

# Анализ свойств локальных моделей в задачах кластеризации точек квазипериодических временных рядов

Грабовой Андрей Валериевич

Московский физико-технический институт  
Факультет управления и прикладной математики  
Кафедра интеллектуальных систем

Научный руководитель д.ф.-м.н. В. В. Стрижов

*Москва,  
2019г*

# Задача кластеризации точек временного ряда

**Цель:** предложить алгоритм поиска характерных квазипериодических сегментов внутри временного ряда.

## Задачи

- 1 Предложить признаковое описание точек временного ряда.
- 2 Предложить функцию расстояния между точками временного ряда в новом признаковом описании, для их дальнейшей кластеризации.

## Исследуемые проблемы

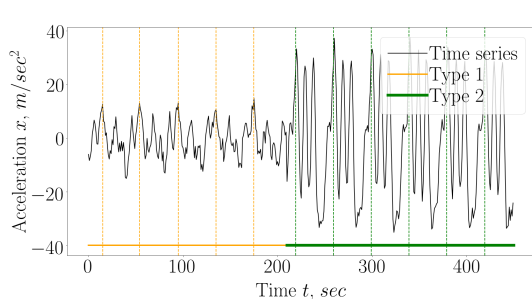
- 1 Построение адекватного признакового описания точек временного ряда низкой размерности.

## Методы решения

Алгоритм поиска характерных сегментов основывается на методе главных компонент для локального снижения размерности сегмента фазовой траектории в окрестности каждой точки временного ряда. Главные компоненты рассматриваются как признаковое описание точек временного ряда.

- *A. P. Motrenko, V. V. Strijov* Extracting fundamental periods to segment biomedical signals // *Journal of Biomedical and Health Informatics*, 2015, 20(6). P. 1466–1476.
- *Y. G. Cinar and H. Mirisae* Period-aware content attention RNNs for time series forecasting with missing values // *Neurocomputing*, 2018. Vol. 312. P. 177–186.
- *A. D. Ignatov, V. V. Strijov* Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer. // *Multimedial Tools and Applications*, 2015.
- *A. Olivares, J. Ramirez, J. M. Gorris, G. Olivares, M. Damas* Detection of (in)activity periods in human body motion using inertial sensors: A comparative study. // *Sensors*, 12(5):5791–5814, 2012.
- *Д. Л. Данилова, А. А. Жигловский* Главные компоненты временных рядов: метод "Гусеница". — СПбУ, 1997.

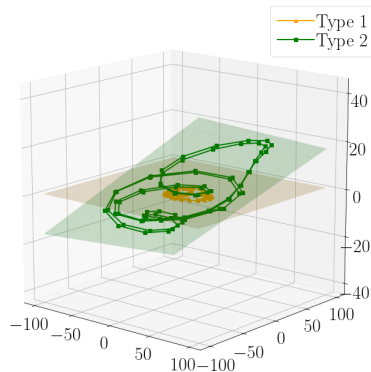
# Постановка задачи кластеризации точек



Задан временной ряд

$$\mathbf{x} \in \mathbb{R}^N, \quad \mathbf{x} = [\mathbf{v}_1, \dots, \mathbf{v}_M], \quad \mathbf{v}_i \in \mathcal{V},$$

где  $\mathcal{V}$  множество сегментов в ряде  $\mathbf{x}$ .



Проекция фазовых траекторий на первые две главные компоненты.

Предположения:

- число различных действий внутри временного ряда известно и равно  $K$ ,
- для всех  $\mathbf{v} \in \mathcal{V}$  выполняется  $|\mathbf{v}| \leq T$ , где  $|\mathbf{v}|$  длина сегмента,
- для всех  $i$  либо  $[\mathbf{v}_{i-1}, \mathbf{v}_i]$  либо  $[\mathbf{v}_i, \mathbf{v}_{i+1}]$  является цепочкой действий.

# Постановка задачи кластеризации точек

Строится отображение

$$a : t \rightarrow \mathbb{Y} = \{1, \dots, K\},$$

где  $t \in \{1, \dots, N\}$  некоторый момент времени, на котором задан временной ряд. Требуется, чтобы отображение  $a$  удовлетворяло следующим свойствам:

$$\begin{cases} a(t_1) = a(t_2), & \text{если в моменты } t_1, t_2 \text{ совершается один тип действий,} \\ a(t_1) \neq a(t_2), & \text{если в моменты } t_1, t_2 \text{ совершаются разные типы действий.} \end{cases}$$

Пусть задана ассессорская разметка точек временного ряда:

$$\mathbf{y} \in \{1, \dots, K\}^N.$$

Ошибка алгоритма  $a$  на временном ряде  $\mathbf{x}$ :

$$S = \frac{1}{N} \sum_{t=1}^N [y_t \neq a(t)],$$

где  $t$  — момент времени,  $y_t$  ассессорская разметка  $t$ -го момента времени для заданного временного ряда.

# Построение признакового описания точек

Фазовая траектория ряда  $\mathbf{x}$ :

$$\mathbf{H} = \{\mathbf{h}_t | \mathbf{h}_t = [x_{t-T}, x_{t-T+1}, \dots, x_t], T \leq t \leq N\},$$

где  $\mathbf{h}_t$  — точка фазовой траектории.

Множество сегментов фазовой траектории:

$$\mathbf{S} = \{\mathbf{s}_t | \mathbf{s}_t = [\mathbf{h}_{t-T}, \mathbf{h}_{t-T+1}, \dots, \mathbf{h}_{t+T-1}], 2T \leq t \leq N - T\},$$

где  $\mathbf{s}_t$  — это сегмент фазовой траектории в окрестности момента времени  $t$ .

Множество базисов, полученных методом главных компонент для каждого сегмента фазовой траектории:

$$\mathbf{W} = \{\mathbf{W}_t | \mathbf{W}_t = [\mathbf{w}_t^1, \mathbf{w}_t^2]\}, \quad \mathbf{\Lambda} = \{\mathbf{\lambda}_t | \mathbf{\lambda}_t = [\lambda_t^1, \lambda_t^2]\},$$

где  $[\mathbf{w}_t^1, \mathbf{w}_t^2]$  и  $[\lambda_t^1, \lambda_t^2]$  это базисные векторы и соответствующие им собственные числа для сегмента фазовой траектории  $\mathbf{s}_t$ . Далее  $\mathbf{W}_t$  и  $\mathbf{\lambda}_t$  рассматриваются как признаковое описание момента времени  $t$ .

# Функция расстояния (общий случай)

Для кластеризации точек временного ряда, вводится расстояние в предложенном признаковом описании данного ряда. Расстояние между элементами  $\mathbf{W}_{t_1}, \mathbf{W}_{t_2}$ :

$$\rho(\mathbf{W}_{t_1}, \mathbf{W}_{t_2}) = \max \left( \max_{\mathbf{e}_2 \in \mathbf{W}_{t_2}} d_1(\mathbf{e}_2), \max_{\mathbf{e}_1 \in \mathbf{W}_{t_1}} d_2(\mathbf{e}_1) \right),$$

где  $\mathbf{e}_i$  это базисный вектор пространства  $\mathbf{W}_i$ , а  $d_i(\mathbf{e})$  является расстоянием от вектора  $\mathbf{e}$  до пространства  $\mathbf{W}_i$ .

# Функция расстояния (двумерный случай)

Расстояние между элементами  $\mathbf{W}_{t_1}, \mathbf{W}_{t_2}$ :

$$\rho(\mathbf{W}_{t_1}, \mathbf{W}_{t_2}) = \max_{\{\mathbf{a}, \mathbf{b}, \mathbf{c}\} \subset \mathbf{W}_{t_1} \cup \mathbf{W}_{t_2}} V(\mathbf{a}, \mathbf{b}, \mathbf{c}),$$

где  $\mathbf{W}_{t_1} \cup \mathbf{W}_{t_2}$  это объединение базисных векторов первого и второго пространства,  $V(\mathbf{a}, \mathbf{b}, \mathbf{c})$  — объем параллелепипеда построенного на векторах  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ , которые являются столбцами матрицы  $\mathbf{W}_{t_1} \cup \mathbf{W}_{t_2}$ .

Расстояние между элементами  $\mathcal{L}$ :

$$\rho(\lambda_1, \lambda_2) = \sqrt{(\lambda_1 - \lambda_2)^T (\lambda_1 - \lambda_2)}.$$

Расстояние между точками временного ряда:

$$\rho(t_1, t_2) = \rho(\mathbf{W}_1, \mathbf{W}_2) + \rho(\lambda_1, \lambda_2).$$

Матрица попарных расстояний:

$$\mathbf{M} = \mathbb{R}_+^{N \times N}.$$



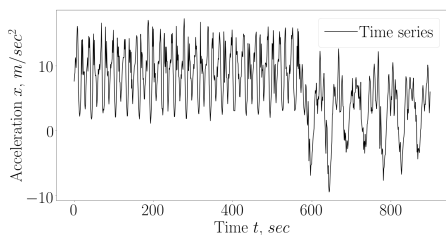
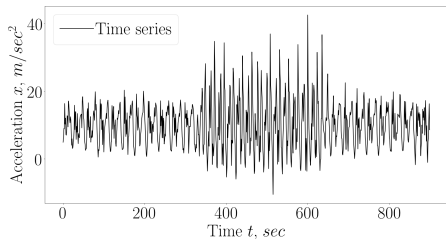
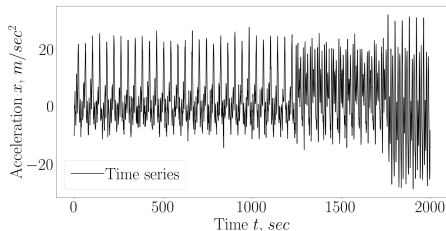
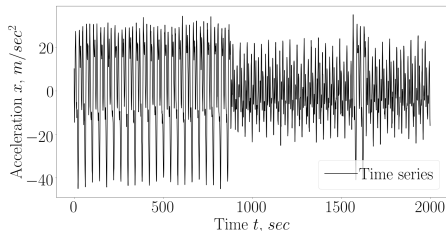
# Описание временных рядов в эксперименте

- Physical Motion — реальные временные ряды, которые получены при помощи мобильного акселерометра.  
Характерные действия: ходьба, бег, приседания.
- Synthetic — синтетические временные ряды, которые были построены при помощи нескольких первых слагаемых ряда Фурье со случайными коэффициентами из стандартного нормального распределения.

Ряд, $x$	Длина, $N$	Сегментов, $K$	Длина, $T$	Ошибка, $S$
Phys. Motion 1	900	2	40	0.06
Phys. Motion 2	900	2	40	0.03
Synthetic 1	2000	2	20	0.04
Synthetic 2	2000	3	20	0.03

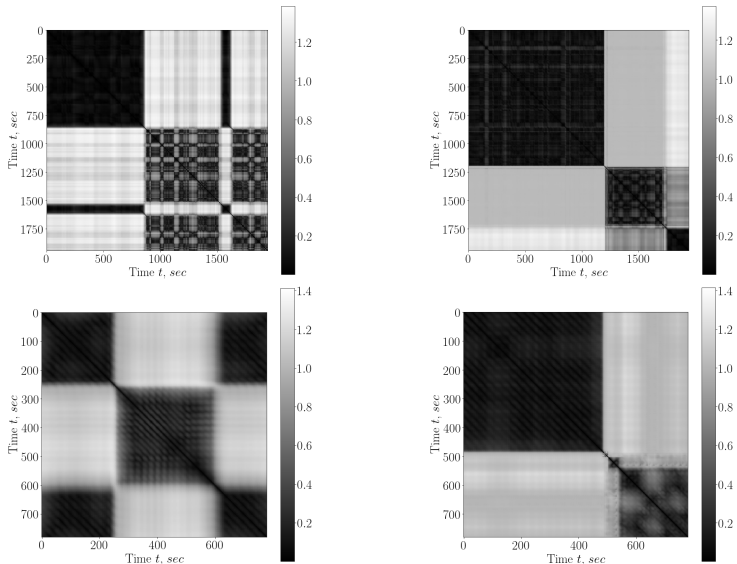
- $N$  — число точек во временном ряде,
- $K$  — число различных действий во временном ряде,
- $T$  — максимальная длина сегмента,
- $S$  — точность кластеризации.

# Пример временных рядов



Временные ряды построенные синтетически, а также при помощи мобильного акселерометра.

# Матрица попарных расстояний $M$



Матрицы попарных расстояний для временных рядов, построенных синтетически, а также при помощи мобильного акселерометра.

# Проекция точек фазовой траектории на плоскость

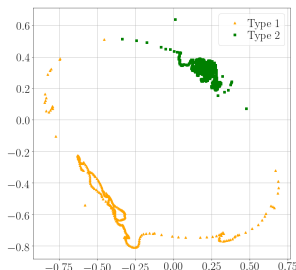
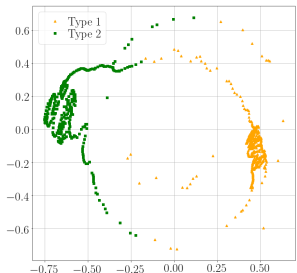
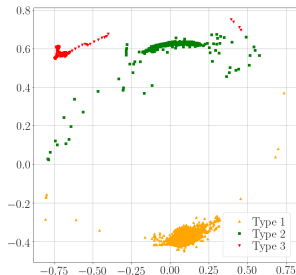
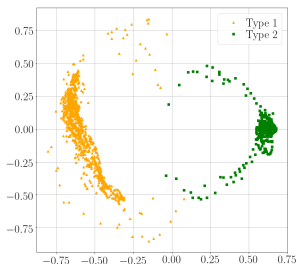
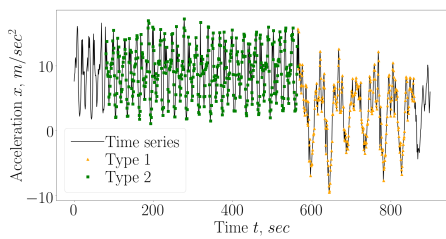
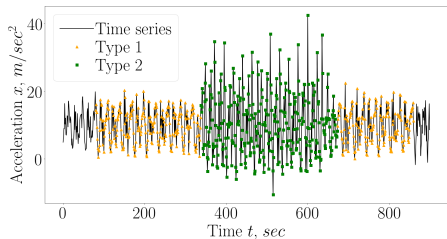
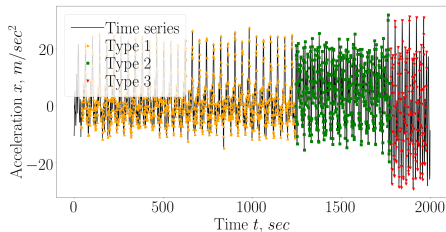
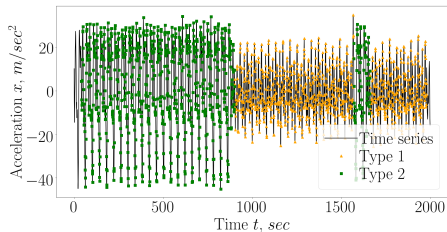


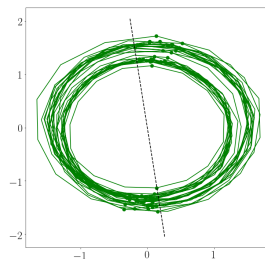
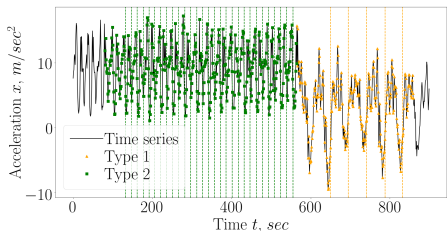
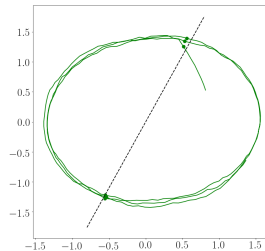
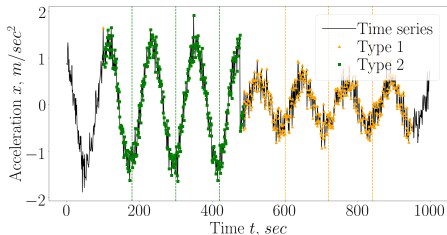
Иллюстрация проекции признаков описания точек временного ряда на плоскости для временных рядов, построенных синтетически, а также при помощи мобильного акселерометра.

# Кластеризация точек временного ряда



Результат кластеризации точек временных рядов, построенных синтетически, а также при помощи мобильного акселерометра.

# Сегментация временных рядов



Результат сегментации временных рядов, в случае двух синусоидальных сигналов в произвольной частотой и амплитудой, а также в случае реальных данных, полученных при помощи акселерометра.

- ❶ Предложен алгоритм поиска характерных сегментов, который основывается на методе главных компонент для локального снижения размерности
- ❷ Введена функция расстояния между локальными базисами в каждый момент времени, которые интерпретировались как признаковое описание точки временного ряда. Данная функция является метрикой.
- ❸ В ходе эксперимента, на реальных показаниях акселерометра, а также на синтетических данных, было показано, что предложенный метод измерения расстояния между базисами хорошо разделяет точки, которые принадлежат различным действиям, что приводит к хорошей кластеризации объектов.
- ❹ Также в эксперименте была проведена полная сегментация временных рядов для каждого кластера по отдельности.

Планируется решить задачу нахождения минимального размера фазового пространства, для которого фазовая траектория не имеет самопересечений.

- ① Грабовой А. В., Стрижов В. В. Анализ свойств локальных моделей в задачах кластеризации квазипериодических временных рядов // (в процессе)
- ② Грабовой А. В., Бахтеев О. Ю., Стрижов В. В. Определение релевантности параметров нейросети // Информатика и ее применения, 2019, 13(2).
- ③ Гадаев Т. Т., Грабовой А. В., Мотренко А. П., Стрижов В. В. Численные методы оценки объема выборки в задачах регрессии и классификации // (в процессе)
- ④ Бучнев Т. Т., Грабовой А. В., Гадаев Т. Т., Стрижов В. В. Ранее прогнозирование достаточного объема выборки для обобщенно линейной модели // (в процессе)
- ① 12 октября 2018. ИОИ-2018. Автоматическое определение релевантности параметров нейросети.
- ② 29 ноября 2019. 61-я Всероссийская научная конференция МФТИ. Поиск оптимальной модели при помощи алгоритмов прореживания.