Enhanced Weighted Dynamic Time Warping for Time Series Classification

Pichamon Anantasech and Chotirat Ann Ratanamahatana

Abstract Dynamic time warping (DTW) has been widely used as a distance measure for time series classification because its matching is elastic and robust in most cases. However, DTW may lead to over compression that could align too many consecutive points from one time series to only one point on another. As a result, important feature information could be overlooked, which can be the cause of misclassification particularly when the shape of time series is an essential feature. In order to fix this problem and improve the classification accuracy, we propose a distance measure called an enhanced weighted dynamic time warping, where weight functions are proposed and applied to the DTW distance measure. Other than being parameter-free, our experiment results have demonstrated to impressively outperform other rival methods by a large margin while having less time complexity than the state-of-the-art approaches.

Keywords Dynamic time warping · Weighted dynamic time warping
Adaptive weight · Time series classification

1 Introduction

Dynamic time warping [1] is a distance measure that has been widely accepted and used for classification and clustering of time series data. It has played an important role in time series analysis and has been extensively applied in many fields, such as finance, biometrics, artificial intelligence [2–4]. The main feature for DTW's popularity is its robustness and elastic ability to handle pattern shifting in the time axis, unlike Euclidean distance metric and its improved variations [5, 6] that only take amplitude information into account. Despite its advantages over many distance

P. Anantasech (✉) · C. A. Ratanamahatana
Department of Computer Engineering, Chulalongkorn University, Bangkok 10330, Thailand
e-mail: Pichamon.An@student.chula.ac.th; pichamon.ana@gmail.com

C. A. Ratanamahatana
e-mail: chotirat.r@chula.ac.th

X.-S. Yang et al. (eds.), *Third International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 797, https://doi.org/10.1007/978-981-13-1165-9_60

measures, DTW still has some drawbacks; one data point could be assigned to align with too many consecutive data points, which may not be perfectly sensible to human intuition.

Recent work [11] has attempted to solve this problem by proposing an idea of giving weights to the DTW distance function and demonstrating that these weights could contribute to an improvement in classification accuracy. To further refine the weight function, in this work, we then propose an "enhanced weighted dynamic time warping" function, which will be demonstrated to help improve the classification accuracy while the time complexity remains the same as the original approach. The remainder of this paper is organized as follows. Section 2 provides background and related works. In Sect. 3, our proposed method and the enhanced weight function are explained. Experiments and results are given in Sect. 4, followed by a conclusion in Sect. 5.

2 Background and Related Works

2.1 Dynamic Time Warping (DTW) Distance Measure

DTW was introduced into classification problems and time series mining by Berndt and Clifford [7], where dynamic programming [8] was used to align sequences with different lengths. The idea of DTW is to find an optimal match between two sequences by allowing a nonlinear mapping of one sequence to another and minimizing the distance between two sequences. Considering two time series A and B as follows.

$A = (a_1, a_2, \ldots, a_n)$ and $B = (b_1, b_2, \ldots, b_m)$, DTW finds an optimal warping path between A and B by using dynamic programming to calculate the minimal cumulative distance $\gamma(n, m)$, where $\gamma(i, j)$ is defined as:

$$\gamma(i, j) = \left(a_i - b_j\right)^2 + \min \begin{cases} \gamma(i-1, j) \\ \gamma(i-1, j-1) \\ \gamma(i, j-1) \end{cases} \tag{1}$$

$$\text{DTW distance}\,(A, B) = \sqrt{\gamma(i, j)} \tag{2}$$

This allows DTW to practically match similar shape sequences together even though the sequences may be shifted or out of phase and this qualification has made DTW viable for multitude of time series classification problems. However, DTW does not work well on every single type of data and in many cases align one or only a few data points in one sequence to too many consecutive data points in another sequence. In 2001, Keogh and Pazzani [9] have considered replacing the value of each data point with the first derivative within the dynamic time warping distance function to resolve such problem. Even though this "derivative dynamic time warping" could seem to fix the problem, it may lose the sight of overall shapes or significant features of the sequences [10].

2.2 Weighted Dynamic Time Warping (WDTW) Distance Measure

In 2011, Jeong et al. [11] have proposed a "weighted dynamic time warping" (WDTW) distance measure that provides different weights for each point in the DTW calculation, which has been demonstrated to significantly improve the overall accuracy. Figure 1 illustrates the alignment results when different weights were applied to the distance function, which is shown to resolve the alignment problem.

Given two time series A and B, weighted dynamic time warping (WDTW) can be calculated as follows:

$$\gamma^*(i, j) = |w_{|i-j|}(a_i - b_j)|^2 + \min \begin{cases} \gamma^*(i-1, j) \\ \gamma^*(i-1, j-1) \\ \gamma^*(i, j-1) \end{cases} \tag{3}$$

$$w_k = \left[\frac{w\max}{1 + \mathrm{e}^{-g(k-m_c)}}\right] \tag{4}$$

$$WDTW(A, B) = \sqrt{\gamma^*(i, j)} \tag{5}$$

where $k = |i - j|$, m_c is a midpoint of a sequence.

This proposed work has been a great initiative in suggesting that optimal weight function could consequently improve the classification accuracy. Unfortunately, WDTW has two parameters [$w\max$ and g in Eq. (4)] that need to be optimized, but the authors were still unclear about an approach to attain optimal parameters.

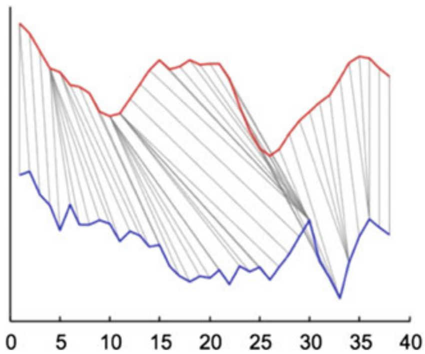

Fig. 1 Alignment results generated by DTW and by WDTW, respectively. (*Source* Jeong et al. [11])

Instead, the paper reports the best possible classification results after trying different sets of parameters.

2.3 Adaptive Cost Dynamic Time Warping (AC-DTW) Distance Measure

As the idea of WDTW has demonstrated to be quite promising; thus, in 2017, Wan et al. [12] proposed new distance measure called "adaptive cost dynamic time warping distance (AC-DTW)" which was developed based on WDTW by Jeong et al. [11]. AC-DTW is calculated as follows:

$$\mathrm{AC}-\mathrm{DTW}(A, B) = \min\begin{cases} c(b_{i-1,j}) \times d_{i,j} + \mathrm{AC}-\mathrm{DTW}(i-1, j) \\ d_{i,j} + \mathrm{AC}-\mathrm{DTW}(i-1, j-1) \\ c\left(a_{i,j-1}\right) \times d_{i,j} + \mathrm{AC}-\mathrm{DTW}(i, j-1) \end{cases} \tag{6}$$

$$d_{i,j} = \left(a_i - b_j\right)^2 \tag{7}$$

$$c(x) = g \cdot r \cdot x + 1 \tag{8}$$

$$r = \frac{\min(n, m)}{\max(n, m)} \tag{9}$$

Nonetheless, the problem of finding optimal g value in Eq. (8) still remains. Moreover, the time complexity of AC-DTW is considerably larger than that of WDTW since their cost function needs to search for an optimal warping path on an $n \times m$ matrix prior to the dynamic programming part. So, the time complexity explodes to $O(N_1 N_2 nm)$ where n and m are the lengths of two time series, N_1 is the number of time series of length m for training, and N_2 is the number of time series of length n for testing.

With our observation regarding the weighted dynamic time warping scheme, in this paper, we propose a parameter free enhanced weighted dynamic time warping (EWDTW) as a competitive choice for a distance measure, as it preserves the same time complexity as the original DTW, and importantly, it has no parameter to tune. Our proposed method restricts the DTW warping through applying proper weights for each data point. As we will demonstrate in Sect. 4, EWDTW can reduce the alignment problem that we mentioned earlier, and as a result could improve the classification accuracy by a large margin.

3 Proposed Algorithm

This section explains our proposed enhanced weighted dynamic time warping (EWDTW) based on the concepts by WDTW [11] and AC-DTW [12] method (EWDTW) which promisingly could reduce the alignment problem.

3.1 Enhanced Weighted Dynamic Time Warping (EWDTW) Distance Measure

Our proposed EWDTW considers a sigmoid function as a weight function since it is an S-shape continuous function, which means that the output would be bounded. As a first step, we need to find an average distance between a reference point and its neighbors. Then, a sum of all the averaged distance is squared in Eq. (12) and is input into the sigmoid function in Eq. (11). From our extensive experiments on the number of neighbors on all datasets, we have considered various number of points around the reference point as its neighbors, and eight (four to the left and four to the right) turned out to be a reasonable value without smoothing out the sequence too much. By applying this weight function to DTW, small values of weight allow EWDTW to warp further than large values of weight. In other words, more average distance between the reference points allows less warping; EWDTW gives priority to the neighbors, whose values are close to the reference point. Given two time series A and B as follows: $A = (a_1, a_2, \ldots, a_n)$ and $B = (b_1, b_2, \ldots, b_m)$, EWDTW distance is defined as follows: A and B as follows: $A = (a_1, a_2, \ldots, a_n)$ and $B = (b_1, b_2, \ldots, b_m)$, EWDTW distance is defined as

$$\gamma^*(i, j) = w_{|i-j|}\left(a_i - b_j\right)^2 + w_i + \min\begin{cases} \gamma^*(i-1, j) \\ \gamma^*(i-1, j-1) \\ \gamma^*(i, j-1) \end{cases} \tag{10}$$

$$w_i = \frac{1}{\left(1 + e^{-k}\right)} \tag{11}$$

$$k = \left(\left(\sum_{i=1}^{n}\left(\sum_{j=-4}^{4}\left|a_i - a_j\right|\right)\right)/8\right)^2 \tag{12}$$

$$\text{EWDTW}(A, B) = \sqrt{\gamma^*(i, j)} \tag{13}$$

3.2 Time Complexity

To compute the EWDTW distance of two time series with the lengths n and m, we need to find a weight for each point. Since its only finds the average distance among neighbors and the reference point, the time complexity is a constant k. To find the optimal warping path on an $n \times m$ matrix, the time complexity is $O(nm)$. As, we disregard the constant, so the time complexity of EWDTW is $O(nm)$, which still is the same as DTW.

4 Experiments

This section explains our proposed enhanced weighted dynamic time warping (EWDTW) based on the concepts by WDTW [11] and AC-DTW [12] method (EWDTW) which promisingly could reduce the alignment problem.

4.1 Datasets

We use 19 datasets from UCR time series classification archive [13] to measure the classification accuracy of our proposed EWDTW method. The number of classes ranges from 2 to 50 classes, and the sequence lengths ranges from 60 to 637 data points. All datasets provided on the archive are already split into training and test sets and are z-normalized.

4.2 Evaluation and Experiment Results

We use one nearest neighbor (1-NN) classifier to measure the performance of EWDTW comparing with Euclidean distance and previously proposed DTW-based methods. The error rate is calculated by Eq. (14), and the results are shown in Table 1. The winning results are underlined and shown in boldface.

$$\text{Error rate} = 1 - \frac{\text{Number of time series correctly classified}}{\text{Total number of time series in the test set}} \tag{14}$$

From Table 1, we can readily see that our proposed EWDTW impressively wins in 16 out of 19 datasets. Moreover, the two datasets—Face All and Wafer have error rates very close to the winning ones. This reconfirms the performance of our proposed method. In addition, to see how well our EWDTW performs, and we also compare our results with those of WDTW reported as best achievable error rates from [11]. Note that the results of WDTW reported in [11] were achieved by trying various

Table 1 Classification error rates of different distance measures on various datasets

Datasets	Classes	Time series length	Training set size	Test set size	Euclidean distance	DTW	Our proposed EWDTW
Synthetic control	6	60	300	300	0.120	0.007	**0.013**
GunPoint	2	150	50	150	0.087	0.093	**0.053**
CBF	3	128	30	900	0.148	0.003	**0.002**
Face all	14	131	560	1690	0.286	**0.192**	0.194
OSULeaf	6	427	200	242	0.479	0.409	**0.297**
Swedish Leaf	15	128	500	625	0.211	0.208	**0.104**
50 words	50	270	450	455	0.369	0.31	**0.187**
Trace	4	275	100	100	0.240	**0**	0.170
Two patterns	4	128	1000	4000	0.090	**0**	**0**
Face four	4	350	24	88	0.216	0.17	**0.102**
Lightning-2	2	637	60	61	0.246	0.131	**0.065**
Lightning-7	7	319	70	73	0.425	0.274	**0.246**
Adiac	37	176	390	391	0.389	0.396	**0.384**
Fish	7	463	175	175	0.217	0.177	**0.166**
Beef	5	470	30	30	**0.333**	0.367	**0.333**
Coffee	2	286	28	28	**0**	**0**	**0**
Olive oil	4	570	30	30	**0.1333**	0.167	**0.133**
Wafer	2	152	1000	1000	0.005	**0.004**	0.006
Yoga	2	426	300	300	0.170	0.165	**0.142**

The winning results are underlined in boldface

sets of parameters (brute force) and the best error rates were selected (Not through parameter tuning). The results are shown in Table 2.

Impressively, our proposed EWDTW could achieve better or the same classification error rates in 12 out of 19 datasets, and the rest do have error rates quite close to the optimal cases of WDTW. These promising results reconfirm that our proposed method has great potentials in improving the classification accuracy.

Table 2 Classification error rates, comparing our proposed EWDTW with the best case of WDTW on all datasets

Datasets	WDTW [11]	Our proposed EWDTW
Synthetic control	**0.002**	0.013
GunPoint	**0.004**	0.053
CBF	**0.002**	**0.002**
Face all	0.257	**0.194**
OSULeaf	0.372	**0.297**
Swedish Leaf	0.138	**0.104**
50words	0.194	**0.187**
Trace	**0**	0.170
Twopatterns	**0**	**0**
FaceFour	0.136	**0.102**
Lightning-2	0.1	**0.065**
Lightning-7	**0.2**	0.246
Adiac	**0.364**	0.384
Fish	**0.126**	0.166
Beef	0.6	**0.333**
Coffee	0.133	**0**
Olive oil	0.188	**0.133**
Wafer	**0.002**	0.006
Yoga	0.165	**0.142**

The winning results are underlined in boldface

In addition, to see if our proposed EWDTW could actually reduce the alignment problem, we plot the some of the alignments between time series sequences of the same class from the Swedish Leaf dataset. Figure 2 illustrates the alignments achieved from DTW distance measure (Left) and the alignments achieved from our proposed EWDTW distance measure (Right). The alignments are apparently improved and match better to human intuition.

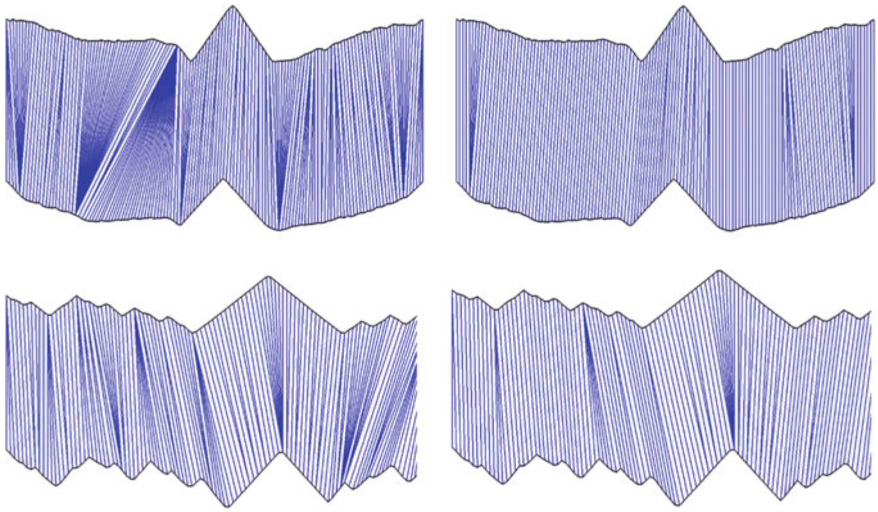

Fig. 2 Alignment between two time series in the same class of Swedish Leaf dataset generated by DTW (Left) and EWDTW (Right)

5 Conclusion

In this paper, we proposed an enhanced weighted dynamic time warping (EWDTW) algorithm as a distance measure for time series classification, by adopting sigmoid function into our weight function. Our EWDTW can reduce the DTW alignment problem, where too many consecutive data points are aligned with one or a few data points. Our algorithm is parameter free and has the same time complexity as the original DTW algorithm. The experiment results demonstrate that our method outperforms other existing methods in almost all the datasets. Our future work will be focus on possibility of other weight functions that able to further improve the accuracy.

References

1. Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. In: Proceedings of the workshop on knowledge discovery in databases, pp 359–370
2. Legrand B, Chang CS, Ong SH et al (2008) Chromosome classification using dynamic time warping. Pattern Recognit Lett 29:215–222
3. Rath TM, Manmatha R (2003) Word image matching using dynamic time warping, In: 2003 IEEE Computer society conference on computer vision and pattern recognition proceedings, vol 2, pp 521–527, IEEE
4. Bashir M, Kempf J (2008) Reduced dynamic time warping for handwriting recognition based on multidimensional time series of a novel pen device. Int J Intell Syst Technol (WASET) 3:1839–1845

5. Agrawal R, Faloutsos C, Swami A (1993) Efficient similarity search in sequence databases, In: Proceeding of Intl Conference on Data Organazation. 730, pp 69–84
6. Struzik ZR, Siebes A (1998) Wavelet transform in similarity paradigm, In: Proceedings of the second Pacific-Asia conference on research and development in knowledge discovery and data mining, pp 295–309
7. Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns, In: Time series, KDD workshop. 10, pp 359–370
8. Berndt DJ, Clifford J (1996) Finding patterns in time series: a dynamic programming approach, In: Advances in knowledge discovery and data mining, American association for artificial intelligence, pp 229–248
9. Keogh EJ, Pazzani MJ (2001) Derivative dynamic time warping. In: Proceedings of the 2001 SIAM international conference on data mining
10. Yu D, Yu X, Hu Q et al (2011) Dynamic time warping constraint learning for large margin nearest neighbor classification. Inform Sci 181:2787–2796
11. Jeong YS, Jeong MK, Omitaomu OA (2011) Weighted dynamic time warping for time series classification. Pattern Recogn 44(9):2231–2240
12. Wan Y, Chen XL, Shi Y (2017) Adaptive cost dynamic time warping distance in time series analysis for classification. J Comput Appl Math 319:514–520
13. Chen Y, Keogh E, Hu B, Begum N, Bagnall A (2015) Abdullah Mueen and Gustavo Batista (2015) The UCR Time Series Classification Archive. URL www.cs.ucr.edu/~eamonn/time_series_data/