

Shape Averaging under Time Warping

Vit Niennattrakul

Chotirat Ann Ratanamahatana

Department of Computer Engineering, Chulalongkorn University
Phayathai Rd., Pathumwan, Bangkok 10330 Thailand
{g49vnn, ann}@cp.eng.chula.ac.th

Abstract

Dynamic Time Warping (DTW) distance measure has increasingly been used as a similarity measurement for various data mining tasks in place of traditional Euclidean distance metric due to its superiority in sequence-alignment flexibility. However, in some tasks where shape averaging is required, e.g., in template matching and k -means clustering problems, current averaging methods are inaccurate in that they produce undesired templates and cluster representatives. In this work, we emphasize the importance of the correctness of this averaging subroutine and propose a novel shape averaging method, called Prioritized Shape Averaging (PSA), using hierarchical clustering approach. In experimental evaluation, our proposed method, PSA, achieves a lower discrepancy distance between an averaged sequence and every original sequence than existing method on various domains.

1. Introduction

After Dynamic Time Warping (DTW) distance has been introduced to time series data mining community, it shows great superiority in accuracy of many tasks, e.g., classification and clustering, over numerous similarity measures [3] including Euclidean distance. However, utilizing DTW distance in some mining tasks imposes some limitations in several situations especially where shape-based templates in template matching or cluster representatives in k -means clustering are required. Although shape averaging method is only a small part in a mining task, we cannot ignore this crucial subroutine. More specifically, inaccurate templates or cluster representatives may significantly lead to meaningless results.

In the past few decades, numerous shape averaging methods have been introduced to signal processing community [1, 5, 2, 6], but only a handful has been adapted to time series data mining domain [9]. For example, Gupta et al.'s work [5] efficiently uses DTW in sequence alignment by

proposing two frameworks, NLAFF1 and NLAFF2, for averaging a set of time series sequences. NLAFF1 assumes that a total number of time series sequences must be in a power of two; therefore, time series sequences are randomly paired and hierarchically averaged. NLAFF2 is additionally proposed to be used under any number of time series sequences by sequentially averaging the series of data.

In this work, we propose a novel framework of shape averaging called Prioritized Shape Averaging (PSA), which uses hierarchical clustering to dynamically create an order of averaging. Additionally, we propose a new DTW averaging function, called Scaled Dynamic Time Warping (SDTW), which increases capability of averaging in time domains. As a result, our PSA is robust to outlier sequences. To demonstrate efficiency of PSA, average discrepancy distances between the averaged sequence and every sequence from the same class are compared with NLAFF, which is the combination method of NLAFF1 and NLAFF2.

The rest of this paper is organized as follows. We provide important background in Section 2, and Section 3 describes our shape averaging method, PSA – Prioritized Shape Averaging. We demonstrate our experiment and applications in Section 4. Finally, we conclude our work in Section 5.

2. Background

In this section, we provide essential background, including Dynamic Time Warping distance measure, Dynamic Time Warping averaging, and hierarchical clustering, before describing Prioritized Shape Averaging in the next section.

2.1. Dynamic Time Warping Distance Measure

Dynamic Time Warping (DTW) distance measure [10] is a well-known shaped-based similarity measurement. It uses a dynamic programming technique to find the optimal warping path between two time series sequences. Suppose we have two sequences, a sequence $X_{1..n} = \langle x_1, x_2, \dots, x_i, \dots, x_n \rangle$ and a sequence $Y_{1..m} =$

$\langle y_1, y_2, \dots, y_j, \dots, y_m \rangle$. The distance is calculated by following the equations.

$$D(X, Y) = d(x_n, y_m) + \min \begin{cases} D(X_{1\dots n-1}, Y_{1\dots m-1}) \\ D(X_{1\dots n}, Y_{1\dots m-1}) \\ D(X_{1\dots n-1}, Y_{1\dots m}) \end{cases} \quad (1)$$

where $D(\emptyset, \emptyset) = 0$, $D(X_{i\dots n}, \emptyset) = D(\emptyset, Y_{j\dots m}) = \infty$, and \emptyset is an empty sequence. Any distance metrics can be used for $d(x_i, y_j)$, including L_1 -norm, $d(x_i, y_j) = |x_i - y_j|$, and L_2 -norm, $d(x_i, y_j) = (x_i - y_j)^2$. For simplicity, we use L_1 -norm to describe our proposed method, but L_2 -norm is used in experimental evaluation to achieve better accuracy.

2.2. Dynamic Time Warping Averaging Function

This shape averaging function [5] exploits DTW calculation to find the appropriate mappings between two sequences before averaging. The function first creates a path matrix in order to store the direction of selected minimum distance. After the optimal distance has been found, the warping path $W = \langle w_1, w_2, \dots, w_k, \dots, w_K \rangle$ is determined by back-tracking the directions of cumulative distances. Each element w_k on warping path stores a pair of indices i and j of data x_i and y_j , denoted as $(i, j)_k$, from sequences X and Y , respectively. Therefore, each new value z_k on an averaged sequence $Z = \langle z_1, z_2, \dots, z_k, \dots, z_K \rangle$ is then calculated using arithmetic mean of x_i and y_j , as shown in Equation 2 at the warping element w_k .

$$z_k = \frac{\lambda_X \cdot x_i + \lambda_Y \cdot y_j}{\lambda_X + \lambda_Y} \quad (2)$$

where λ_X and λ_Y are weights of the sequences X and Y , respectively.

2.3. Hierarchical Clustering

Hierarchical clustering [7] is a method that visualizes relationships among data sequences in hierarchical structure based on distance. Agglomerative clustering [7], a well-known hierarchical clustering method, has been proposed to group data using a bottom-up strategy. The method iteratively merges atomic clusters into a larger one until one single cluster contains all the data. More specifically, in each iteration, a pair of clusters which has minimum distance are merged. To calculate distance between two clusters, many inter-cluster distance functions have been proposed including minimum distance, maximum distance, and average distance.

3. Proposed Method

In this work, we propose a novel averaging method, Prioritized Shape Averaging (PSA), which utilizes hierarchical clustering to efficiently specify a sequence of calculations. In addition, to calculate an averaged sequence between two time series sequences, Scaled Dynamic Time Warping (SDTW) averaging, extended from DTW averaging function described in Section 2.2, is also proposed.

3.1. Scaled Dynamic Time Warping Averaging Function

Scaled Dynamic Time Warping (SDTW) averaging function is extended from original DTW averaging described in Section 2.2. Like DTW averaging, SDTW also uses a warping path from DTW distance calculation as a guideline before computing the average. In addition, SDTW has extra capability in stretching some parts of the warping path so that an expected averaged result is more similar to a sequence with more weight. For example, in Figure 1, suppose we have two sequences, X and Y , with weights of three and one, respectively. If we use original DTW averaging, the result will be as shown in Figure 1 a), but since the sequence X has larger weight, the shape of the averaged result should be more similar to the sequence X than the sequence Y as shown in Figure 1 b). In other words, original DTW averaging weighs only amplitudes of the sequences, not in the time domain.

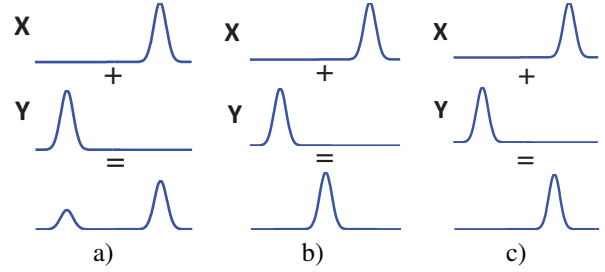


Figure 1. Averaged results of a) amplitude averaging (with weights) comparing with b) DTW averaging and c) SDTW averaging when sequences X and Y have weight of three and one, respectively; therefore, the averaged sequence should be more similar to X than to Y .

Suppose the sequences, X and Y , have weights of λ_X and λ_Y , respectively. SDTW first calculates the weights for each element w_k on the warping path W . In stretching the warping path W , we consider three cases, i.e., where w_{k-1} is $(i-1, j)_{k-1}$, $(i, j-1)_{k-1}$, or $(i-1, j-1)_{k-1}$. If w_{k-1} is

$(i-1, j)_{k-1}$, SDTW will generate λ_Y elements of $(i, j)_k$; on the other hand, if w_{k-1} is $(i, j-1)_{k-1}$, SDTW will generate λ_X elements of $(i, j)_k$. Otherwise, SDTW will generate $\frac{\lambda_X + \lambda_Y}{2}$ elements of $(i, j)_k$. After a new warping path is generated, an amplitude for each new resulted element is calculated according to Equation 2. Since the resulted sequence $Z' = \langle z'_1, z'_2, \dots, z'_k, \dots, z'_K \rangle$ is much longer than the original one, uniform scaling [4] is applied to generate a new resulted sequence $Z = \langle z_1, z_2, \dots, z_j, \dots, z_m \rangle$, where $z_j = z'_{\lfloor j \cdot n/m \rfloor}$ and n and m are the lengths of the sequences Z' and Z , respectively. To be more illustrative, we provide pseudo code of SDTW averaging and a function to determine a warping path in Table 1.

Table 1. Scaled Dynamic Time Warping (SDTW) averaging

FUNCTION $[Z] = \text{SDTW}[X, Y, \lambda_X, \lambda_Y]$	
1	$W = \text{WARPINGPATH}(X, Y)$
2	for each element w_k and w_{k-1} in W
3	if $(k = 1)$
4	add w_k in W' $(\lambda_X + \lambda_Y) / 2$ times
5	else
6	$[i, j] = w_k$
7	if $(w_{k-1} \text{ is } [i, j-1])$
8	add w_k in W' λ_X times
9	else if $(w_{k-1} \text{ is } [i-1, j])$
10	add w_k in W' λ_Y times
8	else
9	add w_k in W' $(\lambda_X + \lambda_Y) / 2$ times
10	endif
11	endif
12	endfor
13	for each element w'_k in W'
14	$z'_k = (\lambda_X \cdot x_i + \lambda_Y \cdot y_j) / (\lambda_X + \lambda_Y)$
15	endfor
16	$Z = \text{UNIFORMSCALING}(Z', \text{length of } X \text{ or } Y)$
17	return Z

3.2. Prioritized Shape Averaging

Since current averaging method averages time series sequences in random order, we propose a novel framework called Prioritized Shape Averaging (PSA) using hierarchical clustering as a heuristic to order the priority. More specifically, to average a set of time series sequences, the set is first clustered using agglomerative clustering, as described in Section 2.3, to get hierarchical relationships among sequences, and then each node is averaged in a bottom-up manner using SDTW averaging.

Therefore, the most similar sequences are averaged first, and the weight of an averaged sequence is calculated from the number of sequences that the averaged sequence is formed. For example, suppose we have five time series sequences which have hierarchical relationships as shown in Figure 2. Initially, all time series sequences have the same weight of one. Sequences A and B are averaged and result in a new sequence F ; therefore, the sequence F has the weight of two. Additionally, if sequences F and H are averaged, a sequence I will have the weight of five since sequences F and H have weight of two and three, respectively. As a result, PSA efficiently averages a set \mathbb{T} of sequences based on shape. For the ease of implementation, we provide recursive pseudo code of PSA, shown in Table 2.

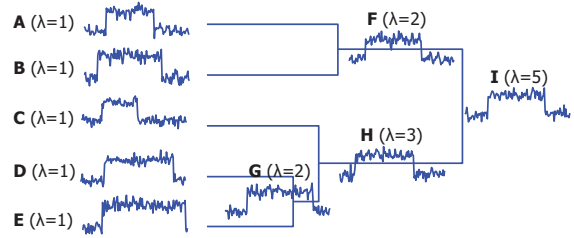


Figure 2. An example of five sequences averaged using our proposed PSA

Table 2. Prioritized Shape Averaging (PSA)

FUNCTION $[Z] = \text{PSA}[\mathbb{T}]$	
1	$\mathbb{C} = \text{AGGLOMERATIVECLUSTERING}(\mathbb{T})$
2	$[Z, \lambda] = \text{AVERAGE}(\text{root of } \mathbb{C})$
3	return Z

FUNCTION $[Z, \lambda] = \text{AVERAGE}[C]$	
1	if $(C \text{ is a leaf node})$
2	return $[C.data, 1]$
3	endif
4	$[Z_1, \lambda_1] = \text{AVERAGE}(C.child_1)$
5	$[Z_2, \lambda_2] = \text{AVERAGE}(C.child_2)$
6	$Z = \text{SDTW}(Z_1, Z_2, \lambda_1, \lambda_2)$
7	return $[Z, \lambda_1 + \lambda_2]$

4. Experiment

To demonstrate superiority of our proposed method over existing averaging approaches, we compare our PSA with NLAFF in terms of averaging discrepancy distances over

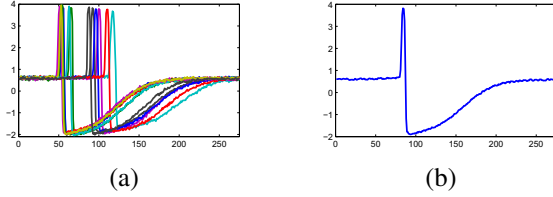


Figure 3. a) Example sequences from Trace dataset and b) the averaged sequence from our PSA

Table 3. Average discrepancy distances between the averaged sequence and each original sequence

Dataset	Sequence length	Dataset size	Avg. distances	
			PSA	NLAFF
Beef	470	60	25.65	25.50
Coffee	286	56	0.72	1.55
ECG200	96	200	9.16	13.30
FaceFour	350	112	33.68	44.34
OliveOil	570	60	0.02	1.30
SyntheticControl	60	600	10.97	12.34
Trace	275	200	1.66	3.72

various datasets from UCR classification/clustering archive [8]. For each dataset, training and test data sequences are merged for each class according to pre-defined class labels.

In the experiment, we average all data sequences which are from the same class, and then calculate average distances between the averaged sequence and every same-class data sequences. This discrepancy distance evaluates the quality of the averaged result. In addition, small average discrepancy distance reflects the good averaging method in generating the averaged result as a representative of a group of data sequences. Figure 3 (a) shows some time series sequences which are from the same class of Trace dataset, and Figure 3 (b) shows the result averaged from our PSA.

As expected, PSA has very little discrepancy comparing with existing methods, as shown in Table 3; in other words, our PSA is more suitable to be used in subroutines of many mining tasks, especially in template matching and k -means clustering.

5. Conclusion

This work proposes a novel shape averaging method called Prioritized Shape Averaging (PSA) which uses hierarchical clustering to arrange the order of averaging. More specifically, PSA first averages a pair of sequences with

high similarity. The key idea of PSA is an efficient averaging function, Scaled Dynamic Time Warping (SDTW) averaging. SDTW averages not only amplitudes of sequences but also alignments on the time domain. In conclusion, PSA outperforms the existing method in terms of having low discrepancy distance.

References

- [1] S. Boudaoud, H. Rix, and O. Meste. Integral shape averaging and structural average estimation: a comparative study. *IEEE Transactions on Signal Processing*, 53(10):3644–3650, Oct. 2005.
- [2] S. Boudaoud, H. Rix, O. Meste, C. Heneghan, and C. O’Brien. Corrected integral shape averaging applied to obstructive sleep apnea detection from the electrocardiogram. *EURASIP Journal on Advances in Signal Processing*, 2007.
- [3] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh. Querying and mining of time series data: Experimental comparison of representations and distance measures. In *Proceedings of the 34th International Conference on Very Large Data Bases (VLDB 2008)*, Auckland, New Zealand, 23 - 28 August 2008.
- [4] A. W.-C. Fu, E. J. Keogh, L. Y. H. Lau, C. A. Ratanamahatana, and R. C.-W. Wong. Scaling and time warping in time series querying. *The VLDB Journal*, 17(4):899–921, 2008.
- [5] L. Gupta, D. Molfese, R. Tammana, and P. Simos. Nonlinear alignment and averaging for estimating the evoked potential. *IEEE Transactions on Biomedical Engineering*, 43(4):348–356, April 1996.
- [6] W. K., B. H., and P. B. Warp-averaging event-related potentials. *Clinical Neurophysiology*, 112:1917–1924(8), October 2001.
- [7] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley, 1990.
- [8] E. Keogh, X. Xi, L. Wei, and C. A. Ratanamahatana. UCR time series classification/clustering page. http://www.cs.ucr.edu/~eamonn/time_series_data.
- [9] C. A. Ratanamahatana and E. J. Keogh. Multimedia retrieval using time series representation and relevance feedback. In *Proceedings of 8th International Conference on Asian Digital Libraries (ICADL2005)*, pages 400–405, Bangkok, Thailand, December 12-15 2005.
- [10] C. A. Ratanamahatana and E. J. Keogh. Three myths about dynamic time warping data mining. In *Proceedings of 2005 SIAM International Data Mining Conference (SDM 2005)*, pages 506–510, Newport Beach, CL, USA, April 21-23 2005.