DTW Mean: Time Series Averaging and k-Means Clustering under Dynamic Time Warping for Time Series with Missing Values

David Schultz

TU Berlin, Germany

Abstract

DTW Mean is a Matlab library that provides implementations of two mean algorithms for computing a sample mean of time series under Dynamic Time Warping (DTW). The time series can be multivariate and of varying length. The algorithms return heuristic solutions to the NP-hard DTW mean problem. Moreover, the library contains a DTW-based k-means implementation, which is suitable for warping invariant time series clustering. Optionally, the sample time series can have missing values. Those will be estimated during the clustering process.

Keywords: Dynamic time warping, Time series averaging, Sample mean, Fréchet function, Subgradient methods

1. The DTW Mean Problem

Time series are data sequences that consist of time-dependent observations. Examples of time series include pollution sensor data, wheather recordings, audio signals and motion gestures. Time series often vary in length and speed. To cope with these temporal variations, the Dynamic Time Warping (DTW) distance is frequently used. The idea of DTW is to align two time series such that the Euclidian distance of the aligned time series is minimal.

Depending on the problem, a Euclidean sample mean (i.e. arithmetic mean) of a sample of equal-length time series may not be appropriate. Particularly, if the sample time series are affected by distortions along the time series, a Euclidean mean often smooths the shape of the time series which can lead to loss of information. This problem is illustrated in Figure 1. In this figure, the two sample time series hav a similar shape, both start flat, the have a peak followed by a valley and go back to a flat region. However, the shapes are affected by distortions along the time axis so that in Euclidean view the peak of the upper time series is aligned to the valley of the lower time series. Hence, the Euclidean mean smooths out these shapes. In contrast, the DTW mean aligns the time series properly and builds the mean based on these alignments.

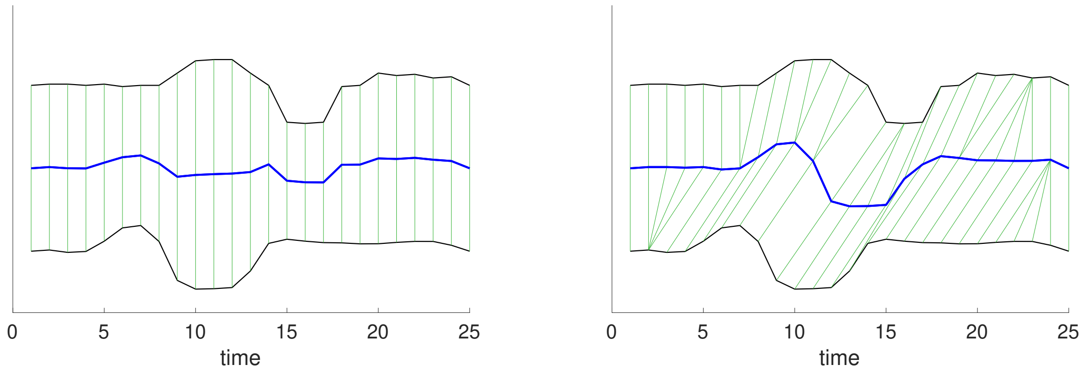
To formalize the DTW Mean problem, suppose that $\mathcal{X} = \left(x^{(1)}, \dots, x^{(N)}\right)$ is a sample of N time series $x^{(i)}$. Then a (sample) mean of $\mathcal{X}$ under DTW is any time series that minimizes the Fréchet function

$$F(x) = \frac{1}{N} \sum_{k=1}^{N} \mathrm{dtw}^2\Big(x, x^{(k)}\Big),$$

where dtw is the DTW distance. Finding a sample mean under DTW is NP-hard [1].

2. Sample mean algorithms

DTW Mean provides two heuristic strategies to optimize the Fréchet function, a stochastic subgradient (SSG) method and a majorize minimize (MM) algorithm, which is also known as



(a) Euclidean mean (b) DTW mean

Figure 1: Two sample time series (black) and their means (blue). The green lines illustrate the alignments.

Algorithm 1 k-means DTW

procedure K-MEANS_DTW($X = (x_1, \dots, x_N), k$)
  Randomly initialize centroids $\mu_1, \dots, \mu_k \in X$
  Initialize clusters $C_1 = \{\}, \dots, C_k = \{\}$
  repeat (iteration t)
    `//*** Cluster assignment step **********************************//`
    for all $i = 1, \dots, k$ **do**
      $C_i^{(t)} = \{x_j \mid \mathrm{dtw}^2(x_j, \mu_i^{(t)}) \leq \mathrm{dtw}^2(x_j, \mu_{i^*}^{(t)})$ for all $i^* = 1, \dots, k\}$
    end for
    `//*** Mean update step ******************************************//`
    for all $i = 1, \dots, k$ **do**
      $\mu_i^{(t+1)} = \text{DTW-Mean}(C_i^{(t)})$
    end for
  until termination
  return $\mu_1, \dots, \mu_k$, C
end procedure

DTW Barycenter Averaging (DBA) [2]. For a technical description of these algorithms see [3]. The file `SSG.m` contains an implementation of the SSG algorithm, `MM.m` is an implementation of the MM algorithm. The input and output parameters are documented in the source code. For usability, both implementations provide default values for all parameters, except the dataset `X`. The file `example_dtw_mean.m` demonstrates the use of the sample mean algorithms.

3. DTW based k-means

The file `kmeans_dtw.m` provides a k-means implementation which uses a DTW mean algorithm for the mean computation step and the DTW distance as cost, cf. Algorithm 1.

The k-means_dtw algorithm is useful for warping invariant clustering, as demonstrated in the file `example_kmeans_dtw.m`. In this example, two classes of data were generated. Class 1 contains 1 peak at a random position, class 2 contains 2 peaks at random positions. Hence the invariances of the classes are time-shifts which are special cases of warping. Examples of the generated data are illustrated in Figure 2.

The Euclidean k-means algorithm with arithmetic mean yields wrong cluster assignments and spurious centroids, as illustrated in Figure 3.

The k-means_dtw algorithm usually yields a high accuracy in cluster assignments and sound centroids, as illustrated in Figure 4.

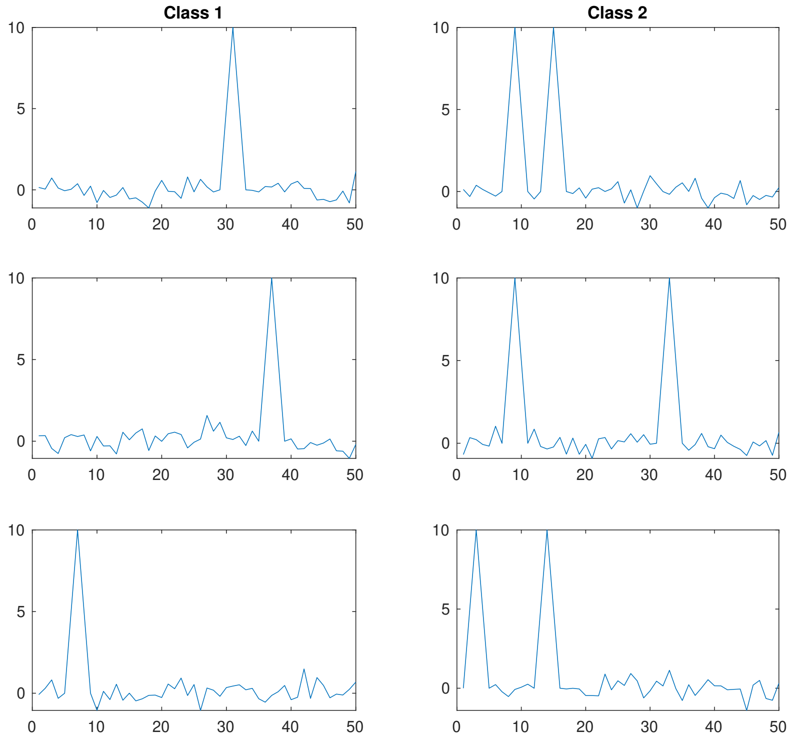


Figure 2: Examples from the generated dataset

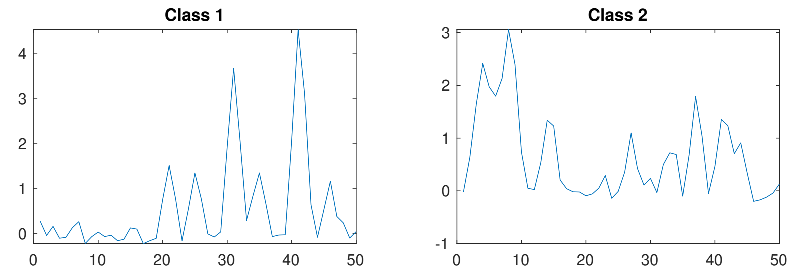


Figure 3: Centroids found by the Euclidean k-means algorithm.

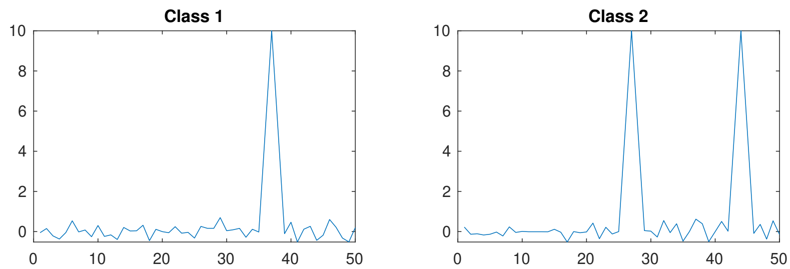


Figure 4: Centroids found by the DTW based k-means algorithm (Algorithm 1).

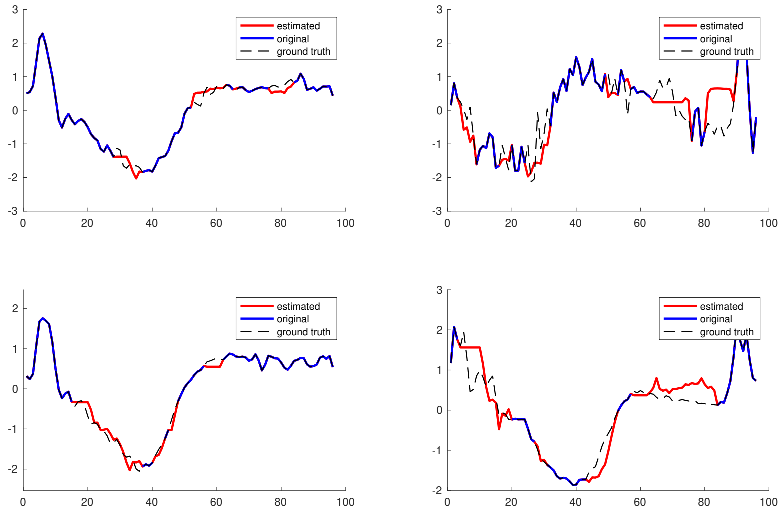


Figure 5: Imputation examples from the DTW based k-means algorithm with data imputation.

4. DTW based k-means for time series with missing values

The file `kmeans_dtw_imputation.m` is a variant of the k-means implementation where the sample time series can have missing values. During the clustering process, the missing values are estimated based on warping the time series with missing values to its cluster centroid. The file `example_kmeans_dtw_imputation.m` provides a demonstration yielding imputation results comparable to those shown in Figure 5.

References

[1] L. Bulteau, V. Froese, and R. Niedermeier. Hardness of consensus problems for circular strings and time series averaging. *arXiv preprint arXiv:1804.02854*, 33(1), 2018.

[2] F. Petitjean, A. Ketterlin, and P. Gancarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition* 44(3):678–693, 2011.

[3] D. Schultz, and B. Jain. Nonsmooth analysis and subgradient methods for averaging in dynamic time warping spaces. *Pattern Recognition*, 74, 340-358, 2018.