

Построение метода динамического выравнивания многомерных временных рядов, устойчивого к локальным колебаниям сигнала.*

И. О. Автор¹, И. О. Соавтор², И. О. Фамилия^{1,2}

kulagin.pyoter@mail.ru; co-author@site.ru; co-author@site.ru

¹Организация, адрес; ²Организация, адрес

Данная работа посвящена построению эффективного алгоритма динамического выравнивания многомерных временных рядов. Для решения данной задачи предлагается использовать функцию расстояния DTW между двумя многомерными временными рядами, согласно которому выравниваются две оси времени, при этом внутри функционала DTW выбирается расстояние между i -м и j -м измерениями такое, что оно устойчиво к локальным “сдвигам” сигнала. В качестве решения будет рассмотрено более продвинутое, основанное на DTW между парой измерений. Для проверки корректности используются как и реальные данные, например измерения активность мозга обезьян, так и искусственно сгенерированные, например движение сигнала в пространстве по часовой и против часовой стрелки.

Ключевые слова: *многомерные временные ряды; DTW; динамическое выравнивание.*

1 Введение

В данной работе исследуется проблема динамического выравнивания многомерных временных рядов, устойчивого к локальным колебаниям сигнала. Временной ряд - собранный в разные моменты времени статистический материал о значении каких-либо параметров (в простейшем случае одного исследуемого процесса). В данном случае рассматривается многомерный случай.

Необходимость сравнивать 2 временных ряда на схожесть возникает из практических задач таких как классификация геномных сигналов [1], распознавание жестов [2].

При этом необходимо учитывать, что в таких задачах как ... возникает необходимо учитывать, что близость временных рядов не ограничивается сдвигами по времени, а также присутствует расположением сигналов.

Для сравнения определения расстояния между временными рядами будет использовать известный алгоритм DTW [3] [4], однако возникает проблема определения расстояния между сигналами в этом алгоритме.

Базовое решение задачи с помощью метрики L2 расстояния между рядами [5] не всегда оказывается эффективным. Таким примером являются 2 временных ряда, полученные при близком расположении датчиков с сигналами, которые могут зафиксировать один и тот же пик. Полученный пик окажет большое влияние на значение метрики L2.

Напротив, использование известного алгоритма DTW, но уже в многомерном случае позволит обойти проблему малого расстояния между датчиками, как более частный метод для решения конкретного типа задач.

Итак, в данной работе рассматриваются 2 варианта расстояния между сигналами L2 и DTW, выравнивание по временной оси будет сглаживать DTW.

Полученные алгоритмы тестировались на реальных данных и искусственно сгенерированных. Полученные результаты показали преимущество использования попарного DTW алгоритма.

*Работа выполнена при финансовой поддержке РФФИ, проекты № №00-00-00000 и 00-00-00001.

27 2 Постановка задачи.

28 Рассматривается алгоритм решения задачи динамического выравнивания многомер-
29 ных рядов.

30 Задача классификации временных рядов имеет вид.

31 Пусть X - множество временных рядов, Y - множество меток классов.

32 $\forall x \in X x \in \mathbb{R}^K$

33 Рассматриваем $Y = \{0, 1\}$, $|Y| = 2$, то есть задачу бинарной классификации.

34 Предполагается существование зависимости $f : X \rightarrow Y$

35 Задача заключается в поиске оптимального алгоритма из семейства $A = \{a_\alpha\}$, пред-
36 сказывающего класс из множества Y .

37 Семейство алгоритмов задаётся

38 Пусть $X^n = \{X_1, X_2 \dots X_n\}$ - обучающая выборка из временных рядов, $Y^n =$
39 $= \{Y_1, Y_2 \dots Y_n\}$ - метки классов.

40 В качестве семейства алгоритмов классификации будем использовать метод k ближай-
41 ших соседей с некоторой

42 метрикой ρ - расстоянием между 2 временными рядами.

43 Оптимизируемый функционал $accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

44 Хотим предоставить $\rho = \arg \max(accuracy)$ - как можно ближе к оптимальному.

45 При этом считаем, что данные представляют собой не произвольной временной ряд, а
46 временной ряд

47 у которого известны местоположения датчиков в пространстве, а сигнал может иска-
48 жаться и улавливаться соседними датчиками.

49 3 Рассматриваемые метрики

50 В данной работе будут рассматриваться и сравниваться 2 метрики, кото-
51 рые будут введены в этом разделе.

52 Введём функцию расстояния между 2 временными рядами с помощью алгоритма
53 DTW.

54 $D(i, j) = d(i, j) + \min(D(i-1, j), D(i, j-1), D(i-1, j-1))$

55 $d(i, j)$ - расстояние между парой сигналов. Определяется 2 разными способами.

56 1) L2 - метрика.

57 $d(i, j) = \sum_{k=1}^K (X(i, k) - X(j, k))^2$

58 2) DTW - расстояние.

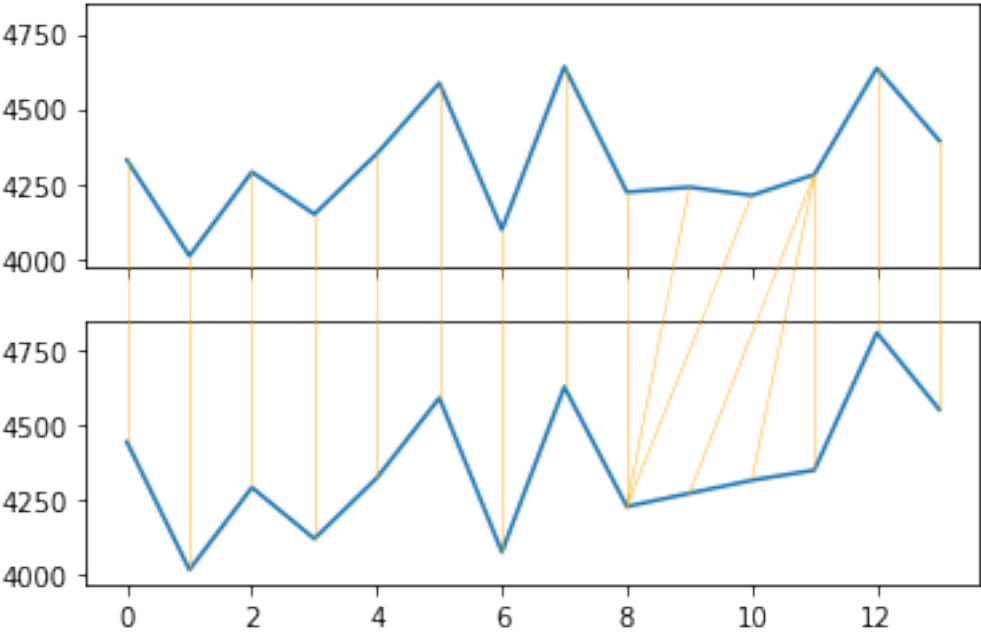
59 $d(i, j, k, t) = (X(i, k) - X(j, t))^2 + \min(d(i, j, k-1, t), d(i, j, k, t-1), d(i, j, k-1, t-1))$

60 $d(i, j) := d(i, j, K, K)$

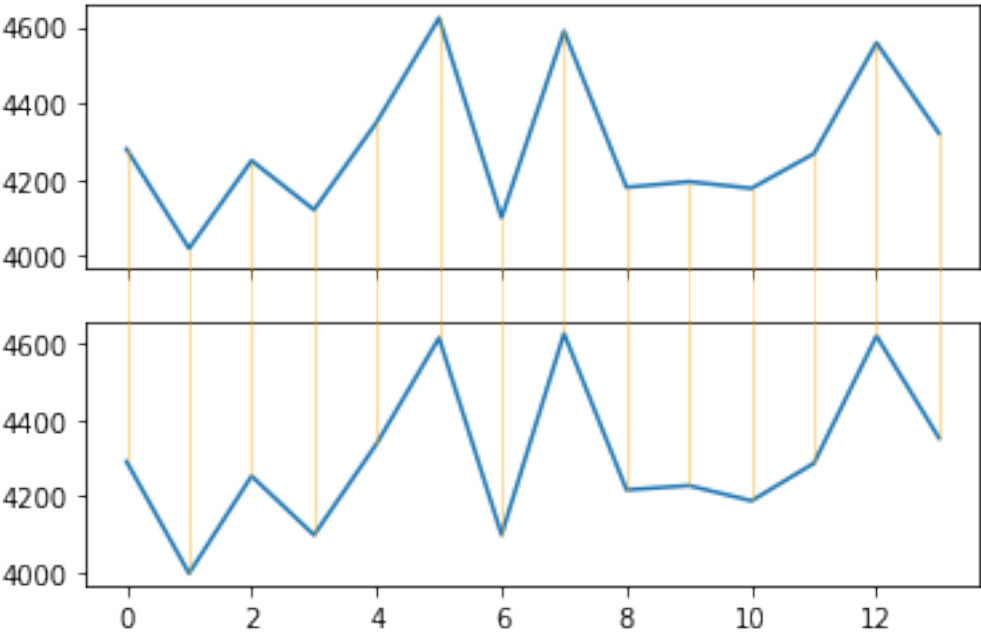
61 Будем исследовать метрики L2 и DTW, задающие различные варианты расстояний
62 между сигналами.

63 Пример выравнивания для DTW временного ряда с 14 измерителями.

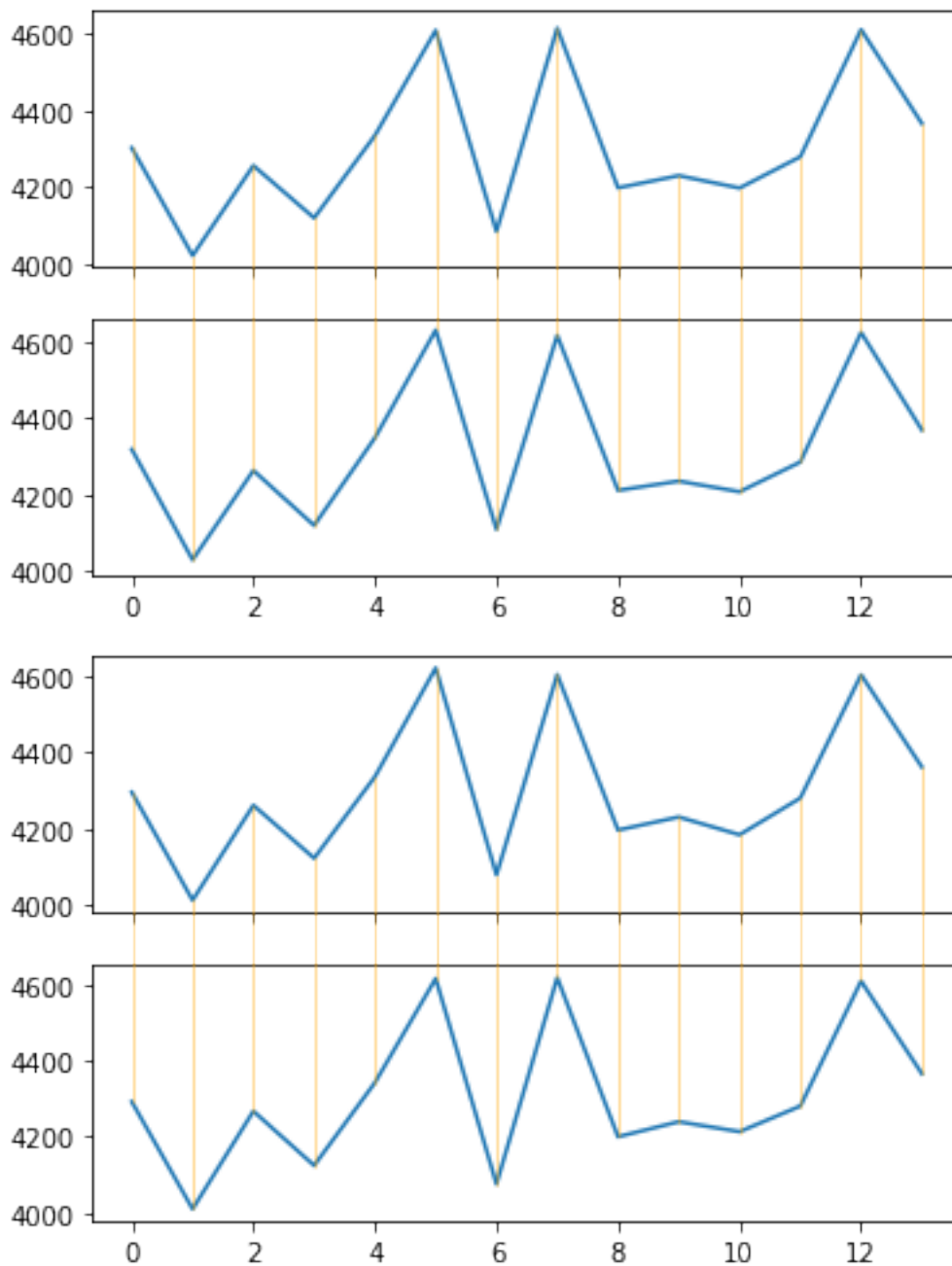
64 Для DTW



65



66



67

68

69 4 Вычислительный эксперимент

70 4.1 Цель эксперимента

71 Данный преследует 3 цели.

72 1) Найти примеры временных рядов, удовлетворяющих свойству пространственности
73 измерителей.

74 2) Измерить время работы 2 методов выравнивания.

75 3) Решить задачу классификации для многомерных временных рядов в предположении
76 о данных сделанное ранее и сравнить точность предсказания.

77 2 пункт важен, так как время работы алгоритмов отличается довольно сильно.

78 4.2 Эксперимент на искусственных данных

79 Рассмотрим примеры данных, полученные искусственной генерацией.

1) Сигнал перемещается по часовой стрелке вокруг одного набора датчиков и против часовой стрелки для другого набора датчиков.

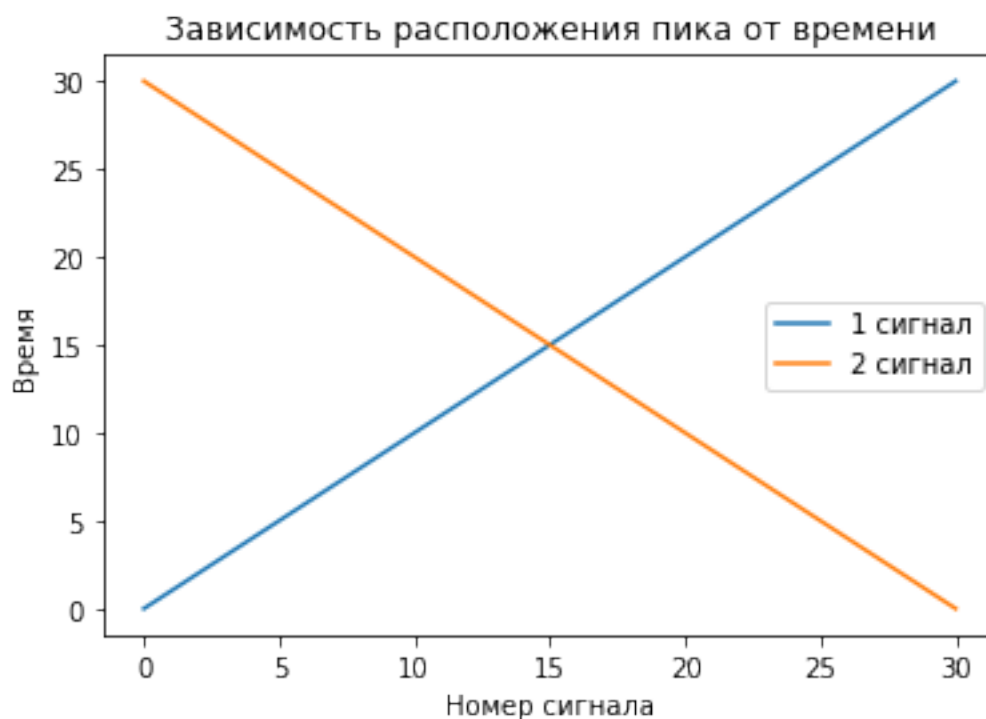
Более формально, рассматриваем 2 временных ряда t_1 и t_2 длины k и размера k .

Условимся все индексы векторов нумеровать с 0.

Тогда для i измерения в ряде t_1 i сигнал будет равен 1, а остальные нули.

В ряде t_2 единицы будут стоять на местах $k - i - 1$ для i измерения.

На графике показано движение сигнала для 2 временных рядов.



2) Сигнал перемещается в 2 раза быстрее вокруг одного из наборов датчиков.

Более формально, рассматриваем 2 временных ряда t_1 и t_2 длины k и размера k .

Условимся все индексы векторов нумеровать с 0.

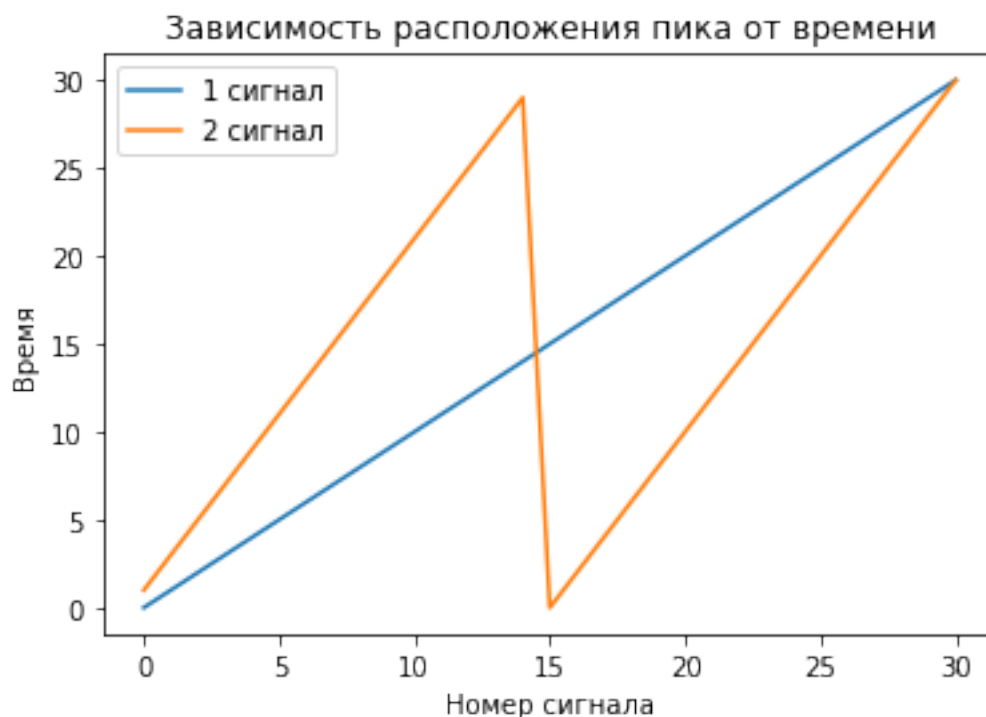
Тогда для i измерения в ряде t_1 i сигнал будет равен 1, а остальные нули.

В ряде t_2 единицы будут стоять на местах $(2 * i + 1) \bmod k$ для i измерения.

Пример хорошо используется только для нечётного k .

На графике также показано движение сигнала для 2 временных рядов.

На графике показано движение сигнала для 2 временных рядов.



Если измерить их по метрике L2 и DTW, то в 1 случае получатся значения 60 и 4, а во втором - 60 и 2. То есть алгоритм DTW имеет большую необходимость в этом случае.

4.3 Временной эксперимент

Временной эксперимент имеет также имеет огромное значение, так как использование алгоритма может быть затруднено в силу низкой скорости обработки, и соответственно малым количеством данных для обучения.

Эксперимент со временем запускался на данных с

размерностью сигнала: 3

количество измерений: 50

количеством временных рядов: 20

В итоге зависимость среднего времени измерения расстояния (использовалось 100 случайных временных рядов) от количества измерений

представлена на следующем графике.

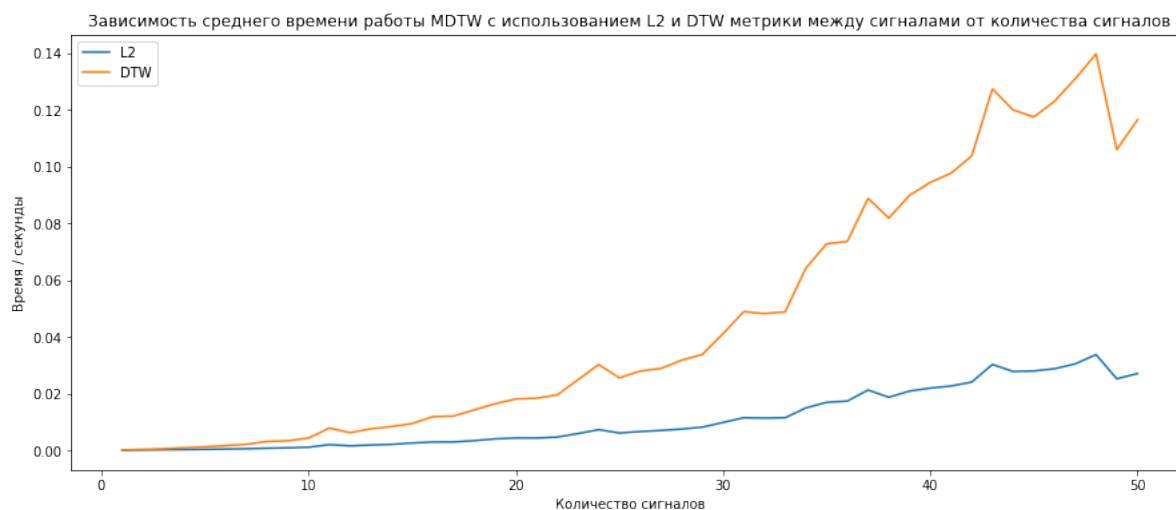
4.4 Эксперимент на реальных данных

Рассматриваем набор данных, соответствующих EEG-тесту на глаза, который по активности участков головного мозга показывает открыты глаза и пациента или нет.

Заметим, что эксперимент удовлетворяет гипотезе о данных, так как измерение происходит по датчикам в головном мозге.

В полученном эксперименте замерялось, в каком проценте случаев DTW-метрика получает лучшие результаты, нежели L2.

Итого, MDTW показала больший ассигасу в 9% случаев, в остальном - 81% - обе метрики показали сходные результаты.



Можно сделать вывод, что DTW работает медленнее и не на всех данных следует его использовать.

5 Заключение

Было продемонстрировано значение метрики DTW для поиска расстояний между сигналами в многомерном случае. Однако затрачиваемое время на порядок медленнее L2 алгоритма.

Поэтому стоит использовать L2 или DTW в зависимости от задачи и временных ограничений.

Литература

- [1] Babula Petr Kizek Rene Provaznik Ivo Skutkova Helena, Vitek Martin.
- [2] Gineke ten Holt, Marcel Reinders, and Emile Hendriks. Multi-dimensional dynamic time warping for gesture recognition. *Annual Conference of the Advanced School for Computing and Imaging*, 01 2007.
- [3] Eamonn J. Keogh and Michael J. Pazzani. Derivative dynamic time warping. In *SIAM International Conference on Data Mining*, 2001.
- [4] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal*, 11(5):561–580, 2007.
- [5] P. Sanguansat. Multiple multidimensional sequence alignment using generalized dynamic time warping. *WSEAS Transactions on Mathematics*, 11:684–694, 08 2012.
- [6] Jörg P. Bachmann and Johann-Christoph Freytag. High dimensional time series generators. *CoRR*, abs/1804.06352, 2018.
- [7] Nurjahan Begum, Liudmila Ulanova, Hoang Anh Dau, Jun Wang, and Eamonn J. Keogh. A general framework for density based time series clustering exploiting a novel admissible pruning strategy. *CoRR*, abs/1612.00637, 2016.

Поступила в редакцию