

Содержание

1. Постановка задачи.
2. Алгоритмы решения.
3. Евклидова мера.
4. DTW-based алгоритмы.
5. Shapelets алгоритмы.
6. Interval алгоритмы.
7. CNN алгоритмы.

Постановка задачи

Имея набор временных рядов с размеченной принадлежностью классов, научиться определять класс для неразмеченных данных - классическая задача классификации.

$D = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N)\}$ - где $Y \in (Y_1, \dots, Y_k)$, где K - количество классов

Проблема

Все временные ряды имеют разную длину, и использовать в качестве признаков сами значения ряда не получится (в классической постановке задачи).

$$|X_i| \neq |X_k|$$

Алгоритмы

- Основанные на расстоянии
 - Евклидово расстояние
 - Dynamic time warping
 - Shapelets
- Основанные на признаках
 - Interval алгоритмы
 - CNN алгоритмы

Евклидово расстояние

Для двух временных рядов X , Y длины N евклидово расстояние определяется как сумма квадратов расстояний от i -ой точки одного ряда до i -ой точки другого ряда.

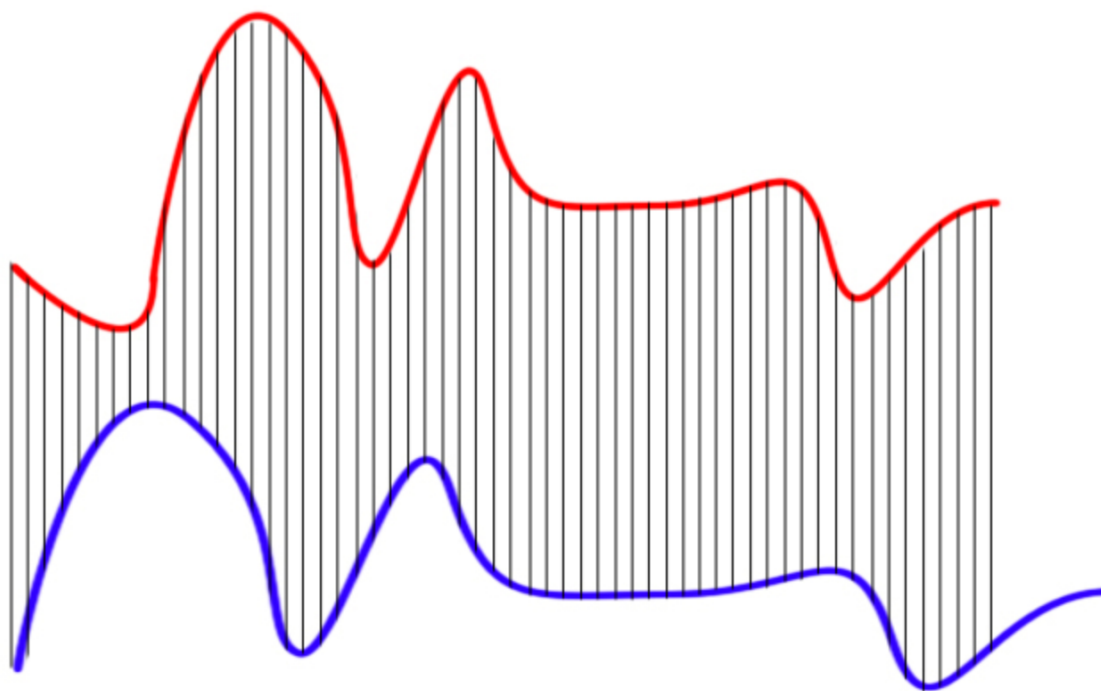
$$D = (X_i - Y_i)^2, i \in (0, N)$$

Алгоритм считает данную меру для всех рядов их обучающей выборки и присваивает ряду класс с наименьшим евклидовым расстоянием - по сути one nearest neighbour алгоритм.

Проблема

Если два одинаковых ряда будут сдвинуты относительно друг друга, евклидово расстояние для них не сработает.

Пример



Euclidean Matching

Решение проблемы - DTW

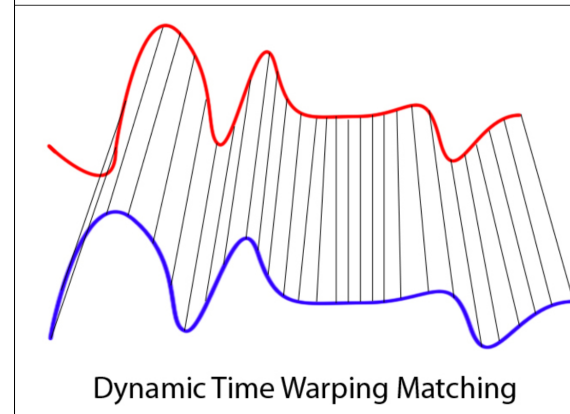
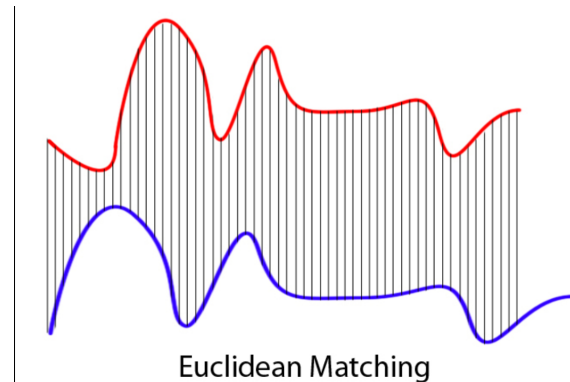
DTW - dynamic time warping

Изначально был придуман для сравнения фраз в анализе звуковых последовательностей.

По-прежнему является очень сильным baseline для задач классификации временных рядов.

Идея DTW

Для ряда X^n найти оптимальный путь из первой точки ряда до последней точки ряда X^m .



Алгоритм

Пусть есть два ряда $X = x_1, x_2, \dots, x_n$ и $Y = y_1, y_2, \dots, y_m$

1. Построим матрицу расстояний d - $d(x_i, y_j) = |x_i - y_j|$

2. Построим матрицу трансформаций D

$$D(i, j) = d_{ij} + \min(D_{i-1, j}, D_{i-1, j-1}, D_{i, j-1})$$

3. Найти минимальный путь $W = (w_1, w_2, \dots, w_k)$, удовлетворяющий следующим условиям:

- * начало пути $w_1 = (1, 1)$, конец пути $w_k = (n, m)$

- * оба индекса i, j могут увеличиваться только на 1 на каждом шаге

- * оба индекса i, j не могут уменьшаться

4. Вычислить DTW по формуле

$$DTW(Q, C) = \min \left\{ \frac{\sum_{k=1}^K d(w_k)}{K} \right\}$$

Алгоритм

Визуализация работы алгоритма

[https://upload.wikimedia.org/wikipedia/commons/8/84/
Animation_Dynamic_Time_Warping.gif](https://upload.wikimedia.org/wikipedia/commons/8/84/Animation_Dynamic_Time_Warping.gif)

DTW - семейство

На основе DTW построено большое количество алгоритмов, в частности:

- Time warp edit distance
- Weighted dynamic time warping
- Derivative transform distance
- Derivative DTW
- И др.

Подробнее - <https://timeseriesclassification.com/algorithm.php>

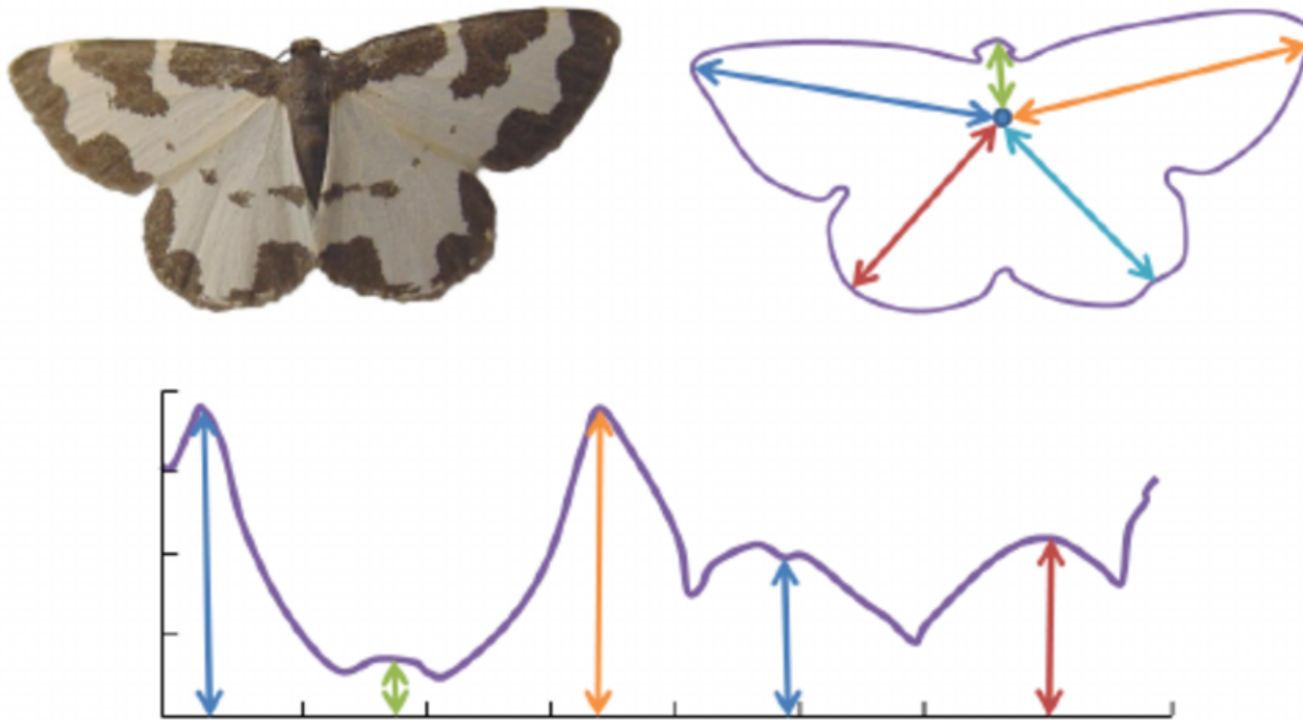
Примечание*

Для классификации изображений можно использовать подход к классификации временных рядов.

Вопрос - как превратить изображение во временной ряд?

Ответ - выбрать некоторую точку на изображении (обычно в центре) и рассчитать расстояния от этой точки до всех точек контура, поворачивая вектор с равным градусом.

Примечание* пример



Shapelets

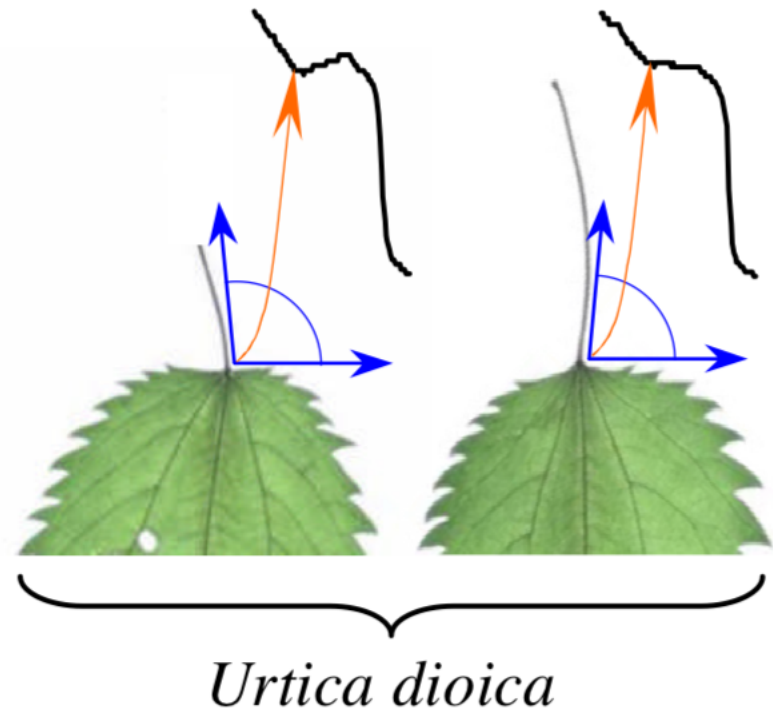
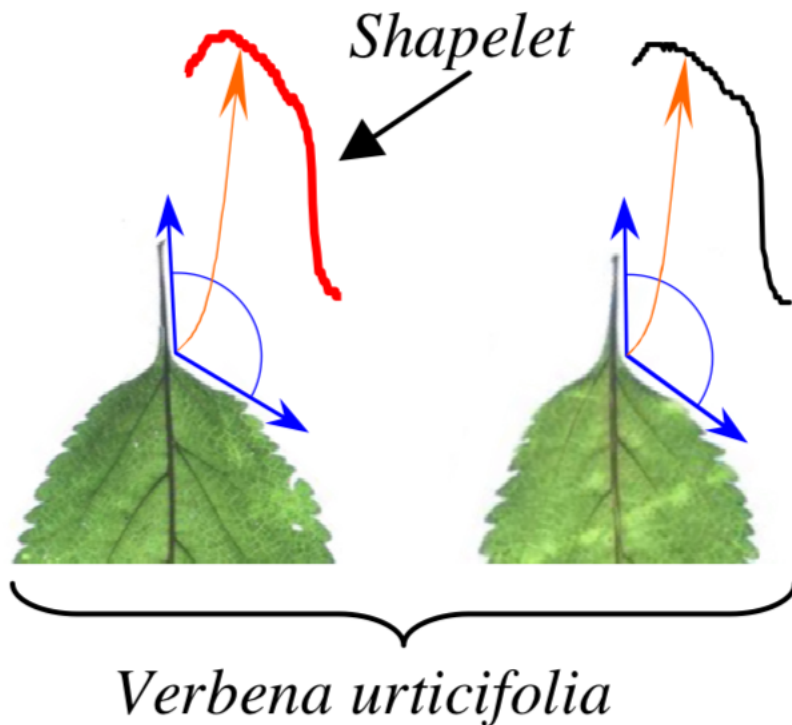
Проблема DTW в том, что он сильно чувствителен к шуму.

Пример ниже - разность в цвете и потертости листьев нивелирует разность в виде листа.



Shapelets

Идея shapelets заключается в том, чтобы выбрать наиболее характерную подпоследовательность временного ряда, которая описывает отличие между классами.



Shapelets

Далее алгоритм будет заключаться в том, что для каждого временного ряда по определенному алгоритму выбирается набор подпоследовательностей и если найдется такая, что даст сходство с выбранным shapelet-ом - ряд классифицируется по данному shapelet-у.

Подробнее в статье по ссылке -

<https://www.cs.ucr.edu/~eamonn/shaplet.pdf>

Feature-based алгоритмы

Основанная идея:

Feature-based алгоритмы основаны на получении вектора признаков для каждого временного ряда.

Далее вектор признаков подается в алгоритм классификации, например RandomForestClassifier.

Interval алгоритмы

Данное семейство алгоритмов основано на генерации признаков из набора подпоследовательностей временного ряда, выбираемых произвольным образом.

Признаками для каждой подпоследовательности X_{sub} могут быть:

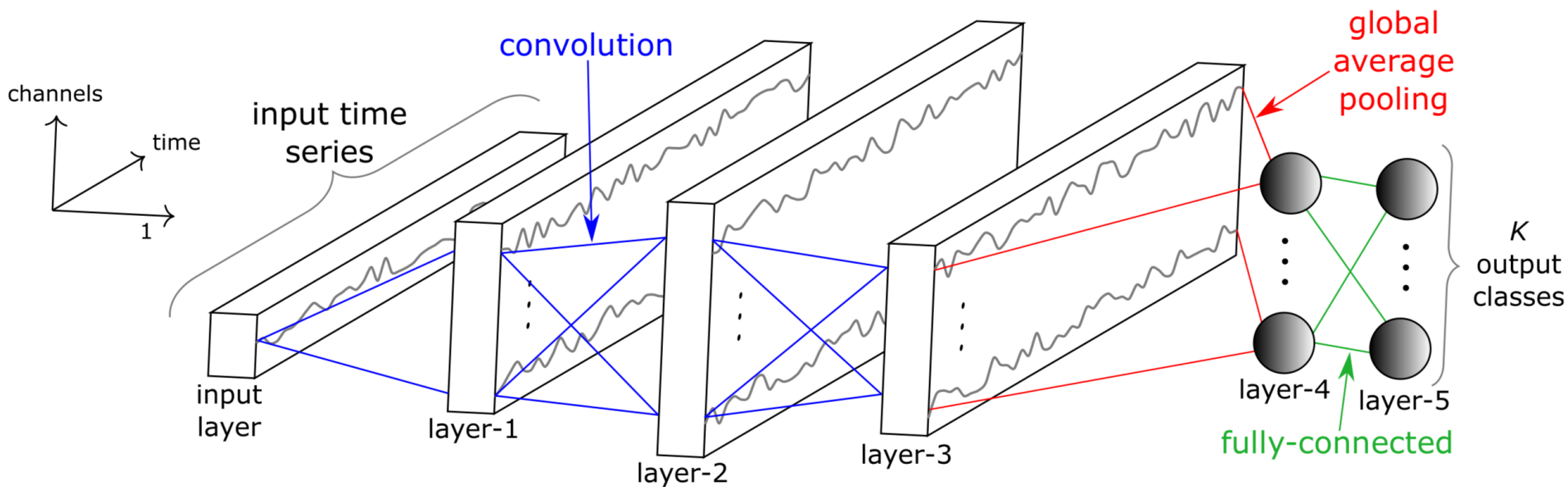
- $VAR(X_{sub})$
- $STD(X_{sub})$
- $FFT(X_{sub})$ - частоты из быстрого преобразования Фурье
- $MEAN(X_{sub})$
- И т.д.

Пример с SVM -

https://link.springer.com/chapter/10.1007/1-84628-102-4_18

CNN алгоритмы

Идея данного метода основана на том, что скользящий фильтр для временного ряда по сути представляет собой одномерную конволюцию, соответственно, можно попробовать использовать CNN на скользящем окне.



Материалы

- <https://timeseriesclassification.com/index.php> - главный источник информации по классификации временных рядов, обзор всех алгоритмов классификации. Есть исходный код, результаты полностью воспроизводимы.
- <https://arxiv.org/pdf/1809.04356.pdf> - CNN для классификации временных рядов.
-