On Clustering Multimedia Time Series Data Using K-Means and Dynamic 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

After the generation of multimedia data turned digital, an explosion of interest in their data storage, retrieval, and processing has drastically increased. This includes videos, images, and audios, where we now have higher expectations in exploiting these data at hands. Typical manipulations are in some forms of video/image/audio processing, including automatic speech recognition, which require fairly large amount of storage and are computationally intensive. In our recent work, we have demonstrated the utility of time series representation in the task of clustering multimedia data using k-medoids method, which allows considerable amount of reduction in computational effort and storage space. However, k-means is a much more generic clustering method when Euclidean distance is used. In this work, we will demonstrate that unfortunately, k-means clustering will sometimes fail to give correct results, an unaware fact that may be overlooked by many researchers. This is especially the case when Dynamic Time Warping (DTW) is used as the distance measure in averaging the shape of time series. We also will demonstrate that the current averaging algorithm may not produce the real average of the time series, thus generates incorrect k-means clustering results, and then show potential causes why DTW averaging methods may not achieve meaningful clustering results. Lastly, we conclude with a suggestion of a method to potentially find the shape-based time series average that satisfies the required properties.

1. Introduction

Multimedia data have evolved around our daily lives for quite some time. Especially, once the data could be stored digitally, Multimedia data processing has become a very active area of research as users have higher expectations in using these data at hands. However, typical multimedia manipulations require considerable amount of storage and are computationally intensive. Generally, we can use various image processing techniques [8][12][14][24] [26] to cluster multimedia data, by measuring similarities among the raw videos or images, using certain features such as color, texture, or shape. However, recent work [17][22] have demonstrated the utility of time series representation as an efficient alternative to the raw multimedia data, whose advantages include time and space complexity reduction on clustering, classification, and other data mining tasks. In clustering multimedia time series data, *k*-medoids algorithm with Dynamic Time warping distance measure is often used. In fact, there are many other distance measures that can be effectively used for time series data, but we will mainly focus on DTW due to its ideal shape-based similarity measurement that can break the limitation of one-to-one mapping in Euclidean distance, the most well-known distance metric. Although *k*-medoids with DTW gives satisfactory results, *k*-means clustering is conceivably much more typical in clustering task, where an averaging algorithm is a crucial subroutine in finding a data representation of each cluster. In general, Euclidean distance metric (or other types of Minkowski metric) is used to find an average of all the data within the clusters. However, its one-to-one mapping nature is unable to capture the average shape of the two time series, in which case the Dynamic Time Warping is more favorable.

However, much of past work involving time series averaging appears to avoid using DTW in spite of their need in the shape-similarity-based calculation [2][5][6][7][10][13][18][19][20][25]. On the other hand, some works employ DTW in their template calculation by hierarchically average each time series pair without realizing that this method does not have

2007 International Conference on Multimedia and Ubiquitous Engineering(MUE'07)
0-7695-2777-9/07 $20.00 © 2007 IEEE

an associative property. In fact, there is no publication thus far proposing a generic time series shape averaging method with a proof of correctness. Instead, those shape averaging algorithms are proposed explicitly for specific domains [3][15][16]. In particular, there is only one by Gupta et al. [9], who introduced the shape averaging approach using Dynamic Time Warping, and their work since then has been the basis for all subsequent work involving shape averaging [1][23]. They claim that their approach can correctly average time series when the number of data is in a power of two, e.g., 2, 4, 8, 16, etc.

In this paper, we will show that their claim is not true and demonstrate how *k*-means with DTW averaging could fail to cluster multimedia time series data when the current shape averaging method is applied, and suggest a remedy to the problems of multimedia time series clustering using *k*-means algorithm with DTW in the future.

The rest of the paper is organized as follows. Section 2 explains some important background involving shape averaging. In section 3, we describe the multimedia time series dataset used in this work. Section 4 empirically reveals the problems with current shape averaging method. Sections 5 and 6 provide some discussion and conclusion, along with a suggestion of possible treatment to *k*-means clustering with DTW.

2. Background

In this section, we briefly explain the three main components of this work, Dynamic Time Warping distance measure, DTW averaging, and *k*-means clustering.

2.1. Distance Measurement

Distance measure is extensively used in finding the similarity/dissimilarity between any two time series. The two well known measures are Euclidean distance metric and DTW distance measure. As a distance metric, it must satisfy the four properties – symmetry, self-identity, non-negativity, and triangular inequality.

A distance measure, however, does not need to satisfy all the properties above, e.g., DTW does not have the triangular inequality property, an important key to the explanation why we have such a hard time in shape averaging/template calculation using the Dynamic Time Warping distance measure.

2.2. Dynamic Time Warping Distance Measure

DTW [21] is a well-known shape-based similarity measure for time series data. Unlike the Minkowski distance function, dynamic time warping breaks the limitation of one-to-one alignment, and also supports non-equal-length time series. It uses dynamic programming technique to find all possible paths, and selects the one that yields a minimum distance between the two time series using a distance matrix, where each element in the matrix is a cumulative distance of the minimum of the three surrounding neighbors. Suppose we have two time series, a sequence $Q = q_1, q_2, \ldots, q_i, \ldots, q_n$ and a sequence $C = c_1, c_2, \ldots, c_j, \ldots, c_m$. First, we create an n-by-m matrix, where every (i, j) element of the matrix is the cumulative distance of the distance at (i, j) and the minimum of the three elements neighboring the (i, j) element, where $0<i \le n$ and $0<j \le m$. We can define the (i, j) element as:

$$e_{ij} = d_{ij} + \min\{e_{(i-1)(j-1)}, e_{(i-1)j}, e_{i(j-1)}\} \quad (1)$$

where $d_{ij} = (c_i + q_j)^2$ and e_{ij} is (i, j) element of the matrix which is the summation between the squared distance of q_i and c_j, and the minimum cumulative distance of the three elements surrounding the (i, j) element. Then, to find an optimal path, we have to choose the path that gives minimum cumulative distance at (n, m). The distance is defined as:

$$D_{DTW}(Q,C) = \min_{\forall w \in P}\left\{\sqrt{\sum_{k=1}^{K} d_{w_k}}\right\} \quad (2)$$

where P is a set of all possible warping paths, and w_k is *(i, j)* at k^{th} element of a warping path and K is the length of the warping path.

It is very important to note that during the DTW calculation in equation (1), there could be some ties in selecting the minimum value from the three surrounding element. In this case, the algorithm could arbitrarily choose any neighbor in the tie, thus producing different optimal warping *paths* even though the warping *distance* will always turn out to be the same.

2.3. Dynamic Time Warping Averaging

In some situations, we may need to find a template or a model of a collection of time series, in which case, shape averaging algorithm is desired for a more accurate/meaningful template. DTW distance measure will be exploited to find appropriate mappings for an average. More specifically, the algorithm needs to create a DTW distance matrix and find an optimal

warping path. After the path is discovered, a time series average is calculated along this path by using the index (i, j) of each data point w_k on the warping path, which corresponds to the data points q_i and c_j on the time series Q and C, respectively. Each data point in the averaged time series is simply the mean of two values on the two time series that index (i, j) maps to. $W = w_1, w_2, \ldots, w_k, \ldots, w_K$ is an optimal warping path, where w_k is the mean value between time series whose indices are i and j.

$$w_k = \frac{(q_i + c_j)}{2} \quad (3)$$

Note that in query refinement, where the two time series may have different weights, α_Q for a sequence Q and α_C for a sequence C, eq. (3) above may then be simply generalized according to the desired weight below

$$w_k = \frac{(\alpha_Q \cdot q_i + \alpha_C \cdot c_j)}{\alpha_Q + \alpha_C} \quad (4)$$

Averaging only two time series using Dynamic Time Warping seems to work well because of its symmetric property and the shape after averaging looks correct by our observation. However, we would like to emphasize the importance of choosing the right distance measure for the shape averaging problem. As shown in Figure 1, what we want from a shape averaging algorithm is illustrated in Figure 1 (a) where DTW is used. If the Euclidean or any one-to-one mapping distance measures were used, we would probably end up with undesirable result, as shown in Figure 1 (b).

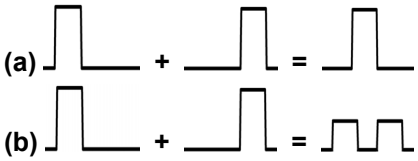

Figure 1. A comparison between (a) shape averaging and (b) amplitude averaging

2.4. *K*-means Clustering

In this work, our main emphasis is on *k*-means clustering algorithm [4] (shown in Table 1) for time series data because time series average/mean is required by the algorithm and must be obtained, as illustrated in step 4 of Table 1. So far, most of the work had chosen Euclidean distance for the averaging method, partly because they were conceivably unaware of any existing shape averaging algorithm for time series data, apart from the reason that Euclidean distance is very simple and fast to compute.

As shown in Table 1 below, the *k*-means algorithm tries to divide N data objects into k partitions or clusters, where each would have one object (mean) as its cluster center, representing all data objects within that cluster. We then assign the rest of the objects to proper clusters and recalculate new centers. We repeat this step until all cluster centers are stable. In general, after each iteration, the quality of the clusters and the means themselves will essentially be improved.

Table 1. *k*-means clustering algorithm

Algorithm *k*-means
1. Decide on a value for *k*.
2. Initialize *k* cluster centers by randomizing data objects.
3. Assign each object to appropriate cluster centers by calculating distance between each particular object and the centers then choosing the nearest center to that object.
4. Calculate new cluster centers by averaging all members containing in each cluster.
5. If none of the *N* objects changes membership, clustering is complete. Otherwise, repeat steps 3 to 5.

Unlike *k*-means clustering, *k*-medoids [11] only differs from *k*-means in the way the cluster centers are chosen and represented (step 4), i.e., it will find new cluster centers by choosing an *existing* data member within each cluster that best represents its cluster center, instead of calculating the cluster members' average.

3. Dataset

To evaluate our hypothesis, we use three classification (labeled) datasets: Leaf, Face, and Gun datasets from the UCR time series data mining archive [http://www.cs.ucr.edu/~eamonn/time_series_data/].

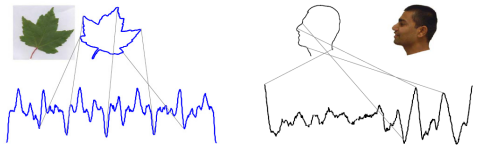

Figure 2. Transformation of raw image shapes to time series

The Leaf dataset contains 6 classes of time series data which are transformed from raw leaf images by using image processing technique (See Figure 2). This dataset contains 442 instances of rescaled lengths of 150 data points. Figure 3 shows some samples of each of the 6 classes of the leaf data--Circinatum, Garryana, Glabrum, Kelloggii, Macrophyllum, and Negundo--

[http://web.engr.oregonstate.edu/~tgd/leaves/]. The original images of the dataset are 512x512 pixels in size, as shown in Figure 4.

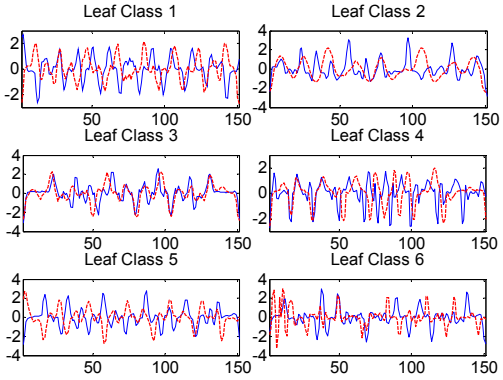

Figure 3. Examples of six species in Leaf dataset

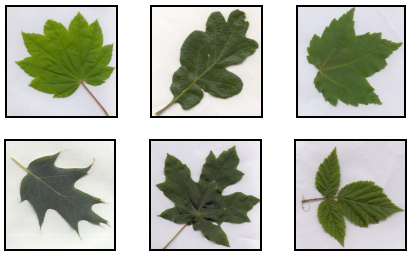

Figure 4. Examples of six-class Leaf images

The Face dataset is derived from four individual head profiles, making different face expressions, as shown in Figure 5. We convert those raw images into time series data of lengths between 107 to 240 data points, 112 instances in total. The transformation technique used in converting from raw images to time series data is similar to that of the leaf dataset above (see Figure 2), and then all instances are normalized and rescaled to 350 data points (See Figure 6).

Lastly, we use the Gun dataset from a video surveillance domain, where the time series data were acquired by plotting movement of hand's position with gun (or finger pointing) in each video frame using image processing techniques (see Figure 7). This dataset contains two classes, i.e., hand with a gun and hand without a gun, with 100 instances each; all are 150 data points long, as shown in Figure 8.

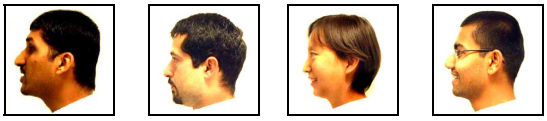

Figure 5. Examples of original Face profiles

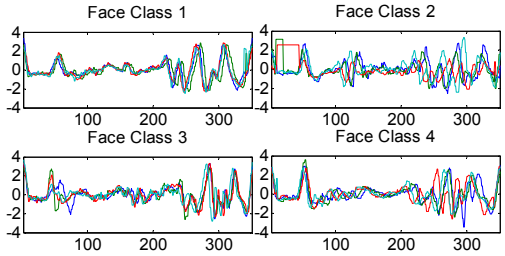

Figure 6. Examples of four different Face profiles after converted into time series

Figure 7. Tracking hand position in each video frame

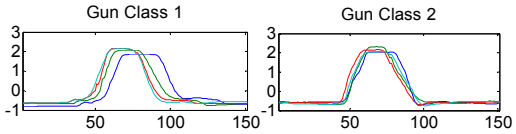

Figure 8. Examples of two classes of Gun data

4. Experiment Evaluation

To validate our claims, we set up a few experiments to uncover the complication in the current shape averaging method. First, we test our overall hypotheses by running *k*-means clustering algorithm and demonstrate its failure in returning meaningful results. The second experiment tests whether reordering of the sequences will have any effect on the averaged result/template. To investigate possible causes of failure, we also run another experiment to test whether averaging two time series using DTW will actually give the correct average.

4.1. *K*-means Clustering with DTW

It has been demonstrated that *k*-medoids clustering for multimedia time series data runs smoothly with DTW. In contrast, this experiment will show that if *k*-means method is instead used in clustering, there is a high probability of failure, comparing to the *k*-medoids algorithm (and that is probably why Euclidean averaging is often used for *k*-means shape averaging despite the use of DTW in cluster membership assignment). We will illustrate this by reporting the average number of iteration up to the point where *k*-means fails, which is when any one of the cluster

centers has no member. Note that this phenomenon would never occur in *k*-means with Euclidean averaging nor in *k*-medoids clustering algorithms. In each dataset, we run the experiment 1,000 times, and *k* is chosen to be its actual numbers of class. For clearer evaluation, we compare the results with the number of iteration obtained using the *k*-medoids methods (which always succeed). Table 2 shows the mean and standard deviation of the number of iteration when *k*-means fails to give meaningful clustering results for each dataset.

Table 2. The mean and standard deviation of number of iteration when *k*-means fails, compares to *k*-medoids successes

	Leaf	Gun	Face
Failure: iteration (*k*-means)	1.32±0.34	5.16±1.71	1.72±0.83
Success: iteration (*k*-medoids)	4.19±0.90	4.06±0.93	3.61±0.72

We can see that the *k*-means clustering often fails prematurely. We then set up some experiments to further examine why this is the case.

4.2. Associative property in DTW averaging

In each iteration of *k*-means clustering, we need to compute the cluster center by averaging the shape of all the cluster members using the Dynamic Time Warping distance measure. We claim that the associative property in DTW averaging does not hold, i.e. different orderings of the data items in an average will give different average result. Therefore, this experiment will demonstrate that a reordering of the sequences of balanced hierarchical averaging will affect the final averaged time series. According to [9], the authors claim the associative property under 2^n data constraint, and explicitly state that no matter how we rearrange the data, it will not make any difference in the final averaged outcome. In this section, we will show that even under the 2^n data constraint, the associative property still does not hold, except when $n = 1$, i.e., having only two time series in the dataset, where the associative property is not defined.

Even when $n = 1$, we cannot guarantee the commutative property of the method, i.e., *DTW_Avg*(*Q*,*C*) may or may not give the same result as *DTW_Avg*(*C*,*Q*), though its symmetric property will give the same the DTW distance. When *n* is larger than one, we would like to test whether shuffling the sequences would affect the (balanced hierarchical) average result. We compute the distance of *every possible pairings*, then reshuffle the data and repeat the computation (100 runs). We then compare whether the distances among all averaged results from each variation using DTW distance are in fact equal, by drawing 64 instances randomly from each dataset to test this discrepancy. It is very surprising to see that the averaged time series from each run do not have the same shape, giving the distances among each of the averaging results from different runs much larger than zero. The mean and standard deviation of discrepancies (distance between two averaging results) are shown in Table 3.

Table 3. Discrepancy distance

	Leaf	Face	Gun
Discrepancy distance	99.32±10.66	24.58±2.49	142.50±13.31

Here, we have confirmed that associative property in DTW averaging does not hold, i.e., the averaged shape solely depends on the order of the time series pair being averaged.

5. Discussion

In search of the remedies, we can categorize the problems into three parts, i.e., a distance measure, an averaging method, and dataset properties. First, since DTW is the distance measure that has no triangular inequality property, the averaged time series may not be the actual mean because DTW cannot guarantee the position of averaged result in Euclidean space. Second, in finding a new the averaging method, we suggest that a new averaging method should satisfy various criteria in our proposed experiments. And third, to satisfy triangular inequalities, it also depends on the properties of the data at hands (generally, only a handful of data within a dataset would violate the triangular inequalities). It is possible to first split the data into groups that triangular inequalities hold within. We can simply find the DTW average for each group, and then finally merge those averages together.

6. Conclusion and Future Works

We have empirically demonstrated some counterexamples and problems to current shape averaging method using DTW distance. From these experiments' findings, we have confirmed that the current DTW averaging is inaccurate and should not be used as a subroutine in *k*-means clustering. In this paper, we intend to make a first attempt in pointing out some interesting problem, which occurs when using *k*-means clustering and DTW. As our future work, from

these findings, we will investigate how these problems can be resolved and come up with possible remedies in accurately averaging shape-based time series data.

7. References

[1] Abdulla, W.H., Chow, D., and Sin, G. (2003). Cross-words reference template for DTW-based speech recognition systems. In Proc. of TENCON, vol.4, pp. 1576- 1579.

[2] Bagnall, A. and Janacek, G. (2005). Clustering Time Series with Clipped Data. Mach. Learn. 58, 2-3 (Feb. 2005), 151-178.

[3] Boudaoud, S., Rix, H., and Meste, O. (2005). Integral shape averaging and structural average estimation: a comparative study. IEEE Trans. on ASSP, vol.53 (10) pp. 3644- 3650.

[4] Bradley, P. S., and Fayyad, U.M. (1998). Refining Initial Points for K--Means Clustering. In Proc. of the 15th ICML. pp. 91-99.

[5] Caiani, E.G., Porta, A., Baselli, G., Turiel, M., Muzzupappa, S., Pieruzzi, F., Crema, C., Malliani, A. and Cerutti, S. (1998) Warped-average template technique to track on a cycle-bycycle basis the cardiac filling phases on left ventricular volume. IEEE Computers in Cardiology. pp.73-76, 13-16 Sep 1998.

[6] Corradini, A. (2001). Dynamic Time Warping for Off-Line Recognition of a Small Gesture Vocabulary. In Proceedings of the IEEE ICCV Workshop--RATFG-RTS. IEEE,.

[7] Chu, S., Keogh, E., Hart, and D., Pazzani, M. (2002). Iterative deepening dynamic time warping for time series. In Proceedings of SIAM International Conference on Data Mining.

[8] Deselaers, T., Keysers, D., and Ney, H. (2003). Clustering Visually Similar Images to Improve Image Search Engines. In Informatiktage 2003 der Gesellschaft für Informatik, Bad Schussenried, Germany, November.

[9] Gupta, L., Molfese, D.L., Tammana, R., and Simos, P.G. (1996). Nonlinear alignment and averaging for estimating the evoked potential. IEEE Transactions on Biomedical Engineering. vol.43, no.4pp.348-356, Apr 1996.

[10] Hu, J. and Ray, B. (2006). An Interleaved HMM/DTW Approach to Robust Time Series Clustering. IBM T.J. Watson Research Center. January 27, 2006.

[11] Kaufman, L. and Rousseeuw, P.J. (1990). Finding Groups in Data. An Introduction to Cluster Analysis. Wiley: New York.

[12] Käster, T., Wendt, V., and Sagerer, G. (2003). Comparing Clustering Methods for Database Categorization in Image Retrieval, LNCS, Vol. 2781, September, pp. 228-235

[13] Keogh, E. and Smyth, P. (1997). An enhanced representation of time series which allows fast classification, clustering and relevance feedback. In Proceedings of the 3rd Conference on Knowledge Discovery in Databases, pp. 24--30, 1997.

[14] Krishnamachari, S. and Abdel-Mottaleb, M. (1999). Image Browsing using Hierarchical Clustering, In Proceeding of 4th IEEE Symposium on Computers and Communications (ISCC).

[15] Lange, D.H., Pratt, H., and Inbar, G.F. (1997). Modeling and estimation of single evoked brain potential components. IEEE Transactions on Biomedical Engineering. Vol.44, pp. 791-- 799.

[16] Mor-Avi, V., Gillesberg, I.E., Korcarz, C., Sandelski, J., Lang, R.M. (1994). Signal averaging helps reliable noninvasive monitoring of left ventricular dimensions based on acoustic quantification. Computers in Cardiology, pp.21-24, 25-28 Sep 1994.

[17] Niennattrakul, V. and Ratanamahatana, C.A. (2006), Clustering Multimedia Data Using Time Series. In Proceedings of International Conference on Hybrid Information Technology. Cheju Island, Korea, Nov 9-11.

[18] Oates, T., Firoiu, L., and Cohen, P.R. (2001). Using Dynamic Time Warping to Bootstrap HMM-Based Clustering of Time Series. In Sequence Learning Paradigms, Algorithms, and Applications. Vol. 1828 of LNCS. Springer Verlag, pp. 35--52.

[19] Prem, E., Hrtnagl, E., and Dorffner, G. (2002). Growing Event Memories for Autonomous Robots. In Proceedings of the Workshop On Growing Artifacts That Live, 7th Int. Conf. on Simulation of Adaptive Behavior, Edinburgh, Scotland.

[20] Rabiner, L. R., Levinson, S. E., Rosenberg, A. E., and Wilpon, J. G. (1990). Speaker-independent recognition of isolated words using clustering techniques. In Readings in Speech Recognition, A. Waibel and K. Lee, Eds. Morgan Kaufmann, pp. 166-179.

[21] Ratanamahatana, C.A. and Keogh, E. (2004). Everything you know about Dynamic Time Warping is Wrong. In Proc. of 3rd SIGKDD Workshop on Mining Temporal and Sequential Data.

[22] Ratanamahatana., C.A. and Keogh, E. (2005). Multimedia Retrieval Using Time Series Representation and Relevance Feedback. In Proceedings of 8th International Conference on Asian Digital Library (ICADL), Bangkok, Thailand.

[23] Salvador, S. and Chan, P. (2004). FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space. In Proc. of KDD Workshop on Mining Temporal and Sequential Data.

[24] Wang, J., Yang, W.-J., and Acharya, R. (1997). Color Clustering Techniques for Color-Content-Based Image Retrieval from Image Databases. In Proc. of the Int'l Conference on Multimedia Computing and Systems, pp. 442-449.

[25] Wilpon, J. and Rabiner, L. (1985). A modified K-means clustering algorithm for use in isolated work recognition. IEEE Transactions Acoustics, Speech, and Signal Processing. vol.33, no.3pp. 587- 594, Jun 1985

[26] Yeung, M.M. and Liu, B. (1995). Efficient Matching and Clustering of Video Shots. In 1995 International Conference on Image Processing, vol. 1, pp. 338-34.