Локальная кластеризация 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

Цель работы

Исследуется

Исследуется задача локального распознавания и локальной разметки человеческой активности в течении некоторого времени.

Требуется

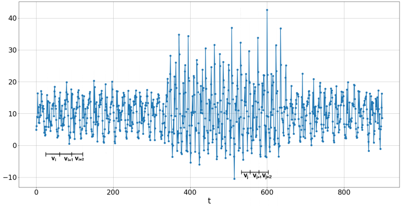
Требуется предложить локальное признаковое описание временного ряда, а также кластеризации точек данного ряда после получения его признакового описания.

Проблемы

Построение признаквого описания временного ряда низкой размерности.

Список литературы

- *И. П. Ивкин, М. П. Кузнецов* Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию. // Машинное обучение и анализ данных, 2015.
- *V. V. Strijov, A. M. Katrutsa* Stresstes procedures for features selection algorithms. // Schemometrics and Intelligent Laboratory System, 2015.
- *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.
- *I. Borg, P. J. F. Groenen* Modern Multidimensional Scaling. — New York: Springer, 2005. 540 p.
- *Д. Л. Данилова, А. А. Жигловский* Главные компоненты временных рядов: метод "Гусеница". — СПбУ, 1997.



Задан временной ряд:
$\mathbf{x} \in \mathbb{R}^{\mathrm{N}}, \quad \mathbf{x} = [\mathbf{v}_1, \cdots, \mathbf{v}_M], \quad \mathbf{v}_i \in \mathcal{V},$
где $\mathcal{V}$ множество возможных сигналов.

Предположения:

- $|\mathcal{V}| = K$,
- $\forall \mathbf{v} \in \mathcal{V} \ \ |\mathbf{v}| \leq T$,
- $\forall i$ выполняется $\mathbf{v}_i = \mathbf{v}_{i-1}$ или $\mathbf{v}_i = \mathbf{v}_{i+1}$,

где $|\mathcal{V}|$ мощность множества сигналов, а $|\mathbf{v}|$ длина сигнала.

Рассмотрим отображение:

$$a : x \to \{1, \cdots, K\}$$

где $x \in \mathbf{X}$ некоторая точка временного ряда.

Отображение должно удовлетворять следующим свойствам:

$$\begin{cases} a(x_1) = a(x_2), & \text{если } \exists \mathbf{v} \in \mathcal{V} : x_1, x_2 \in v \\ a(x_1) \neq a(x_2), & \text{если } \nexists \mathbf{v} \in \mathcal{V} : x_1, x_2 \in v \end{cases}$$

Фазовая траектория ряда $\mathbf{X}$:
$\mathcal{H} = \{\mathbf{h}_t | \mathbf{h}_t = [x_{t-T}, x_{t-T+1}, \cdots, x_t],\ \mathrm{T} \leq t \leq \mathrm{N}\}.$

Фазовые подпространства:
$\mathcal{S} = \{\mathbf{s}_t | \mathbf{s}_t = [h_{t-2T}, h_{t-2T+1}, \cdots, h_t],\ 2\mathrm{T} \leq t \leq \mathrm{N}\}.$

Пространство базисов:

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

где $[\mathbf{w}_t^1, \mathbf{w}_t^2]$ и $[\lambda_t^1, \lambda_t^2]$ это базисные векторы и сингулярные числа метода главных компонент для подпространства s_t.

Кластеризация точек

Расстояние между элементами $\mathcal{W}$:
$\rho\left(\mathbf{W}_1, \mathbf{W}_2\right) = \max_{\{\mathbf{a},\mathbf{b},\mathbf{c}\}\subset\mathbf{W}_1\cup\mathbf{W}_2} V\left(\mathbf{a},\mathbf{b},\mathbf{c}\right)$, где $V\left(\mathbf{a},\mathbf{b},\mathbf{c}\right)$ — объем паралелепипеда на $\mathbf{a}, \mathbf{b}, \mathbf{c}$.

Расстояние между элементами $\mathcal{L}$:
$\rho\left(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\right) = \sqrt{\left(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2\right)^{\mathsf{T}}\left(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2\right)}$.

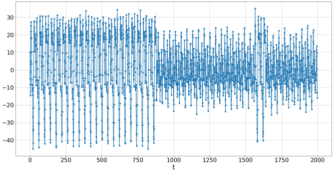
Расстояние между точками временного ряда:
$\rho\left(t_1, t_2\right) = \rho\left(\mathbf{W}_1, \mathbf{W}_2\right) + \rho\left(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\right)$.

Матрица попарных растояний:
$\mathbf{M} = [0, 1]^{N\times N}$.

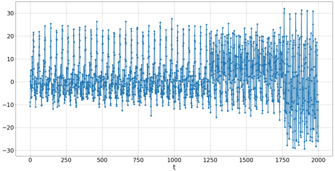
Таблица: Описание выборок

Physical Motion 1	900	2	20
Physical Motion 2	1000	2	20
Synthetic 1	2000	2	20
Synthetic 2	2000	3	20

Синтетические данные



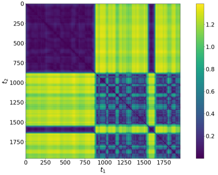
(a) Synthetic 1



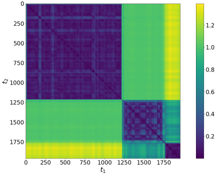
(b) Synthetic 2

Рис.: Пример синтетически построенных временных рядов

Синтетические данные



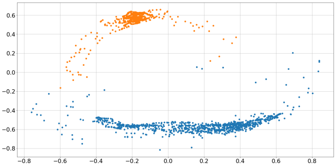
(a) Synthetic 1



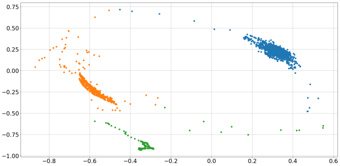
(b) Synthetic 2

Рис.: Матрица попарных расстояний **M** между точками временного ряда

Синтетические данные



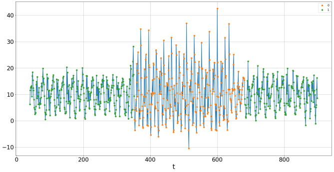
(a) Synthetic 1



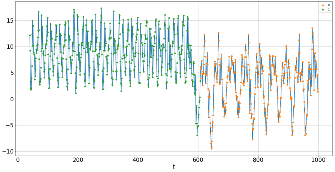
(b) Synthetic 2

Рис.: Проекция точек временного на плоскость при помощи матрицы попарных расстояний **M**

Синтетические данные

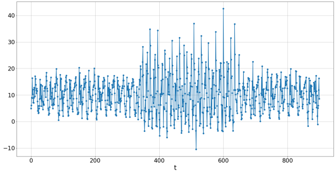


(a) Synthetic 1

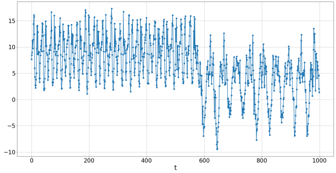


(b) Synthetic 2

Рис.: Кластеризация точек временного ряда

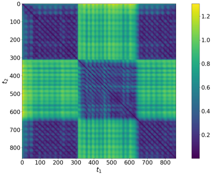


(a) Physical Motion 1

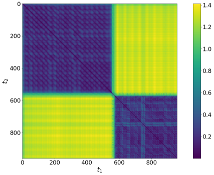


(b) Physical Motion 2

Рис.: Пример временных рядов полученных из мобильного акслерометра

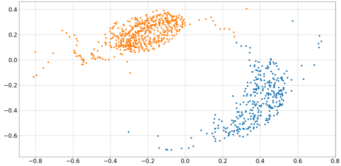


(a) Physical Motion 1

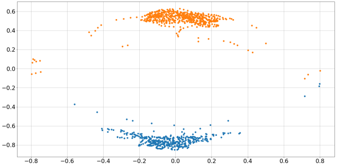


(b) Physical Motion 2

Рис.: Матрица попарных расстояний **M** между точками временного ряда

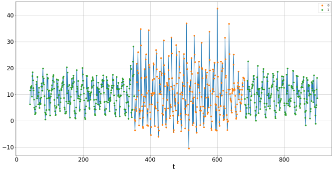


(a) Physical Motion 1

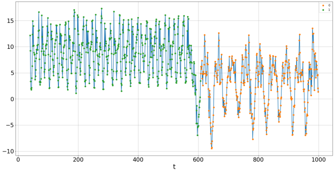


(b) Physical Motion 2

Рис.: Проекция точек временного на плоскость при помощи матрицы попарных расстояний **M**



(a) Physical Motion 1



(b) Physical Motion 2

Рис.: Кластеризация точек временного ряда

Заключение

- Был предложен алгоритм поиска характерных сигналов, который основывается на методе главных компонент для локального снижения размерности.
- Была предложена функция расстояния между локальными базисами в каждый момент времени, которые интерпретировались как признакового описание точки временного ряда.
- Предложенный алгоритм хорошо разделяет точки которые принадлежат разным классам сигналов, что хорошо для кластеризации точек временного ряда.

Публикации

- Грабовой А. В., Стрижов В. В. Локальная кластеризация временных рядов // (в процессе)
- Грабовой А. В., Бахтеев О. Ю., Стрижов В. В. Определение релевантности параметров нейросети // Информатика и ее применения, 2019, 13(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)
- Бучнев Т. Т., Грабовой А. В., Гадаев Т. Т., Стрижов В. В. Ранее прогнозирование достаточного объема выборки для обобщенно линейной модели // (в процессе)