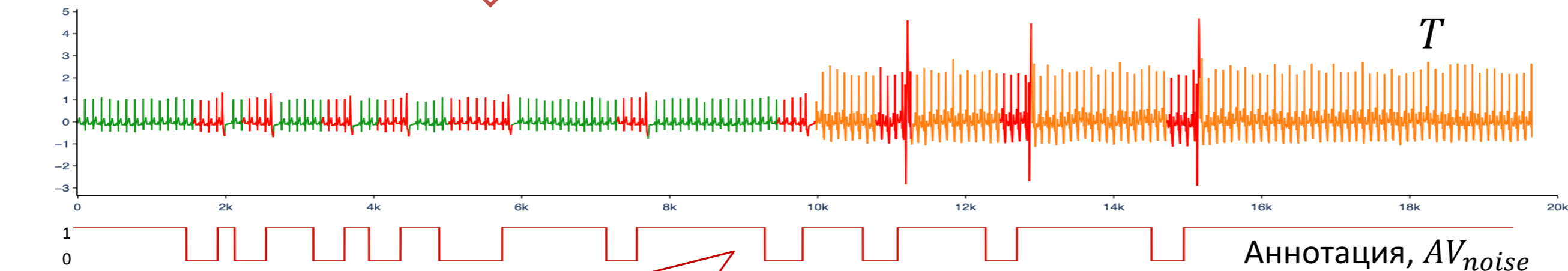
Активность 1



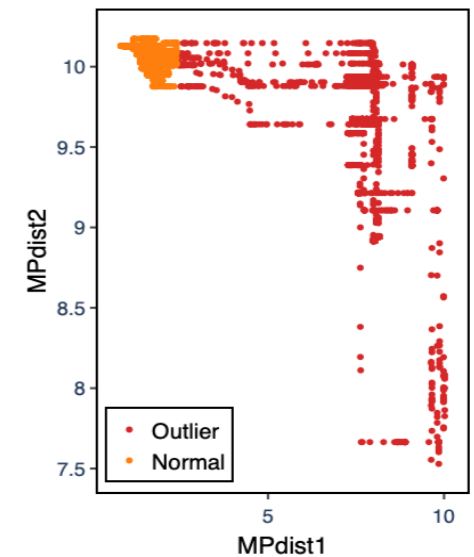
Активность 2



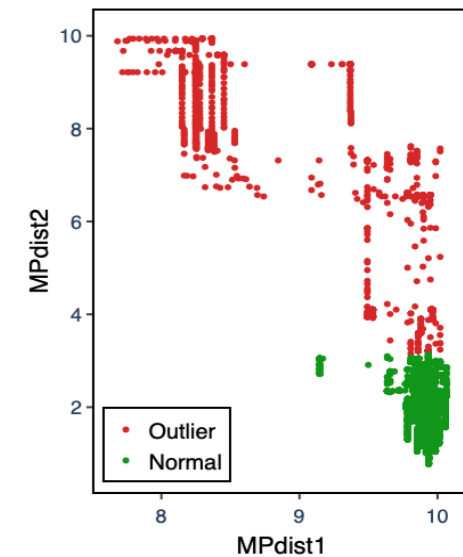
for $i \leftarrow 1$ to $|C_T^m|$ do
 $O \leftarrow \text{IsolationForest}(C_i.\text{profile}) \cup O$



Активность 1

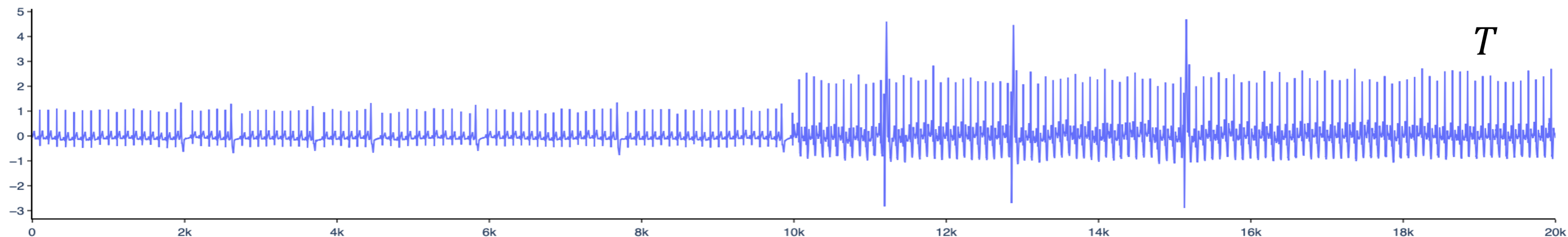


Активность 2

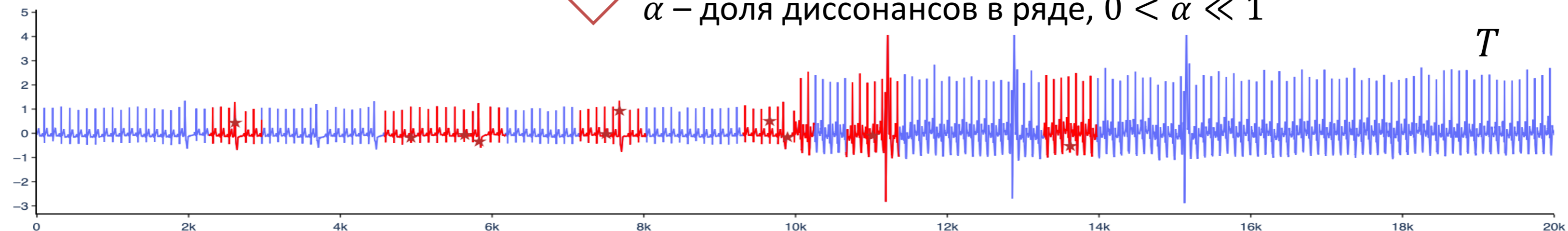


$$AV_{noise}[i] = \begin{cases} 0, & T_{i,m} \in O \\ 1, & \text{otherwise} \end{cases}$$

Параллельный поиск аномальных активностей



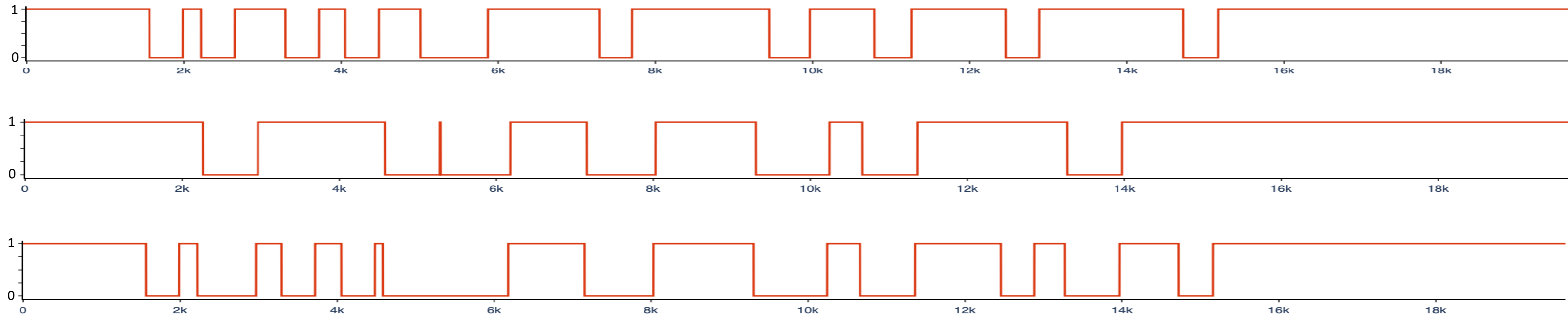
$D \leftarrow \text{PALMAD}(T, m, \alpha),$
 α – доля диссонансов в ряде, $0 < \alpha \ll 1$



$$AV_{discords}[i] = \begin{cases} 0, & T_{i,m} \in D \\ 1, & otherwise \end{cases}$$

Краева Я.А., Цымблер М.Л. Поиск аномалий в больших временных рядах на кластере с GPU узлами. Суперкомпьютерные дни в России: Труды международной конференции 2023. С. 149–160. DOI: [10.29003/m3478.978-5-317-07070-0](https://doi.org/10.29003/m3478.978-5-317-07070-0)

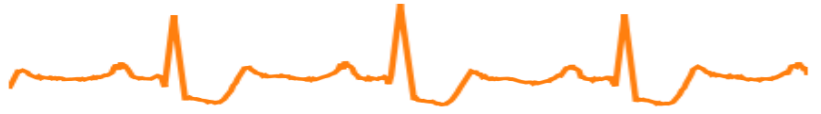
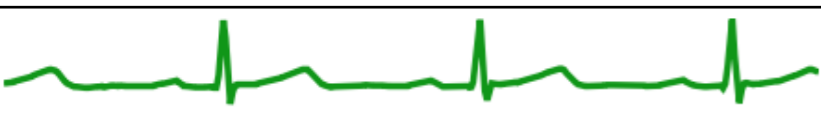

Очистка



$$AV_{noise} \wedge AV_{discords} = AV$$



Множество подпоследовательностей S

Подпоследовательность	Номер класса
	1
	2
...	...
	1

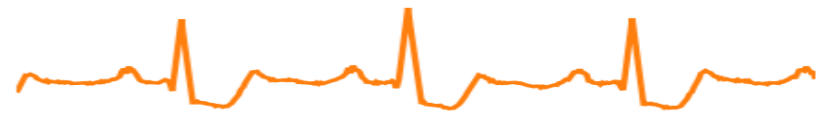

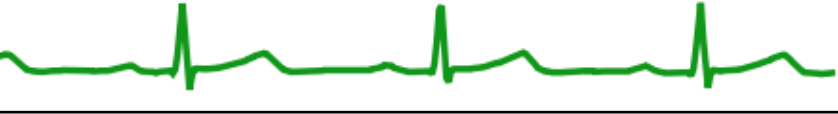
Множество $S_{anomaly}$
 $\{s_i \in S \mid AV(i) = 0\}$

Подпоследовательность

...


$$S \leftarrow S \setminus S_{anomaly}$$

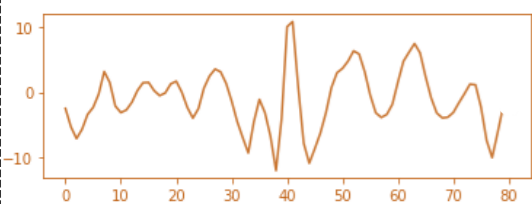
Очищенное множество S

Подпоследовательность	Номер класса
	1
	2
...	...
	2

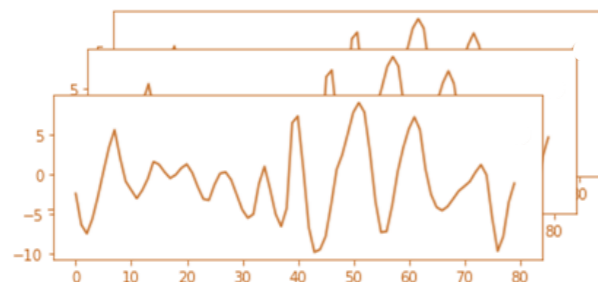
Обучающая выборка \mathcal{L}

Очищенное множество S

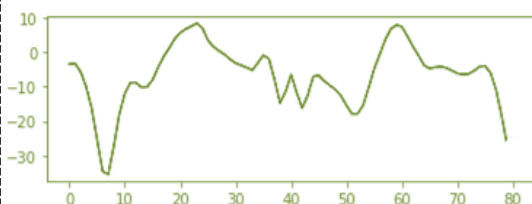
Сниппет C_1



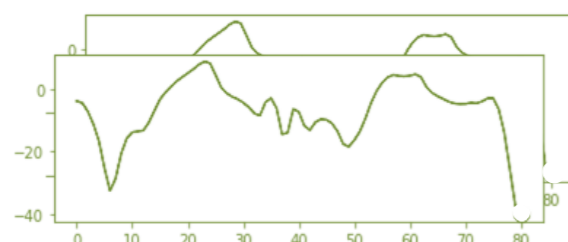
Соседи $C_1.NN$



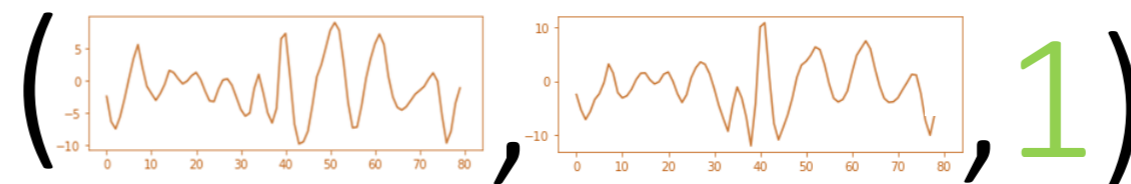
Сниппет C_2



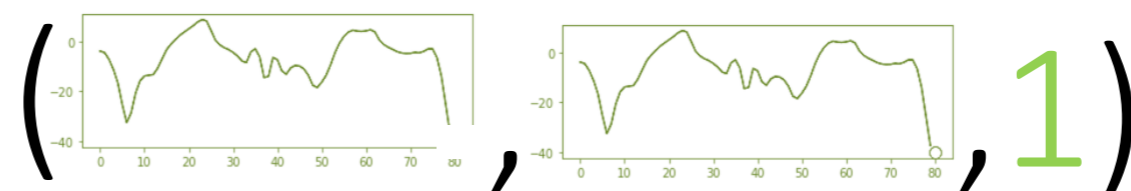
Соседи $C_2.NN$



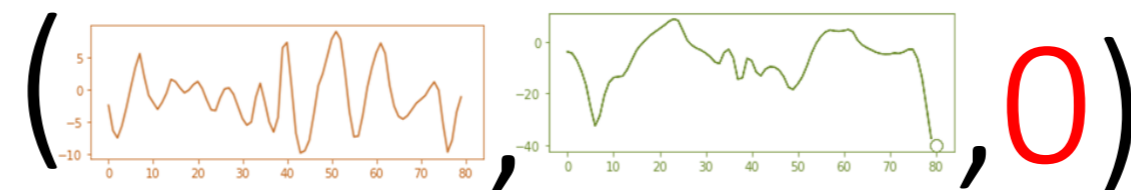
Позитивные прецеденты



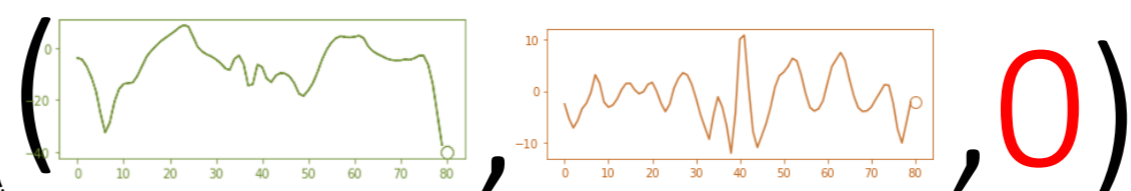
⋮



Негативные прецеденты



⋮



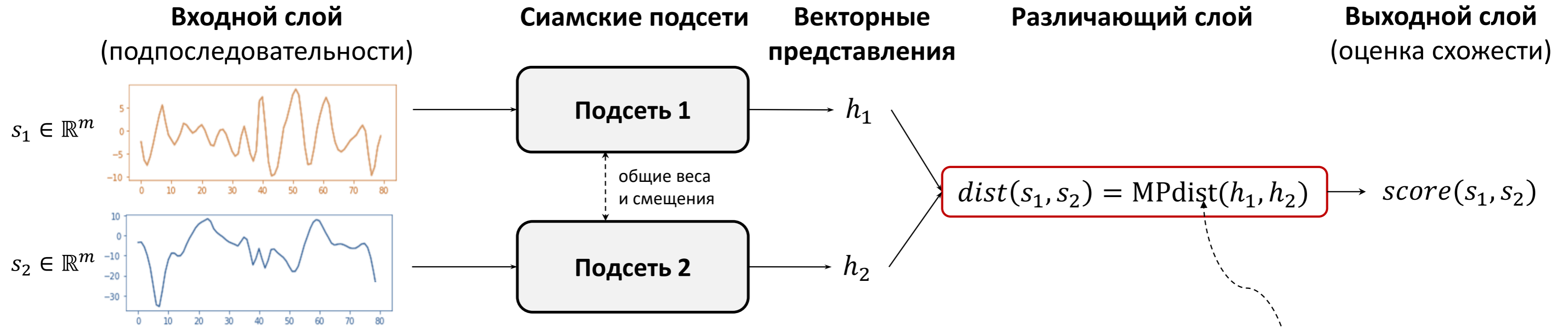
Max # **ПОЗИТИВНЫХ** пар =

$$= \sum_{i=1}^K \binom{2}{|C_i.NN|}$$

Max # **НЕГАТИВНЫХ** пар =

$$= \sum_{i=1}^K \sum_{j=i}^K |C_i.NN| \cdot |C_j.NN|$$

Сиамская нейронная сеть



Модифицированная функция контрастных потерь L :

$$L(s_1, s_2) = \delta_{s_1 s_2} \cdot \text{MPdist}_\ell^k(h_1, h_2) + (1 - \delta_{s_1 s_2}) \cdot \max(\tau - \text{MPdist}_\ell^k(h_1, h_2), 0)$$

$$\delta_{s_1 s_2} = \begin{cases} 1, & s_1, s_2 \in C_i.NN, 1 \leq i \leq K \\ 0, & s_1 \in C_i.NN, s_2 \in C_j.NN, i \neq j, 1 \leq i, j \leq K \end{cases}$$

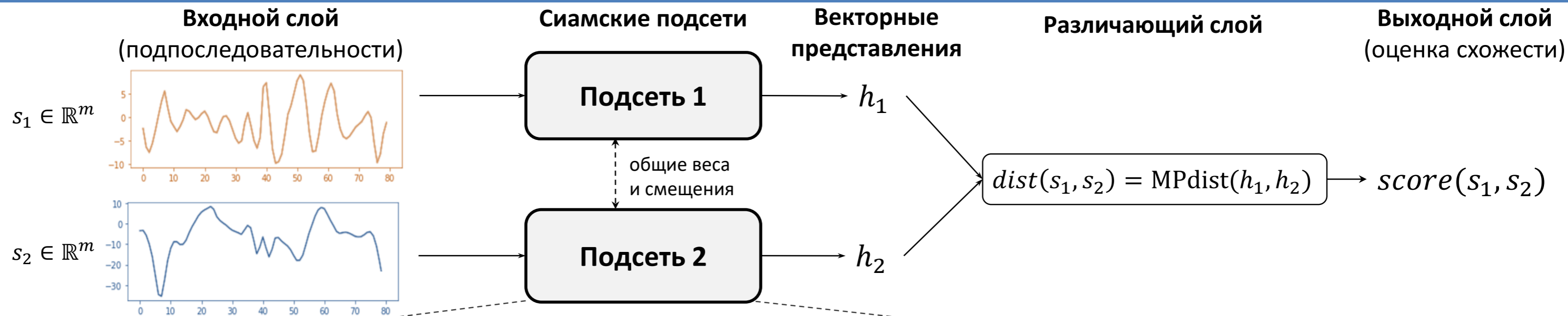
$$\tau = \min_{\{s_1, s_2 \mid \delta_{s_1 s_2} = 0\}} \text{MPdist}_\ell^k(h_1, h_2)$$

Мера MPdist^*

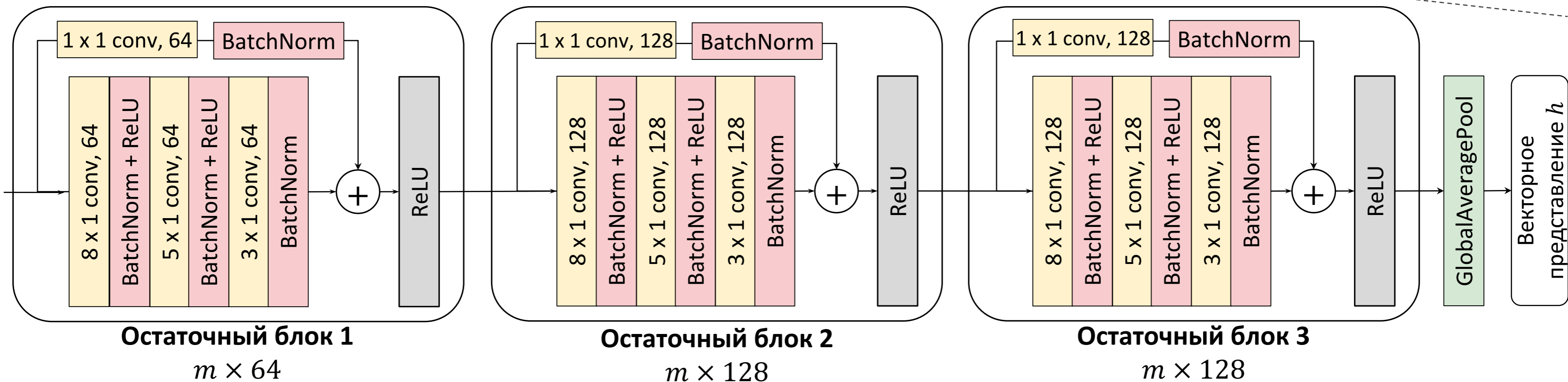
$$\begin{aligned} \text{MPdist}_\ell^k(A, B) &= \text{AscSort}(P_{ABBA})(k), \\ P_{ABBA} &= P_{AB} \cdot P_{BA}, P_{AB} = \{\text{ED}_{\text{norm}}^2(A_{i,\ell}, B_{j,\ell})\}_{i=1}^{m-\ell+1}, \\ B_{j,\ell} &= \arg \min_{1 \leq q \leq m-\ell+1} \text{ED}_{\text{norm}}^2(A_{i,\ell}, B_{q,\ell}), \\ &\text{где } 0 < k < m, k_{\text{default}} = \lceil 0.1m \rceil \end{aligned}$$

* Gharghabi S. 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. P. 1104–1135. DOI: [10.1007/s10618-020-00695-8](https://doi.org/10.1007/s10618-020-00695-8)

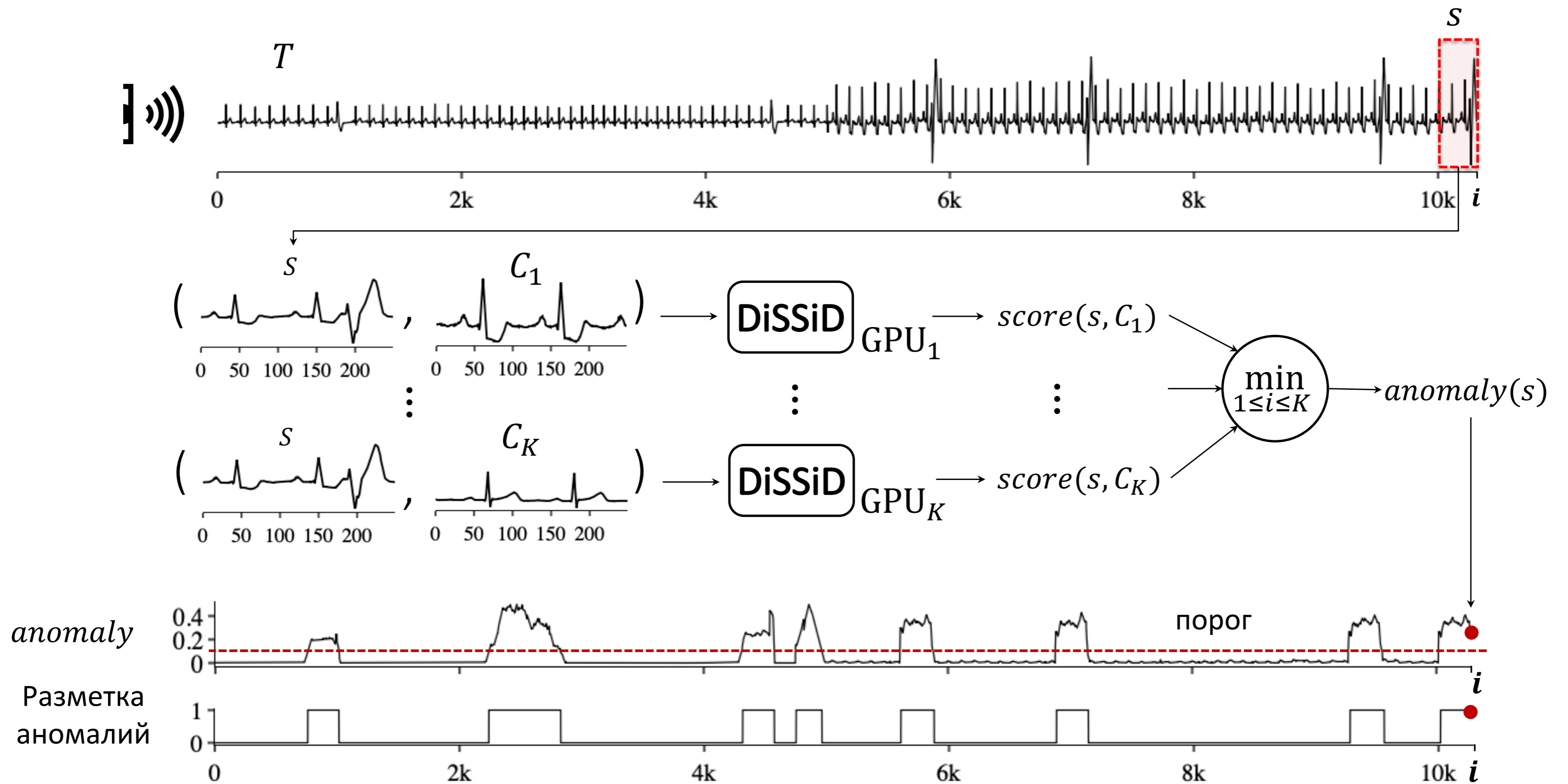
Модифицированная подсеть ResNet



ResNet



Применение нейросетевой модели

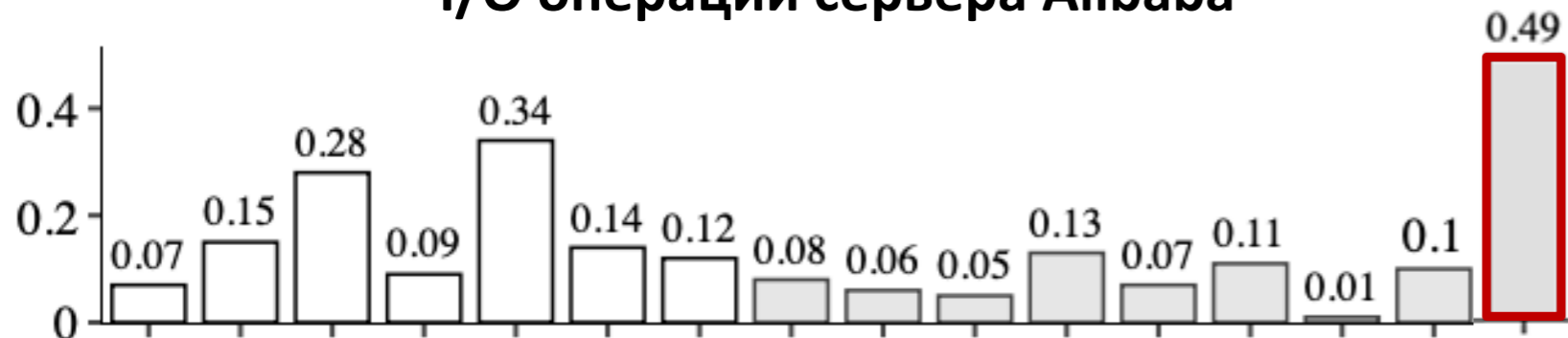


Эксперименты: точность

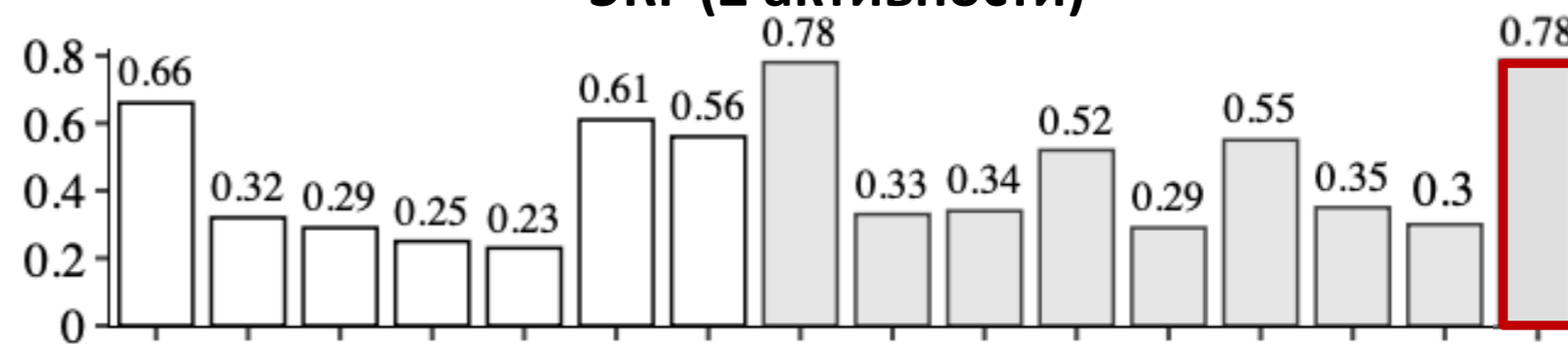
Метрика **VUS-PR** $\in [0, 1]$:
большее значение означает
 более **высокую** точность

Аппаратная платформа: «Нейрокомпьютер» ЮУрГУ, NVIDIA Tesla V100 (5 120 CUDA-ядер, 7 TFLOPS)

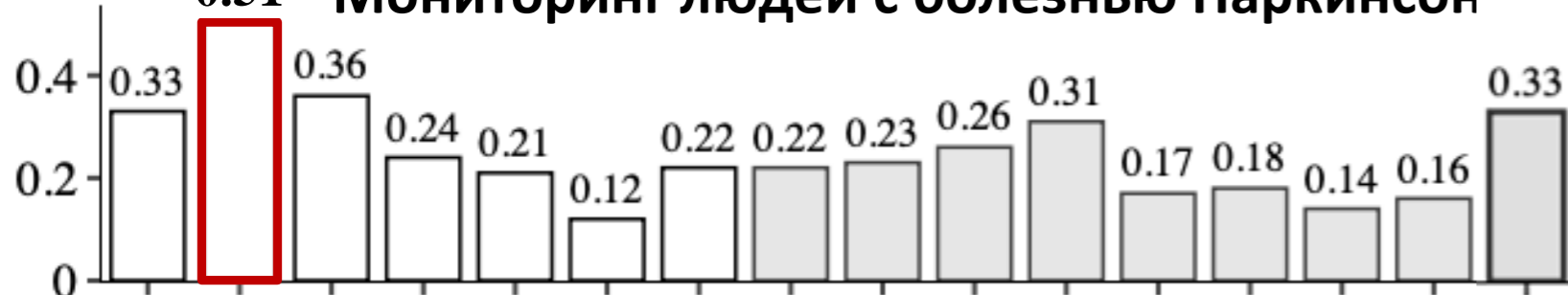
I/O операции сервера Alibaba



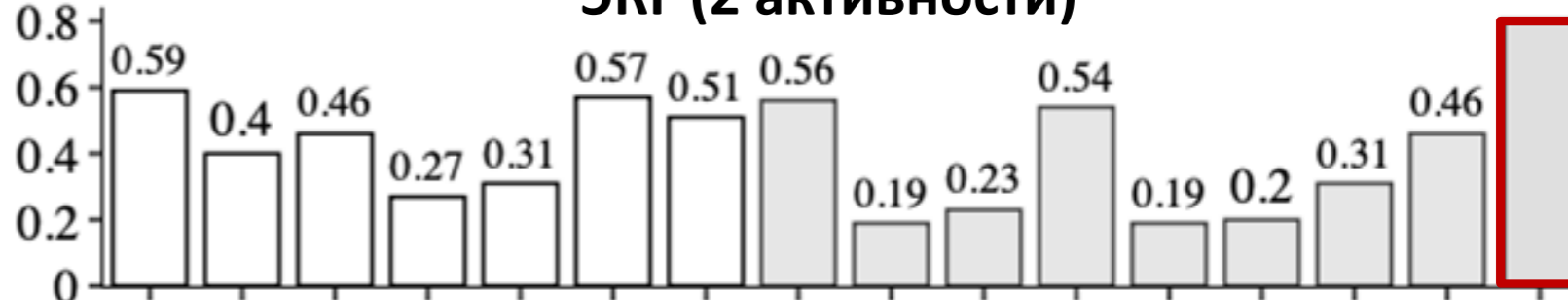
ЭКГ (2 активности)



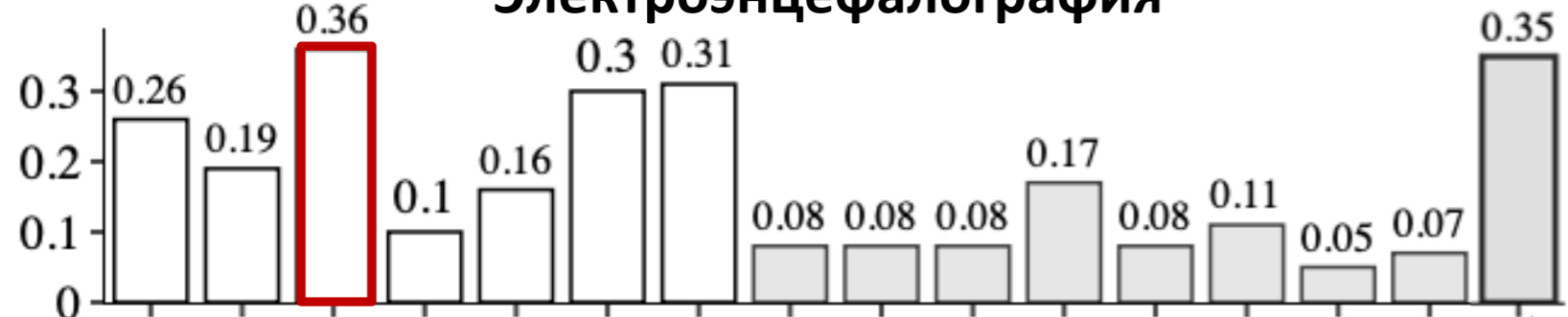
Мониторинг людей с болезнью Паркинсон



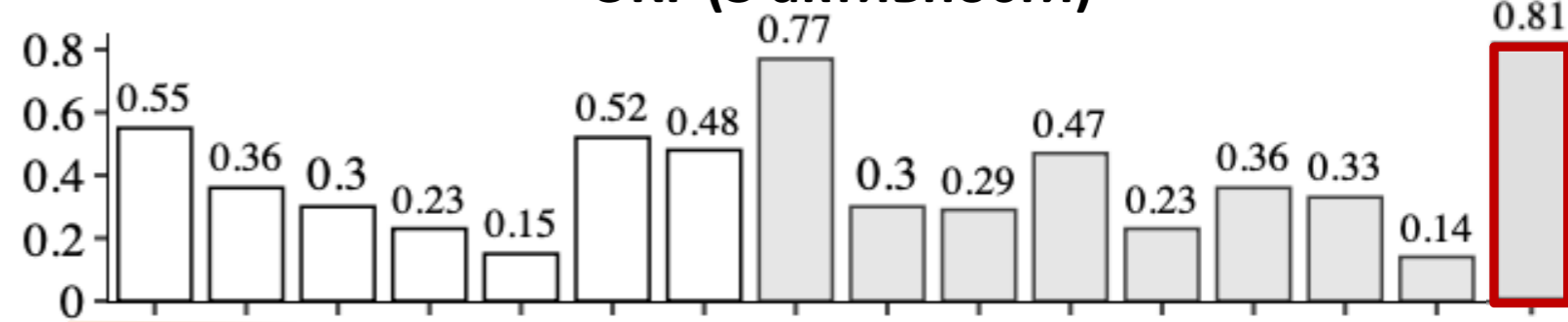
ЭКГ (2 активности)



Электроэнцефалография



ЭКГ (3 активности)



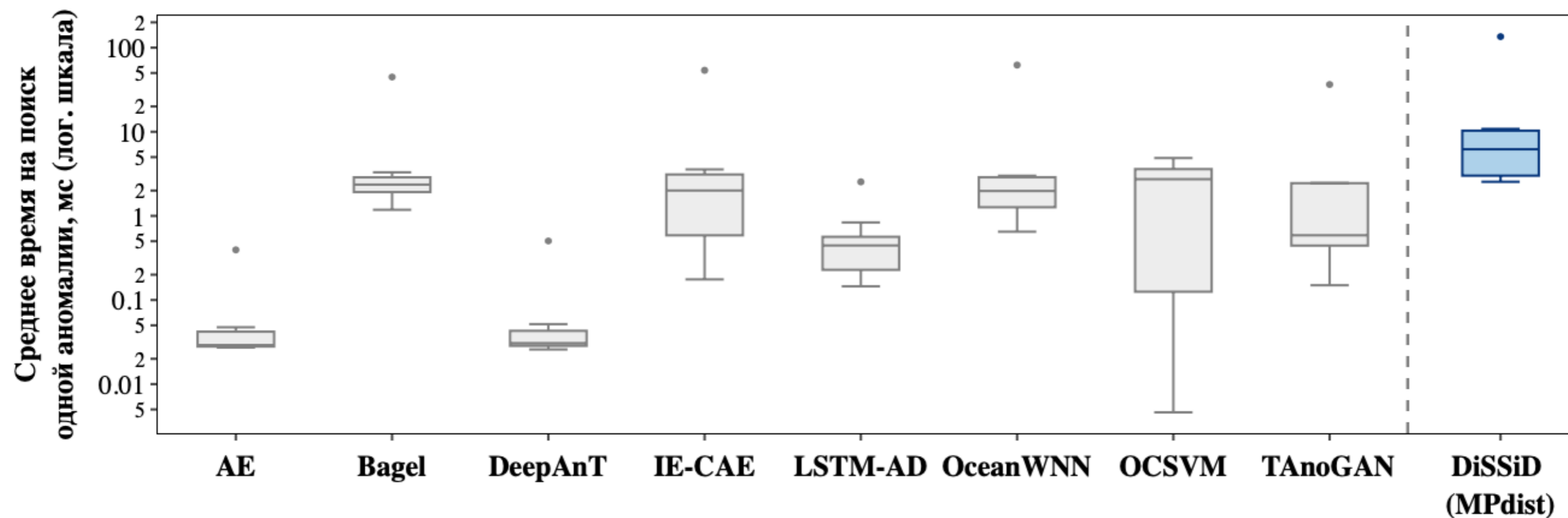
Unsupervised методы

Semi-supervised методы

Unsupervised методы

Semi-supervised методы

Эксперименты: производительность



Модель DiSSiD с функцией $MPdist(\cdot, \cdot)$
применима для поиска аномалий ряда
 в режиме реального времени

Датчик температуры Rosemount 248 Emerson*	Метод DiSSiD, сек	
	Худший случай	Лучший случай
Мин. время обновления данных, сек	1	1
	0.135	0.011

* Rosemount 248 wireless temperature transmitter.

URL: <https://www.emerson.com/documents/automation/product-data-sheet/rosemount-248-wireless-temperature-transmitter-en-87848.pdf>

Заключение

- Предложен новый метод поиска аномалий временного ряда в режиме реального времени, включающий в себя нейросетевую модель и алгоритм построения обучающей выборки для указанной модели
- Проведены вычислительные эксперименты, подтверждающие высокую точность метода
- Будущие исследования: разработка нейросетевой модели для поиска аномалий в многомерных временных рядах в режиме реального времени

Спасибо за внимание! Вопросы?

Яна Александровна Краева

kraevaya@susu.ru