

Progressive and Iterative Approaches for Time Series Averaging

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

Abstract. Averaging a set of time series is a major topic for many temporal data mining tasks as summarization, extracting prototype or clustering. Time series averaging should deal with the tricky multiple temporal alignment problem; a still challenging issue in various domains. This work compares the major progressive and iterative averaging time series methods under dynamic time warping (DTW).

1 Introduction

Time series centroid estimation is a major issue for many temporal data analysis and mining tasks as summarization, extracting temporal prototype or clustering. Estimating the centroid of a set of time series under time warp should deal with 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 the 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 similar 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 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 identifies progressive approaches, is based on combining progressively pairwise time series centroids to estimate the global one. The progressive approaches may suffer of the early error propagation through the set of pairwise centroid combinations. The third approach is iterative, it works similarly to the progressive approach, but mainly reduces the error propagation by repeatedly refining the barycenter and realigning it to the initial time series. In general, the main progressive and iterative approaches are of heuristic nature limited to the dynamic time warping metric, that provide an estimation of the barycenter without guarantee of an optimal solution.

The main contribution of this work is to introduce some major progressive and iterative approaches for time series centroid estimation, prior to present their

characteristics, as well as an extensive comparison between the mentioned methods throughout real and synthetic datasets, where to the best of our knowledge this necessary study is never conducted before.

The rest of this paper is organized as follows: In the next section, different approaches are studied and Section 3 presents the conducted experimentation and discuss the results obtained. Lastly, Section 4 concludes the paper.

2 Progressive and iterative approaches

The progressive and iterative methods for averaging a set of time series are mostly derived from the multiple sequence alignment methods to address the tricky multiple temporal alignment problem. In the following, we review the major progressive and iterative approaches for time series averaging under time warp.

Gupta et al. in [8] used the DTW in the sequence alignment to average a set of time series. The method, called "*NonLinear Alignment and Averaging Filters (NLAAF)*", uses a tournament scheme averaging approach that its simplest averaging ordering consists in pairwise averaging sequences following a tournament scheme. That way, $N/2$ average sequences are created at first step. Then those $N/2$ sequences, in turn, are pairwise averaged into $N/4$ sequences, and so on, until one sequence is obtained. In this approach, the averaging method between two sequence is applied $(N - 1)$ times. NLAAF works by placing each element of the average sequence of two time sequences, as the center of each association created by DTW. Its main drawback lies in growth of its resulting length, because each use of the average method can almost double the length of the average sequence. That is why NLAAF is generally used in conjunction with a process reducing the length of the average, leading to a loss of information and thus to an unsatisfactory approximation. Additionally, the average strongly depends on the order of time series sequences and so, different orders of sequences give different average sequence.

To avoid the bias induced by random selection, Niennattrakul et al. [11, 12] proposed a framework of shape averaging called "*Prioritized Shape Averaging (PSA)*", which uses hierarchical clustering with a new DTW averaging function, labeled "*Scaled Dynamic Time Warping*" with extra capability in stretching some parts of warping path so that the result is more similar to a sequence time series with more weight. Niennattrakul used hierarchical clustering as a heuristic to order the priority. In spite of this hierarchical averaging method aims to prevent the order dependency, the length of average sequences remains a problem. Local averaging strategies like NLAAF or PSA may let an initial approximation error propagate throughout the averaging process. If the averaging process has to be repeated, the effects may dramatically alter the quality of the result. This is why a global approach is desirable, where sequences would be averaged all together, with no sensitivity to their order of consideration.

A direct manner to estimate the centroid proposed by Abdulla et al. [1], called "*Cross-Words Reference Template (CWRT)*", which uses medoid as the

reference time series as follows. First, the time series medoid is selected. The whole time series are then described in the representation space defined by the reference medoid. In the next step, all sequences are aligned by DTW to a single medoid and then the average is computed by averaging the time-aligned time series across each point. Petitjean et al. [3] proposed a global averaging method, called "*Dtw Barycenter Averaging* (DBA)", which consists in iteratively refining an initially average sequence, in order to minimize its distance to the averaged sequence. As a summary, the DBA under temporal warping is a global approach that can average a set of sequences all together.

All the methods define heuristic approaches, although with no guarantee of optimal solutions, the provided approximations are accurate particularly for time series that behave similarly within the set. However these approaches may fail principally for time series with similar global behavior and local temporal differences, as one needs to deploy local instead of global averaging process.

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 [9]. 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. [10]. These data define a favorable case for the averaging task as time series behave similarly within the classes. Then, we consider more complex datasets: BME¹, UMD¹, SPIRAL [4], NOISED SPIRAL¹ and CONSSEASON [10]. They are composed of time series that behave differently within the same classes while sharing several local characteristics. 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 four mentioned methods NLAFF, PSA, CWRT and DBA described in Section 2 is compared together. The performances of these approaches are evaluated through the centroid estimation of each class of the above described datasets.

¹ <http://ama.liglab.fr/~douzal/data>

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	
CHAR. 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	

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 averaged through a bootstrap process, with 10 repetitions. Finally for all reported results, the best one which is significantly different from the rest (t -test at 1% risk) is indicated in bold.

Inertia reduction rate Time series averaging approaches are used to estimate 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 $IRR=1 - \frac{\sum_{i=1}^N D(x_i, c)}{\sum_{i=1}^N D(x_i, m)}$, averaged per dataset; x_1, \dots, x_N being the set of time series, D the metric, c the determined centroid and m the initial medoid. Table 2 shows that the DBA provides the highest IRR for the most datasets. Some negative rates observed indicate an inertia increase.

Table 2: Comparison of Inertia Reduction Rate(IRR)

DATASET	NLAAF	PSA	CWRT	DBA
CBF	8.3%	12.3%	-61.3%	32.1%
CC	9.8%	28.6%	6.8%	34.2%
Digits	26.1%	79.5%	77.6%	82.2%
CHAR. TRAJ.	67.1%	87.7%	85.2%	90.6%
BME	34.9%	43.1%	-11.8%	59.4%
UMD	25.6%	51.1%	-56.2%	48.8%
SPIRAL	59.8%	64.4%	64.2%	65.8%
NOISED SPIRAL	61.4%	66.3%	9.3%	9.8%
CONSSEASON	84.1%	70.5%	4.6%	21.4%

Time and space complexity In Table 3 the studied approaches are compared w.r.t their space and time complexity. The results, averaged per dataset, reveal almost DBA the faster method 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. 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.

Table 3: Comparison of Time/Space complexity

DATASET	NLAAF		PSA		DBA	
	LENGTH	TIME	LENGTH	TIME	LENGTH	TIME(NB-IT.)
CBF	8283	392.32	35042	9999.99	128	42.91 (30)
CC	992	4.15	1677	12.75	60	6.46(40)
DIGITS	313	0.52	530	1.09	85	0.51 (15)
CHAR. TRAJ.	33	0.06	29	0.06	20	0.03 (10)
BME	2027	5.46	2781	11.92	90	3.93 (30)
UMD	2729	10.32	4280	28.87	121	4.75 (30)
SPIRAL	660	1.62	1122	3.33	95	1.19 (10)
NOISED SPIRAL	1699	16.13	9030	269.93	300	34.84(25)
CONSEASON	5741	77.10	32706	3680.81	144	29.79 (35)

3.3 Discussion

From Table 2, we can see that DBA and PSA lead to the highest inertia reduction rates, where the best scores (indicated in bold) are reached by DBA for almost all datasets. However it is significantly lower for some challenging datasets. Finally, CWRT has the lowest inertia reduction rates. The negative rates observed for CWRT indicate an inertia increase. 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).

From Table 3, the results reveal DBA 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 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.

Figure (1) and (2) display the centroids obtained by the mentioned methods respectively for the class "funnel" of CBF and "cyclic" of data set CC. As one can note, for global datasets, almost all the 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.

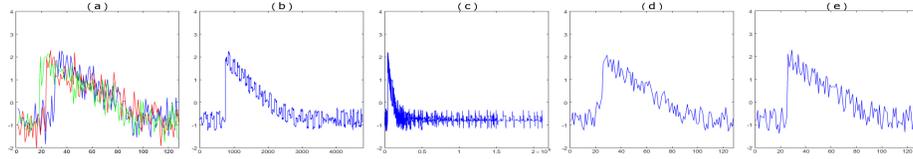


Fig. 1: CBF-"funnel" centroids: (a) ground through, (b) NLAAF, (c) PSA, (d) CWRT, (e) DBA

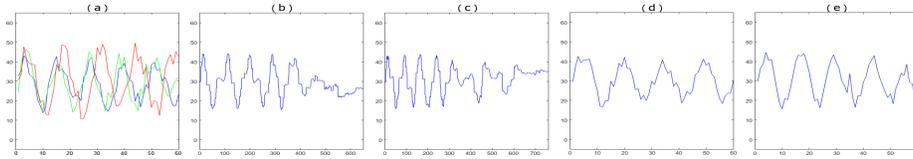


Fig. 2: CC-"cyclic" centroids: (a) ground through, (b) NLAAF, (c) PSA, (d) CWRT, (e) DBA

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 Kmedoid to circumvent time series averaging under DTW and tricky multiple temporal alignments problem. The present study compares the major progressive and iterative time series averaging approaches under dynamic time warping. The experimental validation is based on global datasets in which time series share similar behaviors within classes, as well as on more complex datasets exhibiting time series that share only local characteristics, that are multidimensional and noisy. Both the quantitative evaluation, based on an inertia criterion and time and space complexity, and the qualitative one (consisting in the visualization of the centroids obtained by different methods) show the effectiveness of DBA approach. In particular, the DBA method that iteratively optimizes an inertia criterion, not only, reaches higher values than the non-iterative methods (NLAAF, PSA and CWRT), but also provides a fast time series averaging for global and local datasets.

References

1. Abdulla, W.H. and Chow, D. and Sin, G.: Cross-words reference template for DTW-based speech recognition systems. Proc. TENCON, Pages 1576–1579, Vol. 2 (2003)
2. Hautamaki, V. and Nykanen, P. and Franti, P.: Time-series clustering by approximate prototypes. 19th International Conference on Pattern Recognition, (2008).

3. Petitjean, F. and Ketterlin, A. and GanÇarski, P.: A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, Pages 678-693, Vol. 44 (2011)
4. F. Zhou and F. De la Torre: Generalized time warping for multi-modal alignment of human motion. *IEEE, Computer Vision and Pattern Recognition (CVPR)*, Pages 1282-1289 (2012)
5. Kruskall, J.B. and Liberman, M.: The symmetric time warping algorithm: From continuous to discrete. *Addison-Wesley, Time Warps Journal* (1983)
6. Sakoe, H. and Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Pages 43-49, Vol. 26 (1978)
7. Sankoff, D. and Kruskal, J.B.: *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*. Cambridge University Press, Addison-Wesley, (1983)
8. Lalit Gupta, D. Molfese, R. Tammana, P. Simos: Nonlinear alignment and averaging for estimating the evoked potential. *IEEE T. on Biomedical Engineering*, No. 4, Pages 348-356, Vol. 43 (1996)
9. C. Frambourg, A. Douzal-Chouakria and E. Gaussier: Learning Multiple Temporal Matching for Time Series Classification. In *Advances in Intelligent Data Analysis XII* (pp. 198-209). Springer Berlin Heidelberg. (2013)
10. UCI Machine Learning Repository, "<http://archive.ics.uci.edu/ml/>"
11. V. Niennattrakul, C. Ratanamahatana: On Clustering Multimedia Time Series Data Using K-means and Dynamic Time Warping. *Multimedia and Ubiquitous Engineering, MUE'07. International Conference on IEEE*, Pages 733-738, (2007)
12. V. Niennattrakul, C. Ratanamahatana: Shape Averaging under Time Warping. *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 6th International Conference on IEEE*, Vol. 2, Pages 626-629, May (2009)