DTW MEAN: Time Series Averaging and k-Means Clustering under Dynamic Time Warping

David Schultz
TU Berlin, Germany

Abstract

DTW MEAN is a Matlab library that provides implementations of two mean algorithms for computing a sample mean under Dynamic Time Warping (DTW) of a given time series dataset. A sample mean under DTW is an average curve accounting for temporal distortions in the sample time series. 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.

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, weather 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 have a similar shape: both start flat, then they have a peak followed by a valley and finally they 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 $\mathcal{X} = (x^{(1)}, \ldots, x^{(N)})$ 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

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

where dtw is the DTW distance. Finding a sample mean under DTW is NP-hard [1].
Figure 1: Two sample time series (black) and their means (blue). The green lines illustrate the alignments.

**Algorithm 1** k-means DTW

1: procedure k-MEANS_DTW($X = (x_1, \ldots, x_N), k$)
2: Randomly initialize centroids $\mu_1, \ldots, \mu_k \in X$
3: Initialize clusters $C_1 = \{\}, \ldots, C_k = \{\}$
4: repeat (iteration $t$)
5: //*** Cluster assignment step *****************************************
6: for all $i = 1, \ldots, k$ do
7: $C_i^{(t)} = \{x_j \mid \text{dtw}^2(x_j, \mu_i^{(t)}) \leq \text{dtw}^2(x_j, \mu_{i^*}^{(t)}) \text{ for all } i^* = 1, \ldots, k\}$
8: end for
9: //*** Mean update step ****************************************
10: for all $i = 1, \ldots, k$ do
11: $\mu_i^{(t+1)} = \text{DTW-Mean}(C_i^{(t)})$
12: end for
13: until termination
14: return $\mu_1, \ldots, \mu_k, C$
15: end procedure

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 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.
Figure 2: Examples from the generated dataset

Figure 3: Centroids found by the Euclidean k-means algorithm.
The k-means_dtw algorithm usually yields a high accuracy in cluster assignments and sound centroids, as illustrated in Figure 4.

References