

Спектральный анализ в классификации временных рядов*

Максим Хри스토любов, В. В. Стрижов

Московский физико-технический институт

В работе решается задача классификации моментов временных рядов на пересекающиеся классы. По временным рядам, снятых с показаний акселерометра, определяется движение, совершаемое человеком. Исследуется возможность декомпозиции сложных движений на элементарные действия. Для порождения информативных признаков, в пространстве которых происходит классификация и декомпозиция, используется алгоритм SSA "Гусеница".

Ключевые слова: временной ряд; классификация; SSA "Гусеница"; локально-аппроксимирующая модель, порождение признаков

1 Введение

Временные ряды являются объектами сложной структуры, требующие предварительной обработки и представления их в удобном для классификации виде. Необходимо отобрать исходный временной ряд в пространство признаков. Например, в [3] временной ряд аппроксимируется моделью авторегрессии, а признаками являются ее параметры. В [?] качестве аппроксимирующей модели используется модель сингулярного спектра, а признаками являются k наибольших собственных чисел траекторной матрицы участка временного ряда.

В статье изучается задача классификации движений человека по временным рядам и декомпозиция его движений на элементарные действия. В работе исследуется признаковое описание временных моментов ряда. В качестве признаков для классификации используются собственные числа траекторной матрицы участка ряда, предшествующего моменту времени. Новизна подхода заключается в том, чтобы не просто брать несколько наибольших собственных чисел траекторной матрицы, а учитывать частоты, которым соответствуют эти собственные числа, при классификации. На основе спектрального разложения производится построение пространства описаний элементарных действий.

2 Постановка задачи

Задан исходный временной ряд $\mathbf{d} = \{d_i\}_{i=1}^M \in \mathbb{R}^M$. Каждому моменту времени d_i соответствует множество меток классов $\mathbf{y}_i \subset Y$, где $|Y|$ — количество всех классов. Считается, что каждое движение человека представляется, как сумма периодических рядов, причем известен максимальный период T .

Требуется решить задачу классификации точек ряда:

$$R: \mathcal{I} \rightarrow 2^Y,$$

где $\mathcal{I} = \{1, \dots, M\}$ — моменты времени, на котором задан временной ряд, а Y — множество меток классов.

Каждый классифицируемый момент времени $i \in \{2T, \dots, M\} \subset \mathcal{I}$, отобразим с помощью $h: \mathcal{I} \rightarrow \mathbb{R}^{2T}$ в временной сегмент \mathbf{x}_i длины $2T$, содержащий локальную информацию о поведении ряда:

$$h(i) = \mathbf{x}_i = \{d_k\}_{k=i-2T+1}^i. \quad (1)$$

Полученные сегменты $\mathbf{x} \in \mathbb{X} \subseteq \mathbb{R}^{2T}$ — объекты сложной структуры, представленные временными рядами. Рассматривается задача классификации, а именно восстановление зависимости

$$\mathbf{y} = f(\mathbf{x}),$$

где $\mathbf{y} \in Y$. Тогда исходная задача классификации представляет собой $R = f \circ h$.

Данная постановка задачи естественным образом обобщается на случай, когда движения человека описываются S временными рядами с общими моментами времени \mathcal{I} . Этот случай отличается только тем, что компонентами временного ряда \mathbf{x}_i являются векторы $\{x_i^{(s)}\}_{s=1}^S$

Классификация временных рядов \mathbf{x} производится в пространстве признаков, порожденных с помощью алгоритма многомерной "Гусеницы" MSSA-L. Задача алгоритма MSSA-L состоит в представлении временного ряда в виде суммы интерпретируемых компонент. Поставим в соответствие временному ряду \mathbf{x} его траекторную матрицу Ганкеля $\mathbf{X} \in \mathbb{R}^{T \times ST}$:

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_S], \quad (2)$$

где T — ширина окна, $\mathbf{X}_s \in \mathbb{R}^{T \times T}$ — матрица Ганкеля для ряда $\mathbf{x}^{(s)}$

$$\mathbf{X}^{(s)} = \begin{pmatrix} x_1^{(s)} & x_2^{(s)} & \dots & x_T^{(s)} \\ x_2^{(s)} & x_3^{(s)} & \dots & x_{T+1}^{(s)} \\ \vdots & \vdots & \ddots & \vdots \\ x_T^{(s)} & x_{T+1}^{(s)} & \vdots & x_{2T}^{(s)} \end{pmatrix}.$$

Ее сингулярное разложение

$$\mathbf{X} = \mathbf{U} \text{diag} \mathbf{h} \mathbf{V} = \sum_{j=1}^d h_j \mathbf{u}_j \mathbf{v}_j^\top, \quad (3)$$

где $h_1 \dots h_T$ — сингулярные числа матрицы \mathbf{X} , которые будут использоваться как признаковое описание временного ряда \mathbf{x} .

По каждой компоненте в сингулярном разложении определяется его период φ_j :

TODO

Таким образом, каждому временному ряду \mathbf{x} сопоставляется вектор

$$\mathbf{h}(\mathbf{x}) = \{h_j\}_{j=1}^{|\Phi|} \in \mathbb{H}$$

, где Φ — множества всех периодов компонент $\Phi = \{\exists i \in \{2T, \dots, M\} \subset \mathcal{I} : \varphi \in \text{index}(\mathbf{h}_i)\}$

В качестве гипотезы порождения данных рассматривается следующее предположение: каждой метке класса $y \in Y$ соответствует вектор $\mathbf{h}_y \in \mathbb{H}$, а признаки временного ряда \mathbf{x} представляются как сумма

$$\mathbf{h}(\mathbf{x}) = \sum_{y \in f(\mathbf{x})} \mathbf{h}_y + \varepsilon, \quad \varepsilon = N(0, \sigma^2 I_{|Phi|})$$

. Составим из \mathbf{h}_y матрицу $\mathbf{H} = (\mathbf{h}_{y1}, \dots, \mathbf{h}_{yp})$, тогда гипотеза переформулируется в виде

$$\mathbf{h}(\mathbf{x}) = \mathbf{H} \mathbf{y}.$$

Искомая функция f запишется в виде $f(x, H) = \mathbf{H}^{-1}\mathbf{h}(\mathbf{x})$

Пусть после вышеописанных преобразований моментов временного ряда $i \in \mathcal{I}$ получена выборка $\mathfrak{D} = \{(\mathbf{h}_i, \mathbf{y}_i)\}_{i=2T}^M$. Задача состоит в нахождении функции матрицы H , минимизирующие суммарные потери на выборке \mathfrak{D} , при заданной функции потерь вида

$$\mathcal{L}(\mathfrak{D}, \mathbf{H}) = \sum_{i=2T}^M \mathcal{L}(\mathbf{h}_i, \mathbf{H}\mathbf{y}_i)$$

, характеризующей суммарную ошибку классификации на элементах выборки \mathfrak{D} .

В качестве базовой функции для минимизации рассматривается Средне Квадратичная Ошибка $\mathcal{L} = MSE$.

3 Вычислительный эксперимент

References

- [1] McNames J.. 1999. Innovations in local modeling for time series prediction // *Ph.D. Thesis, Stanford University*
- [2] Zhuravlev U.I., Ryazanov V. V., Senko O. V.. 2005. Recognition. Mathematical methods. Software system. Practical applications. // *Fazis, Moscow*
- [3] N. P. Ivkin, M. P. Kuznetsov. 2015. Time series classification algorithm using combined feature description. *Machine Learning and Data Analysis* (11):1471–1483.
- [4] Strijov V.V., Motrenko A.P.. 2016. Extracting fundamental periods to segment human motion time series. *Journal of Biomedical and Health Informatics* 20(6):1466 – 1476.
- [5] Grabovoy A.V., Strijov V.V. 2020. Quasiperiodic time series clustering for human activity recognition *Lobachevskii Journal of Mathematics*

Received