

# Динамическое выравнивание многомерных временных рядов

Моргачев Г., Смирнов В., Липницкая Т.,  
Руководитель: Гончаров А.

Moscow Institute of Physics and Technology

22 апреля 2019 г.

# Цели исследования

## Цель работы

Исследовать влияния выбора внутренней функции расстояния на работу алгоритма DTW.

## Проблема

При обобщении выравнивания временных рядов на многомерный случай остается открытым вопрос определения расстояния между парами векторов.

## Метод решения

Получение оптимальной функции расстояния путем проведения эксперимента на задачах поиска паттернов и кластеризации

# Постановка задачи

Есть множество временных рядов  $\mathbb{S} \subset \mathbb{R}^{l \times n}$ , где  $l$  - количество каналов,  $n$  - длина ряда.

Задано множество функций расстояния между векторами

$$\mathbb{R} = \{\rho : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+\}$$

$$DTW_\rho : \mathbb{S} \times \mathbb{S} \rightarrow \mathbb{R}^+$$

## Кластеризация

$\forall s_i \in \mathbb{S} \subset \mathbb{S}$  задано  $y_i \in \mathbb{Y}$  - множество меток классов. Матрица попарных расстояний:

$$D(DTW_\rho(\mathbb{S})) = \|D_{ij}\|, \quad D_{ij} = DTW_\rho(s_i, s_j), \quad s_i, s_j \in \mathbb{S}$$

Кластеризатор:  $f : D \rightarrow Z^N$ ,  $Z$  - множество меток кластеров

## Функции качества кластеризации

$$Q_1(f(D), S) = \frac{1}{|Z|} \sum_{z \in Z} \max_y \frac{N_z^y}{N_z}$$

$$Q_2(f(D), S) = \frac{1}{|Z|} \sum_{z \in Z} \max_y \frac{(N_z^y)^2}{N_z N^y}$$

## Поиск паттернов

Задан временной ряд  $\mathcal{A}$  длины  $n$ , содержащий подряды класса  $P$ .  
 $P$  - временные ряды длины  $m \ll n$ .

Известны представители класса  $P$ , необходимо найти участки  $\mathcal{A}$ , соответствующие данному классу.

$T = \{t_1, \dots, t_j\}$  - множество начал таких событий.

Участок найден, если пересечение с предполагаемым более 80% от  $m$ .

## Функции качества поиска шаблонов

$$Q(DTW_{\rho}, A, P_k, T) = \frac{\sum_{i=1}^j [t_i - \text{найден}]}{j}$$

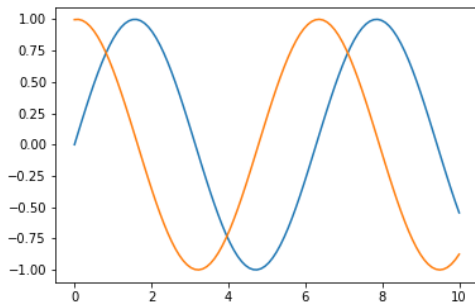
## Общая постановка задачи

$$\rho_i = \operatorname{argmax}_{\rho} Q_i(\rho)$$

# Сравнение рядов

## Проблемы

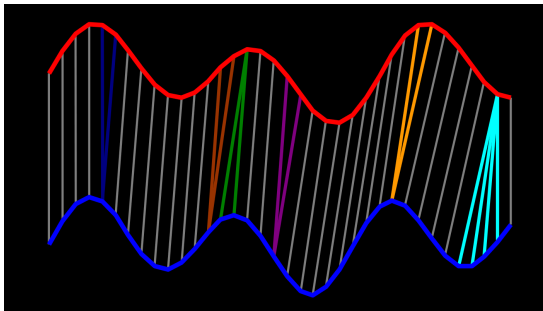
- Растяжение
- Сдвиги



# DTW

## DTW

- Выравнивание рядов друг относительно друга
- Позволяет задать функцию расстояния
- Использует матрицу попарных расстояний между точками рядов



# Многомерное DTW

## Особенность

Необходимость выбора функции расстояния между соответственными точками рядов

## Постановка задачи

Зависимость качества кластеризации временных рядов от выбора функции расстояния между ними



## Кластеризация

**Иерархическая** с функциями расстояния между кластерами:

- 1 *complete*:  $d(A, B) = \max_{a \in A, b \in B} (dist(a, b))$
- 2 *weighted*:  $d(A, B) = \frac{(dist(S, B) + dist(T, B))}{2}$ , где кластер  $A = S \cup T$
- 3 *weighted*:  $d(u, v) = \sum_{a \in A, b \in B} \frac{d(a, b)}{(|A| * |B|)}$

# Эксперимент

## Данные: класстеризация

- Размеченные данные ускорений акселерометра телефона: из 6 состояния человека, 3 канала, разбит по 50 точек.

## Данные: поиск паттернов

- Данные ECG: 4 состояния человека, 3 канала, разбиты на ряды по 206 точек
- Написание букв: 20 символов, 3 канала, разбиты по 182 точки

# Результаты: поиск паттернов

$\rho$	average	characters			epi		
		$Q$	$t$	$t_{\text{no optim}}$	$Q$	$t$	$t_{\text{no optim}}$
$L_1$	DBA	0.857	2.123	11.767	0.744	14.335	13.064
	mean	0.894	2.361	11.614	0.744	13.541	13.912
$L_2$	DBA	0.818	1.551	11.499	0.687	12.342	13.205
	mean	0.854	1.527	10.164	0.687	14.199	12.738
ED	DBA	0.08	17.511	17.511	0.172	1.620	1.620
	mean	0.09	17.645	17.645	0.172	1.540	1.540

# Результаты: кластеризация

$\rho$	$N_{clust}$	$Q_1$			$Q_2$		
		<i>compl.</i>	<i>aver.</i>	<i>weight.</i>	<i>compl.</i>	<i>aver.</i>	<i>weight.</i>
$L_1$	24	0.506	0.585	0.638	0.273	0.376	0.449
	36	0.533	0.620	0.616	0.299	0.425	0.414
	48	0.556	0.639	0.631	0.330	0.443	0.431
$L_2$	24	0.488	0.622	0.626	0.270	0.417	0.425
	36	0.498	0.646	0.643	0.270	0.455	0.449
	48	0.534	0.648	0.653	0.270	0.455	0.462

# Результаты

## Выводы

Спасибо за внимание!