

Динамическое выравнивание многомерных временных рядов*

Гончаров А. В., Моргачев Г. И., Смирнов В., Липницкая Т.

morgachev.gi@phystech.edu, smirnov.vs@phystech.edu, tanya.lipnizky@yandex.ru

МФТИ

Данная работа посвящена задаче получения расстояния между многомерными временными рядами с помощью алгоритма DTW и его оптимизаций. Траектория точки в 3х-мерном пространстве является примером многомерного временного ряда. При решении задачи нахождения расстояний исследуется наиболее оптимальный вид функции расстояния между двумя измерениями временного ряда. Поскольку нет полного обзора и исследований методов работы с временными рядами высокой размерности, необходимо модифицировать алгоритм и произвести исследование зависимости качества решения от расстояний между измерениями. В качестве прикладной задачи исследуется кластеризация данных мозговой активности обезьян. Качество кластеризации на основе полученной функции расстояния сравнивается с кластеризацией, полученной на основе авторегрессионной модели.

Ключевые слова: *временные ряды, многомерные временные ряды, DTW, авторегрессионная модель.*

1 Введение

В ряде задач возникает проблема анализа данных временных рядов. Для нахождения их сходства применяется функция расстояния, однако стандартный поточечный подход не является информативным [1]. Одним из способов решения этой проблемы является выравнивание временных рядов (DTW) [2] и его модификаций [3]. В работе рассматривается обобщение DTW на случай многомерных временных рядов. В работах [4], [5] предлагаются подходы к решению этой задачи. В работе [4] предлагается способ выравнивания многомерных рядов, основанный на нормализации исходных данных и нахождении векторной нормы. В работе [5] рассматривается алгоритм, позволяющий выполнить выравнивание временных рядов между координатами. В данной работе исследуется возможность обобщения DTW для выравнивания многомерных временных рядов относительно общей временной шкалы, а также рассматривается возможность выравнивания значений координат друг относительно друга.

Особенность этой работы заключается в использовании различных функций расстояния для определения их влияния на точность выравнивания. Используются такие меры как L_p , RBF, косинусное расстояние и прочие.

Перед работой над сравнением временных рядов рекомендуется провести их предварительную кластеризацию. Качество полученной функции расстояния будет оцениваться как качество кластеризации временных рядов метрическими методами.

С целью оценки качества полученных расстояний предлагается сравнивать качество кластеризации с кластеризацией на основании коэффициентов авторегрессионной модели для различных функций расстояния между выравненными рядами.

Расстояния между выравненными рядами получаются как сумма поточечных расстояний и используются для кластеризации методом k -средних.

* Работа выполнена при финансовой поддержке РФФИ, проект № 00-00-00000. Научный руководитель: Гончаров А. В. Задачу поставил: Гончаров А. В. Консультант: Гончаров А. В.

Данными для задачи служат данные мозговой активности обезьян и её зависимости от положения глаз. Данные представляют собой электрокортикограмму - зависимости потенциалов в 128 точках мозга и положения глаз от времени. В работе рассматривается кластеризация данных мозговой активности, что соответствует кластеризация 128-мерных временных рядов.

2 Постановка задачи

Пусть $s_i \in \mathbb{R}^n$ - временные ряды, являющиеся объектами некоторой выборки, ρ - некоторые выбранные метрики, Q - критерий качества (данные с соизмеримым количеством точек), являющийся ответом.

Требуется построить такую функцию выравнивания, что:

$$f = \operatorname{argmin}(Q(s_i, \rho))$$

3 Описание основных методов

Для построения функции выравнивания и проверки её качества используются модель DTW (и её оптимизации), кластеризация и авторегрессионная модель.

3.1 Описание функции расстояния между объектами

В данной работе в качестве метрического расстояния между объектами предлагается использовать строитость *пути наименьшей стоимости* между объектами.

Dynamic time warping - измерение расстояния между двумя временными рядами.

Задано два временных ряда, Q длины n и C длины m .

$$Q = q_1, q_2, \dots, q_i, \dots, q_n$$

$$C = c_1, c_2, \dots, c_j, \dots, c_m$$

Требуется построить матрицу размера $n \times m$ с элементами $D_{ij} = d(q_i, c_j)$, где d - выбранная метрика.

Чтобы найти наибольшее соответствие между рядами нужно найти выравнивающий путь W , который минимизирует расстояние между ними. W - набор смежных элементов матрицы D , $w_k = (i, j)_k$.

$$W = w_1, w_2, \dots, w_k, \dots, w_K$$

$\max(n, m) \leq K \leq m + n + 1$, где K -длина выравнивающего пути

Выравнивающий путь должен удовлетворять следующим условиям:

$$w_1 = (1, 1), w_K = (n, m)$$

$$w_k = (a, b), w_{k-1} = (a', b') : a - a' \leq 1, b - b' \leq 1$$

$$w_k = (a, b), w_{k-1} = (a', b') : a - a' \geq 0, b - b' \geq 0$$

Оптимальный выравнивающий путь должен минимизировать выравнивающую стоимость пути:

$$DTW(Q, C) = \sum_{k=1}^K w_k$$

Путь находится рекуррентно:

$\gamma(i, j) = d(q_i, c_j) + \min(\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1))$, где $\gamma(i, j)$ суммарное расстояние, $d(q_i, c_j)$ расстояние в текущей клетке.

3.2 Описание алгоритма классификации

todo

3.3 Описание модели авторегрессии

todo

4 Название секции

4.1 Название параграфа.

4.2 Теоретическую часть работы

5 Заключение

Литература

- [1] Hui Ding, Goce Trajcevski, Peter Scheuermann, Xiaoyue Wang, and Eamonn Keogh. *Querying and mining of time series data: Experimental comparison of representations and distance measures*, volume 1, pages 1542–1552. 2 edition, 8 2008.
- [2] Eamonn J. Keogh and Michael J. Pazzani. Derivative dynamic time warping. In *In SIAM International Conference on Data Mining*, 2001.
- [3] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal*, 11(5):561–580, 2007.
- [4] G.A. ten Holt, M.J.T. Reinders, and E.A. Hendriks. Multi-dimensional dynamic time warping for gesture recognition. In *Thirteenth annual conference of the Advanced School for Computing and Imaging*, 2007.
- [5] Parinya Sanguansat. Multiple multidimensional sequence alignment using generalized dynamic time warping. 2012.