

# Классификация временных рядов\*

*Карасиков М. Е.*

karasikov@phystech.edu

Московский физико-технический институт

Работа посвящена задаче классификации временных рядов. Исследуются способы решения задачи классификации, основанные на выделении из временных рядов признаков. Временной ряд рассматривается как последовательность некоторых повторяющихся независимых сегментов, каждый из которых порождает свое признаковое описание. Для решения задачи классификации в качестве описаний временных рядов предлагается использовать восстановленные распределения признаков их сегментов. Проведено экспериментальное сравнение на реальных данных качества классификации временных рядов предложенными алгоритмами и наиболее распространенным методом ближайшего соседа с DTW метрикой.

**Последние изменения:** 13:36, 17 марта 2015 г.

**Ключевые слова:** *временные ряды, признаковая классификация.*

## Time series feature-based classification\*

*Karasikov M. E.*

Moscow Institute of Physics and Technology

### Введение

Временные ряды являются результатом проведения любых повторяющихся во времени измерений. Дальнейший анализ полученных данных диктуется целью конкретной задачи. Обзор по методам и проблемам анализа временных рядов дается в [1]. Одними из основных методов анализа временных рядов являются прогнозирование [], обнаружение аномалий [], сегментация [], кластеризация [] и классификация []. Последние годы связаны с ростом интереса к данной области, проявляющимся в предложениях новых методов анализа временных рядов — метрик [], алгоритмов сегментации [], кластеризации [], и др.

Временным рядом  $\mathcal{T}$  будем называть конечную упорядоченную последовательность точек в  $\mathbb{R}^n$ :

$$\mathcal{T} = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}, \text{ где } \mathbf{v}_t \in \mathbb{R}^n, t \in \{1, \dots, k\}.$$

В данной работе рассматривается задача классификации временных рядов, которая возникает во многих приложениях (медицинская диагностика по ЭКГ [], классификация физической активности по данным с акселерометра [] и т. д.) и заключается в определении неизвестных классов временных рядов рассматриваемого множества.

Формально задача классификации в общем виде может быть поставлена следующим образом. Пусть  $X = \{x_1, \dots, x_\ell\}$  — множество описаний объектов произвольной природы,  $Y = \{y_1, \dots, y_N\}$  — конечное множество меток классов. Существует целевая функция — отображение  $y : X \rightarrow Y$ , значения которого известны только на объектах обучающей выборки

$$\mathfrak{D} = \{(x^1, y^1), \dots, (x^m, y^m)\} \subset X \times Y.$$

Требуется построить алгоритм  $a : X \rightarrow Y$  — отображение, приближающее целевую функцию  $y$  на множестве  $X$ .

Таким образом, в рассматриваемой задаче классификации временных рядов под множеством объектов  $X$  понимается множество временных рядов  $T = \{\mathcal{T}_1, \dots, \mathcal{T}_\ell\}$ .

Этап построения информативного пространства признаков, позволяющего добиться заданной точности классификации, является одним из важнейших этапов решения задачи классификации.

Одним из способов построения пространства признаков является задание функции расстояния  $\|\cdot\|$  между временными рядами, позволяющего в качестве признаков взять расстояния до опорных объектов. Данный метод чрезвычайно распространен в силу того, что позволяет свести исходную задачу классификации временных рядов к задаче выбора метрики — функции расстояния. При удачном выборе метрики дальнейшая классификация может происходить при помощи простейших метрических алгоритмов классификации, например, методом ближайшего соседа  $\|\cdot\|$ .

Другой способ состоит в извлечении из каждого временного ряда набора признаков — его информативного описания, позволяющего строить точные классификаторы с хорошей обобщающей способностью. Здесь возникает задача выбора модели для описания временного ряда  $\|\cdot\|$ . При этом признаками могут быть буквально произвольные функции исходных объектов  $\mathcal{T}_i$ . В работе  $\|\cdot\|$  предлагается использовать в качестве признаков статистические функции (среднее, отклонения от среднего, коэффициенты эксцесса и др.). В работе  $\|\cdot\|$  в качестве признаков предлагается использовать коэффициенты дискретного преобразования Фурье (DFT). В  $\|\cdot\|$  предлагается использовать дискретное вейвлет-преобразование (DWT), которое сравнивается с предыдущими методами. Особенно эффективно DFT и DWT работают на квазипериодических рядах, где проявляется периодическая структура.

Под квазипериодичностью временного ряда будем понимать возможность выделения в нем характерных сегментов — периодов, то есть возможность представления каждого временного ряда  $\mathcal{T}$  последовательностью в определенном смысле похожих его сегментов  $\mathcal{S}_1, \dots, \mathcal{S}_p$ :

$$\mathcal{T} = \bigcup_{i=1}^p \mathcal{S}_i.$$

В нашей работе предлагается алгоритм классификации квазипериодических временных рядов, на основе восстановления распределения параметров сегментации временных рядов. Общий подход к классификации квазипериодических временных рядов изложен в разделе 1. В разделе 1 приводятся представлены эксперименты на реальных данных по сравнению предложенного подхода с подходом, предложенным в  $\|\cdot\|$ .

## Постановка задачи

## Вычислительный эксперимент

## Заключение

## Литература

- [1] *Esling P., Agon C.* Time-series data mining // *ACM Comput. Surv.* — 2012. — December. — Vol. 45, no. 1. — Pp. 12:1–12:34. <http://doi.acm.org/10.1145/2379776.2379788>.