Parameter Free Dynamic Time Warping

V.S. SIYOU FOTSO, E. Mephu Nguifo, P. Vaslin

Laboratoire d'Informatique, de Modelisation et d'Optimisation des Systemes
Clermont Auvergne University

ROADEF- 22/02/2017

Summary

Taking into account the temporal distortion

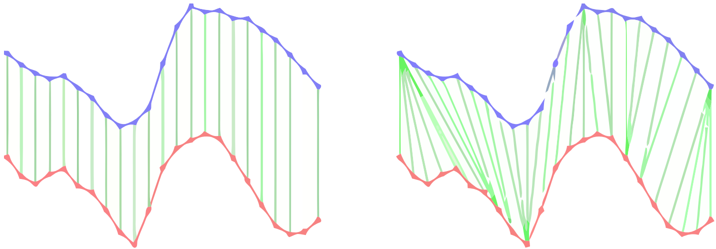


Figure: Euclidean distance (left) - DTW (right)

The dynamic time warping algorithm

Three constraints must be met to align two time series taking into account the time distortion

- The **boundary** condition: The first (respectively last) point of both time series must be aligned.
- The **monotony** condition: during alignment there is no return to a point which has already been used.
- The **continuity** condition: when aligning all data points are considered

The dynamic time warping algorithm

Alignment example

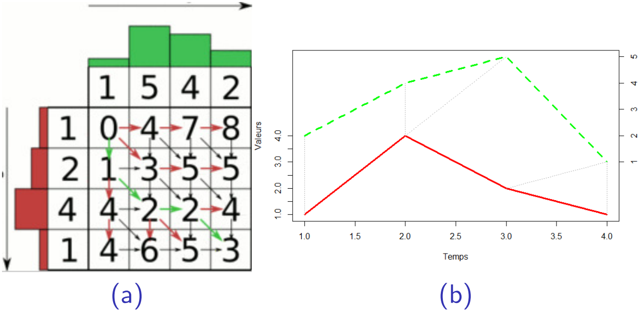


Figure: Alignment example with DTW

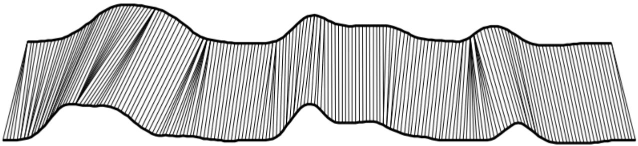
DTW's Problems

- Sensitivity to uncertainty
- Time consuming computation

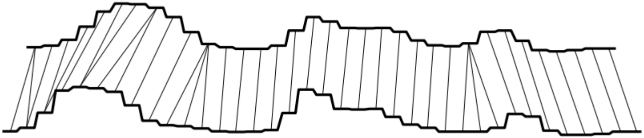
Piecewice DTW to speed-up DTW

1. Piecewice aggregate the time series to reduce their length
2. Applied DTW

Piecewice DTW to speed-up DTW



(a) DTW



(b) PDTW

Figure: Keogh et Al, KDD 2000

Piecewice DTW to speed-up DTW

How to choose the number of segments ?

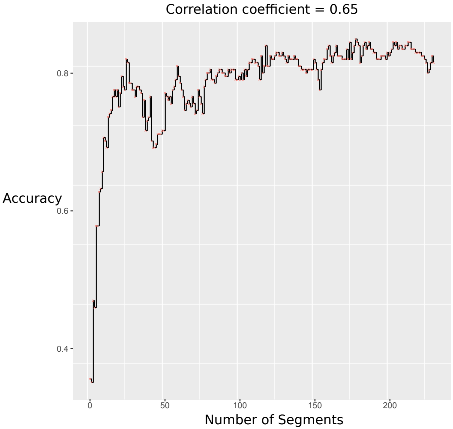


Figure: Relation between Accuracy and the number of segment on FISH dataset

Lignes directrices

Iterative Deepening Dynamic Time Warping

Considers the number of segments which are powers of two

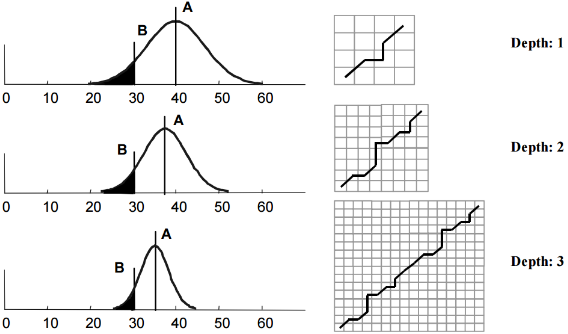


Figure: Shu et Al, SDM 2002

Lignes directrices

FDTW: Definitions

Definition

$X = x_1, \cdots, x_n$is a sequence of numerical values representing the evolution of a specific quantity during the time. x_n is the most recent value.

Definition

A segment X_i of length l of the time series X of length n ($l < n$) is a sequence constituted by l consecutive variables of X starting at the position i and ending at the position $i + l - 1$. We have:
$X_i = x_i, x_{i+1}, ..., x_{i+l-1}$

FDTW: Definitions

Definition

The arithmetic average of the data points of a segment X_i of length l is noted $\bar{X}_i$ and is defined by:

$$\bar{X}_i = \frac{1}{l}\sum_{j=0}^{l-1} x_{i+j}$$

FDTW: Definitions

Definition

Let T be the set of time series. The Piecewise Aggregate Approximation (PAA) is defined as follows:

$$PAA : T \times \mathbb{N}^{\star} \rightarrow T$$

$$(X, N) \mapsto PAA(X, N) \quad = \quad \begin{cases} \bar{X}_1, \cdots, \bar{X}_N \textit{ if } N < |X| \\ X \textit{ otherwise} \end{cases}$$

Definition

Let $N \in \mathbb{N}^{\star}$, X and Y be two time series.

$$PDTW(X, Y, N) = DTW(PAA(X, N), PAA(Y, N)).$$

FDTW : Heuristic

1. We choose N_c candidates distributed in the space of possible values small, large, medium

If the length of time series is $n = 12$ and the number of candidates is $N_c = 4$, we are going to select the candidates 12, 9, 6, 3.

Example

1, 2, [3], 4, 5, [6], 7, 8, [9], 10, 11, [12]

FDTW: Heuristic

1. We choose N_c candidates distributed in the space of possible values small, large, medium

If the length of time series is $n = 12$ and the number of candidates is $N_c = 4$, we are going to select the candidates 12, 9, 6, 3.

Example

1, 2, [3], 4, 5, [6], 7, 8, [9], 10, 11, [12]

FDTW: Heuristic

2. We select the candidate that has the minimal classification error with 1NNPDTW

In our example, we may suppose that we get the minimal value with the candidate 6 : it is thus the best candidate at this step.

Example

1, 2, 3, 4, 5, [6], 7, 8, 9, 10, 11, 12

FDTW: Heuristic

3. We respectively look between the predecessor and successor of the best candidate for a number of segments with a lower classification error. This number of segments corresponds to a local minimum.

In our example, we are going to test the values 4, 5, 7 and 8 to see if there is a local minimum.

Example

1, 2, **3**, [4], [5], **6**, [7], [8], **9**, 10, 11, 12

FDTW: Heuristic

4. We loop

We restart at step one, while choosing differents candidates during each iteration to ensure that we return a good local minimum. We fix the number of iterations to $\lfloor log(n) \rfloor$.

FDTW

Fact

In summary, in the worst case, we test the N_c first candidates to find the best one. Then, we test $\frac{2n}{N_c}$ other candidates to find the local minimum. We finally perform $nb(N_c) = N_c + \frac{2n}{N_c}$ tests. The minimal number of tests is done when the number of candidates $N_c = \sqrt{2n}$.

FDTW

Corollary

For a given a dataset d_i $FDTW(d_i) \leq 1NNDTW(d_i)$. The quality of the alignment of our heuristic is better than that of DTW.

Proof.

$1NNDTW(d_i) = 1NNPDTW(d_i, n)$. $1NNDTW(d_i)$is then one of the candidate considered by the heurisitic $FDTW$. Since $FDTW$returns the minimal classification error from all candidates, the classification error of $1NNDTW$is always greater than or equal to $FDTW$. □

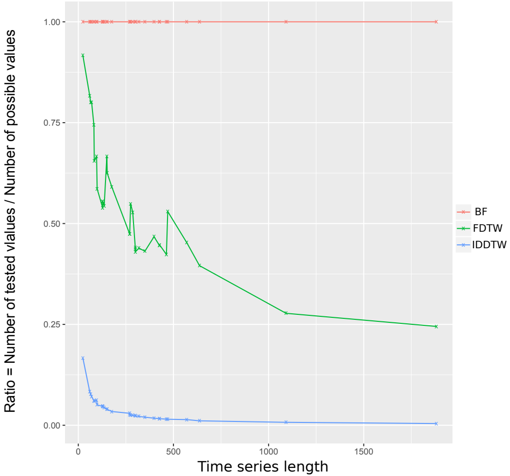
FDTW

Corollary

For a given dataset d_i that has c_i classes, $c_i \in \mathbb{N}^*$,
$acc_{DTW} \geq \frac{1}{c_i} \implies \frac{1}{c_i} \times acc_{max} \leq acc_{FDTW} \leq acc_{max}$

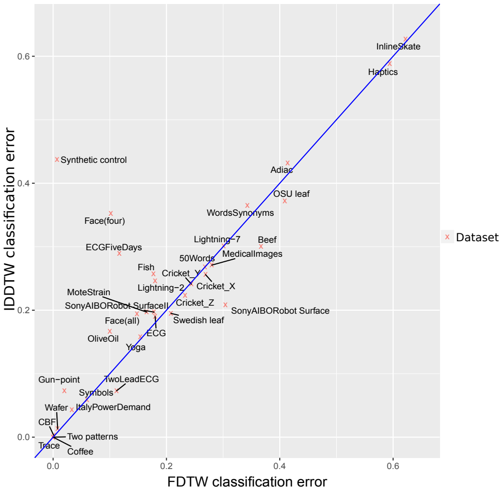
Lignes directrices

Number of candidates tested



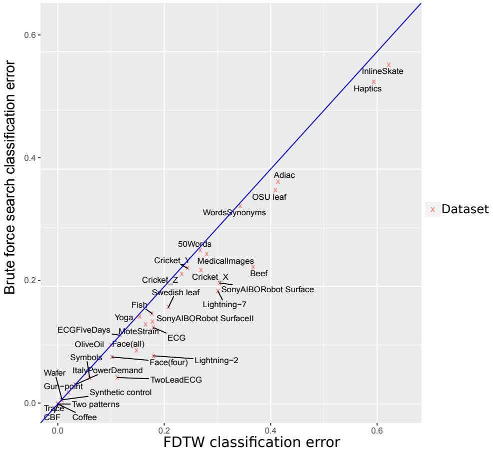
Pairwise comparison

Comparison of the classification error of FDTW in x-axis and IDDTW in y-axis. The points above the diagonal represent the datasets for which FDTW is better than IDDTW



Pairwise comparison

Comparison of the classification error of FDTW in x-axis and BF in y-axis. The points on the diagonal represent the datasets for which FDTW has found the optimal value



Conclusion

- FDTW allows to reduce the storage space and the processing time of time series classification without decreasing the alignment quality.
- Number of segments to be considered for symbolic representations of time series like SAX, ESAX, SAX-TD.

Questions ?

We thank the Ministry of Higher Education and Research for funding this work.