



Всероссийская научная конференция с международным участием
«Параллельные вычислительные технологии (ПаВТ) 2026»



Тьюториал

Параллельные алгоритмы поиска аномалий во временных рядах

<https://kraevaya.github.io/tutorial/>



Яна Александровна Краева, Михаил Леонидович Цымблер

[kraevaya](mailto:kraevaya@susu.ru), [mzym](mailto:mzym@susu.ru)@susu.ru

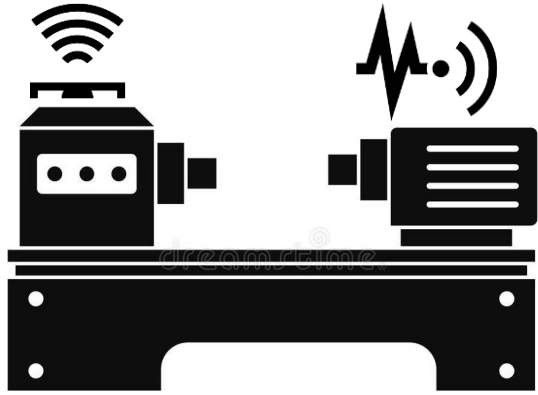
Южно-Уральский государственный университет, Челябинск, Россия

7-9 апреля 2026 г., Уфимский университет науки и технологий, Уфа

Содержание

- **Часть 1. Введение в задачу поиска аномалий (30 мин.)**
 - Теория
 - Понятие временного ряда
 - Задача поиска аномалий, типы аномалий
 - Понятие диссонанса
 - Практика
- **Часть 2. Поиск диссонансов фиксированной длины (30 мин.)**
- **Часть 3. Поиск диссонансов произвольной длины (30 мин.)**

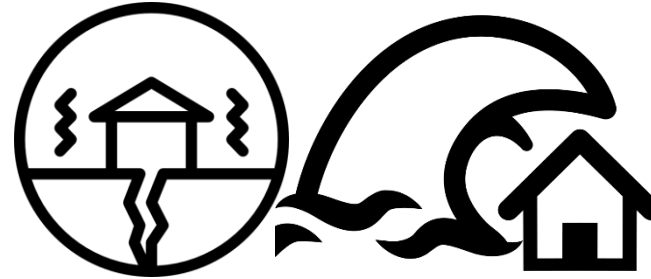
Временные ряды окружают нас повсюду



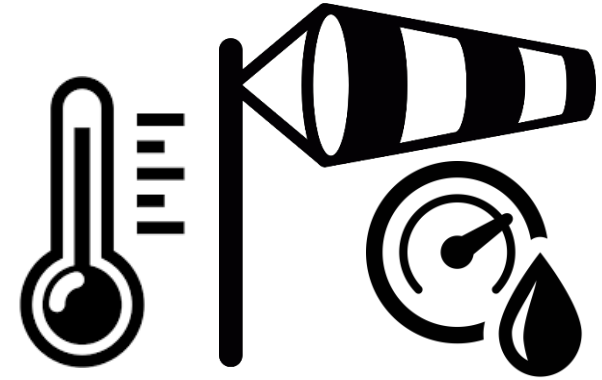
Умное производство,
предиктивное ТО



Интернет
вещей



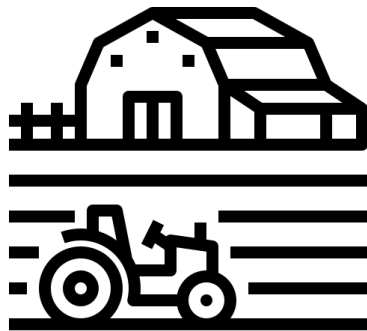
Предсказание
природных катаклизмов



Прогноз погоды,
моделирование климата



Персональная
медицина



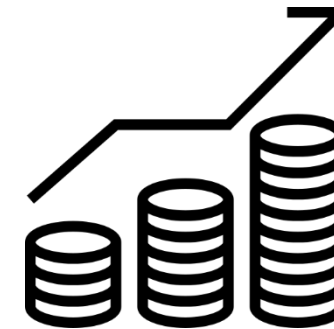
Сельское хоз-во,
животноводство



Борьба
с преступностью



Био- и хемо-
информатика

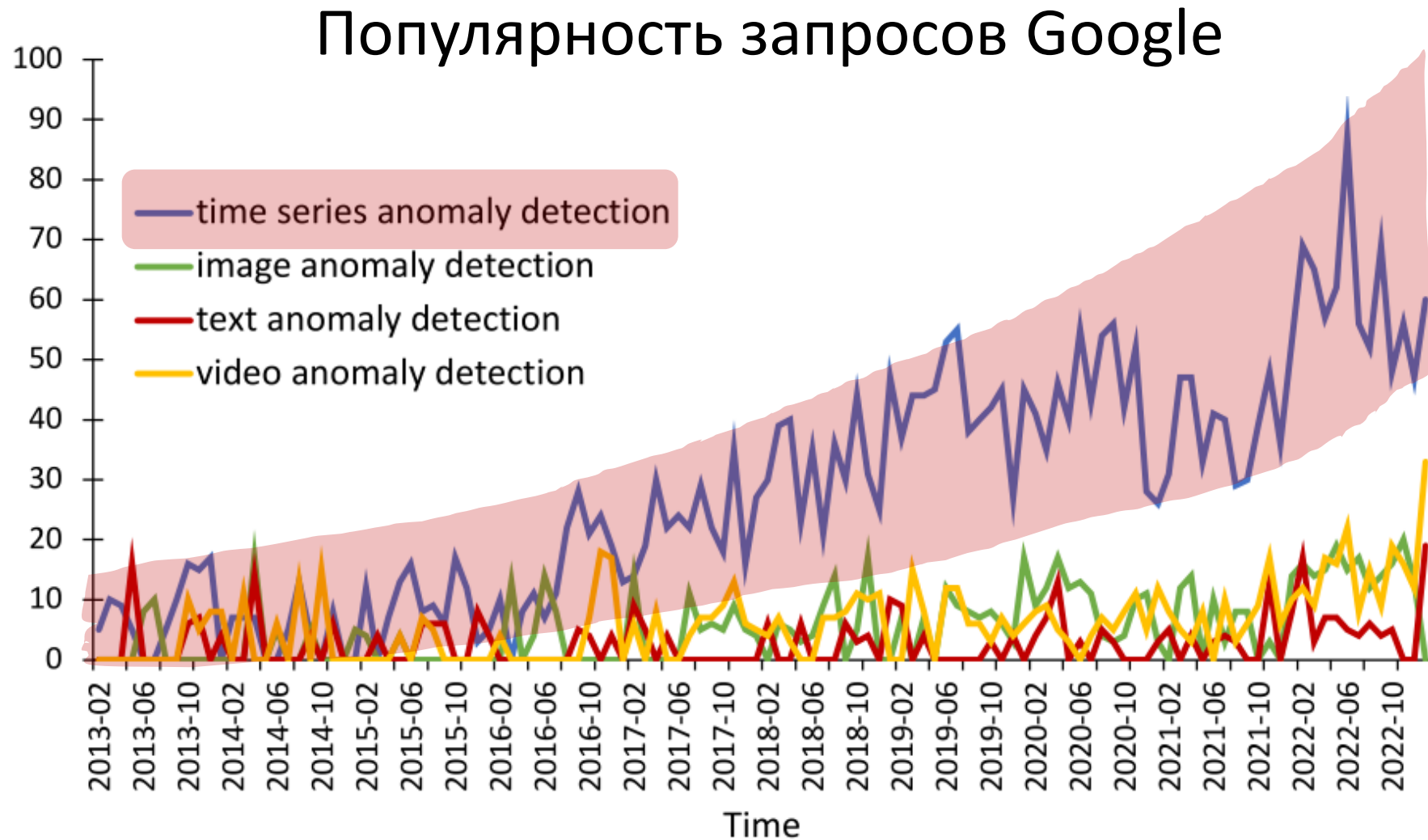


Экономика, бизнес,
финансы



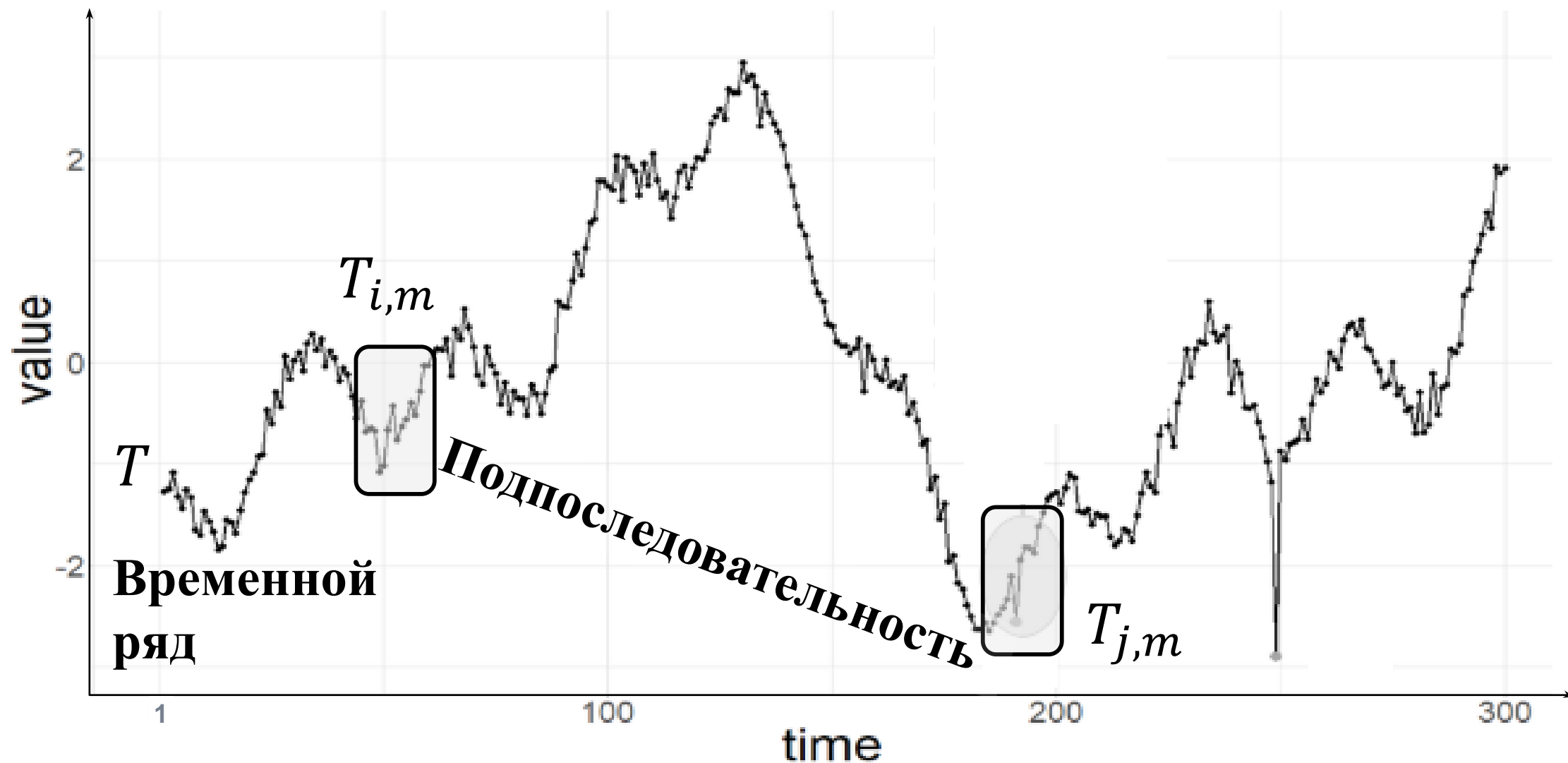
Системы
электронного обучения

Поиск аномалий во временных рядах – «горячая» тема



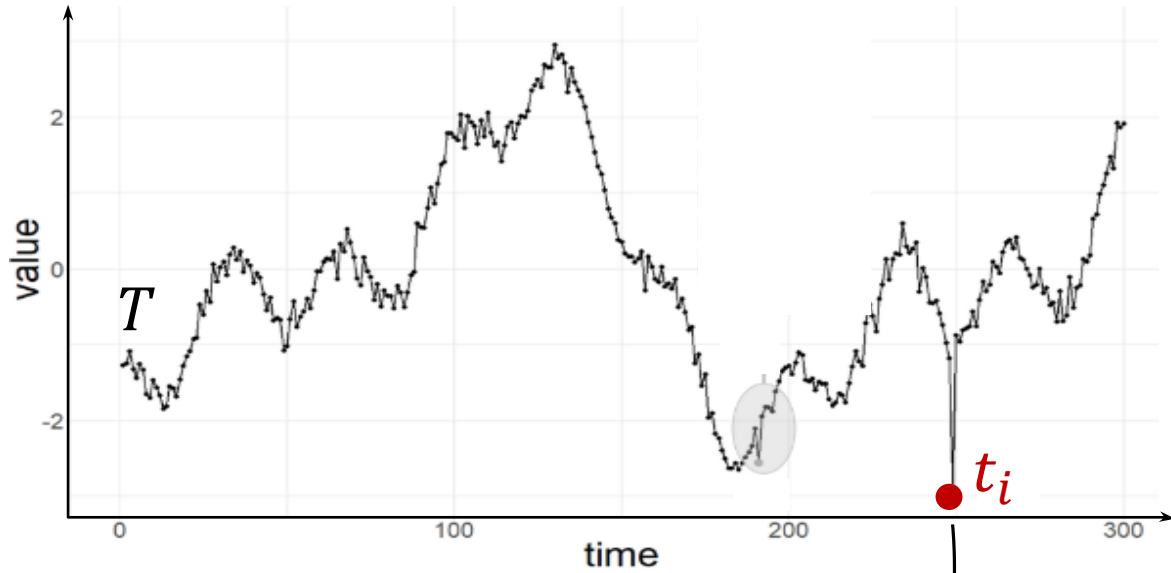
* Boniol P. *et al.* New trends in time-series anomaly detection. EDBT'2023. 847-850 (2023). DOI: [10.48786/edbt.2023.80](https://doi.org/10.48786/edbt.2023.80)

Нотация временных рядов

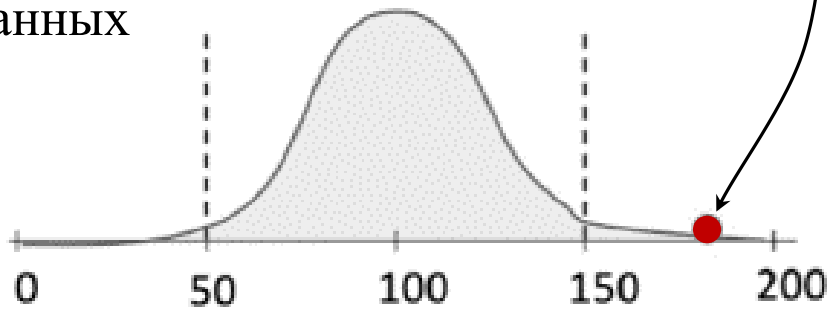


Аномалии временных рядов

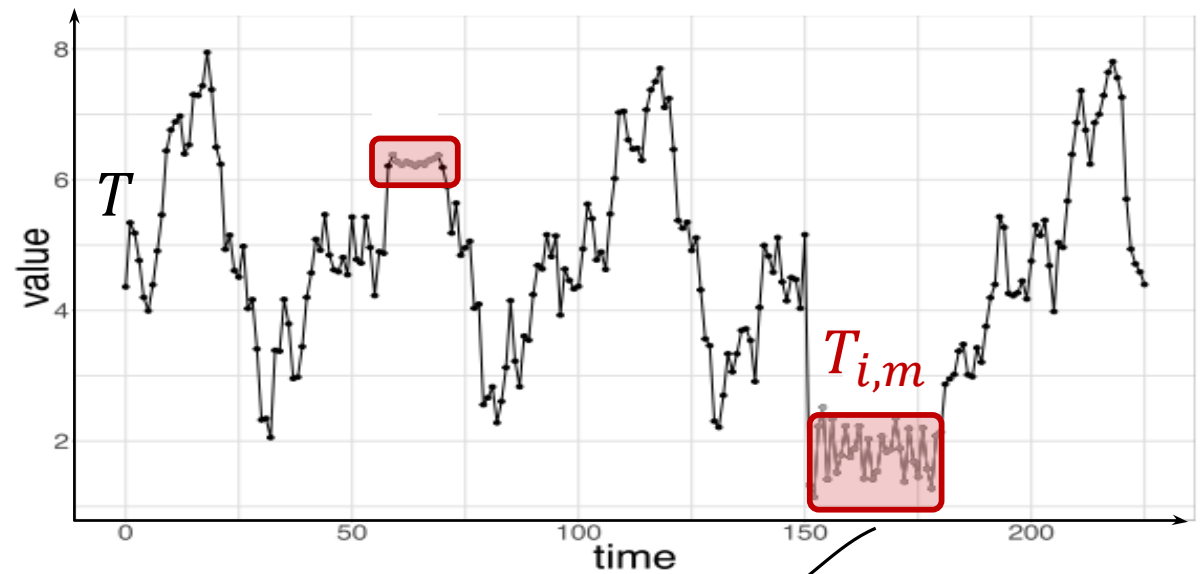
Точечная аномалия



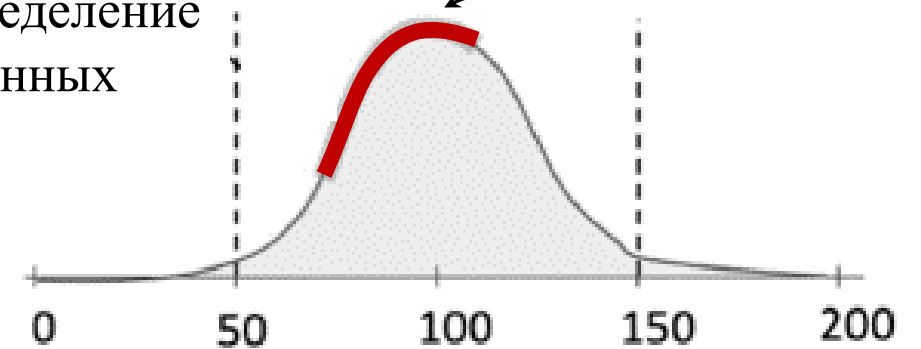
Распределение данных



Аномальная подпоследовательность



Распределение данных



Аномалия временного ряда



Homer



Marge



Bart



Selma



Patty



Barney



Quasimodo

Аномалия – наблюдение, которое настолько сильно отличается от других наблюдений, что вызывает подозрения в том, что оно было создано иным механизмом.

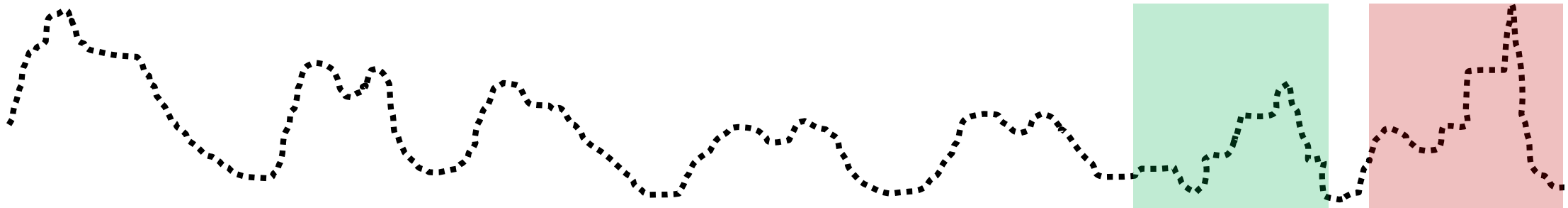
Hawkins D.M. Identification of outliers. Monographs on applied probability and statistics. Springer, 1980. DOI: [10.1007/978-94-015-3994-4](https://doi.org/10.1007/978-94-015-3994-4).



Диссонанс* – формализация аномалии

Диссонанс (discord) – подпоследовательность заданной длины с максимальным расстоянием до ближайшего соседа

Ближайший сосед – подпоследовательность, которая наиболее похожа на данную и имеет с ней не более половины общих точек

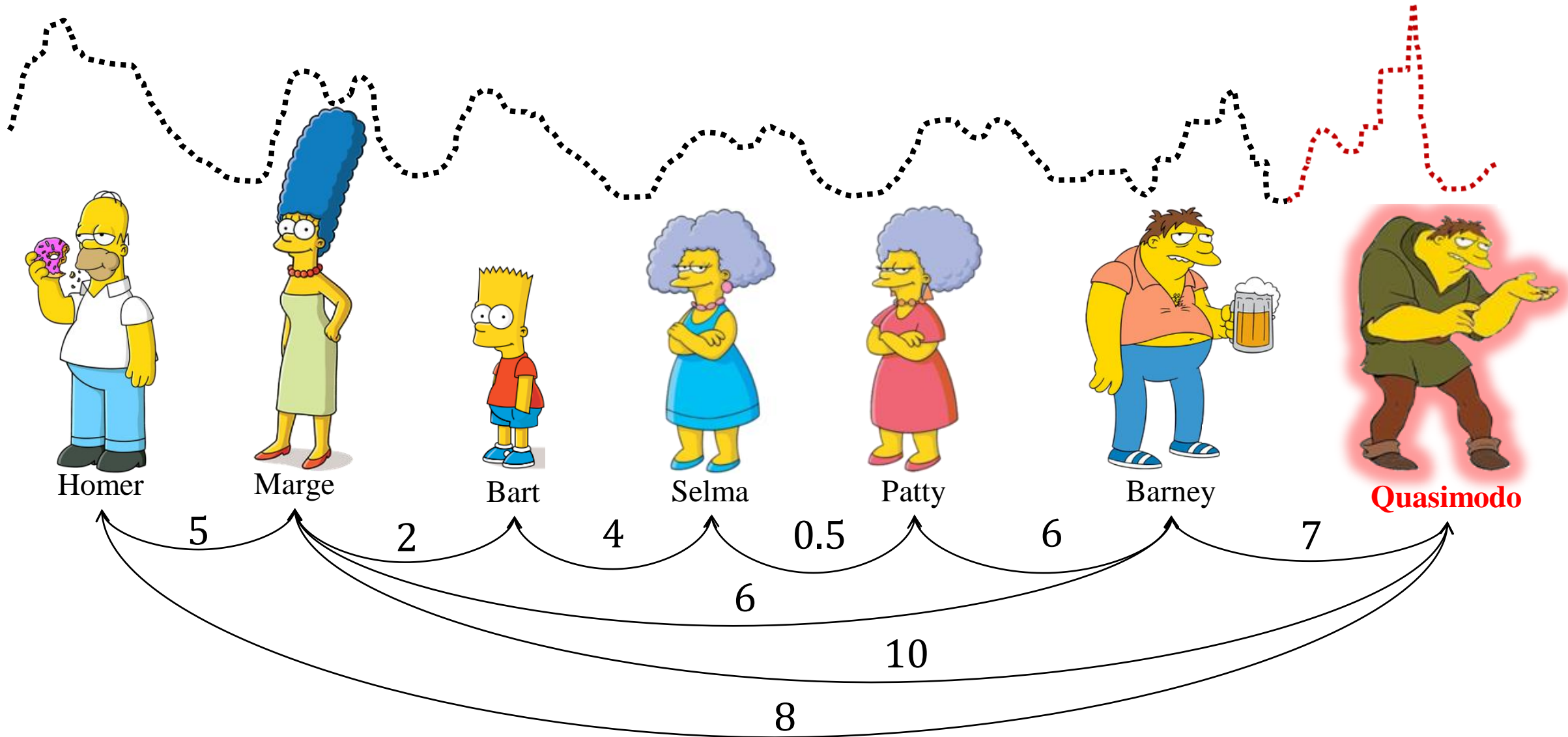


Агностическое и точное понятие:

независимость от предметной области и обучения, результат однозначен и воспроизводим

* Keogh E. *et al.* HOT SAX: Efficiently finding the most unusual time series subsequence. ICDM 2005. 226-233 (2005). DOI: [10.1109/ICDM.2005.79](https://doi.org/10.1109/ICDM.2005.79)

Пример поиска диссонанса



Пример поиска

Homer



Marge



Bart



Selma



Patty



Barney



Quasimodo



Матрица расстояний:
чем ближе соседи,
тем более они схожи

	Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
Homer	0						
Marge		0					
Bart			0				
Selma				0			
Patty					0		
Barney						0	
Quasimodo							0

Пример поиска

Homer



Marge



Bart



Selma



Patty



Barney



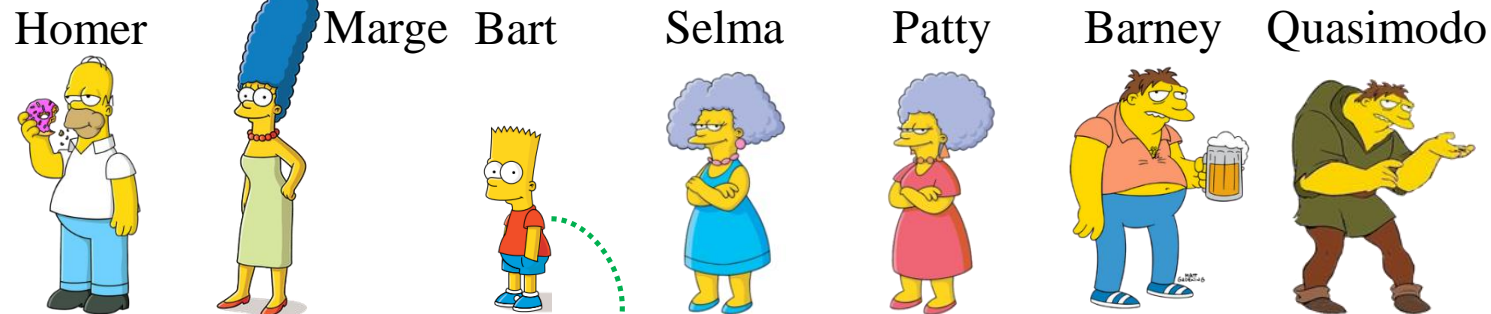
Quasimodo



Матрица расстояний
с вычисленными
расстояниями

	Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
Homer	0	5	2	4	4	6	8
Marge	5	0	2.5	3	3	6	10
Bart	2	2.5	0	4	4	6	9
Selma	4	3	4	0	0.5	5	8
Patty	4	3	4	0.5	0	5	8
Barney	6	6	6	5	5	0	7
Quasimodo	8	10	9	8	8	7	0

Пример поиска



	Homer	Marge	Bart	Selma	Patty	Barney	Quasimodo
Homer	0	5	2	4	4	6	8
Marge	5	0	2.5	3	3	6	10
Bart	2	2.5	0	4	4	6	9
Selma	4	3	4	0	0.5	5	8
Patty	4	3	4	0.5	0	5	8
Barney	6	6	6	5	5	0	7
Quasimodo	8	10	9	8	8	7	0

Матрица расстояний
с расстояниями
до ближайших соседей
(т.е. минимумы
по столбцам)

Пример поиска

Homer



Marge



Bart



Selma



Patty



Barney



Quasimodo



Матрица расстояний
с **наибольшим**
расстоянием
до ближайшего соседа
(т.е. максимум
минимумов по столбцам)

0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	7
8	10	9	8	8	7	0

Пример поиска

Homer



Marge



Bart



Selma



Patty



Barney



Quasimodo

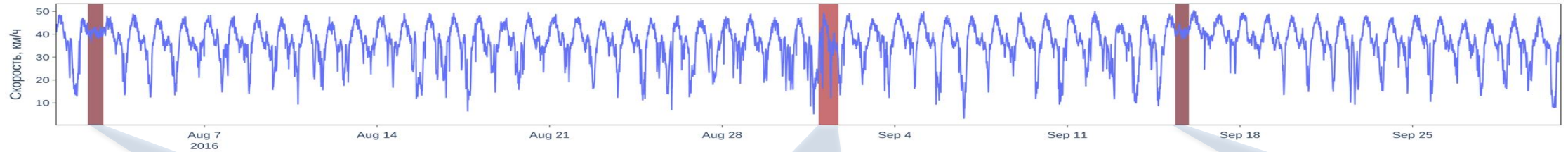


0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	7
8	10	9	8	8	7	0

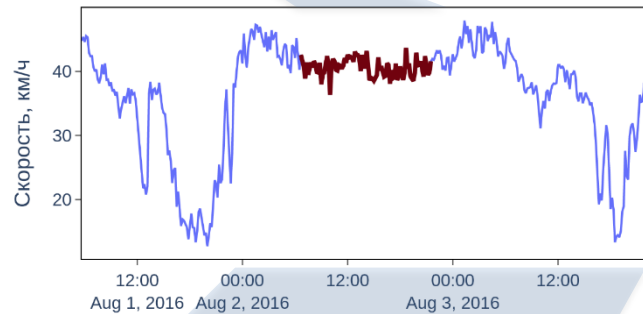
Диссонанс – объект
с самым далеким
ближайшим соседом
(т.е. аргумент
максимума минимумов
по столбцам)

Диссонанс отражает аномалию в реальной жизни, ...

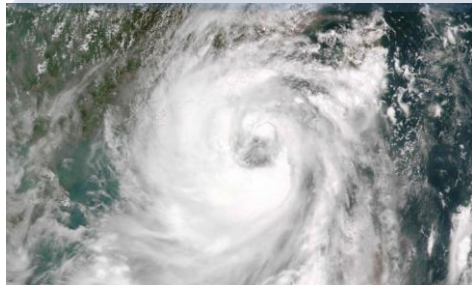
Средняя скорость городского трафика в Гуанчжоу, Китай*



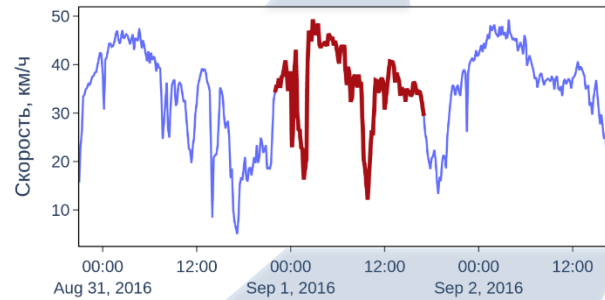
Топ-2 диссонанс



Тайфун Нида



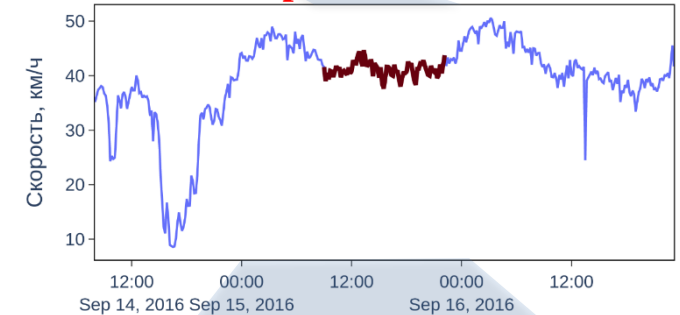
Топ-3 диссонанс



День Победы над Японией



Топ-1 диссонанс

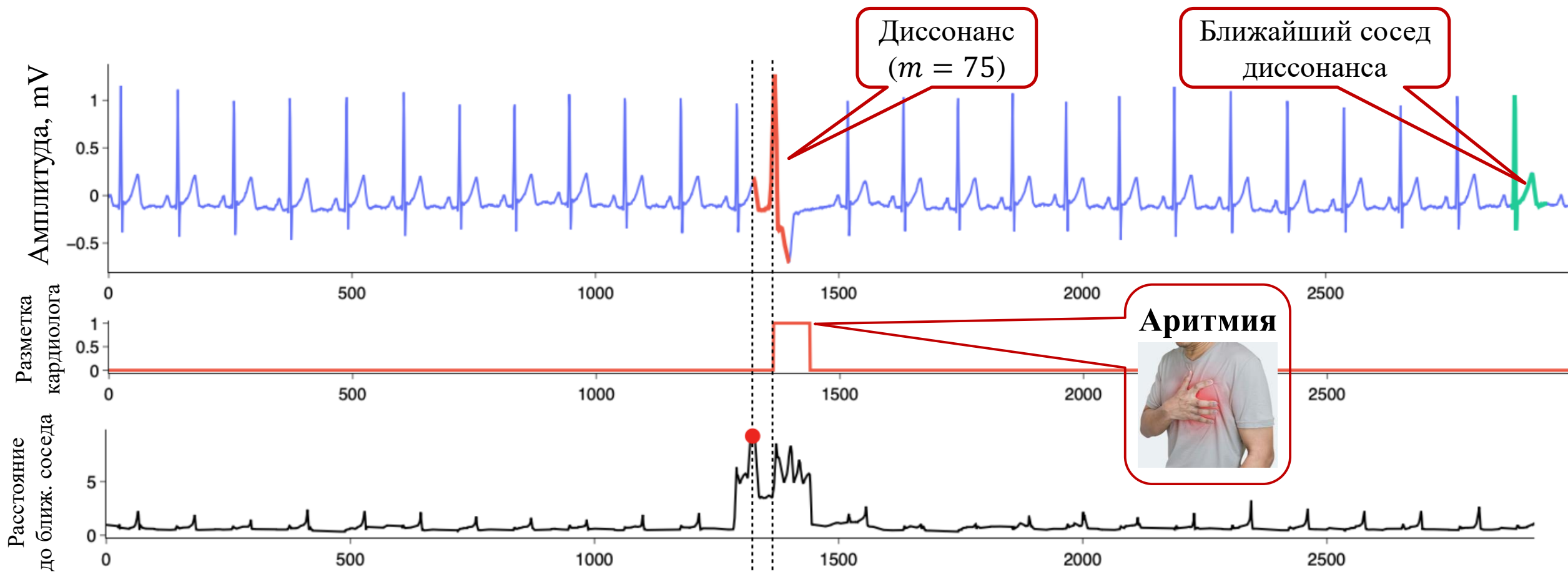


Фестиваль Луны



* Chen X, Chen Y, He Z. Urban traffic speed dataset of Guangzhou, China. 2018. DOI: [10.5281/zenodo.1205229](https://doi.org/10.5281/zenodo.1205229).

...., но диссонанс не идентичен аномалии



Содержание

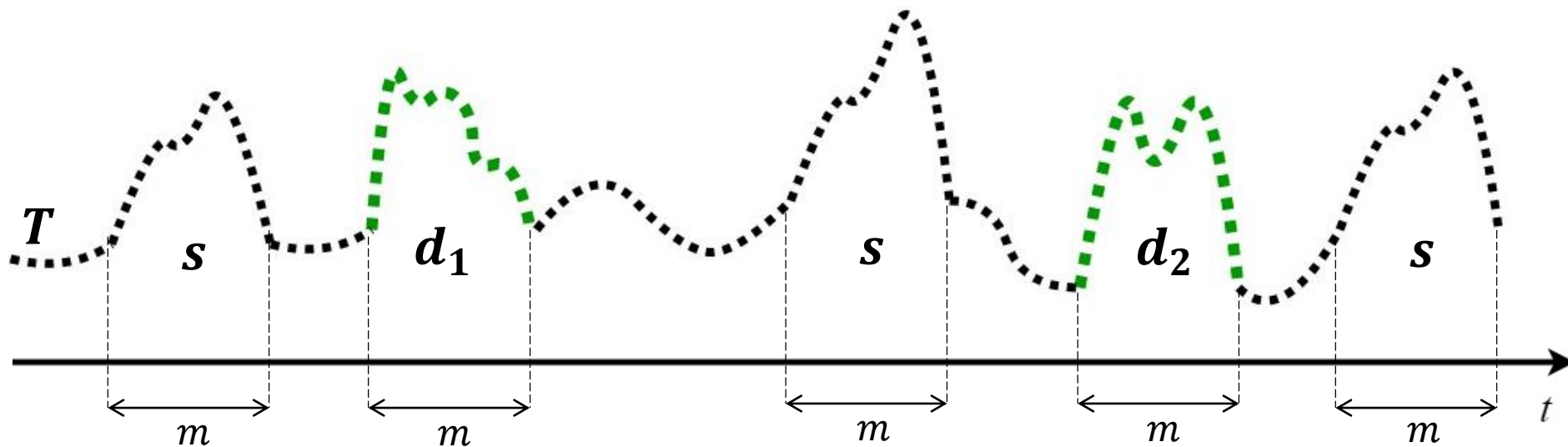
- **Часть 1. Введение в задачу поиска аномалий (30 мин.)**
 - Теория
 - **Практика**
 - Подготовка среды выполнения
 - Загрузка и визуализация ряда NYC Taxi
 - Поиск диссонансов
- **Часть 2. Поиск диссонансов фиксированной длины (30 мин.)**
- **Часть 3. Поиск диссонансов произвольной длины (30 мин.)**

Содержание

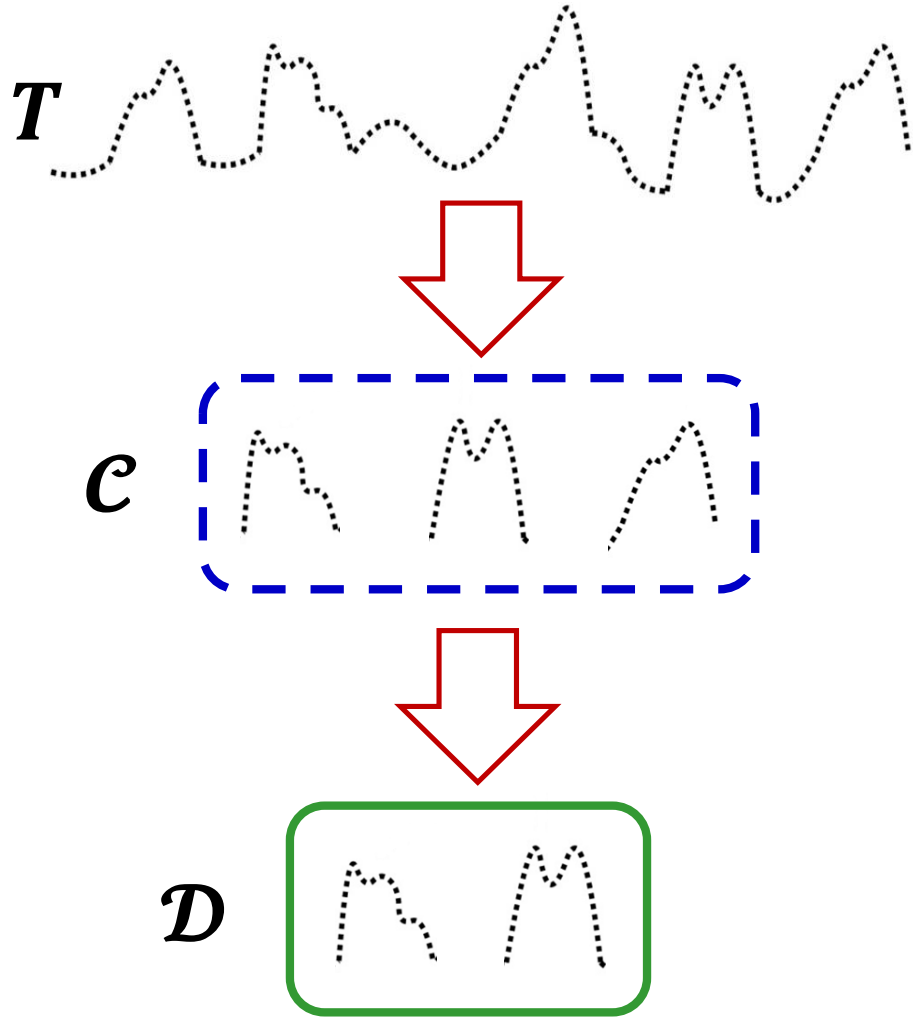
- Часть 1. Введение в задачу поиска аномалий (30 мин.)
- **Часть 2. Поиск диссонансов фиксированной длины (30 мин.)**
 - Теория
 - **Понятие диапазонного диссонанса**
 - **Последовательный алгоритм DRAG**
 - **Параллельный алгоритм PD3**
 - Практика
- Часть 3. Поиск диссонансов произвольной длины (30 мин.)

Диапазонный диссонанс (range discord)

- *Диапазонный диссонанс* – подпоследовательность ряда, расстояние от которой до ее ближайшего соседа не ниже заданного порога
- Дано: ряд T , длина диссонанса m , порог r
- Найти: $\mathcal{D} = \{d_1, d_2, \dots\}$ $d_i \in \mathcal{D} \Leftrightarrow \min_{s \in M_{d_i}} \text{Dist}(d_i, s) \geq r$



Алгоритм DRAG (Discord Range Aware Gathering)



1. Отбор

За одно сканирование ряда сформировать **множество кандидатов** в диссонансы

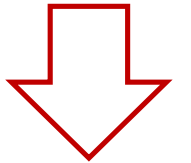
2. Очистка

За одно сканирование ряда **отбросить кандидатов**, которые являются **ложными диссонансами**

DRAG: Отбор кандидатов

```
 $\mathcal{C} := \{T_{1,m}\}$   
while not end of  $T$   
  get next subsequence  $s$   
   $isCandidate := \text{TRUE}$   
  for each  $c_i \in \mathcal{C}$  and  $s \cap c_i = \emptyset$   
    if  $\text{Dist}(s, c_i) < r$  then  
       $\mathcal{C} := \mathcal{C} \setminus c_i$ ;  $isCandidate := \text{FALSE}$   
  if  $isCandidate = \text{TRUE}$  then  $\mathcal{C} := \mathcal{C} \cup s$ 
```

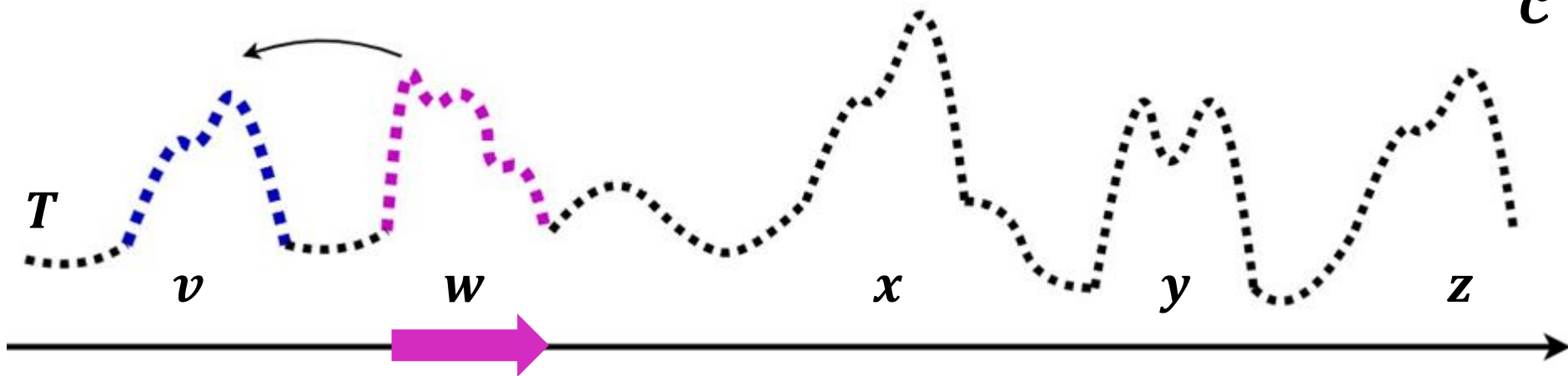
$\mathcal{C} = \{v\}$



$\text{Dist}(w, v) \geq r$



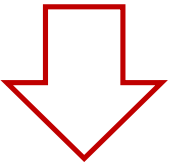
$\mathcal{C} = \{v, w\}$



DRAG: Отбор кандидатов

```
 $\mathcal{C} := \{T_{1,m}\}$   
while not end of  $T$   
  get next subsequence  $s$   
   $isCandidate := \text{TRUE}$   
  for each  $c_i \in \mathcal{C}$  and  $s \cap c_i = \emptyset$   
    if  $\text{Dist}(s, c_i) < r$  then  
       $\mathcal{C} := \mathcal{C} \setminus c_i$ ;  $isCandidate := \text{FALSE}$   
  if  $isCandidate = \text{TRUE}$  then  $\mathcal{C} := \mathcal{C} \cup s$ 
```

$$\mathcal{C} = \{v, w\}$$

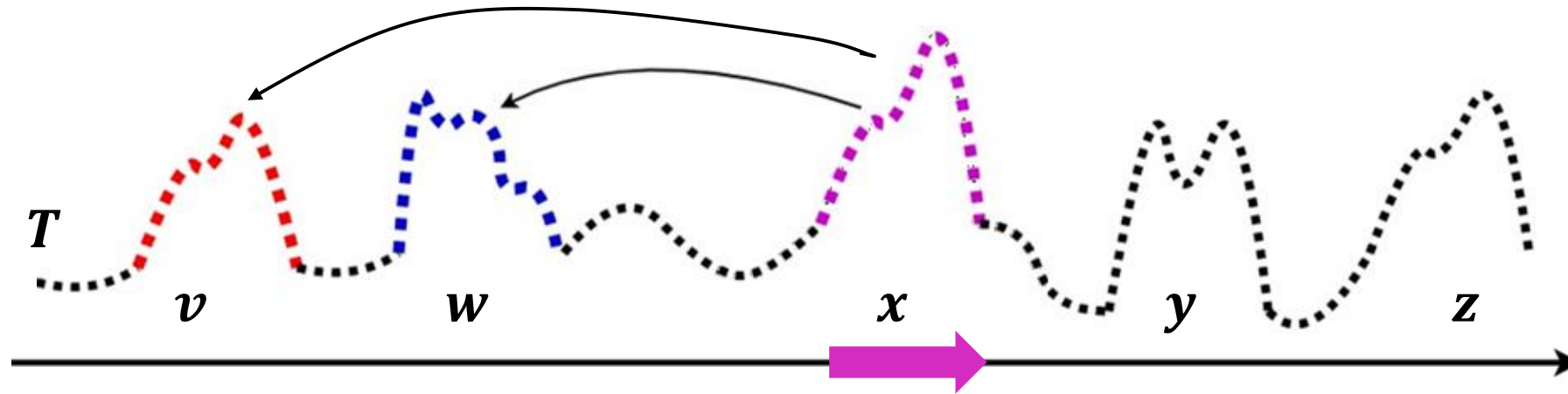


$$\text{Dist}(x, v) < r$$

$$\text{Dist}(x, w) \geq r$$



$$\mathcal{C} = \{w\}$$

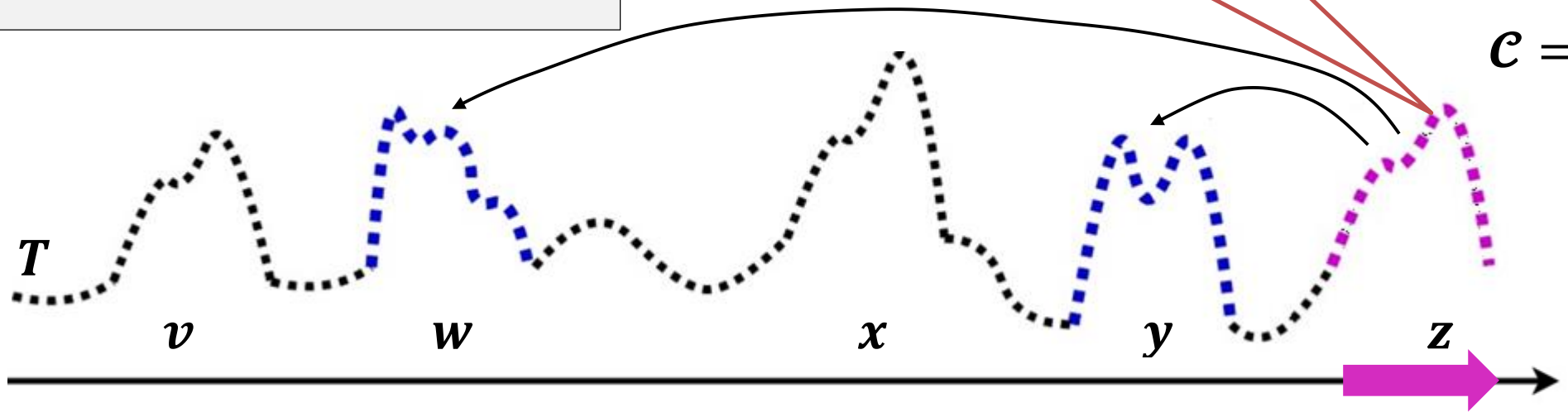


DRAG: Отбор кандидатов

```
 $\mathcal{C} := \{T_{1,m}\}$   
while not end of  $T$   
  get next subsequence  $s$   
   $isCandidate := \text{TRUE}$   
  for each  $c_i \in \mathcal{C}$  and  $s \cap c_i = \emptyset$   
    if  $\text{Dist}(s, c_i) < r$  then  
       $\mathcal{C} := \mathcal{C} \setminus c_i$ ;  $isCandidate := \text{FALSE}$   
  if  $isCandidate = \text{TRUE}$  then  $\mathcal{C} := \mathcal{C} \cup s$ 
```

z – ЛОЖНЫЙ
ДИССОНАНС, т.к.
 $\text{Dist}(z, v) < r$
 $\text{Dist}(z, x) < r$.
Но v и x
были удалены!

$\mathcal{C} = \{w, y\}$
↓
 $\text{Dist}(z, w) \geq r$
 $\text{Dist}(z, y) \geq r$
↓
 $\mathcal{C} = \{w, y, z\}$



DRAG: Очистка кандидатов

```
 $\mathcal{D} := \mathcal{C}$   
while not end of  $T$   
  get next subsequence  $s$   
  for each  $d_i \in \mathcal{D}$  and  $s \cap d_i = \emptyset$   
    if  $\text{Dist}(s, d_i) < r$  then  
       $\mathcal{D} := \mathcal{D} \setminus d_i$ 
```

$$\mathcal{D} = \{w, y, z\}$$



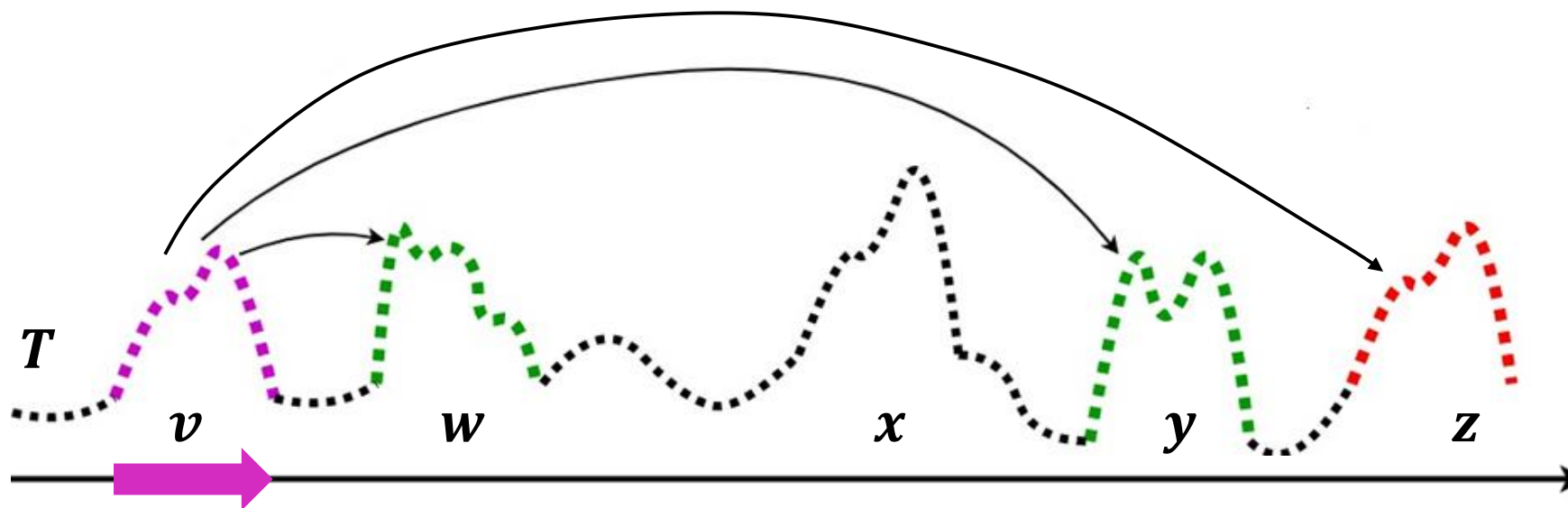
$$\text{Dist}(v, w) \geq r$$

$$\text{Dist}(v, y) \geq r$$

$$\text{Dist}(v, z) < r$$



$$\mathcal{D} = \{w, y\}$$



DRAG: Очистка кандидатов

```
 $\mathcal{D} := \mathcal{C}$   
while not end of  $T$   
  get next subsequence  $s$   
  for each  $d_i \in \mathcal{D}$  and  $s \cap d_i = \emptyset$   
    if  $\text{Dist}(s, d_i) < r$  then  
       $\mathcal{D} := \mathcal{D} \setminus d_i$ 
```

$\mathcal{D} = \{w, y\}$

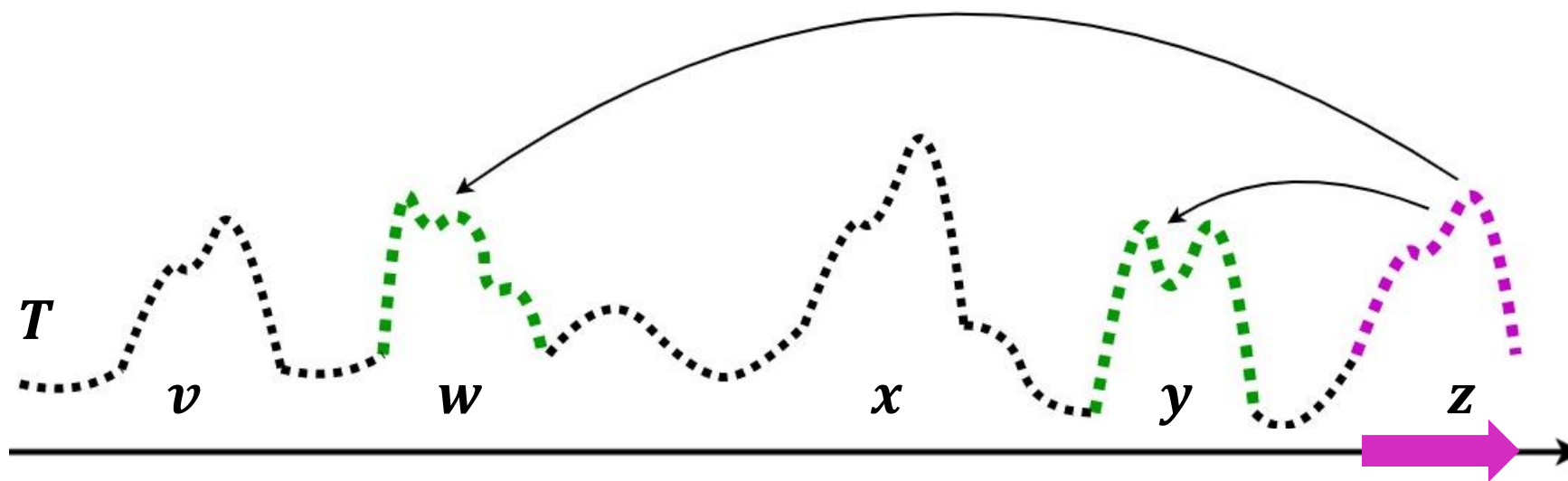


$\text{Dist}(z, w) \geq r$

$\text{Dist}(z, y) \geq r$



$\mathcal{D} = \{w, y\}$

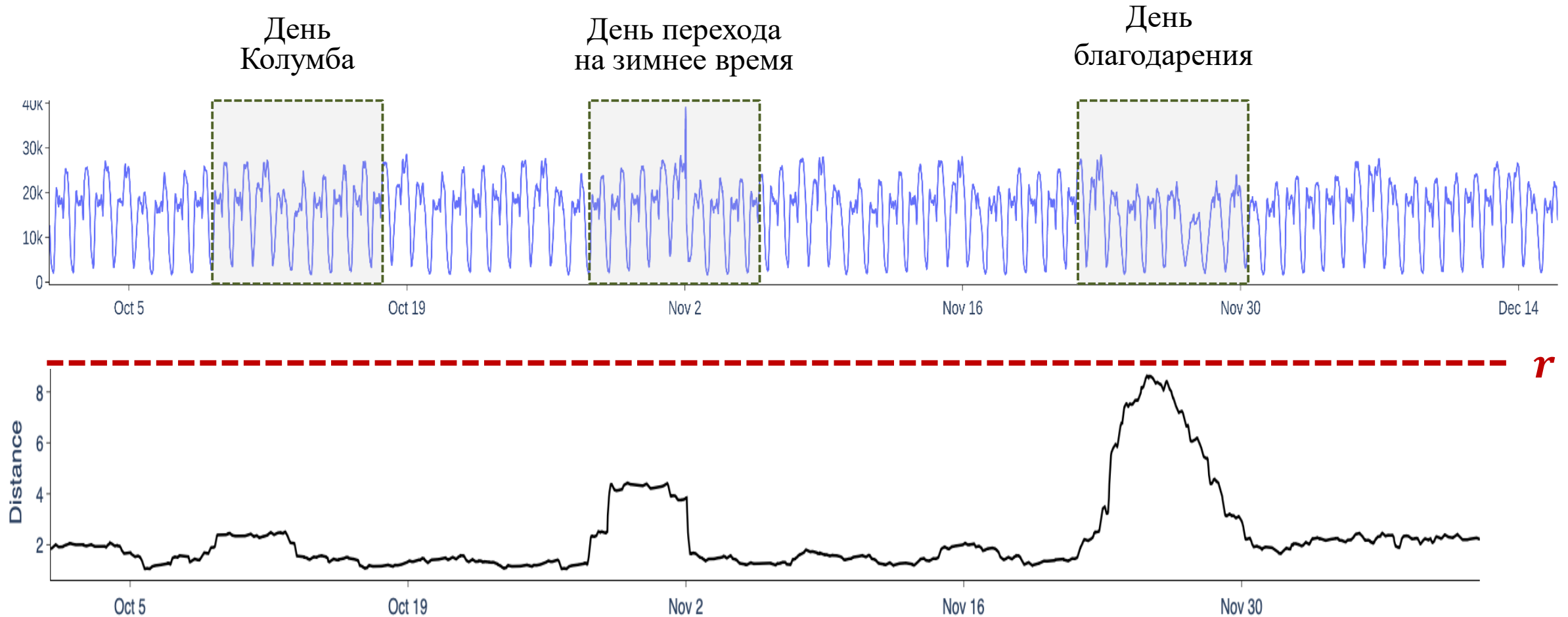


Что не так с DRAG: Ручной подбор порога r

Слишком большой порог – нет диссонансов,
слишком маленький порог – много ложных диссонансов

Ручной подбор порога: $r \rightarrow +\infty \Rightarrow$ нет диссонансов

Среднее число пассажиров NY такси осенью 2014 г.



Ручной подбор порога: $r \rightarrow 0 \Rightarrow$ ложные аномалии

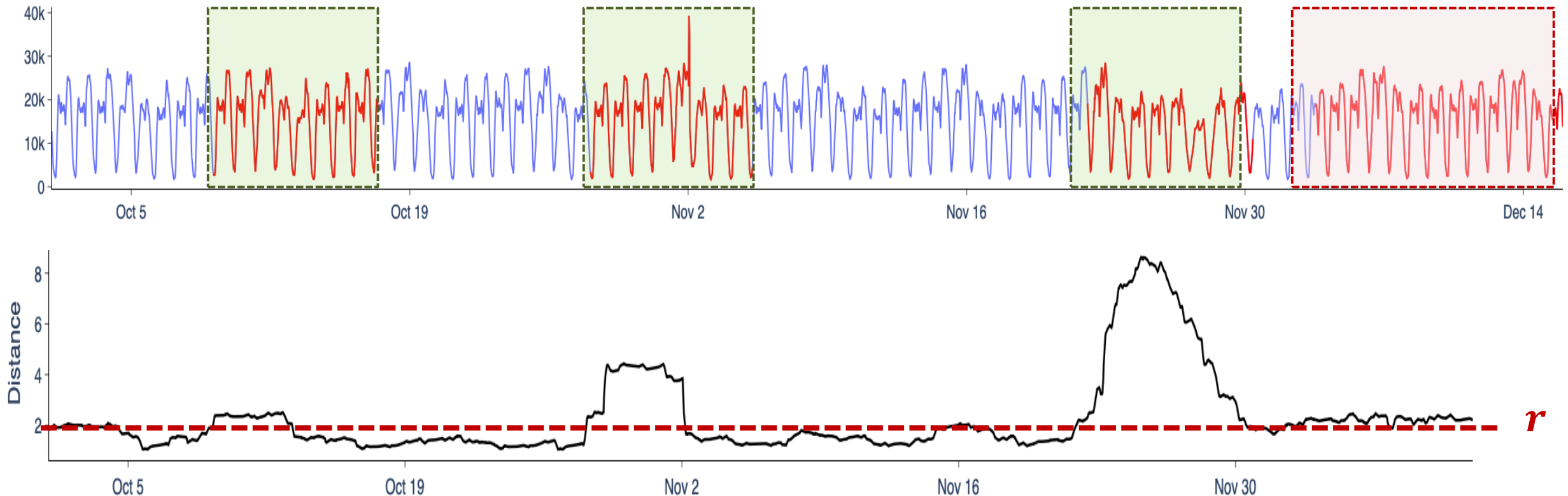
Среднее число пассажиров NY такси осенью 2014 г.

День Колумба

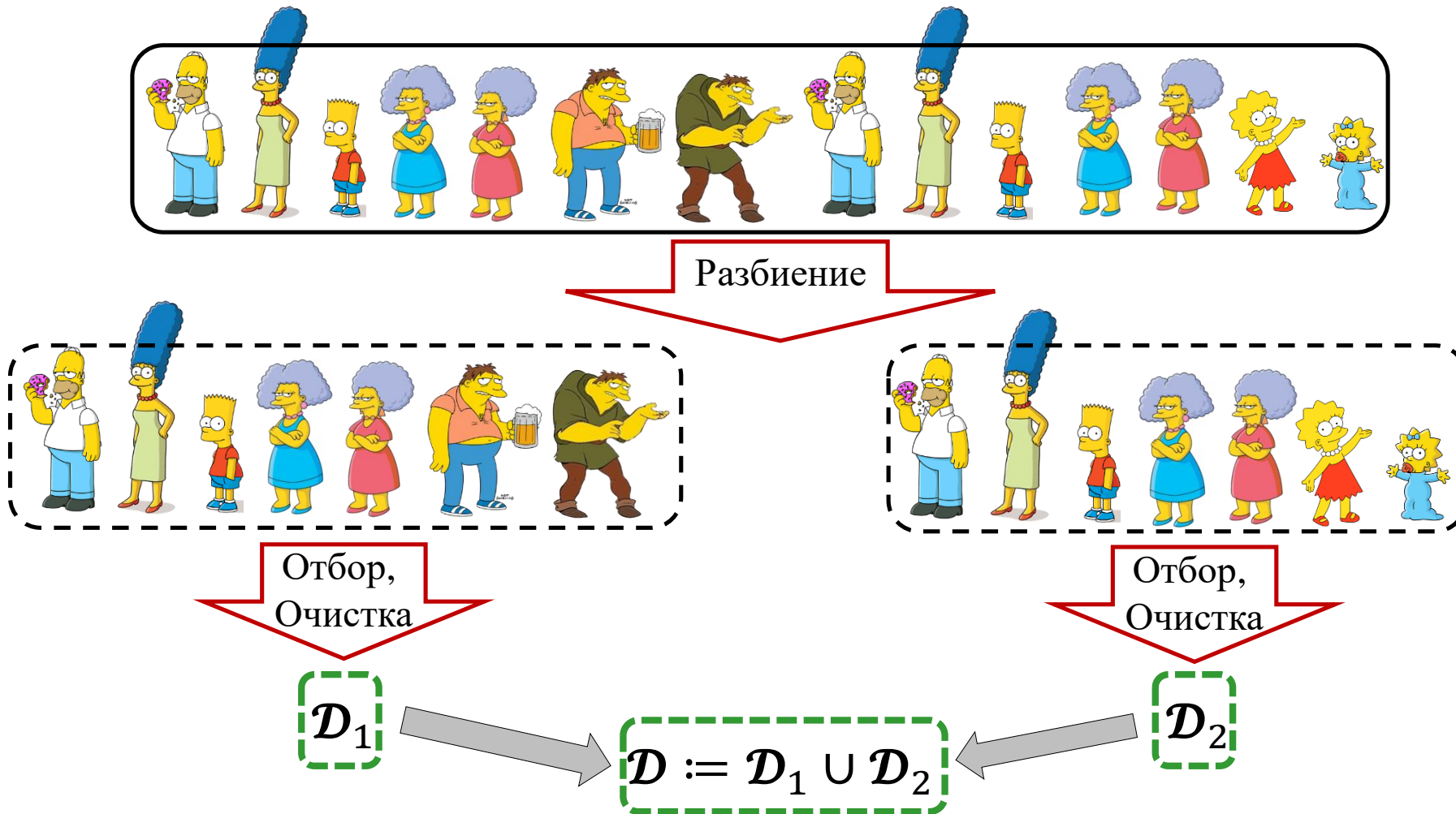
День перехода на зимнее время

День благодарения

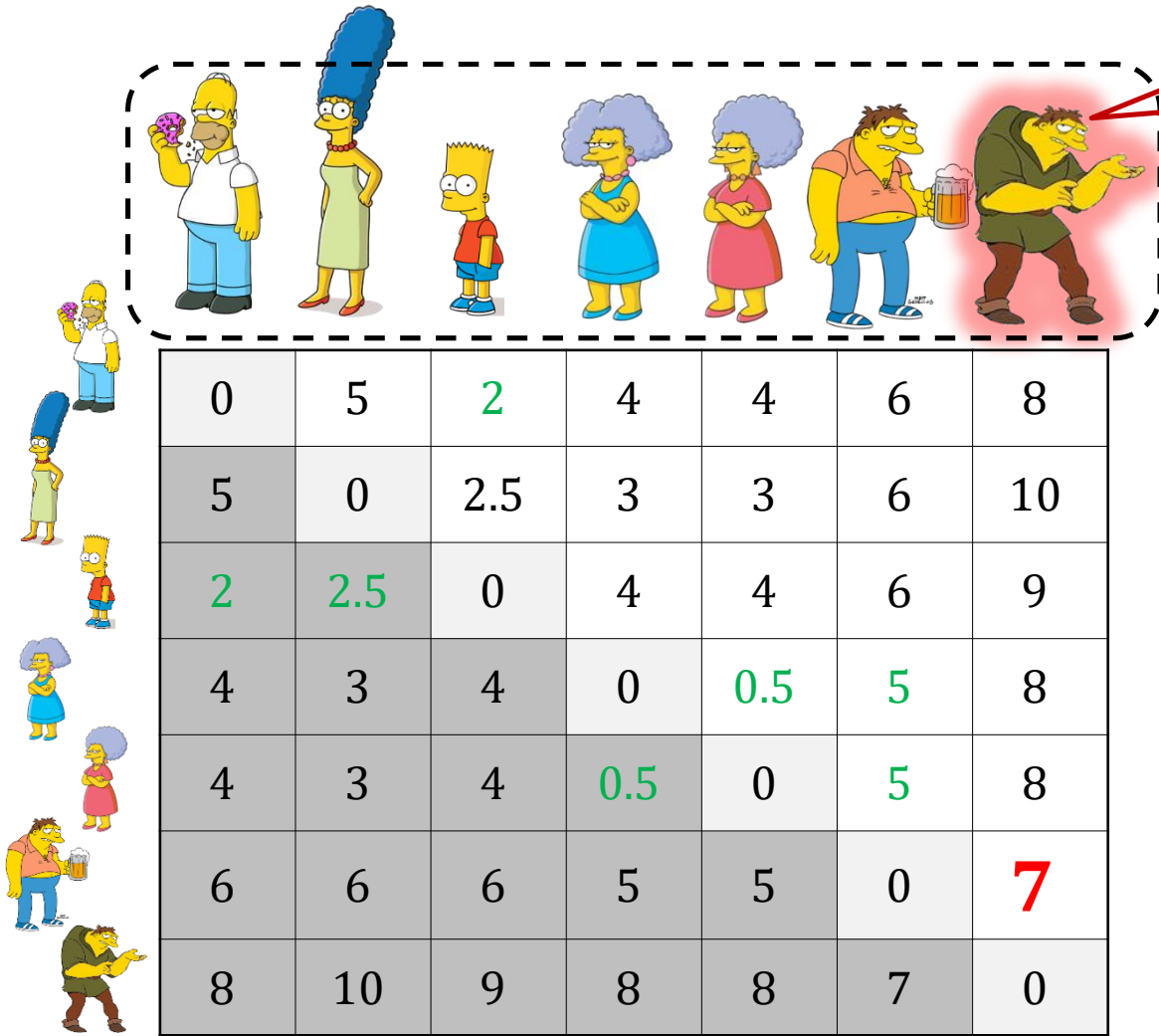
Ложная аномалия



Наивное распараллеливание ...

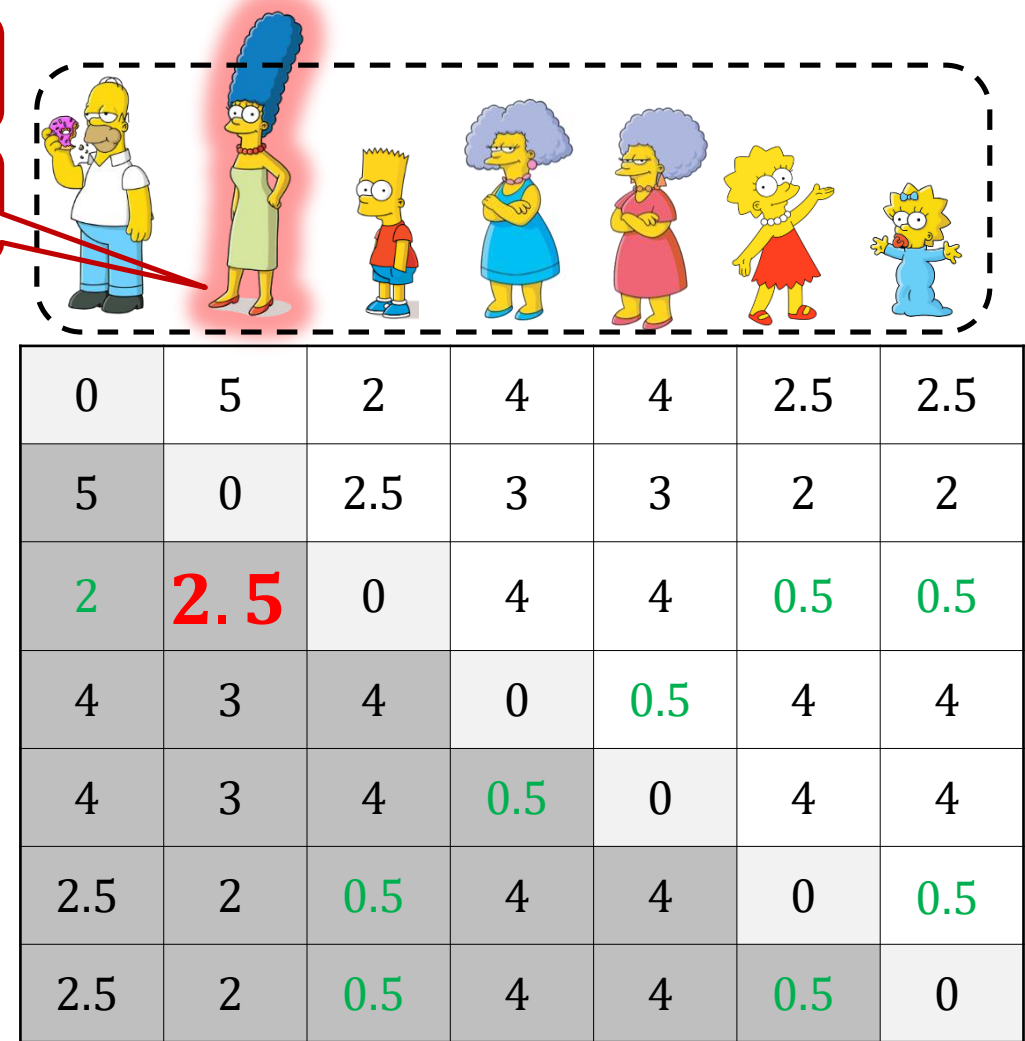


Наивное распараллеливание ...

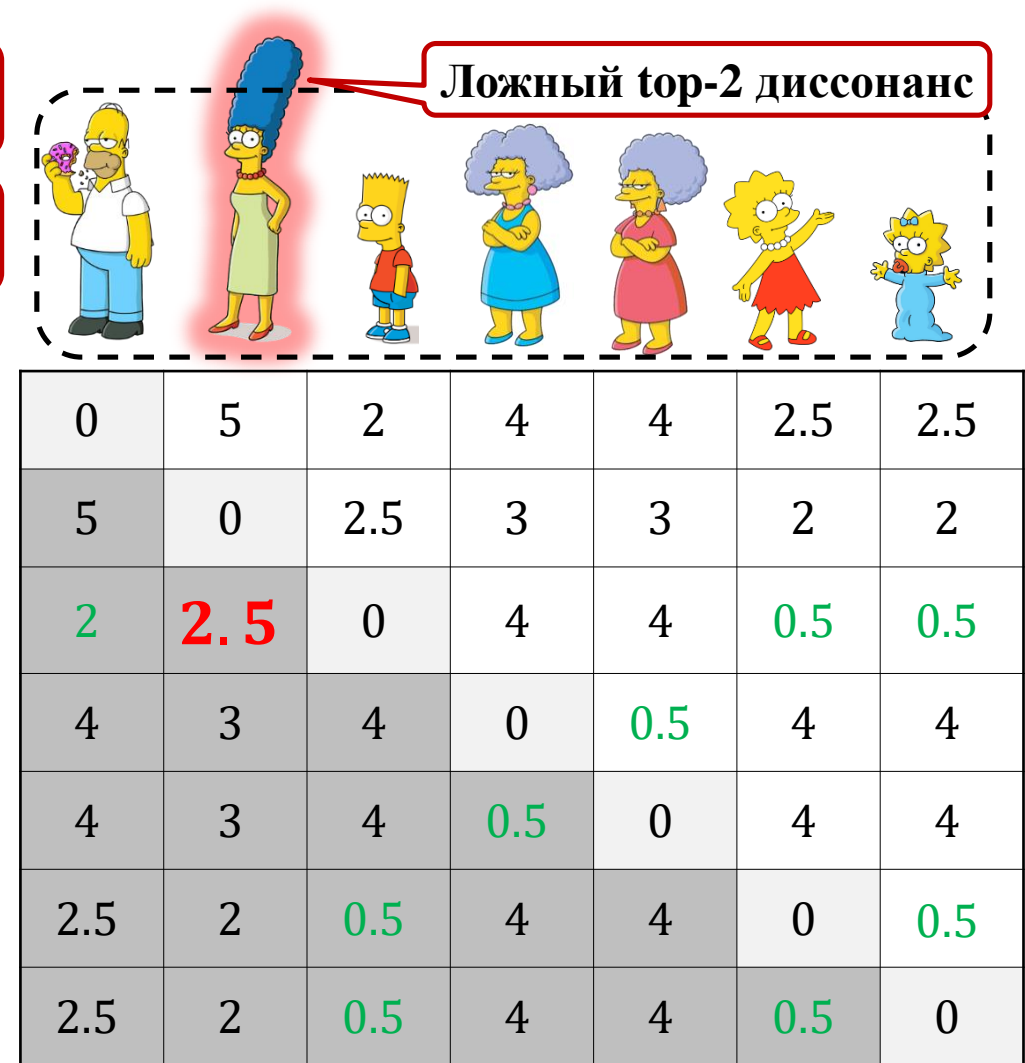
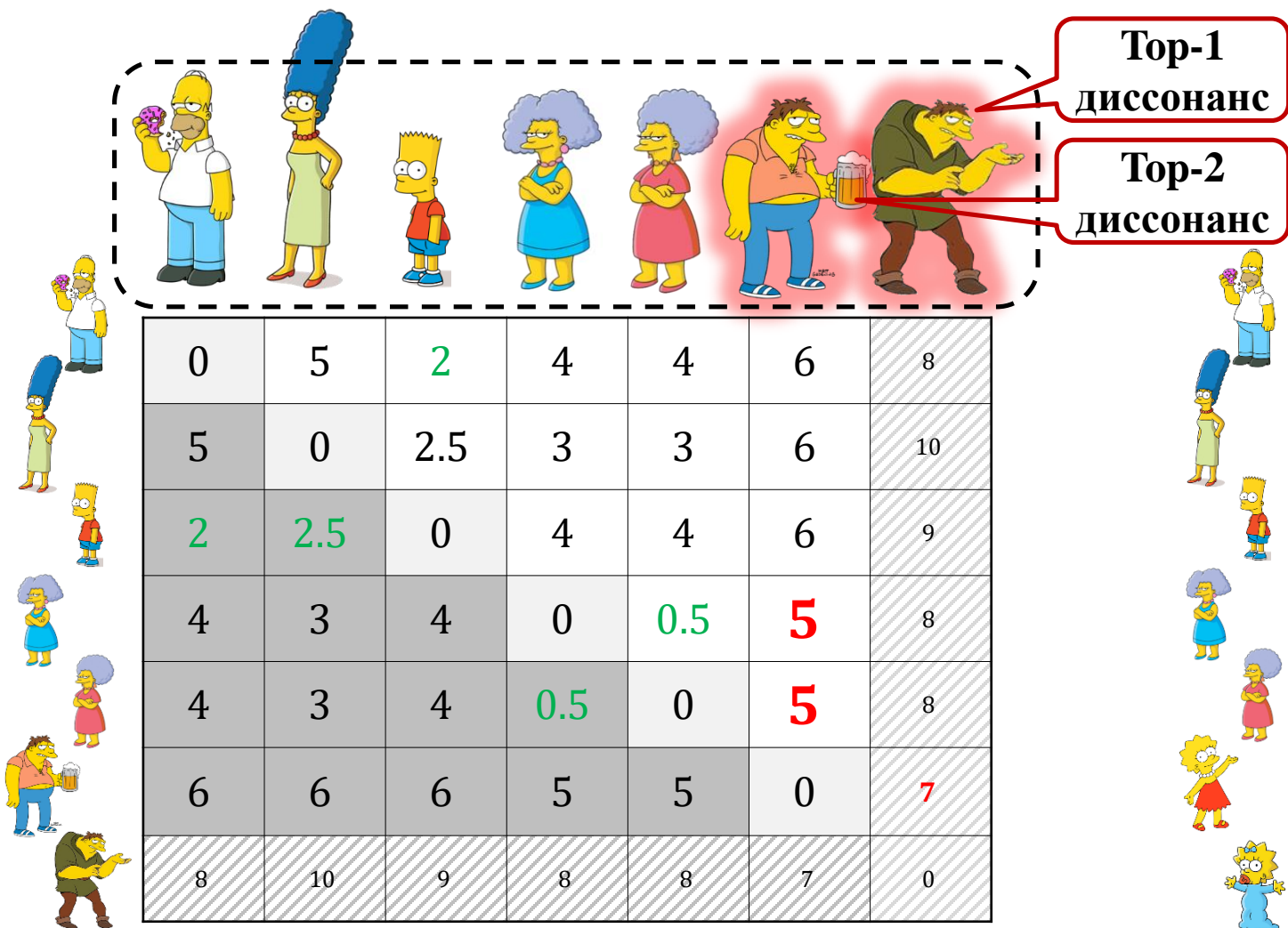


Топ-1
диссонанс

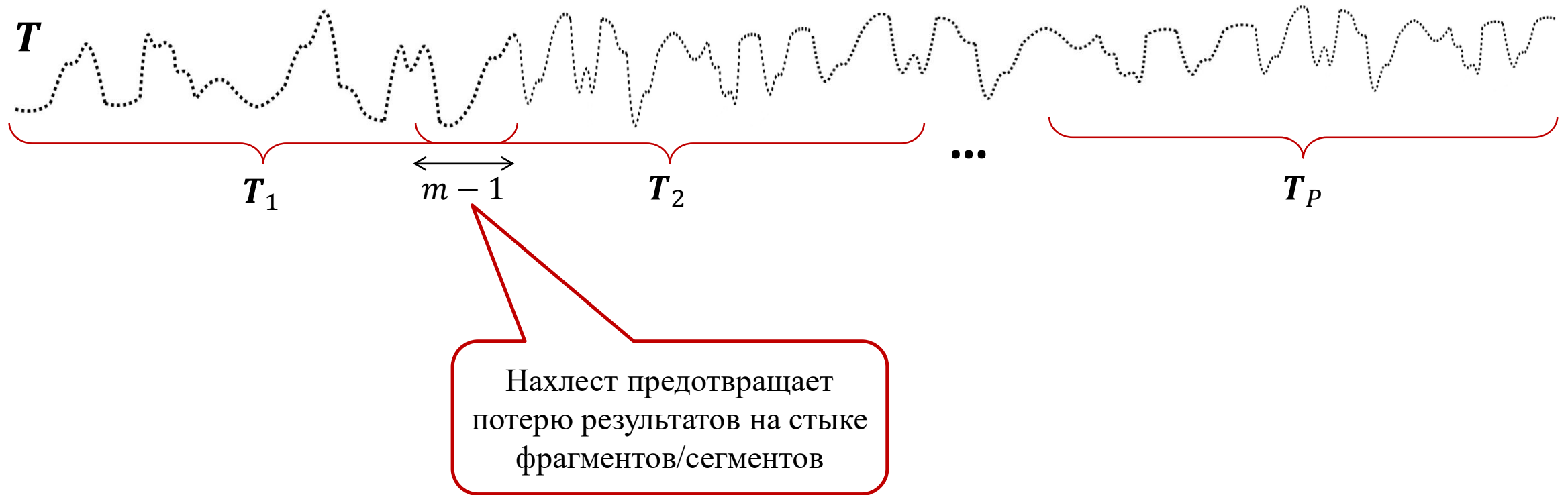
Топ-2
диссонанс



... не работает

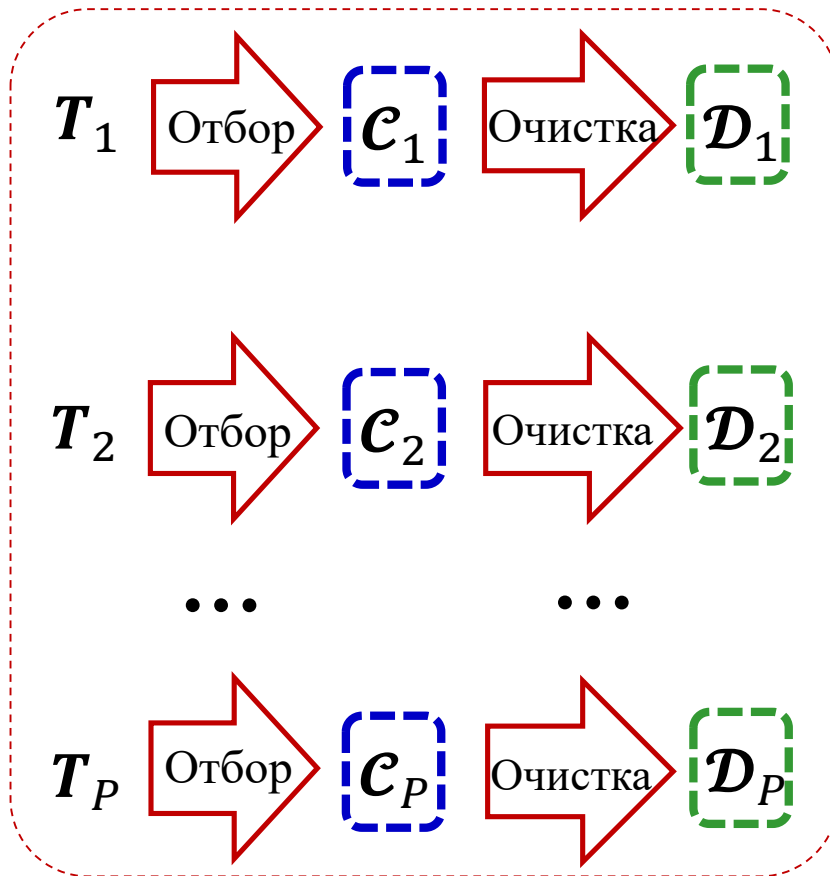


PD3 (Parallel DRAG-based Discord Discovery): Сегментация ряда



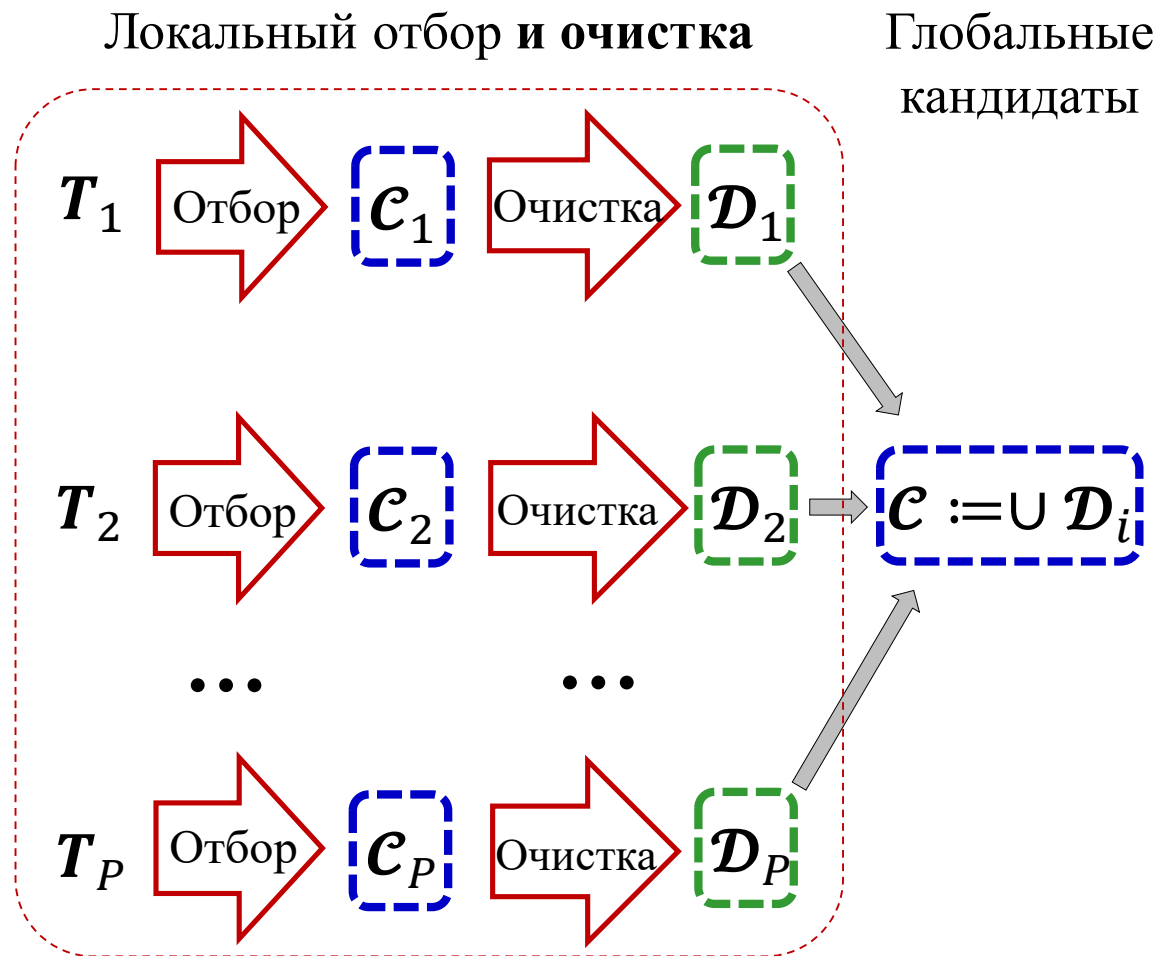
Алгоритм PD3: Распараллеливание

Локальный отбор **и очистка**

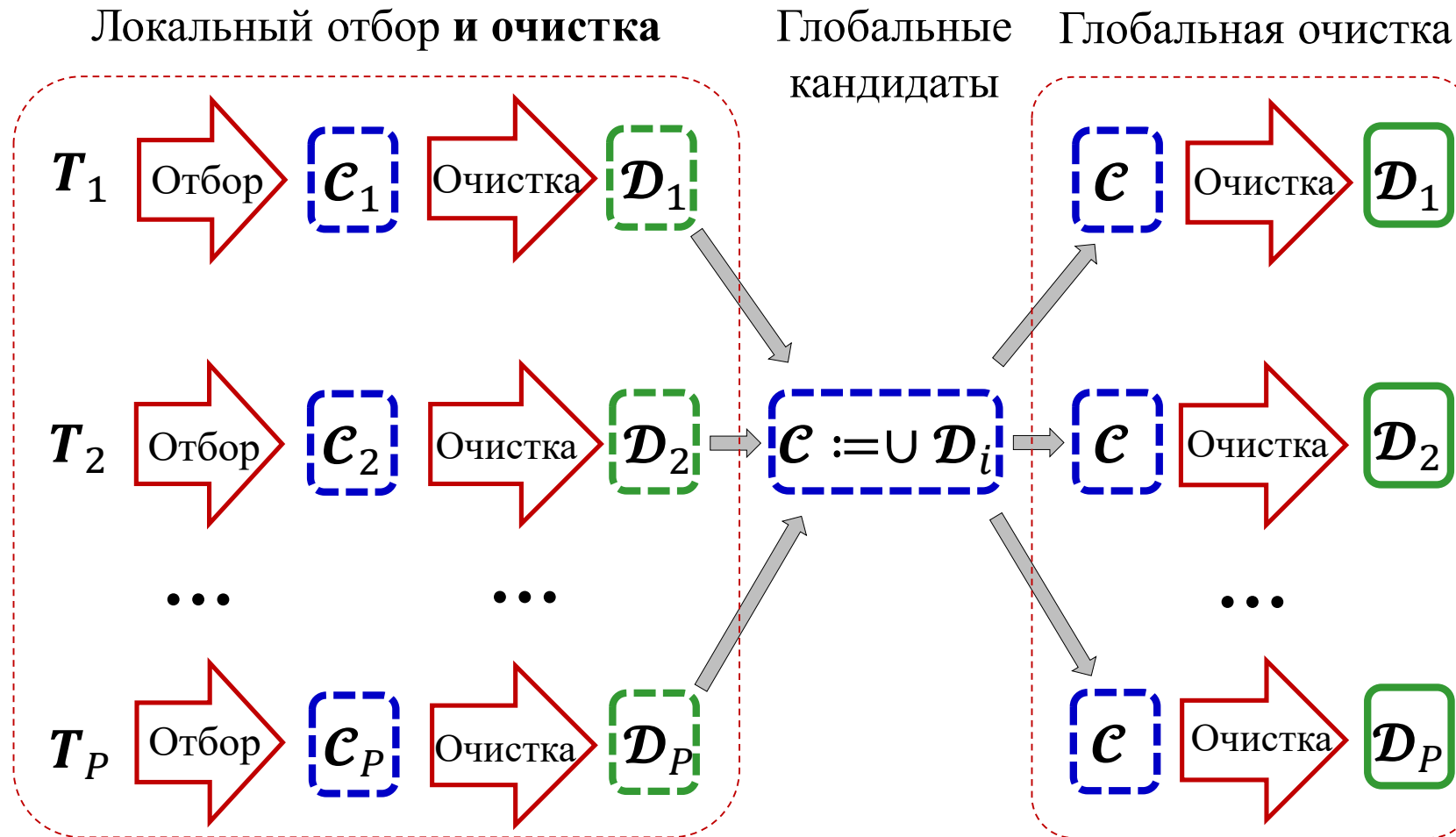


Локальный кандидат не может быть глобальным диссонансом, если он отброшен при очистке хотя бы одного фрагмента/сегмента

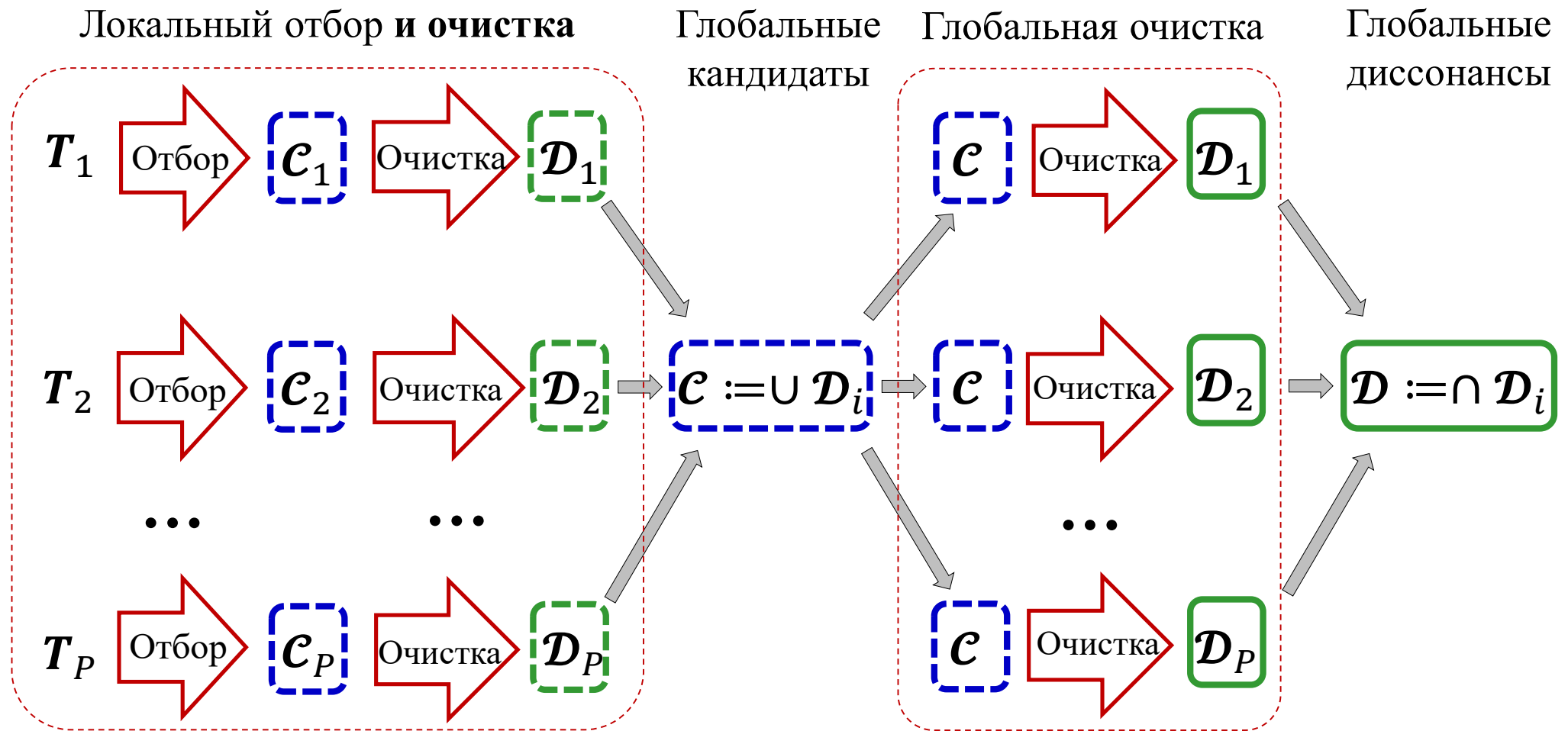
Алгоритм PD3: Распараллеливание



Алгоритм PD3: Распараллеливание



Алгоритм PD3: Распараллеливание



Содержание

- Часть 1. Введение в задачу поиска аномалий (30 мин.)
- **Часть 2. Поиск диссонансов фиксированной длины (30 мин.)**
 - Теория
 - Практика
 - Поиск диссонансов с помощью DRAG и PD3
 - Исследование влияния порога r
- Часть 3. Поиск диссонансов произвольной длины (30 мин.)

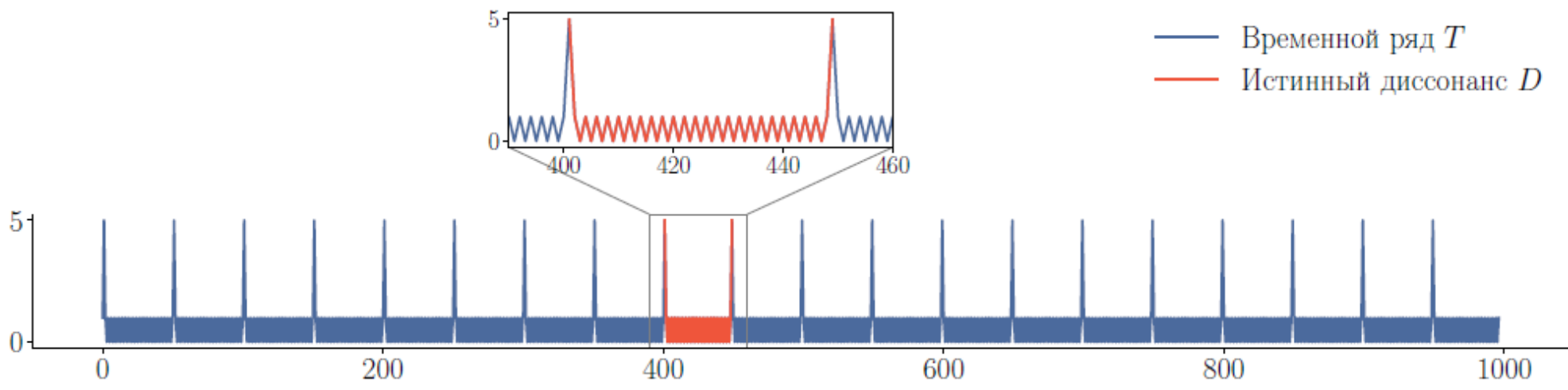
Содержание

- Часть 1. Введение в задачу поиска аномалий (30 мин.)
- Часть 2. Поиск диссонансов фиксированной длины (30 мин.)
- Часть 3. Поиск диссонансов произвольной длины (30 мин.)
 - Теория
 - Почему нужен поиск диссонансов произвольной длины
 - Последовательный алгоритм MERLIN
 - Параллельный алгоритм PALMAD
 - Тепловая карта диссонансов
 - Практика

Что не так с PD3 (и DRAG): Ручной подбор t

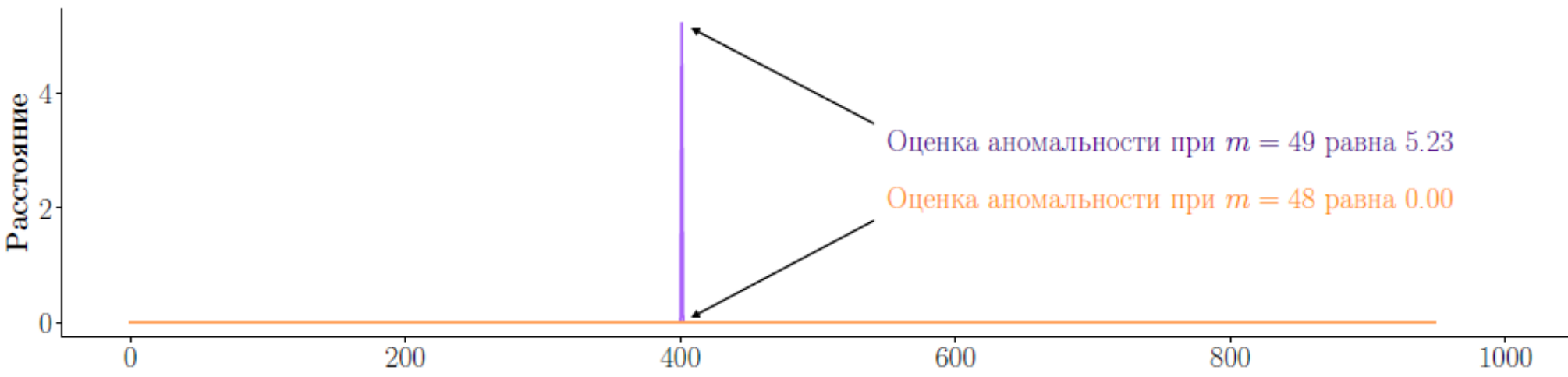
Не всегда заранее известна длина аномалии.
Запуск DRAG/PD3 для всех возможных длин
вычислительно неосуществим

Чтобы найти все аномалии, нужно проверить все значения m



```
// Создание временного ряда
for i ← 1 to 1000 do
  if i mod 50 ≠ 0 then
    ti ← i mod 2
  else
    ti ← 5
  end if
end for
// Создание диссонанса
t430 ← 0; t431 ← 0
```

— Расстояния до ближайших соседей при $m = 49$ — Расстояния до ближайших соседей при $m = 48$

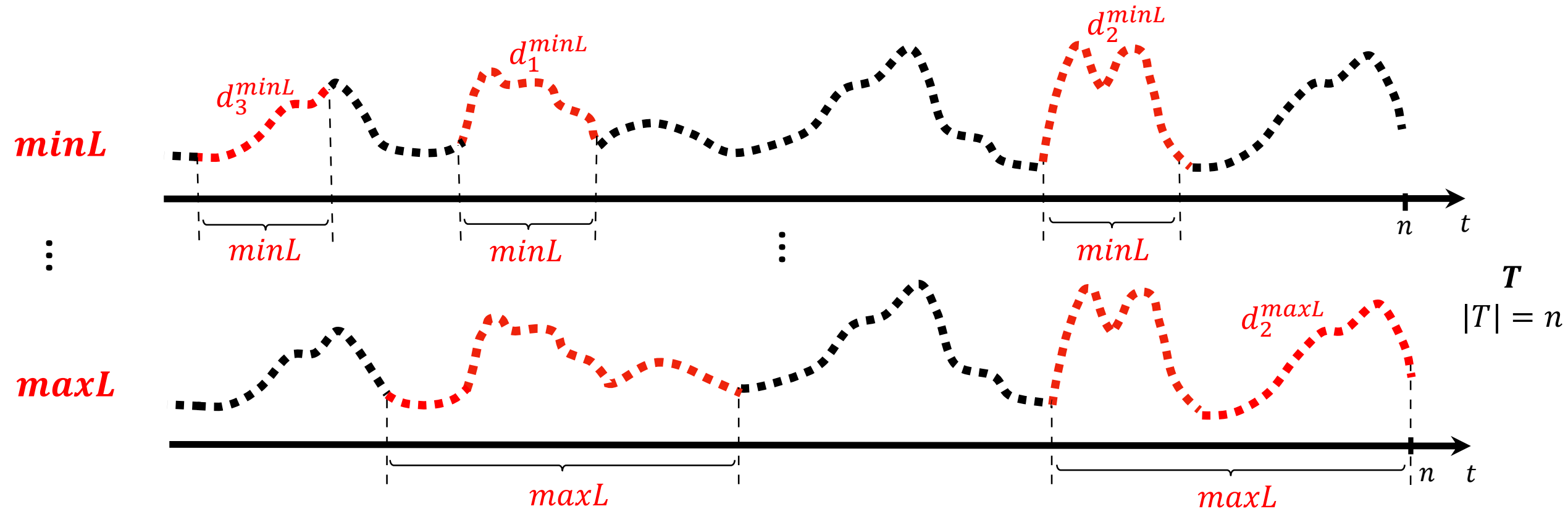


Оценка аномальности при $m = 49$ равна 5.23
 Оценка аномальности при $m = 48$ равна 0.00

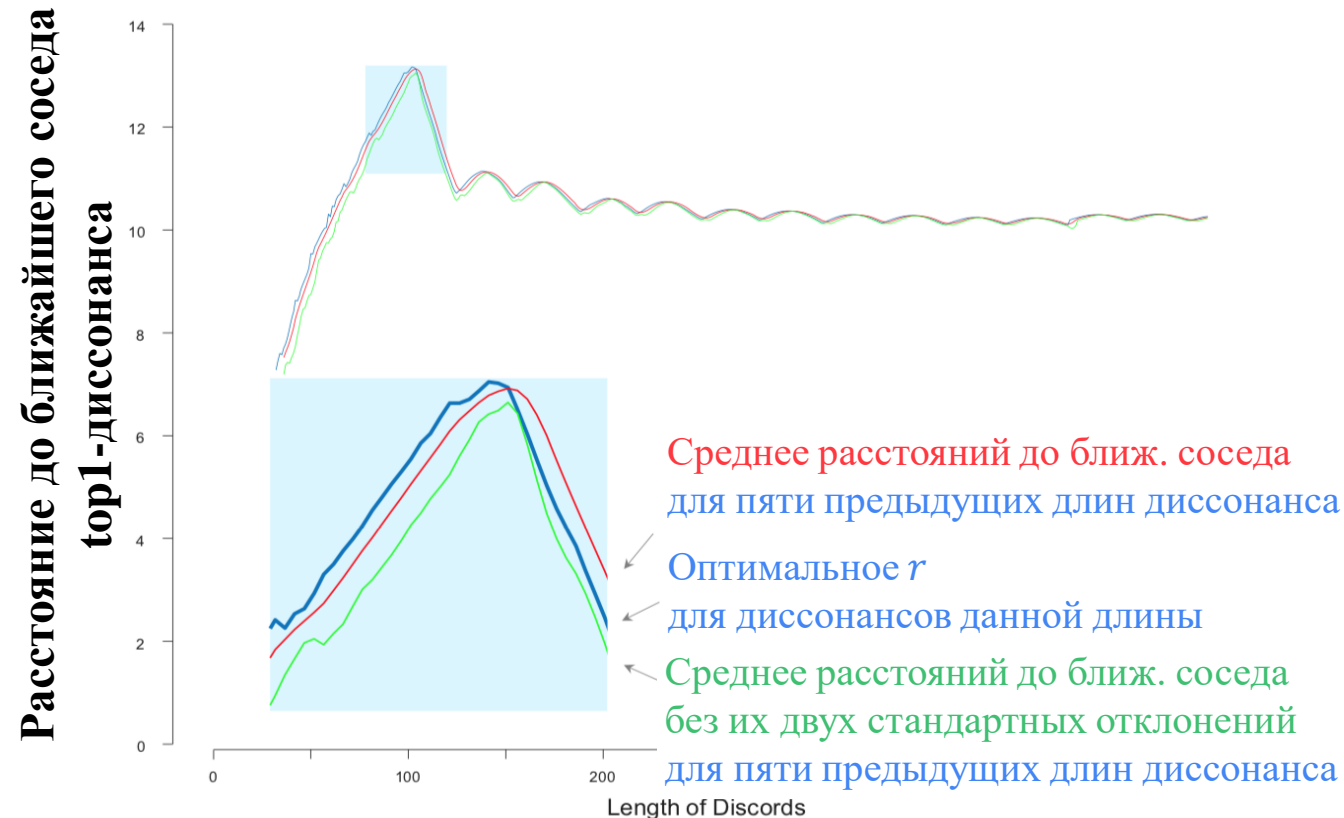
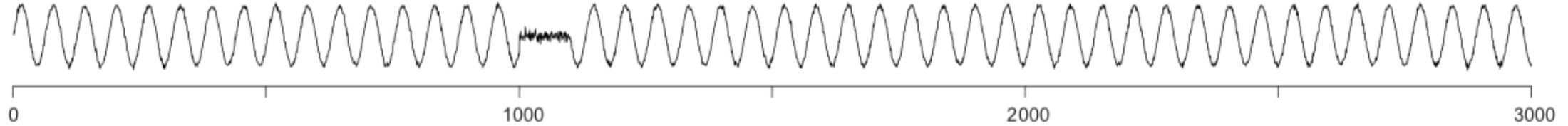
Если $T_{i,m}$ — диссонанс, то не факт, что $T_{i,m \pm 1}$ — диссонанс

Поиск диссонансов произвольной длины

- Дано: временной ряд T , диапазон длины диссонансов $minL \dots maxL$
- Найти: $\mathcal{D} = \bigcup_{m=minL}^{maxL} D_m$, $D_m = \{d_1^m, d_2^m, \dots\}$, где $\exists r_m \min_{s \in M_{d_i^m}} \text{Dist}(d_i^m, s) \geq r_m$



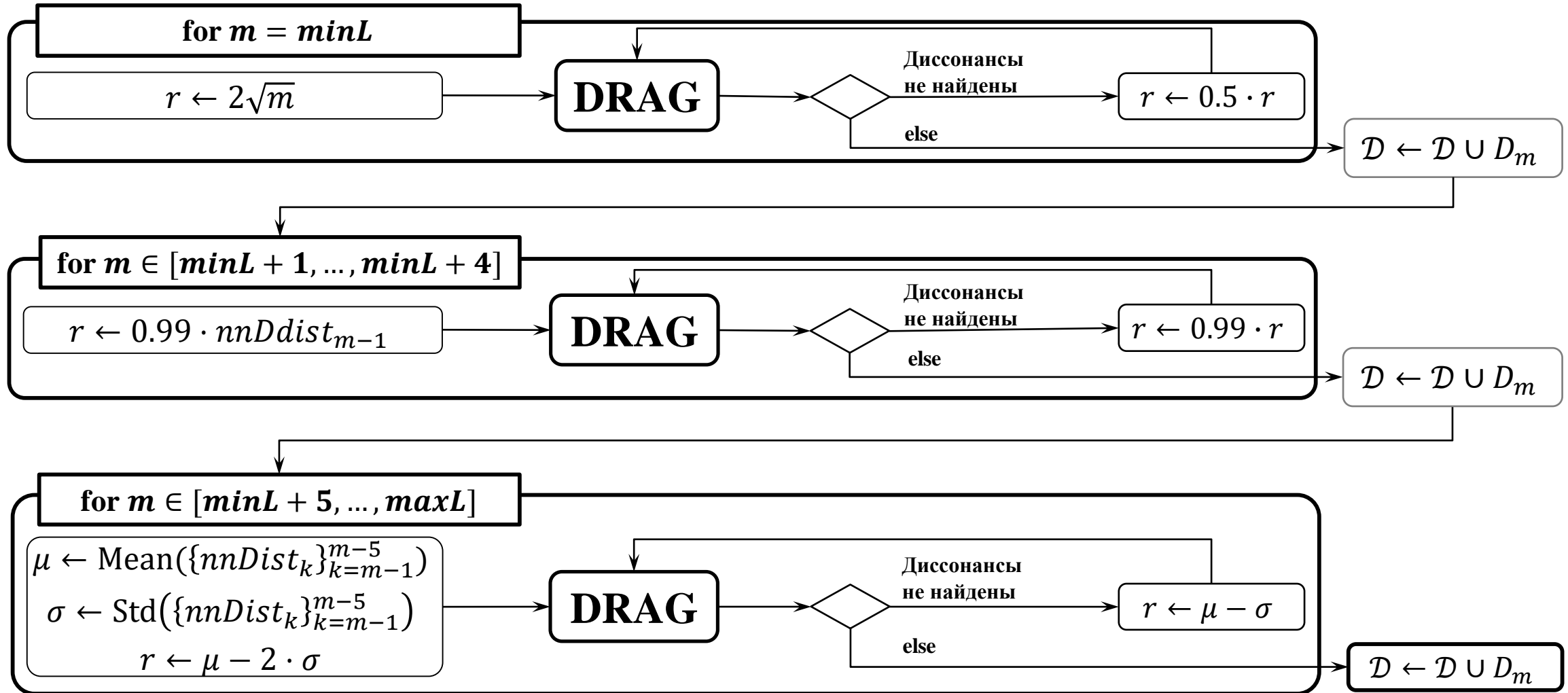
Алгоритм MERLIN: адаптивное вычисление порога r



Порог r нужно вычислять по-разному для разных длин m

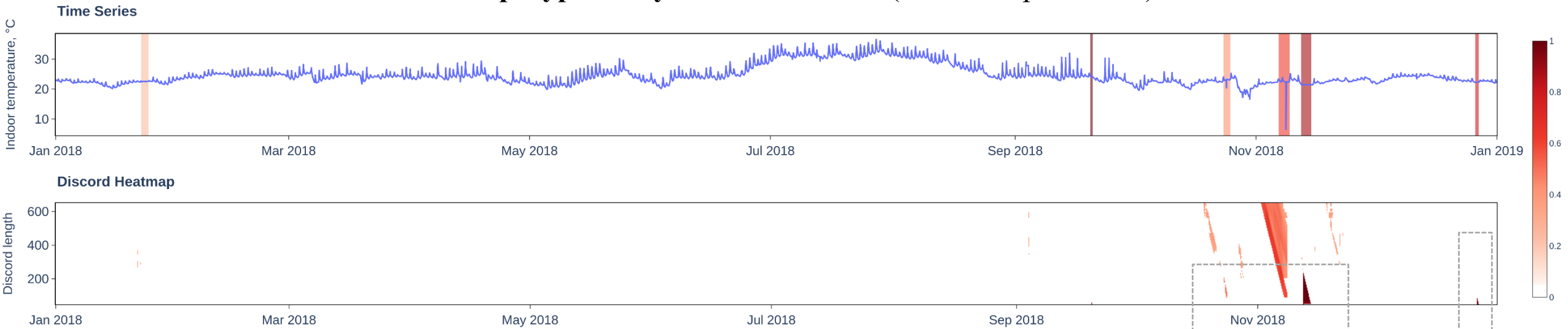
Длина диссонанса, m	Порог, r
$minL$	$r = 2\sqrt{minL}$
$minL + 1, \dots, minL + 4$	$r = 0.99 \cdot nnDist_{m-1}$
$minL + 5, \dots, maxL$	$r = \mu - 2\sigma$

Алгоритм MERLIN



Тепловая карта диссонансов

Температура воздуха в помещении (частота 4 раза в час)

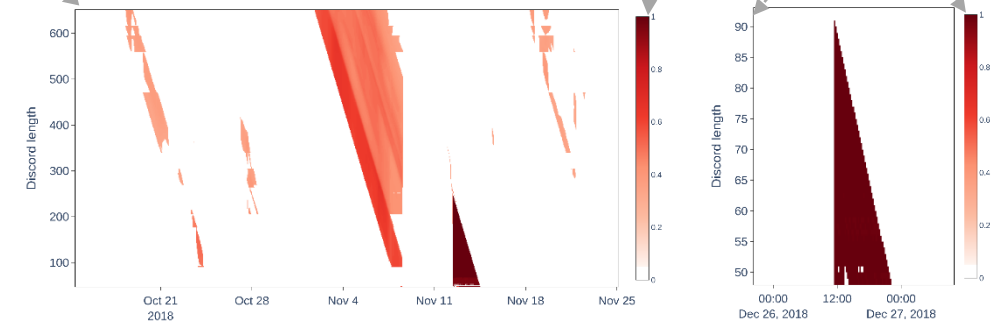


- $heatmap \in \mathbb{R}^{(maxL-minL+1) \times (n-minL)}$,
пиксела (m, i) соответствует диссонансу $T_{i,m} \in D_m$

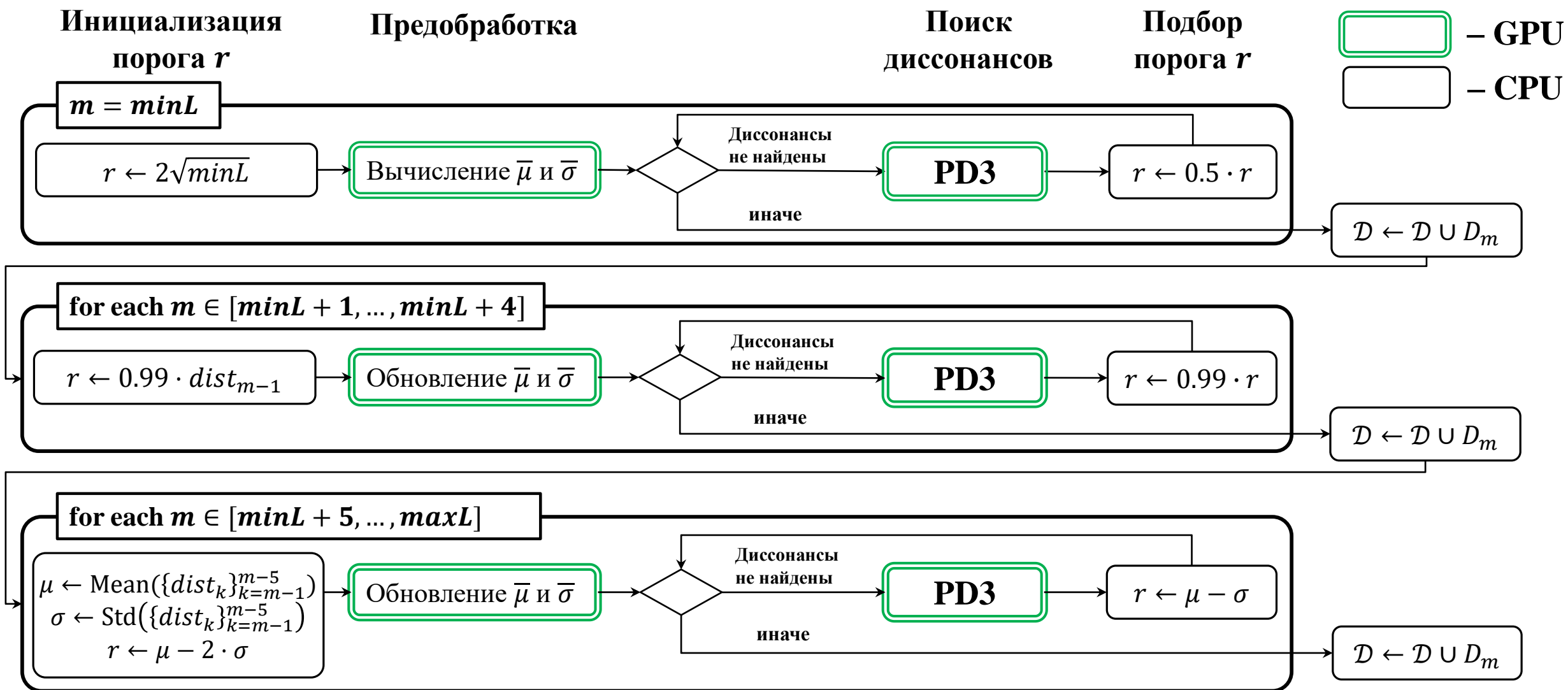
$$heatmap(m, i) = 0.5mT_{i,m} \cdot nnDist$$

- Top-1 диссонанс:

$$\arg \max_{1 \leq i \leq n-m+1} \max_{minL \leq m \leq maxL} heatmap(m, i)$$



PALMAD (Parallel Arbitrary Length MERLIN-based Anomaly Detector)



Содержание

- Часть 1. Введение в задачу поиска аномалий (30 мин.)
- Часть 2. Поиск диссонансов фиксированной длины (30 мин.)
- Часть 3. Поиск диссонансов произвольной длины (30 мин.)
 - Теория
 - Практика
 - Загрузка и визуализация ряда PolyTER
 - Поиск диссонансов с помощью MERLIN и PALMAD
 - Построение тепловой карты диссонансов

Заключение

- **Диссонанс** формализует понятие аномалии временного ряда
- Поиск диссонансов **независим от предметной области, не требует обучения, его результаты однозначны и воспроизводимы**
- Последовательные алгоритмы поиска диссонансов
 - DRAG (фиксированная длина)
 - MERLIN (произвольная длина)
- **Параллельные алгоритмы поиска диссонансов для GPU**
 - PD3 (фиксированная длина)
 - PALMAD (произвольная длина)



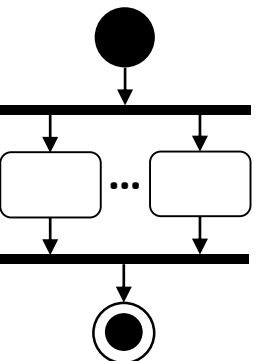
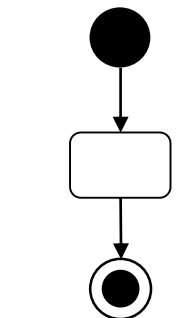
**Михаил Леонидович
Цымблер**, д.ф.-м.н.
mzym@susu.ru
<https://mzym.susu.ru>



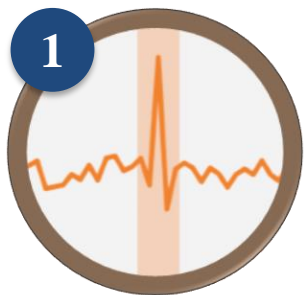
**Яна Александровна
Краева**, к.ф.-м.н.
kraevaya@susu.ru

Приложение: Эволюция алгоритмов поиска диссонансов

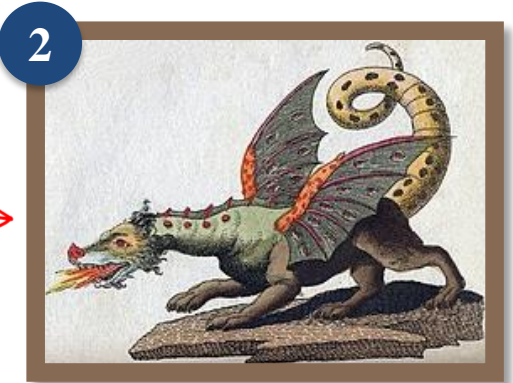
Последовательные алгоритмы



Параллельные алгоритмы



HOTSAX¹⁾
(2005)



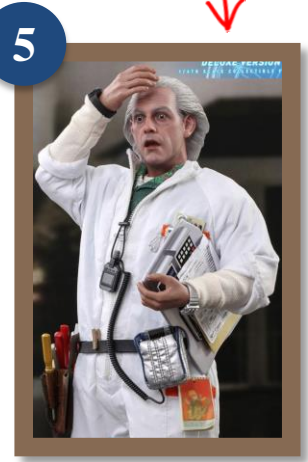
DRAG²⁾
(2007)



MERLIN³⁾
(2020)



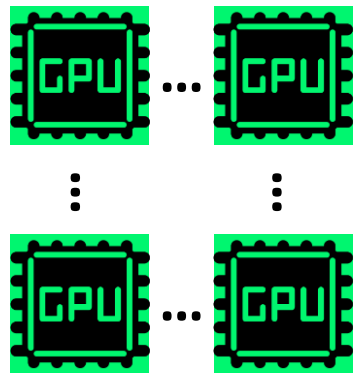
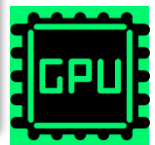
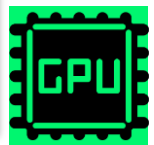
PD3⁴⁾ (2022)



PALMAD⁵⁾ (2023)



PADDi⁶⁾ (2024)



¹⁾ Keogh E. *et al.* HOTSAX: Efficiently finding the most unusual time series subsequence. ICDM 2005. 226-233. DOI: [10.1109/ICDM.2005.79](https://doi.org/10.1109/ICDM.2005.79)

²⁾ Yankov D. *et al.* Disk aware discord discovery: finding unusual time series in terabyte sized datasets. ICDM 2007. 381-390. DOI: [10.1109/ICDM.2007.61](https://doi.org/10.1109/ICDM.2007.61).

³⁾ Nakamura T. *et al.* MERLIN: parameter-free discovery of arbitrary length anomalies in massive time series archives. ICDM 2020. 1190-1195. DOI: [10.1109/ICDM50108.2020.00147](https://doi.org/10.1109/ICDM50108.2020.00147).

⁴⁾ Kraeva Y., Zymbler M. A parallel discord discovery algorithm for a graphics processor. PRIA. 2023. 33(2). 101-112. DOI: [10.1134/S1054661823020062](https://doi.org/10.1134/S1054661823020062).

⁵⁾ Zymbler M., Kraeva Y. High-performance time series anomaly discovery on graphics processors. Mathematics. 2023. 11(14). art. 3193. DOI: [10.3390/math11143193](https://doi.org/10.3390/math11143193).

⁶⁾ Kraeva Y., Zymbler M. PADDi: Highly scalable parallel algorithm for discord discovery on multi-GPU clusters. LJM. 2025. 46(4). 1480-1494. DOI: [10.1134/S1995080225606198](https://doi.org/10.1134/S1995080225606198).