Analysis and Applications of the Minimum First Method for Dynamic Time Warping*

Shan-Ru Liu[1], Kuo-Tsung Tseng[2] and Chang-Biau Yang[1(✉)]
[1] Department of Computer Science and Engineering
National Sun Yat-sen University, Kaohsiung, Taiwan
✉ cbyang@cse.nsysu.edu.tw
[2] Department of Shipping and Transportation Management
National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

Abstract

In 2018, Chen *et al.* proposed the *minimum first method* (MFM) [1] to speed up the calculation of *dynamic time warping* (DTW), which is a well-known and essential step for solving the *time series classification* (TSC) problem. By simply rearranging the calculation order, MFM returns the optimal answer as the original DTW does with less calculation time. MFM is effective in most of experimental datasets, but it may be worse than the original DTW in some other datasets. In this paper, we present two quantitative indicators, including standard deviation of variations and wave oscillation, to automatically determine that which datasets are suitable for MFM. The most of suitability prediction accuracies are higher than 80%. Furthermore, we apply MFM with other DTW related methods to design hybrid methods, including DTW with band constraints and AWarp: warping distance for sparse time series and discuss their time efficiencies of those hybrid methods. In all experiments, our hybrid methods save different amount of time, from 4% to 62%.

Keywords: time series, classification problem, dynamic time warping, minimum first method, quantitative indicator

1. Introduction

Nowadays, many various forms of data are generated. Therefore, how to quickly transform these data into useful information is a critical issue for study. Time series is a sequence of numerical data in chronological order, such as stock indices, exchange rates and electrocardiograms (ECG). In our daily life, there are a large amount of time series data. How to classify these data quickly and correctly is an important issue. Thus, the *time series classification* (TSC) problem was arisen.

To solve the TSC problem, it is essential to define the similarity between two given time series. The *dynamic time warping* (DTW) [2] is one of the most common ways to calculate the distance. The time complexity of DTW is O(n^2), where n is the longer length of the two given series. Many researchers have proposed various methods in order to reduce the execution time of DTW. In general, they reduced the execution time of DTW by shrinking the whole solution space, so the answer may not be optimal. Execution time and approximate result seem to be a trade-off in those methods.

To maintain the original (optimal) calculated distance, Chen *et al.* proposed the *minimum first method* (MFM) [1] to speed up the distance calculation of two time series with DTW by rearranging the calculation order, instead of shrinking its solution space. In most of experimental datasets, MFM takes less time than the original DTW. But in few kinds of datasets, MFM needs more time than the original DTW. In order to automatically determine which datasets are suitable for MFM, we shall find the proper indicators to do it in this paper.

Moreover, we apply MFM to other DTW related methods, such as DTW with band constraints [3, 4] and AWarp: warping distance for sparse time series [5]. The calculated answer is the same as the original method, while the execution time may be reduced.

The rest of this paper is organized as follows. In Section 2, we show the solution spaces of DTW related methods, and Chen's MFM. We discuss how we find our indicators for determining MFM suitability in Section 3. In Section 4, suitability prediction accuracies and time efficiencies of our experiments are presented. Finally, we give our conclusions and future works in Section 5.

2. Preliminaries

2.1 Solution Spaces of Some Known DTW Methods

It is clear to see from (1) that the solution space size of the original *dynamic time warping* (DTW) distance [2] is $O(mn)$, where $m,\ n$ are the lengths of the two given time series respectively.

$$D(i,j)=\begin{cases} 0 & \text{if } i=0 \text{ and } j=0,\\ \infty & \text{if } i=0 \text{ and } j\neq 0,\\ \infty & \text{if } i\neq 0 \text{ and } j=0,\\ d(a_i,b_j)+\min\begin{cases} D(i-1,j)\\ D(i,j-1)\\ D(i-1,j-1)\end{cases} & \text{if } 1\le i\le m \text{ and } 1\le j\le n.\end{cases} \tag{1}$$

In Fig. 1, we give an example to demonstrate the calculation of DTW. Note that the solution space includes the whole matrix in Fig. 1 and the yellow cells form the *warping path* W, which is defined as $W=\{w_1, w_2, \cdots, w_k, \cdots, w_K\}$, where $max(m,n)\le K< m+n-1$.

*This research work was partially supported by the Ministry of Science and Technology of Taiwan under contract MOST 108-2221-E-110-031.

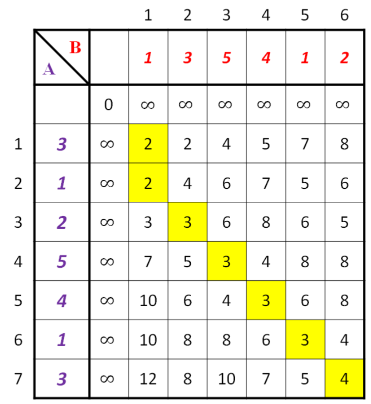

Figure 1: The calculation of DTW for two time series $A = \{3, 1, 2, 5, 4, 1, 3\}$ and $B = \{1, 3, 5, 4, 1, 2\}$. The yellow cells form the warping path W.

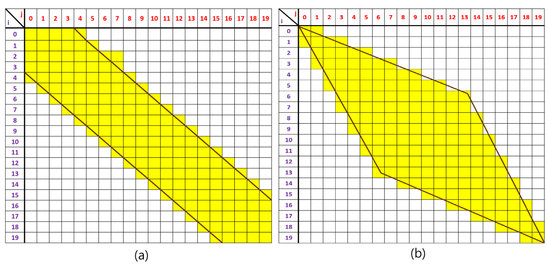

Figure 2: (a) SCBand, where $|i - j| < r = 5$. (b) ITABand, where slopes are -2 and $-\frac{1}{2}$.

In addition , there are two simple and well-known methods for DTW with band constraints, including *Sakoe-Chiba band* [4] and *Itakura Parallelogram band* [3]. The examples of them are shown in Fig. 2. By regulating parameters like the radius or slopes, Sakoe-Chiba band or Itakura Parallelogram band may easily control the size of solution space of the problem, which are yellow cells in Fig. 2. It is worth mentioning that the answer of the Sakoe-Chiba band or Itakura Parallelogram band may not be optimal since some possible solutions (white cells in Fig. 2) are omitted.

As for sparse time series (time series that contain many zeros), instead of shrinking the solution space, AWarp [5] combines those consecutive zeros to make a more compact matrix to reach the goal of speeding up calculations. If the original DTW method is applied, it wastes too much time for calculating zeros. In Fig. 3(a), DTW has to calculate $14 \times 11 = 154$ cells in the matrix, while in Fig. 3(b), AWarp calculates only $8 \times 5 = 40$ cells. AWarp is quite efficient if the given time series is sparse. Note that AWarp gives the optimal answer since the solution space is not shrunk but combined.

2.2 *The Minimum First Method for Dynamic Time Warping*

The *minimum first method* (MFM) [1] combines the concepts of DTW with dynamic windows and the *minimum first search*. When calculating the DTW matrix, MFM always expands the cell with the minimum distance first scheme by using the *priority queue* data structure.

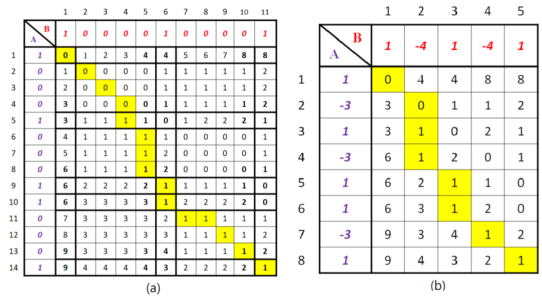

Figure 3: (a) An example for DTW on sparse time series. (b) An example of AWarp.

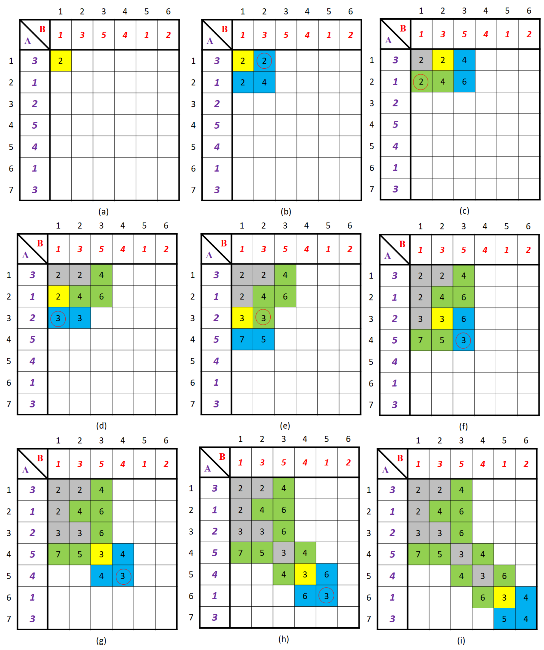

Figure 4: An example for DTW with MFM (DTW-M), where $A = \{3, 1, 2, 5, 4, 1, 3\}$ and $B = \{1, 3, 5, 4, 1, 2\}$.

Fig. 4 shows an example of MFM for DTW (shortened as DTW-M). It stops when the optimal answer is founded, so the calculation of the rest cells are omitted since they cannot be better. The calculated answer is the same as the answer returned by the original DTW method in Fig. 1. Note that in the worst case, it is theoretically possible for MFM to expand the whole matrix to get the optimal answer.

2.3 *Experimental Datasets*

Our experimental datasets for MFM suitability determination are obtained from the UCR time series classification archive [6]. It was first released in 2002 and added to 128 datasets in 2018. Totally 117 UCR datasets form our experimental datasets by removing the 11 datasets which have the missing values.

3. Finding Indicators

To find indicators for MFM suitability determination, we first apply MFM to our 117 experimental datasets and compare the ratio of time with the original DTW

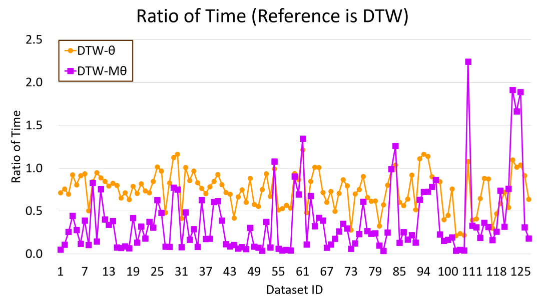

Figure 5: The ratio of time for the 117 datasets with DTW-θ and DTW-Mθ.

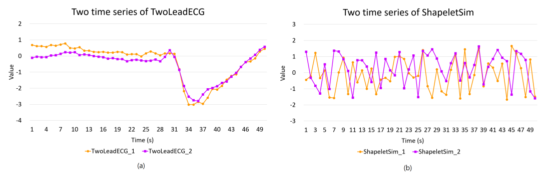

Figure 6: (a) Two time series of the dataset TwoLeadECG. (b) Two time series of the dataset ShapeletSim.

method with threshold. For ease of explanation, we shorten the names of the original DTW method with threshold as DTW-θ, and MFM for DTW with threshold as DTW-Mθ.

Fig. 5 shows the time ratio for the 117 datasets in UCR with DTW-θ and DTW-Mθ. Here, the ratios mean $\frac{\text{DTW-}\theta}{\text{DTW}}$ and $\frac{\text{DTW-M}\theta}{\text{DTW}}$, respectively. We find that DTW-Mθ takes more time than DTW-θ in 11 datasets. So, we want to know the characteristics of these datasets.

3.1 *Observations of Time Series*

Fig. 6(a) shows two time series in TwoLeadECG, which is suitable for DTW-Mθ. We could find that these time series have the similar trends and variations, but fewer wave oscillations. So, when DTW is performed, the warping path is almost along the diagonal direction in these datasets. Therefore, DTW-Mθ expands fewer cells and quickly reaches the bottom-right cell to get the optimal answer (distance).

Fig. 6(b) shows two time series in ShapeletSim, which is not suitable for DTW-Mθ. As one may see, these time series have many wave oscillations, the ups and downs are dramatic. They do not have the similar trends and variations. Therefore, DTW-Mθ needs to expand more cells and takes much time to get the optimal answer.

Based on the observation in Fig. 6, we try to calculate the variation sequence of a time series. Next, we calculate the standard deviation of the variation sequence for each time series. The correlation coefficient of the standard deviation of the variations between time ratio with DTW-Mθ is 0.25. Thus, it may be a good quantitative indicator.

Also, we observe that the number of wave oscillations has a great influence on DTW-Mθ. Thus, we calculate the number of peaks plus the number of valleys in a time series, defined as the number of *oscillations*. The correlation coefficient of the number of oscillations between

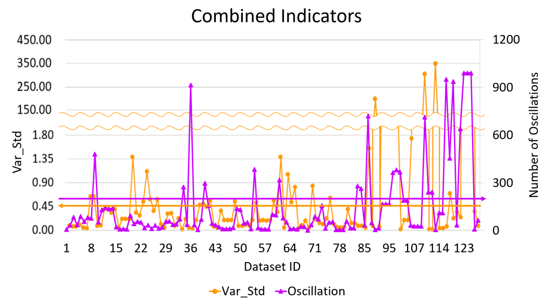

Figure 7: The var_std (orange, left coordinate) and oscillation (purple, right coordinate).

Table 1: The shortened names of the methods.

Shortened name	Meaning of the method
DTW-M	MFM for DTW
DTW-θ	DTW with threshold
DTW-Mθ	MFM for DTW with threshold
SCBand	Sakoe-Chiba band for DTW
SCBand-M	Sakoe-Chiba band with MFM for DTW
SCBand-θ	Sakoe-Chiba band for DTW with threshold
SCBand-Mθ	Sakoe-Chiba band with MFM for DTW with threshold
ITABand	Itakura Parallelogram band for DTW
ITABand-M	Itakura Parallelogram band with MFM for DTW
ITABand-θ	Itakura Parallelogram band for DTW with threshold
ITABand-Mθ	Itakura Parallelogram band with MFM for DTW with threshold
AWarp-M	MFM for AWarp

time ratio with DTW-Mθ is 0.67. Thus, it may be a good quantitative indicator, too.

3.2 *Combined Indicators*

We combine the two indicators mentioned above together and shows in Fig. 7. If we set manually var_std > 0.45 and oscillation > 200, then 7 unsuitable datasets could be found out. Thus, we know that the suitable ratio of all experimental datasets should be $110/117 = 94.02\%$. In fact, there are 4 unsuitable datasets cannot be found out by the above straightforward indicators. To get more precise threshold values for our combined indicators, we use the decision tree to train the combined indicators. We shall present the experimental results in Section 4.

MFM is simple and could be easily implemented with DTW related methods. Thus, not only the original DTW method but also other DTW related methods, we apply MFM to those methods and calculate the MFM suitable ratios with all 117 experimental datasets. In order to make it easy for us to introduce the methods, we list the shortened names of these methods in Table 1.

In Table 2, we list MFM suitability ratios for DTW related methods with 117 experimental datasets. In the worst case, 72% datasets are suitable for MFM-M. In average, 92% datasets are suitable for MFM in the six methods. Thus, MFM is an elegant and useful method to speed up the DTW-related methods.

Table 2: The correlation coefficients of Var_Std and Oscillation for the six methods, and the ratio of MFM suitability.

Method	Name / Ratio	Var_Std	Oscillation	Ratio of suitable datasets
DTW-M	Time	0.15	0.81	84/117=72%
	Cell	0.14	0.32	
DTW-Mθ	Time	0.25	0.67	106/117=91%
	Cell	0.15	0.32	
SCBand-M	Time	0.07	0.78	117/117=100%
	Cell	0.14	0.19	
SCBand-Mθ	Time	0.19	0.47	117/117=100%
	Cell	0.15	0.16	
ITABand-M	Time	0.11	0.75	104/117=89%
	Cell	0.15	0.19	
ITABand-Mθ	Time	0.19	0.50	116/117=99%
	Cell	0.15	0.19	

Table 3: The suitability prediction accuracies of the self-training and self-testing for all of the 117 UCR datasets.

Tree Depth / Hybrided by	2	∞
DTW, DTW-M	85.47%	100.00%
DTW-θ, DTW-Mθ	95.73%	100.00%
ITA, ITA-M	96.58%	100.00%
ITA-θ, ITA-Mθ	100.00%	100.00%

Table 4: The average and standard deviation of the suitability prediction accuracies for 100 times with half-training and half-testing for all of the 117 UCR datasets.

Depth / Hybrided by	2		∞	
	Average	Standard Deviation	Average	Standard Deviation
DTW, DTW-M	79.82%	4.89%	82.32%	5.41%
DTW-θ, DTW-Mθ	91.58%	3.45%	90.46%	3.35%
ITA, ITA-M	92.68%	3.07%	91.60%	2.73%
ITA-θ, ITA-Mθ	98.40%	0.75%	98.45%	0.79%

Table 5: The sum of time and average of time ratio for the self-training and self-testing for all of the 117 UCR datasets.

Depth / Hybrided by	2		∞	
	Sum of Time	Average of Time Ratio	Sum of Time	Average of Time Ratio
DTW	809876	1.00	809876	1.00
DTW, DTW-M	754542	0.72	752443	0.68
DTW-θ	412922	1.00	412922	1.00
DTW-θ, DTW-Mθ	324340	0.43	280398	0.41
ITA	600697	1.00	600697	1.00
ITA, ITA-M	554433	0.63	553958	0.62
ITA-θ	333832	1.00	333832	1.00
ITA-θ, ITA-Mθ	158839	0.38	158839	0.38

4. Suitability Prediction Accuracies and Time Efficiencies with MFM

4.1 MFM Suitability Prediction Accuracies

Table 2 shows the ratios of suitable datasets for the six methods. To increase the suitability prediction accuracy, we use the decision tree to train the combined indicator for DTW-M, DTW-Mθ, ITABand-M and ITABand-Mθ. Note that for SCBand related methods, all datasets are suitable to apply MFM, so they are not necessary to be trained.

In Section 3.2, we use 117 datasets to find the combined indicator. The suitability prediction accuracy of the straightforward method is 94.02%. The way of the straightforward method can be viewed as a self-training and self-testing method.

In order to get higher suitability prediction accuracies, we apply the decision tree to the training of the indicators. First, we apply the self-training and self-testing to all of the 117 UCR datasets, with tree depth 2 and with unlimited tree depth (depth $= \infty$). The suitability prediction accuracies are shown in Table 3.

Next, we do the suitability prediction experiments with half-training and half-testing for 100 times randomly. In other words, the 117 UCR datasets are divided randomly into two parts, one for training and the other for testing. The depths of the decision tree are set to 2 and ∞. Table 4 shows the average and standard deviation of the suitability accuracy for 100 times. Obviously, the decision tree training indeed helps us to determine the suitability of DTW-M, as compared to Table 2.

In Tables 3 and 4, most of our suitability prediction accuracies are higher than 80%, and we could take advantages of that to design quicker hybrid methods.

4.2 Time Efficiencies with MFM

After the suitability has been determined for a dataset, we could use either the original method or MFM accordingly to perform computation in the dataset. We call such a way as the *hybrid method*. In Table 5, the sum of time means the total execution time of all datasets (some are sampled). It is similar to the concept of weighted amount, since some datasets need much execution time and some other datasets need less time. The time ratio for each dataset in the hybrided by DTW, DTW-M means $\frac{\text{DTW-M or DTW}}{\text{DTW}}$, and the average of time ratio is got by averaging the time ratios of the 117 UCR datasets. This is similar to the concept of unweighted amount, since each dataset has the same weight in averaging the time ratios.

In Table 5, we could see that all of our hybrid methods need less time than the original methods. Compared to Table 3, as one may see, if the suitability prediction accuracy is higher, then the sum of time and average of time ratio is lower. Thus, the self-training successfully reduces the required time by determining whether the dataset is suitable for MFM or not in advance.

We illustrate the sum of time and average of time ratio (obtained from the average in 100 experiments with half-training and half-testing) in Table 6. However, we could see the performances of time and time ratio are very different. If the very large datasets, such as Phoneme, PigAirwayPressure and Rock, are not determined correctly in our hybrid method, the execution time would be much more than the original method. In a fair view of each dataset, the average time ratio (viewed as an unweighted sum of time ratio) may be a good measurement. In the table, we could see that the average time ratio of the hybrid methods is better than the original methods. Also, compared to Table 4, it is likely that methods with higher suitability prediction accuracies save more time. Thus, the half-training successfully reduces the time by determining whether the dataset is suitable or not suitable for MFM in advance.

As for sparse time series, since there are only two sparse time series datasets in Mueen's Website [7], we generate the testing datasets of sparse time series randomly. They are represented in the run-length encoding format. The encoding method represents the length of consecutive zeros in the series as a negative number.

We show the experimental results of sparse datasets

Table 6: The average sum of time and average of time ratio for 100 times by half-training and half-testing with the 117 UCR datasets.

Hybrided by \ Depth	2		∞	
	Sum of Time	Average of Time Ratio	Sum of Time	Average of Time Ratio
DTW	365530	1.00	365530	1.00
DTW, DTW-M	811674	0.95	696122	0.96
DTW-θ	211553	1.00	211553	1.00
DTW-θ, DTW-Mθ	162103	0.51	168562	0.56
ITA	301137	1.00	301137	1.00
ITA, ITA-M	376475	0.66	370469	0.70
ITA-θ	177192	1.00	177192	1.00
ITA-θ, ITA-Mθ	82405	0.38	82405	0.38

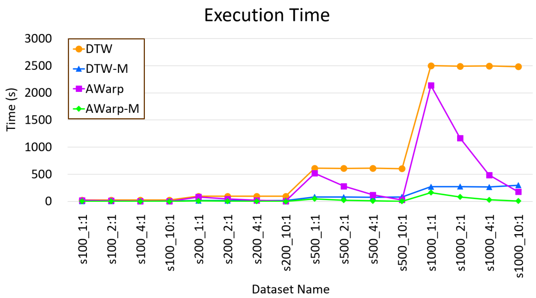

Figure 8: The execution time for sparse datasets with DTW, DTW-M, AWarp and AWarp-M.

for DTW, DTW-M, AWarp and AWarp-M in Fig. 8.

The execution time of DTW and DTW-M are independent of the sparse ratio with a fixed sequence length in Fig. 8. However, AWarp and AWarp-M deeply depend on the sparse ratio. When the ratio of zeros increases, the execution time is reduced.

5. Conclusion

In this paper, we present two indicators, including standard deviation of variations and wave oscillation. They help us to determine if a dataset is MFM suitable or not, with most of prediction accuracies higher than 80%. Then, we design hybrid methods by taking advantages of these indicators, and all of these methods save different amount of time, from 4% to 62%, comparing to their original methods.

It is a pity that presented indicators cannot determine MFM suitability with 100% accuracy. It is a challenging task for us to conquer and we shall try to apply MFM to more DTW related methods in the future.

References

[1] B.-X. Chen, K.-T. Tseng, and C.-B. Yang, "A minimum-first algorithm for dynamic time warping on time series," in *Proceedings of the 23rd International Computer Symposium* , Yunlin, Taiwan, Dec. 2018. Also in *New Trends in Computer Technologies and Applications, Communications in Computer and Information Science*, vol. 1013, pp. 449–456, Singapore: Springer, 2018.

[2] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and Information Systems*, vol. 7, pp. 358–386, Mar. 2005.

[3] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, pp. 67–72, Feb. 1975.

[4] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, pp. 43–49, Feb. 1978.

[5] A. Mueen, N. Chavoshi, N. Abu-El-Rub, H. Hamooni, and A. Minnich, "AWarp: Fast warping distance for sparse time series," in *Proceedings of the 16th International Conference on Data Mining*, (Barcelona, Spain), pp. 350–359, Dec. 2016.

[6] H. A. Dau, E. Keogh, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, Y. Chen, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, "The UCR time series classification archive." `https://www.cs.ucr.edu/~eamonn/time_series_data_2018/`, 2018.

[7] A. Mueen, N. Chavoshi, N. Abu-El-Rub, H. Hamooni, and A. Minnich, "AWarp: Fast warping distance for sparse time series." `https://www.cs.unm.edu/~mueen/Projects/AWarp/`, 2016.