

# Порождение признаков с помощью локально-аппроксимирующих моделей

Садиев Абдурахмон

21 марта, 2019.

# Общее описание исследования

## Задача

Классификация видов деятельности человека по измерениям фитнес-браслетов

## Проблема

Временные ряды - объектсложной структуры. Сложность придумать грамотного признакового описания

## Предлагаемое решение

Требуется построить набор локально-аппроксимирующих моделей и выбрать наиболее адекватные. Найти оптимальный способ сегментации и оптимальное описание временного ряда.

[1] [2]



Anastasia Motrenko and Vadim Strijov.

Extracting fundamental periods to segment biomedical signals.

*IEEE J. Biomedical and Health Informatics*, 20(6):1466–1476, 2016.



В. В. Стрижов М. Е. Карасиков.

Классификация временных рядов в пространстве параметров порождающих моделей.

*Информ. и её примен.*, 10(4):121–131, 2016.

# Формальная постановка задачи

## Данные

Пусть задана выборка :

$$\mathcal{D} = \{(\mathbf{s}_i, y_i) \mid i = 1, \dots, m; \mathbf{s}_i = [\mathbf{s}_i(1), \dots, \mathbf{s}_i(T)] \in \mathbf{S} \subset \mathbb{R}^{n \times m}\},$$

где  $\mathbf{s}_i(t) \in \mathbb{R}^n$ ,  $y_i \in Y$  - пространство ответов,  $|Y| = K \in \mathbb{N}$ ,  $m$  - количество элементов в выборке.

## Модель

Модель будет приближать отображение  $f : \mathbf{S} \rightarrow Y$  и будем искать ее в виде суперпозиции:

$$f'(\mathbf{s}) = g(h(\mathbf{s}), \mathbf{w}) \quad (1)$$

где  $h : \mathbf{S} \rightarrow \Phi$ .  $\Phi \subset \mathbb{R}^p$  - пространство признаков,  $\mathbf{w}$  - вектор параметров модели.

### Локально-аппроксимирующие модели

Модели  $h_j \in \mathcal{H}$ , где  $j \in \{1, \dots, r\}$ , а  $r$  - количество моделей в наборе  $\mathcal{H}$ .  
Оптимальный набор моделей находится решением оптимизационной задачи

$$\mathcal{P}_{opt} = \arg \min_{\mathcal{P} \subset \mathcal{H}} \min_{\mathbf{w} \in \mathbb{R}^p} \mathcal{L}[g(\mathcal{P}, \mathbf{w})] \quad (2)$$

# Алгоритм решения задачи

## Нулевой этап

Сегментируем временной ряд  $s_i$ . Выбираем, как это сделать.

## Первый этап

Для каждого сегмента временной ряда  $s_i$  строим признаковое описание. Для этого есть множество разных алгоритмов.

## Второй этап

После получения признакового описания сегментов временного ряда решаем задачу многоклассовой классификации.

## Выборка

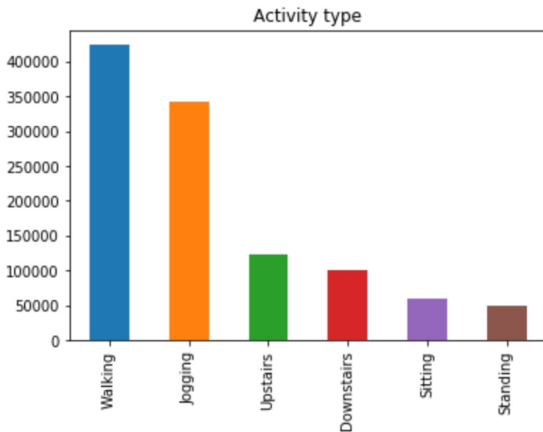


Рис.: Количество измерений для каждого класса

## Выборка

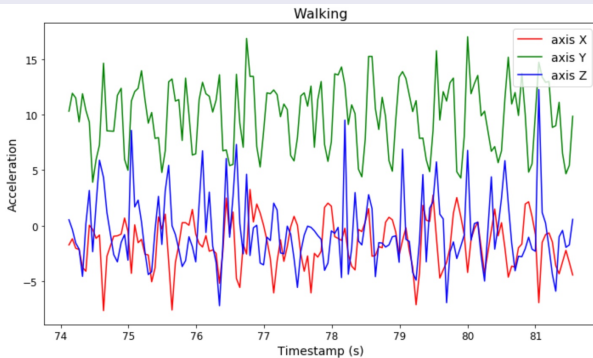


Рис.: Пример ряда из выборки



## Результат

|                 | all      | Standing | Upstairs | Walking  | Sitting  | Jogging  | Downstairs |
|-----------------|----------|----------|----------|----------|----------|----------|------------|
| <b>lr_fft_</b>  | 0.829887 | 0.993201 | 0.995094 | 0.908670 | 0.866615 | 0.924371 | 0.971823   |
| <b>rf_fft_</b>  | 0.850424 | 0.996215 | 0.996495 | 0.912035 | 0.891568 | 0.925913 | 0.978622   |
| <b>svm_fft_</b> | 0.857293 | 0.995304 | 0.995935 | 0.914067 | 0.904465 | 0.922128 | 0.982687   |

|                 | all      | Standing | Upstairs | Walking  | Sitting  | Jogging  | Downstairs |
|-----------------|----------|----------|----------|----------|----------|----------|------------|
| <b>lr_fft_</b>  | 0.828135 | 0.989276 | 0.994673 | 0.904184 | 0.878251 | 0.916240 | 0.973645   |
| <b>rf_fft_</b>  | 0.850775 | 0.995234 | 0.996215 | 0.913927 | 0.890797 | 0.926193 | 0.979183   |
| <b>svm_fft_</b> | 0.854630 | 0.995374 | 0.996075 | 0.911544 | 0.899979 | 0.924932 | 0.981356   |

Рис.: Результат работы

## Резюме

- Мы научились классифицировать виды деятельности человека.
- Если сегментировать более разумно, то можно получить более высокие результаты.
- Классификации на наборе признаков, созданных разными моделями, имеет высше качество.