A Comparison of Progressive and Iterative Centroid Estimation Approaches Under Time Warp

Saeid Soheily-Khah[✉], Ahlame Douzal-Chouakria, and Eric Gaussier

Université Grenoble Alpes, CNRS - LIG/AMA, Grenoble, France
{Saeid.Soheily,Ahlame.Douzal,Eric.Gaussier}@image.fr

Abstract. Estimating the centroid of a set of time series under time warp is a major topic for many temporal data mining applications, as summarization a set of time series, prototype extraction or clustering. The task is challenging as the estimation of centroid of time series faces the problem of multiple temporal alignments. This work compares the major progressive and iterative centroid estimation methods, under the dynamic time warping, which currently is the most relevant similarity measure in this context.

Keywords: Centroid estimation · Multiple temporal alignment · Dynamic time warping · Time series

1 Introduction

Time series centroid estimation is a major issue for many temporal data analysis and mining applications. Estimating the centroid of a set of time series under time warp however faces the tricky multiple temporal alignment problem [1–4]. Temporal warping alignment of time series has been an active research topic in many scientific disciplines. To estimate the centroid of two time series under temporal metrics, as the dynamic time warping [5–7], one standard way is to embed the time series into a new Euclidean space defined by their temporal warping alignment. In this space, the centroid can be estimated as the average of the linked elements. The problem becomes more complex where the number of time series is more than two, as one needs to determine a multiple alignment that links simultaneously all the time series on their commonly shared elements.

A first manner to determine a multiple alignment is to search, by dynamic programming, the optimal path within an N-dimensional grid that crosses the N time series. The complexity of this approach nevertheless prevents its use, as it constitutes an NP-complete problem with a complexity of $O(T^N)$ that increases exponentially with the number of time series N and the time series length T. A second way, that characterizes progressive approaches, is based on combining progressively pairs of time series centroids to estimate the global centroid. Such progressive approaches may suffer from the error propagation problem through the set of pairwise centroid combinations. The third approach

is iterative. It works similarly to the progressive approach but reduces the error propagation by repeatedly refining the centroid and realigning it to the initial time series.

The main contribution of this work is to present some major progressive and iterative approaches for time series centroid estimation, prior to present their characteristics. It also reviews an extensive comparison between the approaches through public real and synthetic datasets. To the best of our knowledge, such a comparison has never been conducted before.

The remainder of the paper is organized as follows: In the next section, different related progressive and iterative approaches are presented. Section 3 presents the experiments conducted for comparison purposes and discuss the results obtained. Lastly, Sect. 4 concludes the paper.
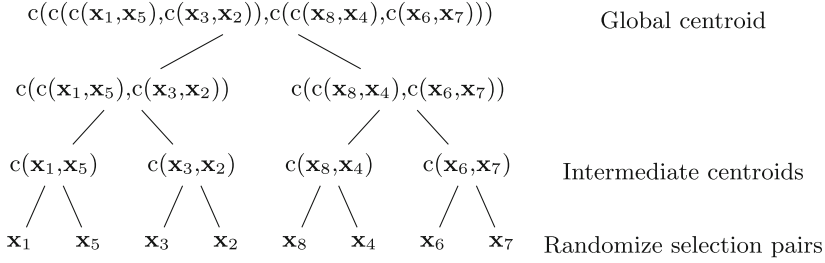
2   Progressive and Iterative Approaches

The progressive and iterative approaches for time series centroid estimation are mainly derived from the multiple sequence alignment methods to address the challenging problem of aligning more than two time series [8–11]. To estimate the centroid of more than two time series, several heuristic approaches have been proposed. Here, we review some major progressive and iterative approaches for time series averaging under the dynamic time warping.

2.1   Progressive Approaches

NonLinear Alignment and Averaging Filters (NLAAF). In the past decades, many averaging methods have been introduced, but only a few of them have been adapted to time series averaging, clustering and mining. For instance, Gupta et al. [12], proposed a time series averaging method based on a tournament scheme, called "*NonLinear Alignment and Averaging Filters* (NLAAF)". First, pairs of time series are selected randomly, and then aligned according to the dynamic time warping. That way, $(N/2)$ averaged sequences are created. The same process is iterated on the centroids estimated, until one sequence is obtained as a global centroid. In this approach, the averaging method between two time series is applied $(N-1)$ times, as illustrated in Fig. 1, where $\mathbf{c}(\mathbf{x}_i, \mathbf{x}_j)$ refers to the estimated centroid of time series $\mathbf{x}_i$ and $\mathbf{x}_j$.

In NLAAF, each element of a centroid is computed as the mean of each linked elements in the DTW alignment. The main drawback of NLAAF approach lies in the growth of its resulting length, because each use of averaging method can almost double the length of the average sequence. As classical datasets comprise hundreds of time series, with each one including hundreds of data points, simply storing the resulting average may be impossible. This length problem is moreover worsened by the complexity of DTW, that grows bi-linearly with the lengths of the sequences. That is why NLAAF is generally used in conjunction with a process reducing the length of the average, unfortunately leading to information loss and unsatisfactory approximation. Additionally, the average strongly depends on the random selection of sequences and different choices lead to different results.

$$c(c(c(\mathbf{x}_1,\mathbf{x}_5),c(\mathbf{x}_3,\mathbf{x}_2)),c(c(\mathbf{x}_8,\mathbf{x}_4),c(\mathbf{x}_6,\mathbf{x}_7)))$$ Global centroid

$$c(c(\mathbf{x}_1,\mathbf{x}_5),c(\mathbf{x}_3,\mathbf{x}_2))$$ $$c(c(\mathbf{x}_8,\mathbf{x}_4),c(\mathbf{x}_6,\mathbf{x}_7))$$

$$c(\mathbf{x}_1,\mathbf{x}_5)$$ $$c(\mathbf{x}_3,\mathbf{x}_2)$$ $$c(\mathbf{x}_8,\mathbf{x}_4)$$ $$c(\mathbf{x}_6,\mathbf{x}_7)$$ Intermediate centroids

$$\mathbf{x}_1 \quad \mathbf{x}_5 \quad \mathbf{x}_3 \quad \mathbf{x}_2 \quad \mathbf{x}_8 \quad \mathbf{x}_4 \quad \mathbf{x}_6 \quad \mathbf{x}_7$$ Randomize selection pairs

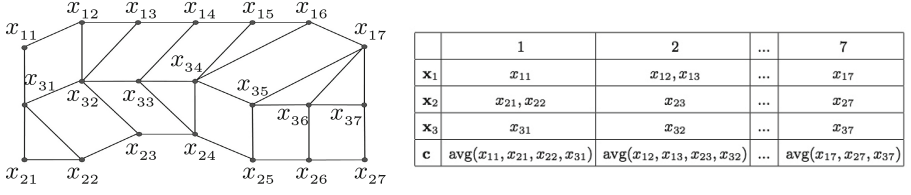**Fig. 1.** Centroid estimation by random pairwise centroid combination.

Prioritized Shape Averaging (PSA). To avoid the bias induced by random selection, Niennattrakul et al. among others [8,16,17] proposed a framework of shape averaging called "*Prioritized Shape Averaging* (PSA)" based on hierarchical clustering. The pairwise time series centering is guided by the dendrogram obtained through hierarchical clustering strategy.

The PSA uses hierarchical clustering as a way to identify priorities between time series. In particular, to estimate a global centroid, the set is first clustered using the agglomerative clustering to get a hierarchical relationship among the whole time series. The simple or complete linkage is considered in general to fasten the dendrogram build, where almost the average linkage or centroids are the best-performed methods. Subsequently, the pairwise time series centroids are combined respectively to their clustering order in the dendrogram. Each parent node is averaged in a bottom-up manner using a weighted DTW averaging. Therefore, the most similar time series are averaged first. Note that the weight of an averaged sequence is calculated from the number of time series upon which the averaged sequence is formed. Initially, all time series have the same weight of one.

Figure 2 describes an example of averaging six sample time series using PSA. According to the dendrogram, first the time series $\mathbf{x}_2$ and $\mathbf{x}_3$ are averaged.
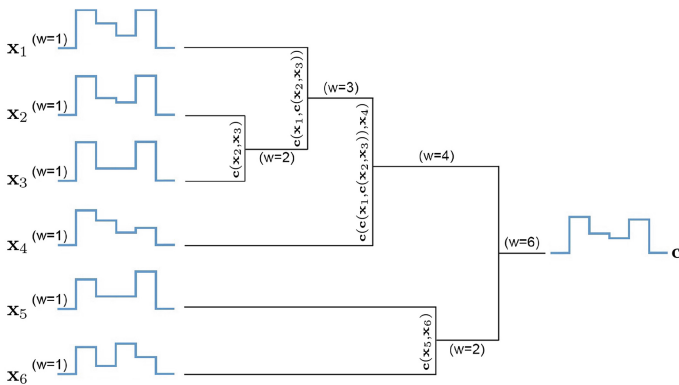
Fig. 2. Example of six time series sequence averaging using PSA

	1	2	...	7
$\mathbf{x}_1$	x_{11}	x_{12}, x_{13}	...	x_{17}
$\mathbf{x}_2$	x_{21}, x_{22}	x_{23}	...	x_{27}
$\mathbf{x}_3$	x_{31}	x_{32}	...	x_{37}
$\mathbf{c}$	avg($x_{11}, x_{21}, x_{22}, x_{31}$)	avg($x_{12}, x_{13}, x_{23}, x_{32}$)	...	avg(x_{17}, x_{27}, x_{37})

Fig. 3. Centroid estimation based on a reference time series. The DTW is performed between $\mathbf{x}_1$, $\mathbf{x}_2$ and the reference time series $\mathbf{x}_3$ (left). $\mathbf{x}_1$ and $\mathbf{x}_2$ are embedded in the space defined by $\mathbf{x}_3$ (right) where the centroid is estimated, and 'avg' is the standard mean function.

The average sequence denoted by $\mathbf{c}(\mathbf{x}_2, \mathbf{x}_3)$, has the weight of two. The intermediate centroid $\mathbf{c}(\mathbf{x}_1, \mathbf{c}(\mathbf{x}_2, \mathbf{x}_3))$ is then computed by averaging the time series $\mathbf{x}_1$ and the average sequence $\mathbf{c}(\mathbf{x}_2, \mathbf{x}_3)$. The intermediate centroid $\mathbf{c}(\mathbf{x}_1, \mathbf{c}(\mathbf{x}_2, \mathbf{x}_3))$ will have the weight of three, since the time series sequences $\mathbf{x}_1$ and $\mathbf{c}(\mathbf{x}_2, \mathbf{x}_3)$ have weight of one and two, respectively. The process goes on till one obtains a global centroid.

Although this hierarchical averaging method aims to remove the bias induced by random selection, growth length of the average sequence remains a problem. Furthermore, local averaging strategies like NLAAF or PSA may let an initial approximation error propagate throughout the averaging process. If averaging process has to be repeated (e.g., during k-means iterations), the effects may dramatically alter the quality of the result. This is why a global approach is desirable, where time series would be averaged all together, with no sensitivity to their order of consideration.

Cross-Word Reference Template (CWRT). A direct manner to estimate the centroid is proposed in Abdulla et al. [1], where a dynamic time warping between each time series and a reference one, generally the time series medoid, is first performed. Each time series is then described in the representation space defined by the reference medoid by resampling, stretching and shortening operations, as in Fig. 3. Finally the global centroid is computed by averaging the time-aligned time series across each point. The method is called "*Cross-Words Reference Template* (CWRT)".

The global estimated centroid has the same length as the medoid, and the result does not depend on the order in which time series are processed.

2.2   Iterative Approaches

Dtw Barycenter Averaging (DBA). Petitjean et al. [3] proposed a global averaging method, called "*Dtw Barycenter Averaging* (DBA)". The method consists in iteratively refining an initially average sequence, in order to minimize its distance to the set of time series. The aim is to minimize the sum of squared DTW distances from the average sequence to the set of time series. Technically,

DBA works in two steps for each refinement. First, computing DTW between each time series and the temporary average sequence, and secondly, updating each element of the average sequence with the barycenter of the elements aligned to it during the first step. In a nutshell, the DBA under temporal warping is a global approach that can average a set of time series all together. The global estimated centroid has the same length as the initial average sequence, and like CWRT, the result is not depending on the order in which time series are processed. However the time complexity of DBA is smaller than NLAAF and PSA [3], but the time complexity problem remains.

Weighted DTW Averaging (WDTW). To circumvent the tricky multiple temporal alignments and the above mentioned limitations, we proposed a tractable and fast centroid estimation that captures both global and local temporal features under weighted time warp measures [20]. It formalizes the multiple time series averaging problem as an optimization problem and propose a solution yielding a local optimum.

For that, we propose to estimate both the time series centroid $\mathbf{c}$ and the weight vector $\mathbf{w}$ that measures the representativeness of the centroid's regions. In addition, we introduce a weighted warping function $f(w_t)$ that guides the learned alignments according to the importance of the centroid elements to capture the shared global and local temporal features.

Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ be a set of time series, and WDTW the weighted dissimilarity between $\mathbf{x}_i$ and the weighted centroid $(\mathbf{c}, \mathbf{w})$. The averaging problem, as formalized in [20], is defined as:

$$\underset{\mathbf{c},\ \mathbf{w}}{\operatorname{argmin}} \sum_{i=1}^{N} \mathrm{WDTW}(\mathbf{x}_i, (\mathbf{c}, \mathbf{w}))$$

with

$$\mathrm{WDTW}(\mathbf{x}, (\mathbf{c}, \mathbf{w})) = \min_{\boldsymbol{\pi} \in \mathcal{A}} \underbrace{\frac{1}{|\boldsymbol{\pi}|} \sum_{(t', t) \in \boldsymbol{\pi}} f(w_t)\, \varphi(x_{t'}, c_t)}_{C(\boldsymbol{\pi})}$$

where $f : (0, 1] \to \mathbb{R}^+$ is a non-increasing function (e.g. $f(w_t) = w_t^{-\alpha}$) and $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ is a positive, real-valued, dissimilarity function. The cost function C computes the sum of the weighted dissimilarities φ between $\mathbf{x}$ and $(\mathbf{c}, \mathbf{w})$ through the alignment $\boldsymbol{\pi}$. When the weights are uniform (or f is a constant function) and when φ is the Euclidean distance, corresponds to the well known Dynamic Time Warping (DTW) [5,7].

The problem given above, can be solved by computing the partial derivatives of Lagrangian with respect to $\mathbf{c}$ to 0 and solving for $\mathbf{c}$, and with respect to $\mathbf{w}$ to 0 and solving for $\mathbf{w}$, described more in details in [20].

Let us summarize the main characteristics of the above approaches. In both NLAAF and PSA, the length of the global centroid increases with the number of

time series to average, inducing an increase of the space and time complexity that is particularly critical under the DTW. The length of the centroids estimated by CWRT, DBA and WDTW averaging is however the same as the reference time series length. Furthermore, all the progressive approaches as well as DBA method are heuristic, with no guarantee of optimality. Even, if the provided approximations are accurate for globally similar time series, they are in general poor for time series that share local characteristics with distinctive global behaviors. In this view, WDTW is a tractable fast and accurate averaging method that captures both global and local temporal features, as shown in [20].

3   Experimental Study

The experiments are conducted to compare the above approaches on classes of time series composing various datasets. The datasets can be divided into two categories. The first one is composed of time series that have similar global behavior within the classes, where the time series of the second category may have distinct global behavior, while sharing local characteristics [19]. For the comparison, the induced inertia reduction rate and the required run time are evaluated as well as the qualitative comparison of the centroids obtained by a visualization. In the following, we first describe the datasets used, then specify the validation process and discuss the obtained results.

3.1   Data Description

The experiments are first carried out on four well known public datasets CBF, CC, DIGITS and CHARACTER TRAJ. [14,15]. These data define a favorable case for the averaging task as time series behave similarly within the classes, as illustrated in Fig. 4.

We then consider more complex datasets: BME[1], UMD (see Footnote 1) [19], SPIRAL [4], NOISED SPIRAL (see Footnote 1) and CONSSEASON [14]. They are composed of time series that behave differently within the same classes while sharing several local characteristics.
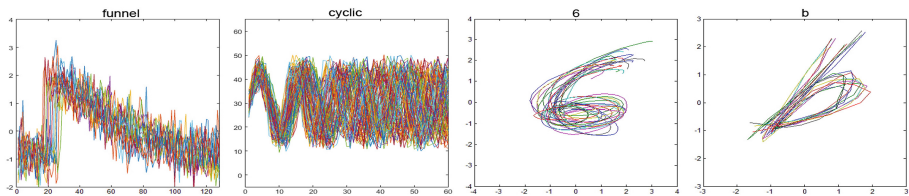
Fig. 4. The time series behaviors with the classes "Funnel", "Cyclic", "6" and "b" of the datasets CBF, CC, DIGITS and CHARACTER TRAJ., respectively

---

[1] http://ama.liglab.fr/douzal/data.
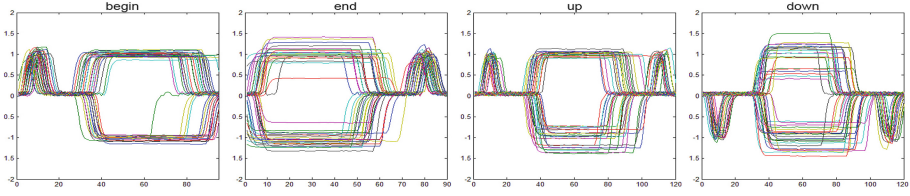
Fig. 5. The time series behaviors with the classes "Begin" and "End" of the dataset BME (left), "Up" and "Down" of the dataset UMD (right)

BME includes two challenging classes BEGIN and END (Fig. 5-left), which characterized by a small bell arising at the initial and final period respectively. The overall behavior may be different depending on whether the large bell is pointing upward or downward. UMD introduces more complexity with the classes UP and DOWN (Fig. 5-right) characterized by a small bell that may occur at different time stamps.

The SPIRAL data, proposed in [18], consists of 3-D spatio-temporal time series (2-D spatial and 1-D temporal) generated from latent time series:

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{U}_i^T(\mathbf{Z} + \mathbf{b}_i\mathbf{1}_l^T)\mathbf{M}_i \\ \mathbf{e}_i^T \end{bmatrix} \in \mathbb{R}^{3*n_i}$$

where the canonical time series $\mathbf{Z} \in \mathbb{R}^{2*l}$ is a curve in two dimensions (x, y). $\mathbf{U}_i \in \mathbb{R}^{2*2}$ and $\mathbf{b}_i \in \mathbb{R}^2$ are randomly generated projection matrix and translation vector respectively. The binary matrix $\mathbf{M}_i \in \{0, 1\}^{l*n_i}$ is generated by randomly choosing $n_i \leq l$ columns from $\mathbf{I}_l$ for temporal distortion. The spatial dimension $\mathbf{e}_i \in \mathbb{R}^{n_i}$ is generated with zero-mean Gaussian noise. The latent time series $\mathbf{Z}$ and three generated time series are visualized in Fig. 6.

SPIRAL2 extends SPIRAL data to more challenging time series that are highly noisy and globally behave differently while sharing a three dimensional latent time series that may appear randomly at different time stamps. The latent time series $\mathbf{Z}$ and three generated time series are visualized in Fig. 7.

Finally, CONSSEASON data provides the electric power consumption recorded in a personal home over almost one year. CONSSEASON is composed of time series distributed in two season classes (*Warm* and *Cold*) depending on whether the
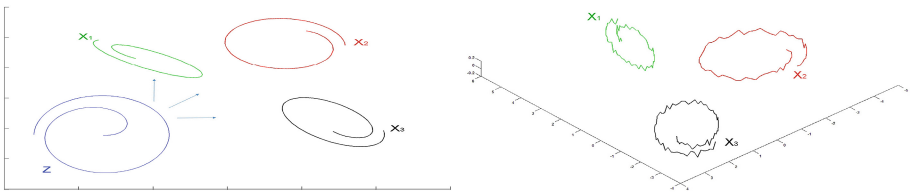
Fig. 6. Latent curve Z and three induced instances X_1, X_2, X_3 without noise (left), and with noise e_i (right) - SPIRAL dataset
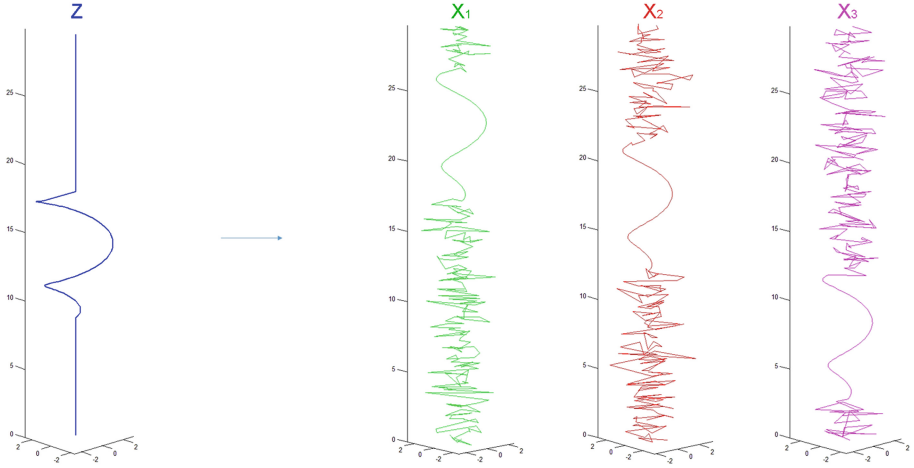
Fig. 7. Latent curve Z and 3 induced instances X_1, X_2, X_3 sharing local characteristics for the SPIRAL2 dataset

Table 1. Data description

DATASET	NB. CLASS	NB. TS.	NB. ATT.	TS. LENGTH	TYPE
CBF	3	930	1	128	GLOBAL
CC	6	600	1	60	
DIGITS	10	220	2	85	
CHARACTER TRAJ.	20	200	3	20	
BME	3	150	1	90	LOCAL
UMD	3	150	1	121	
SPIRAL	1	50	3	95	
NOISED SPIRAL	1	50	3	300	
CONSSEASON	2	365	1	144	

power consumption is recorded during the warm (from April to September) or cold (from October to March) season. Note that the electric power consumption profiles differ markedly within classes.

Table 1 indicates for each data set: the number of classes it includes (NB. CLASS), the number of instances (NB. TS.), the number of attributes (NB. ATT.), the time series length (TS. LENGTH) and the global or local nature of similarity within the classes (TYPE).

3.2   Validation Process

The five mentioned methods NLAAF, PSA, CWRT, DBA and WDTW described in Sect. 2 are compared together. The performances of these approaches are

evaluated through the centroid estimation of each class of the above described datasets. Particularly, the efficiency of each approach is measured through: (a) the reduction rate of the inertia criterion; the initial inertia being evaluated around the time series medoid that minimizes the distances to the rest of time series and (b) the space and time complexity. The results reported hereafter are averages of 10 repetitions of the corresponding algorithms. Finally for all reported results, the one which is significantly different from the rest (two-sided t-test at 5 % risk) is indicated in bold.

Inertia Reduction Rate. Time series averaging approaches are used to esti-mate centroid of the time series classes described above, then the inertia w.r.t. the centroids is measured. Lower is the inertia higher representative is the extracted centroid. Table 2, gives the obtained inertia reduction rates (IRR), averaged per dataset, as:

$$IRR = 1 - \frac{\sum_{i=1}^{N} D(\mathbf{x}_i, \mathbf{c})}{\sum_{i=1}^{N} D(\mathbf{x}_i, \mathbf{m})}$$

where $\mathbf{x}_1, ..., \mathbf{x}_N$ are the set of time series, D is the distance metric, $\mathbf{c}$ is the determined centroid and $\mathbf{m}$ the initial medoid. The alternative of a centroid is a medoid. Medoid is a time series in a set that minimizes sum of the distances to all other time series within the same set. Note that, we use the medoid in the criteria defined above to make the results comparable. In the case, we compare the obtained centroid of each method with a specific time series in the set (i.e. medoid) as its alternative one. Table 2 shows that the WDTW provides the highest IRR for the most datasets, followed by DBA and PSA. Notice that, the results presented by PSA is obtained through the centroid linkage, which outperformed the other linkage (e.g. simple linkage, complete linkage), mostly. Some negative rates observed indicate an inertia increase.

Table 2. The mean of inertia reduction rate (in %) and standard deviations ($\pm\sigma$)

DATASET	NLAAF	PSA	CWRT	DBA	WDTW
CBF	$8.3^{\pm4.0}$	12.3	$-61.3^{\pm9.1}$	$32.1^{\pm1.4}$	$\mathbf{82.6}^{\pm1.8}$
CC	$9.8^{\pm6.6}$	28.6	$6.8^{\pm5.0}$	$34.2^{\pm0.9}$	$70.5^{\pm0.7}$
DIGITS	$26.1^{\pm7.2}$	$\mathbf{79.5}$	$\mathbf{77.6}^{\pm2.4}$	$\mathbf{82.2}^{\pm1.7}$	$\mathbf{87.6}^{\pm1.3}$
CHARACTER TRAJ.	$67.1^{\pm8.6}$	$\mathbf{87.7}$	$\mathbf{85.2}^{\pm5.1}$	$\mathbf{90.6}^{\pm2.3}$	$\mathbf{93.1}^{\pm2.0}$
BME	$34.9^{\pm5.4}$	43.1	$-11.8^{\pm9.2}$	$59.4^{\pm2.7}$	$\mathbf{91.2}^{\pm3.1}$
UMD	$25.6^{\pm9.7}$	51.1	$-56.2^{\pm9.4}$	$48.8^{\pm1.9}$	$78.5^{\pm1.5}$
SPIRAL	$\mathbf{59.8}^{\pm2.5}$	$\mathbf{64.4}$	$\mathbf{64.2}^{\pm1.1}$	$\mathbf{65.8}^{\pm1.1}$	$72.8^{\pm0.9}$
NOISED SPIRAL	$61.4^{\pm4.2}$	66.3	$9.3^{\pm14.7}$	$9.8^{\pm3.2}$	$\mathbf{98.9}^{\pm0.4}$
CONSSEASON	$\mathbf{84.1}^{\pm3.2}$	70.5	$4.6^{\pm12.7}$	$21.4^{\pm1.6}$	$\mathbf{95.5}^{\pm1.7}$

Table 3. Comparison of Time/Space complexity

DATASET	NLAAF		PSA		DBA		WDTW	
	LENGTH	TIME	LENGTH	TIME	LENGTH	TIME(NB-IT.)	LENGTH	TIME(NB-IT.)
CBF	8283	392.32	35042	9999.89	128	42.91(30)	128	**33.42**(25)
CC	992	**4.15**	1677	12.75	60	6.46(40)	60	**3.91**(20)
DIGITS	313	**0.52**	530	1.09	85	**0.51**(15)	85	**0.47**(10)
CHAR. TRAJ.	33	0.06	29	0.06	20	**0.03**(10)	20	**0.02**(05)
BME	2027	5.46	2781	11.92	90	3.93(30)	90	**3.04**(15)
UMD	2729	10.32	4280	28.87	121	4.75(30)	121	**3.43**(15)
SPIRAL	660	1.62	1122	3.33	95	1.19(10)	95	**0.83**(05)
NOISED SPIRAL	1699	**16.13**	9030	269.93	300	34.84(25)	300	**15.74**(10)
CONSSEASON	5741	77.10	32706	3680.81	144	29.79(35)	144	**21.85**(20)

Time and Space Complexity. In Table 3 the studied approaches are compared w.r.t their space and time complexity. Are reported the length of the extracted centroid (LENGTH), the time consumption in seconds (TIME) and for the iterative methods the request number of iterations (ITER NB.). The results, averaged per dataset, reveal almost WDTW the faster method, followd by DBA, and PSA the slowest one. The CWRT approach is not comparable to the rest of the methods as it performs directly an euclidean distance on the time series once the initial DTW matrix evaluated. Remark that for NLAAF and PSA the centroid lengths are very large making these approaches unusable for large time series, while the centroid lengths for the remaining methods are equal to the length of the initial medoid. The higher time consumptions observed for NLAAF and PSA are mainly explained by the progressive increase of the centroid length during the pairwise combination process.

3.3 Discussion

From Table 2, we can see that WDTW lead to the highest inertia reduction rates for almost all datasets, where the best scores (significantly different) indicated in bold. As expected, the DBA method that iteratively optimizes an inertia criterion, in general, reaches higher values than the non-iterative methods (NLAAF, PSA and CWRT). Finally, CWRT has the lowest inertia reduction rates. The negative rates observed for CWRT indicate an inertia increase.

From Table 3, the results reveal WDTW the fastest method and the PSA the slowest one. For NLAAF and PSA the estimated centroids have a drastically large dimension (i.e. a length around 10^4) making these approaches unusable for large time series datasets. The NLAAF and PSA methods are highly time consuming, largely because of the progressive increase of the centroid length during the pairwise combination process. The centroid lengths for the remaining methods are equal to the length of the initial medoid (Table 3). Finally, PSA appears
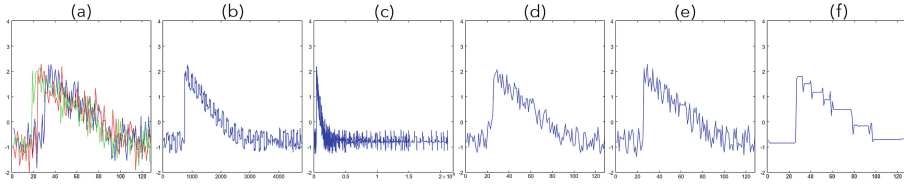
Fig. 8. CBF-"funnel" centroids: (a) ground through, (b) NLAAF, (c) PSA, (d) CWRT, (e) DBA, (f) WDTW
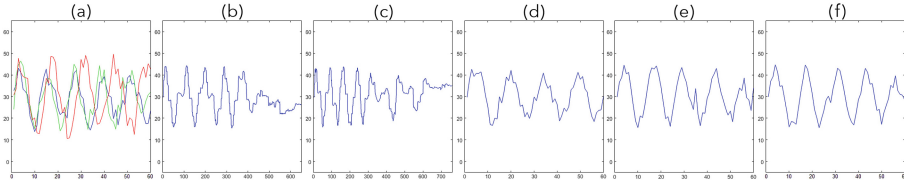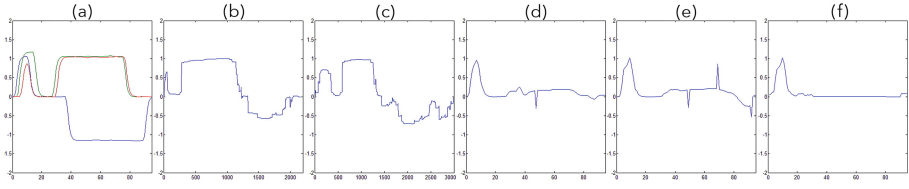

Fig. 9. CC-"cyclic" centroids: (a) ground through, (b) NLAAF, (c) PSA, (d) CWRT, (e) DBA, (f) WDTW
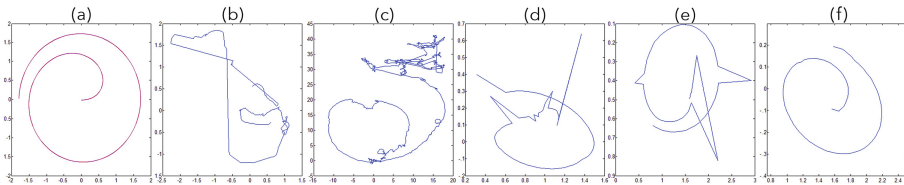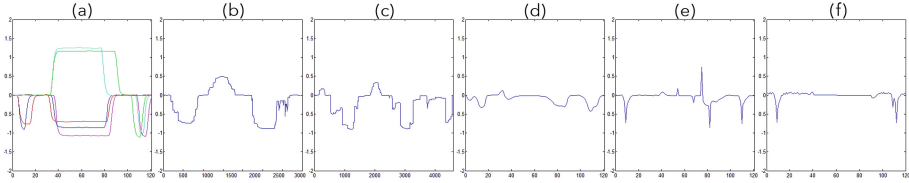

Fig. 10. SPIRAL-1 centroids: (a) ground through, (b) NLAAF, (c) PSA, (d) CWRT, (e) DBA, (f) WDTW

greatly slower than NLAAF; this is due to the hierarchical clustering on the whole time series.

We finally visualize here some of the centroids obtained by the different methods to compare their shape to the one of the time series they represent. Figures 8, 9, 10, 11 and 12 display the centroids obtained by the mentioned methods for the class "*funnel*" of CBF, "*cyclic*" of data set CC, the SPIRAL1, "*begin*" of BME and "*down*" of data set UMD, respectively. As one can note, for global datasets, almost all approaches succeed in obtaining centroids more or less similar to the initial time series. However, we observe generally less representative centroids for NLAAF and PSA, with a drastically large centroid's length of about 10^4 elements vs. 10^2 for the other methods. For the more complex (e.g. SPIRAL), Fig. 10 shows the ability of the DBA method, in obtaining centroids more or less similar to the initial time series, but one should circumvent the noise problem. Finally, as shown in the figures, the WDTW provide the most representative centroid for all datasets. For complex dataset, one can see the ability of the WDTW to circumvent the noise problem and to reveal the locally shared signature.

Fig. 11. BME-"begin" centroids: (a) ground through, (b) NLAAF, (c) PSA, (d) CWRT, (e) DBA, (f) WDTW

Fig. 12. UMD-"down" centroids: (a) ground through, (b) NLAAF, (c) PSA, (d) CWRT, (e) DBA, (f) WDTW

4   Conclusion

The DTW is among the most frequently used metrics for time series in several domains as signal processing, temporal data analysis and mining or machine learning. However, for time series clustering, approaches are generally limited to k-medoid to circumvent time series averaging under dynamic time warping and tricky multiple temporal alignments. The present study compares the major progressive and iterative time series averaging methods under the dynamic time warping. The experimental validation is based on standard datasets in which time series share similar behaviors within classes, as well as on more complex datasets. They are multidimensional, noisy and share only local characteristics. Both the quantitative evaluation, based on an inertia criterion (i.e. IRR), time and space complexity, and the qualitative one (consisting in the visualization of the centroids obtained by different methods) show the effectiveness of WDTW method to provide fastly accurate time series averaging for standard and complex datasets. The centroids obtained through WDTW are more representative of the set than the centroids obtained by the other methods, and the time requirements are lower than the rest. Following WDTW approach, the DBA, the second best method that iteratively optimizes an inertia criterion, reaches higher values than the non iterative methods (NLAAF, PSA and CWRT).

References

1. Abdulla, W.H., Chow, D., Sin, G.: Cross-words reference template for DTW-based speech recognition systems. In: Proceedings of TENCON, vol. 2, pp. 1576–1579 (2003)

2. Hautamaki, V., Nykanen, P., Franti, P.: Time-series clustering by approximate prototypes. In: 19th International Conference on Pattern Recognition (2008)
3. Petitjean, F., Ketterlin, A., Gançarski, P.: A global averaging method for dynamic time warping, with applications to clustering. Pattern Recogn. **44**, 678–693 (2011)
4. Zhou, F., De la Torre, F.: Generalized time warping for multi-modal alignment of human motion. In: Computer Vision and Pattern Recognition (CVPR), pp. 1282–1289. IEEE (2012)
5. Kruskall, J.B., Liberman, M.: The symmetric time warping algorithm: from continuous to discrete. Time Warps Journal. Addison-Wesley (1983)
6. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech, Signal Process. **26**, 43–49 (1978)
7. Sankoff, D., Kruskal, J.B.: Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison. Cambridge University Press, Addison-Wesley, Reading (1983)
8. Thompson, J.D., Higgins, D.G., Gibson, T.J.: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. J. Comput. Biol. **22**(22), 4673–4680 (1994)
9. Notredame, C., Higgins, D.-G., Heringa, J.: T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. - J MOL BIOL **302**(1), 205–217 (2000)
10. Sze, S.-H., Lu, Y., Yang, Q.: A polynomial time solvable formulation of multiple sequence alignment. J. Comput. Biol. **13**(2), 309–319 (2006)
11. Carrillo, H., Lipman, D.: The multiple sequence alignment problem in biology. SIAM J. Appl. Math. **48**, 1073–1082 (1988). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA
12. Gupta, L., Molfese, D., Tammana, R., Simos, P.: Nonlinear alignment and averaging for estimating the evoked potential. IEEE Trans. Biomed. Eng. **43**(4), 348–356 (1996)
13. Soheily-Khah, S., Douzal-Chouakria, A., Gaussier, E.: Progressive and iterative approaches for time series averaging. In: ECML-PKDD (Advanced Analytics and Learning on Temporal Data) (2015)
14. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml/
15. UCR Time Series Classification Archive. http://www.cs.ucr.edu/~eamonn/
16. Niennattrakul, N., Ratanamahatana, C.: On clustering multimedia time series data using K-means and dynamic time warping. In: International onference on IEEE Multimedia and Ubiquitous Engineering, MUE 2007, pp. 733–738 (2007)
17. Niennattrakul, N., Ratanamahatana, C.: Shape averaging under time warping. In: ECTI-CON 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. IEEE, vol. 2, pp. 626–629, May 2009
18. Zhou, F., De la Torre, F.: Canonical time warping for alignment of human behavior. Adv. Neural Inf. Process. Syst. **22**, 2286–2294 (2009)
19. Frambourg, C., Douzal-Chouakria, A., Gaussier, E.: Learning multiple temporal matching for time series classification. In: Tucker, A., Höppner, F., Siebes, A., Swift, S. (eds.) IDA 2013. LNCS, vol. 8207, pp. 198–209. Springer, Heidelberg (2013)
20. Soheily-Khah, S., Douzal-Chouakria, A., Gaussier, E.: Generalized *k*-means-based clustering for temporal data under weighted and kernel time warp. J. Pattern Recogn. Lett. **75**, 63–69 (2016). doi:10.1016/j.patrec.2016.03.007