Information Sciences 656 (2024) 119921

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A novel distance measure based on dynamic time warping to improve time series classification

Yutao Liu [a,*], Yong-An Zhang [b], Ming Zeng [b], Jie Zhao [a]

[a] The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China
[b] Harbin Institute of Technology, Harbin 150001, Heilongjiang, China

ARTICLE INFO

Keywords:
Dynamic time warping
Similarity measure
Time series classification
One nearest neighbor classification

ABSTRACT

Dynamic time warping (DTW) is the most widely used method to evaluate the similarity between time series. However, the DTW distance only takes into account the difference in amplitude, but does not reflect the time distortion information between them. In this paper, we propose a novel time similarity metric, called the time distortion coefficient, based on the DTW warping path to quantify the time distortion between time series. It is able to characterize the type and degree of time distortion between two time series at each point. By summing the absolute values of the time distortion coefficients, the overall time distortion is introduced to quantify time distortion between two time series. For the Nearest Neighbor (NN) based time series classification, a fusional similarity measure combining the DTW distance and the overall time distortion measure is proposed, which is able to evaluate the similarity in both amplitude and time domains. The experimental results conducted on the UCR time series classification archive datasets demonstrate that the proposed fusional similarity measure can significantly improve the classification accuracy of the 1-NN classifier with only a small amount of additional computational cost compared to the DTW distance and other metrics.

1. Introduction

With the rapid development of data acquisition and storage technology, time series data has become ubiquitous in various domains, such as agriculture [1,2], climate [3], finance [4], chemistry [5], clinical medicine [6], biology [7] and more. Time series classification is an increasing research topic due to its wide range of applications including medical diagnosis, gesture recognition, speech recognition, handwriting recognition and sign language recognition. Time series classification methods can be divided into three categories: feature-based, model-based, and distance-based methods [8]. This work focuses on distance-based time series classification. For distance-based classification algorithms, the recognition rate depends heavily on the quality of the similarity measure of the time series [9,10].

Unlike classical data, time series are sequences of time-dependent observations. In addition to differences in amplitude, time distortions, such as shift, stretch, compression, absence and redundancy, are inevitable between time series data. In this paper, we focus on the time series with uniform sampling frequency. Traditional distance measures, such as Manhattan distance and Euclidean distance, are not suitable for measuring the similarity between time series because they adopt a time-rigid alignment and are sensitive to the time distortion. Furthermore, they are unable to deal directly with the unequal length time series without some pre-processing.

* Corresponding author.
E-mail addresses: liuyutao190160@126.com (Y. Liu), zhangyongan76@163.com (Y.-A. Zhang), zengming@hit.edu.cn (M. Zeng), zhaojie@zzu.edu.cn (J. Zhao).

https://doi.org/10.1016/j.ins.2023.119921
Received 14 March 2022; Received in revised form 28 September 2023; Accepted 15 November 2023

Available online 20 November 2023
0020-0255/

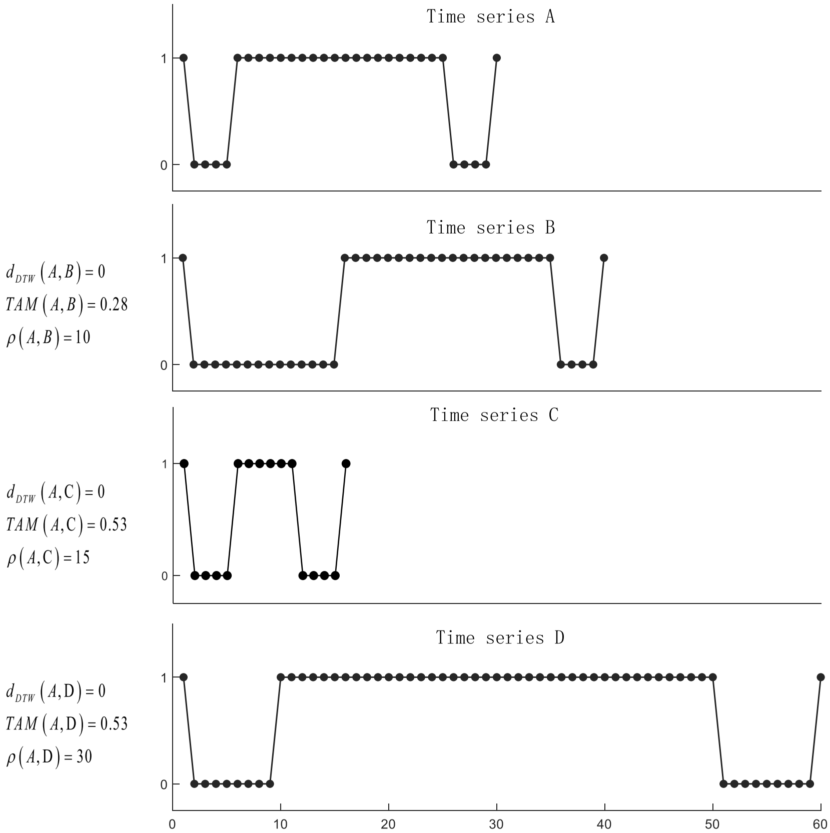


Fig. 1. Similarity measure for time series where the temporal differences dominate. The time series $\boldsymbol{A}$ is a binary series, and the other time series are obtained by local stretching and compression of the time series $\boldsymbol{A}$. An annotation is provided on the left to show the DTW distance, the TAM distance and the time distortion measure.

To overcome the disadvantages of these lock-step distances, a number of elastic distances such as Dynamic Time Warping (DTW) [11], Longest Common Subsequence (LCSS) [12], Edit Distance on Real Sequence (EDR) [13] and Edit Distance with Real Penalty (ERP) [14], have been developed. These elastic distances support flexible alignment to match similar subsequences with different phases, which can handle the time distortions well.

Among these elastic distance measures, DTW is the most widely used method for evaluating the similarity between time series. By warping the time axis of one time series under certain constraints, DTW achieves an optimal alignment with a minimum cumulative cost. This minimum cumulative cost, as known as the DTW distance, is used as a similarity metric between two time series. Recent studies have demonstrated that the One Nearest Neighbor (1-NN) classifier with DTW distance is highly effective and hard to beat [15,16]. However, DTW distance has a fatal limitation in that it only captures differences in the amplitude and does not to reflect differences in the time domain [17]. Fig. 1 clearly illustrates this drawback of the DTW distance, where the time series $\boldsymbol{A}$ is a binary series, and the other time series are obtained by locally stretching and compressing the time series $\boldsymbol{A}$. The DTW distances between time series $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ are all zero. Obviously, the DTW distance fails to measure the similarity for the time series where the temporal differences dominate. Therefore, there is an urgent demand for a metric that characterises the type and degree of time distortion between time series.

Comparing the lengths of time series is a simple and straightforward approach to measure the temporal differences between time series. However, this method provides only a rough result and cannot be applied to the time series of equal length. As another output of DTW, the DTW warping path records the flexible time warping relationship between two time series. Therefore, the time warping information hidden in the warping path can be further exploited to quantify the difference between two time series in the time domain [17]. Unfortunately, the potential of the DTW warping path remains largely untapped in the current literature. In most of the literature, the warping path is only utilized for visual comparison of alignments generated by DTW and its variants [18,19]. A number of metrics, such as the mean deviation, mean absolute deviation [20,21], maximum deviation and error rate [22], have been developed based on the deviation between the ground truth alignment path and any given test alignment path to evaluate the quality of the alignment. In [22], a metric called singularity score was defined based on the number of observations in each observed group to measure the severity of the pathological alignment. To the best of our knowledge, only a few studies have focused on exploiting the time distortion information embedded in the DTW warping path. In [23], the ratio of the length between the DTW warping path and the time series was used to quantify the time distortion between time series. This metric provides only a rough measure of the time distortion between time series. According to the direction of the DTW warping path, a temporal distance, namely Time

Alignment Measurement (TAM), has been developed to measure the similarity of time series in the temporal domain [17]. However, the TAM distance only defines the type of time distortion (in phase, advance or delay) for each pair of alignment points, and does not quantify the degree of time distortion. Therefore, the TAM distance may fail to measure the temporal similarity in some cases. As shown in Fig. 1, the length of the time series $\boldsymbol{A}$ is 30, and the lengths of the test time series $\boldsymbol{C}$ and $\boldsymbol{D}$ are 16 and 60, respectively. Intuitively, the time series with large differences in length tend to have a large time distortion. However, the TAM distance gives a contradictory result, i.e. $TAM(\boldsymbol{A},\boldsymbol{C}) = TAM(\boldsymbol{A},\boldsymbol{D})$.

In this paper, we propose a novel time distance, the time distortion coefficient, to quantify the similarity between time series in the time domain by exploring the warping path. First, according to the time warping information in the warping path, the alignment between two time series can be divided into several independent groups. Based on the number of points in these two time series in each group of alignments, the time distortion coefficient is defined to characterize the type and degree of time distortion between two time series at each point. Then, the overall time distortion between two time series is then calculated by summing the absolute values of the time distortion coefficients. Since the time distortion coefficient is calculated directly from the DTW warping path, its computational complexity is proportional to the length of the DTW warping path. Incidentally, the time distortion coefficient is also applicable to derivatives of DTW and other flexible sequence alignment algorithms.

The 1-NN classifier is a typical distance-based classification algorithm. In order to improve the accuracy of the 1-NN classifier, we develop a fusional similarity measure, which is a weighted sum of the DTW distance and the total time distortion. Considering that the DTW distances and the total time distortions are two different variables, these two distances are first pre-processed through the max-min normalization procedure. The evaluation experiment is conducted on the UCR Time Series Classification Archive datasets over the DTW and other four variants, Constrained DTW (cDTW) [24], Weighted DTW (wDTW) [25], Derivative DTW (dDTW) [23] and shapeDTW [21]. In addition to the DTW distance, other four metrics, such as average warping distance [23], warping path length [23], TAM distance and Singularity Score (SS) [22] are utilized for comparison. The experimental results show that the proposed fusional similarity distance can significantly improve the accuracy of the 1-NN classifier with an almost negligible additional computational cost compared to other distance metrics.

The remainder of this paper is organized as follows. A brief overview of DTW is presented in section 2. In the following section, the characteristics of DTW alignment between time series are first analyzed. Then, based on the DTW warping path, the calculation of the time distortion coefficient is introduced. In section 4, a fusional similarity measure is proposed for nearest neighbor based time series classification. The experimental results are presented and discussed in section 5. Finally, we present the conclusion and perspective of this work in section 6.

2. Review of dynamic time warping

Given two time series $\boldsymbol{P}$ and $\boldsymbol{Q}$, of length $m \in \mathbb{N}$ and $n \in \mathbb{N}$ respectively, where:

$$\boldsymbol{P} = \left(p_1, p_2, \dots, p_m\right)$$

$$\boldsymbol{Q} = \left(q_1, q_2, \dots, q_n\right)$$

Suppose $\boldsymbol{P}$ is the reference time series, and $\boldsymbol{Q}$ is the test time series.

The goal of DTW is to find an optimal alignment between two time series with the minimum cumulative cost by locally stretching and compressing one of the time series. The alignment between two time series can be represented by an $L \times 2$ matrix $\boldsymbol{W}$, also called the warping path. The length of the warping path indicates the number of pairwise alignments between two time series. Each row vector $\boldsymbol{w}_k = (i,j) \in [1:m] \times [1:n]$ for $k \in [1:L]$ of the matrix $\boldsymbol{W}$ represents a pairwise of alignment, where i and j denote the index of the points in $\boldsymbol{P}$ and $\boldsymbol{Q}$, respectively.

Considering the time-order of the time series, the alignment is typically subject to the following constraints:

- Boundary constraint: the first and last points of the time series must be aligned, i.e. $\boldsymbol{w}_1 = (1,1)$ and $\boldsymbol{w}_L = (m,n)$. In other words, the alignment refers to the whole time series.
- Step size constraint: the transition from one point to the next point must be along the following directions: diagonal, up and right, i.e. $\boldsymbol{w}_{k+1} - \boldsymbol{w}_k \in \{(1,0),(0,1),(1,1)\}$ for $k \in [1:L-1]$. This means that the warping path cannot go back in time, cross, and make any jumps.

Based on these constraints, there are various warping paths satisfying the above constraints, and the length L of a warping path satisfies the following condition:

$$\max(m,n) \le L < m + n - 1$$

In order to compare the similarity of these pairs of aligned points, a local cost measure is required. Generally, the Euclidean distance or the squared Euclidean distance is utilized as the local cost measure, i.e.

$$d\left(p_i, q_j\right) = \sqrt{\left(p_i - q_j\right)^2} \quad \text{or} \quad d\left(p_i, q_j\right) = \left(p_i - q_j\right)^2$$

In fact, any cost function can be utilized. For various of feature-based DTW algorithms, the context features based on the neighborhoods of each point, such as shape context [26,27], raw-subsequence [17], discrete wavelet transform, derivative [17,23], piecewise aggregate approximation, histogram of oriented gradients [21], are employed to replace the raw value.

According to the warping path, the cumulative cost of a warping path is defined as the sum of the local costs between all pairs of aligned points,

$$d_{\boldsymbol{W}}(\boldsymbol{P},\boldsymbol{Q}) = \sum_{k=1}^{L} d\left(p_{w_{k,1}}, q_{w_{k,2}}\right) \tag{1}$$

where $p_{w_{k,1}}$ and $q_{w_{k,2}}$ are k-th pairwise of aligned points in $\boldsymbol{P}$ and $\boldsymbol{Q}$ respectively.

The warping path having the smallest cumulative cost is referred to as the DTW warping path $\boldsymbol{W}^*$. The cumulative cost of the DTW warping path $\boldsymbol{W}^*$ is defined as the DTW distance:

$$\begin{aligned} d_{DTW}(\boldsymbol{P},\boldsymbol{Q}) &= d_{\boldsymbol{W}^*}(\boldsymbol{P},\boldsymbol{Q}) \\ &= \min\{d_{\boldsymbol{W}}(\boldsymbol{P},\boldsymbol{Q}) | \boldsymbol{W} \in \mathbb{W}_{m,n}\} \end{aligned} \tag{2}$$

where $\mathbb{W}_{m,n}$ is the space of all possible warping paths of two time series of length m and n.

There are exponentially many warping paths that satisfy the above conditions, but we are only interested in the warping path that minimizes the warping cost. To improve the efficiency, the dynamic programming technique is utilized to find the optimal warping path. First, a local cost matrix $\boldsymbol{d}$ is computed. Then, an $m \times n$ cumulative cost matrix $\boldsymbol{D}$ is constructed, where the element $\boldsymbol{D}(i,j) \in \mathbb{R}^{i \times j}, 1 \le i \le m, 1 \le j \le n$ can be calculated through the following recurrence formula:

$$D(i,j) = \begin{cases} d(i,j) & i=1, j=1 \\ d(i,j) + D(i,j-1) & i=1, j>1 \\ d(i,j) + D(i-1,j) & i>1, j=1 \\ d(i,j) + \min\begin{cases} D(i-1,j-1) \\ D(i-1,j) \\ D(i,j-1) \end{cases} & i>1, j>1 \end{cases} \tag{3}$$

Algorithm 1 DTW warping path.

Input: Cumulative cost matrix $\boldsymbol{D}$.
Output: DTW warping path: $\boldsymbol{W}^*$.
Initialization: $i = m, j = n, \boldsymbol{W}^* = [\quad]$;
while $(i > 1) \| (j > 1)$ **do**
  if $i == 1$ **then**
    $j = j - 1$;
  else if $j == 1$ **then**
    $i = i - 1$;
  else
    $i = i - (\boldsymbol{D}(i-1,j) \le \boldsymbol{D}(i,j-1) \| \boldsymbol{D}(i-1,j-1) \le \boldsymbol{D}(i,j-1))$;
    $j = j - (\boldsymbol{D}(i,j-1) \le \boldsymbol{D}(i-1,j) \| \boldsymbol{D}(i-1,j-1) \le \boldsymbol{D}(i-1,j))$;
  end if
  $\boldsymbol{W}^* = [\boldsymbol{W}^*; i, j]$;
end while
return $\boldsymbol{W}^*$.

Based on Eq. (3), the matrix $\boldsymbol{D}$ is filled from left to right and from bottom to top. As a result, DTW has a quadratic time and space complexity $O(m \cdot n)$. The upper-right cell $D(m,n)$ gives the total cumulative cost, i.e. $d_{DTW}(\boldsymbol{P},\boldsymbol{Q}) = D(m,n)$. Although the DTW distance has already been determined, the DTW warping path is still hidden in the cumulative cost matrix. Using the backtracking procedure, the DTW warping path can be revealed from the cumulative cost matrix based on the opposite direction of the step constraint. Starting from the upper-right cell $D(m,n)$, the three backward neighbors of the current position are checked at each stage until the lower-left cell $D(1,1)$ is reached. The pseudo code of the DTW warping path calculation is presented in Algorithm 1.

There are several remarks about the DTW algorithm:

- The DTW distance is unique, but the DTW warping path is not unique.
- The DTW distance is non-negative if the local cost measure is non-negative.
- $d_{DTW}(\boldsymbol{P},\boldsymbol{Q}) = 0$ is a necessary and insufficient condition for $\boldsymbol{P} = \boldsymbol{Q}$.
- The DTW distance is symmetric if the local cost distance is symmetric, i.e. $d_{DTW}(\boldsymbol{P},\boldsymbol{Q}) = d_{DTW}(\boldsymbol{Q},\boldsymbol{P})$.
- The DTW distance does not satisfy the triangle inequality even if $d(\cdot)$ is a metric.

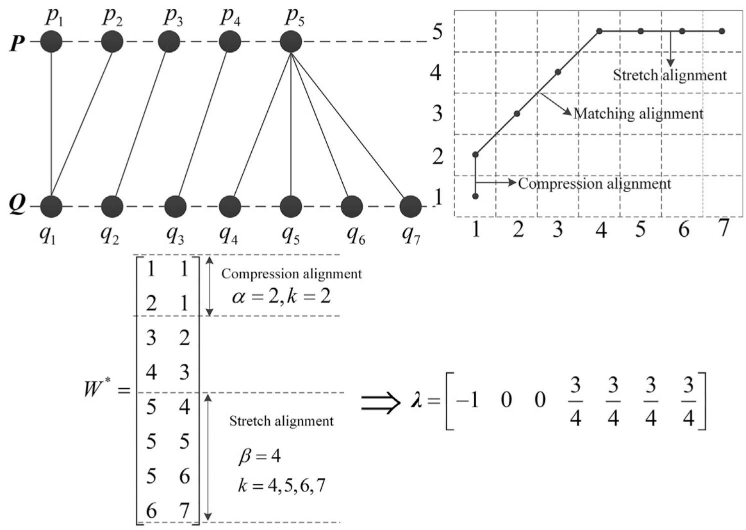


Fig. 2. Example of DTW alignment and time distortion coefficient between two time series.

3. Time distortion measure between time series

The DTW distance is a commonly used similarity metric. However, it only takes into account the differences in amplitude between two time series after elastic warping, but does not reflect the time distortion information between them. As mentioned above, DTW provides not only a similarity measure between two time series, but also an elastic alignment between them. The DTW warping path that contains rich temporal information should be fully exploited. In the following section, we first present a further interpretation of the DTW alignment. Then, based on the DTW warping path, a novel time similarity metric, called the time distortion coefficient, is introduced to evaluate the differences between time series in the time domain.

3.1. The characteristics of DTW alignment

As a type of elastic measure, DTW supports elastic alignment. In addition to one-to-one alignment, DTW also allows one-to-many alignment and many-to-one alignment. The DTW alignment can be segmented into a number of independent groups based on the time warping between two time series. According to the number of points of the test time series in each group of alignment, the DTW alignment can be classified into three classes: matching alignment, compression alignment, and stretch alignment. Taking Fig. 2 as an example, the details of these three types of alignment are discussed below:

- Matching alignment: matching alignment is a one-to-one alignment, where the points of two time series are perfectly aligned one-to-one. The warping path follows the diagonal direction, i.e. $\boldsymbol{w}_{k+1} - \boldsymbol{w}_k = (1, 1)$.
- Compression alignment: compression alignment is a one-to-many alignment, where a single point q_i in the test time series $\boldsymbol{Q}$ aligns with α consecutive points $\{p_j, p_{j+1}, \cdots, p_{j+\alpha-1}\}$ in the reference time series $\boldsymbol{P}$. Obviously, $\boldsymbol{Q}$ is compressed with respect to $\boldsymbol{P}$ at point q_i. The warping path follows the vertical direction, i.e. $\boldsymbol{w}_{k+1} - \boldsymbol{w}_k = (1, 0)$. In other words, the index i of point q_i appears α times in the second column of the DTW warping path, i.e. $\boldsymbol{w}_2(k) = \boldsymbol{w}_2(k+1) = \cdots = \boldsymbol{w}_2(k+\alpha-1)$.
- Stretch alignment: stretch alignment is a many-to-one alignment, where β consecutive points $\{q_i, q_{i+1}, \cdots, q_{i+\beta-1}\}$ in $\boldsymbol{Q}$ align with single point p_j in $\boldsymbol{P}$. In this case, $\boldsymbol{Q}$ is stretched with respect to $\boldsymbol{P}$ at this segment. The warping path follows the horizontal direction, i.e. $\boldsymbol{w}_{k+1} - \boldsymbol{w}_k = (0, 1)$. The index j of point p_j appears β times in the first column of the DTW warping path, i.e. $\boldsymbol{w}_1(k+1) = \boldsymbol{w}_1(k+1) = \cdots = \boldsymbol{w}_1(k+\beta-1)$.

3.2. Calculation of time distortion coefficient

Following these three types of DTW alignment, we define a time distortion coefficient to characterize the type and degree of time distortion between time series. In the matching alignment, the points in two time series are perfectly aligned one to one. Obviously, there is no time distortion in the matching alignment. Both compression alignment and stretch alignment suffer from time distortion. Therefore, we introduce a scheme to define the time distortion coefficient for these two types of alignment. We transform the compression alignment or stretch alignment into the matching alignment by adding or merging operations on the test time series. The adding and merging operations correspond to the negative and positive signs of the time distortion coefficient, respectively. The number of points in the adding or merging operation indicates the value of the time distortion coefficient.

For the stretch alignment, $\alpha - 1$ points are added to the test time series near the point q_i to obtain α one-to-one alignments. Thus, the time distortion coefficient of q_i is defined as $\lambda(k) = 1 - \alpha$. On the contrary, β points of the test time series are merged into one point for the compression alignment. In other words, $\frac{\beta-1}{\beta}$ of each point in this segment is deleted. Thus, the time distortion coefficient is defined as $\lambda(k) = 1 - \frac{1}{\beta}, k = i, i+1, \cdots, i+\beta-1$. In summary, the time distortion coefficient of the test time series with respect to the reference time series is defined as follows:

$$\lambda(k) = \begin{cases} 0 & \text{For matching alignment} \\ 1-\alpha & \text{For compression alignment} \\ 1-\frac{1}{\beta} & \text{For stretch alignment} \end{cases} \tag{4}$$

Since the time distortion of a matching alignment is zero, we will mainly consider the compression alignment and the stretch alignment. As mentioned above, the second column of the warping path is equal for a compression alignment. Therefore, the compression alignment is found by searching for the segments containing consecutive identical values in the second column of the DTW matching path. The length of these segments determines the value of the time distortion coefficient. Similarly, the segments containing consecutive identical values in the first column of the DTW matching path are retrieved to determine the time distortion coefficient of the stretch alignment. As a result, the computational complexity of the time distortion coefficient is equal to the length of the DTW warping path, i.e. $O(L)$. Algorithm 2 summarizes the above process. Fig. 2 shows an example of calculation of the time distortion coefficient.

Algorithm 2 Time distortion coefficient.

Input: DTW warping path $\boldsymbol{W}^*$.
Output: Time distortion coefficient: λ.

1: Initialization: $\lambda = \text{zeros}(n, 1)$;
Time distortion coefficient of compression alignment:
2: $start = i \Leftarrow \boldsymbol{w}_2(i-1) \neq \boldsymbol{w}_2(i)$ && $\boldsymbol{w}_2(i) \neq \boldsymbol{w}_2(i+1)$;
3: $stop = j \Leftarrow \boldsymbol{w}_2(j-1) = \boldsymbol{w}_2(j)$ && $\boldsymbol{w}_2(j) \neq \boldsymbol{w}_2(j+1)$;
4: $\alpha = stop - start + 1$;
5: $\lambda\left(\boldsymbol{w}_2(start)\right) = 1 - \alpha$;
Time distortion coefficient of stretching alignment:
6: $start = i \Leftarrow \boldsymbol{w}_1(i-1) \neq \boldsymbol{w}_1(i)$ && $\boldsymbol{w}_1(i) \neq \boldsymbol{w}_1(i+1)$;
7: $stop = j \Leftarrow \boldsymbol{w}_1(j-1) = \boldsymbol{w}_1(j)$ && $\boldsymbol{w}_1(j) \neq \boldsymbol{w}_1(j+1)$;
8: $\beta = stop - start + 1$;
9: $\lambda(k) = 1 - \frac{1}{\beta}, k \in \left[\boldsymbol{w}_2(start), \boldsymbol{w}_2(stop)\right]$;
10: **return** λ.

Lemma 1. *Suppose $\boldsymbol{P}$ and $\boldsymbol{Q}$ are two time series of length m and n, where $\boldsymbol{P}$ is the reference time series, and $\boldsymbol{Q}$ is the test time series. λ is the time distortion coefficient of $\boldsymbol{Q}$ with respect to $\boldsymbol{P}$. The sum of λ is equal to the difference in length between these two time series, i.e.,*

$$\sum_{k=1}^{n} \lambda(k) = n - m \tag{5}$$

Proof. Assume that the DTW alignment between $\boldsymbol{P}$ and $\boldsymbol{Q}$ contains a groups of matching alignments, b groups of compression alignments and c groups of stretch alignments. For the i-th $i \in [1 : b]$ compression alignment, α_i points of $\boldsymbol{P}$ align with one point in $\boldsymbol{Q}$. For the j-th $j \in [1 : c]$ stretch alignment, β_j points of $\boldsymbol{Q}$ align with one point in $\boldsymbol{P}$. Based on Eq. (4), the sum of λ is calculated as follows:

$$\begin{aligned} \sum_{k=1}^{n} \lambda(k) &= \sum_{i=1}^{b} \left(1-\alpha_i\right) + \sum_{j=1}^{c} \beta_j \left(1 - \frac{1}{\beta_j}\right) \\ &= \sum_{i=1}^{b} \left(1-\alpha_i\right) + \sum_{j=1}^{c} \left(\beta_j - 1\right) \end{aligned} \tag{6}$$

According to the number of points of two time series in each group of alignment, the length of two time series satisfies the following relationship:

$$m = a + \sum_{i=1}^{b} \alpha_i + c \tag{7}$$

$$n = a + b + \sum_{j=1}^{c} \beta_j \tag{8}$$

Subtracting Eq. (7) from Eq. (8) gives

$$n - m = \sum_{i=1}^{b} \left(1-\alpha_i\right) + \sum_{j=1}^{c} \left(\beta_j - 1\right) \tag{9}$$

Combining Eq. (6) and Eq. (9), we can prove that

$$\sum_{k=1}^{n} \lambda(k) = n - m \quad \square$$

For two time series of equal length, the sum of the time distortion coefficients is zero, i.e. $\sum_{k=1}^{n} \lambda(k) = 0$. In other words, the number of compression alignments and stretch alignments must be equal.

In order to make an overall evaluation of the time distortion between two time series, we define the overall time distortion between two time series as the sum of the absolute value of the time distortion coefficient at each point, i.e.,

$$\rho(\boldsymbol{P}, \boldsymbol{Q}) = \sum_{k=1}^{n} |\lambda(k)| \tag{10}$$

Obviously, the overall time distortion between two time series is non-negative, i.e. $\rho(\boldsymbol{P}, \boldsymbol{Q}) \geq 0$. Even a zero overall time distortion between two time series does ensure that they are identical, i.e. $\rho(\boldsymbol{P}, \boldsymbol{Q}) = 0 \nRightarrow \boldsymbol{P} = \boldsymbol{Q}$. Similar to the DTW distance, it also does not satisfy the triangle inequality. Since the concept of compression and stretch between two time series is reciprocal, it satisfies the symmetry, i.e. $\rho(\boldsymbol{P}, \boldsymbol{Q}) = \rho(\boldsymbol{Q}, \boldsymbol{P})$.

4. Fusional similarity measure for nearest neighbor based time series classification

Time series classification plays an important role in various applications. Recent works have been demonstrated that the nearest neighbor classifier using DTW distance has comparable and superior performance to many complex classification algorithms [16,28, 29]. As mentioned above, the performance of nearest neighbor classifier is highly dependent on the quality of the distance measure. However, the DTW distance fails to evaluate the differences between time series in both temporal and amplitude domains. Therefore, a fusional similarity measure that combines the DTW distance and the overall time distortion is proposed to improve the performance of the nearest neighbor classifier.

Suppose $\boldsymbol{T} = \{\boldsymbol{T}_1, \boldsymbol{T}_2, \cdots, \boldsymbol{T}_n,\}$ are n training time series, $\boldsymbol{x}$ is the test time series. $d_{DTW}\left(\boldsymbol{T}_k, \boldsymbol{x}\right)$ and $\rho\left(\boldsymbol{T}_k, \boldsymbol{x}\right), k = 1, 2, \cdots, n$ are the DTW distances and overall time distortions between the test time series and n training time series, respectively. Since the DTW distance and the overall time distortion are two measures with different dimensions, they cannot be used directly to calculate a weighted sum. The Max-min normalization is a commonly used dimensionless approach that can map data of different dimensions to $[0, 1]$ through a linear transformation:

$$\tilde{y}_i = \frac{y_i - y_{min}}{y_{max} - y_{min}} \tag{11}$$

where y_i and $\tilde{y}_i$ are the original and normalized data. y_{max} and y_{min} are the maximum and minimum values of $\boldsymbol{y}$.

Suppose $\tilde{d}_{DTW}\left(\boldsymbol{T}_k, \boldsymbol{x}\right)$ and $\tilde{\rho}\left(\boldsymbol{T}_k, \boldsymbol{x}\right)$ are the normalized DTW distance and overall time distortion. The fusional similarity measure between $\boldsymbol{T}_k$ and $\boldsymbol{x}$ is defined as the weighted sum of the normalized DTW distance and overall time distortion:

$$d_F\left(\boldsymbol{T}_k, \boldsymbol{x}\right) = \theta \cdot \tilde{d}_{DTW}\left(\boldsymbol{T}_k, \boldsymbol{x}\right) + (1 - \theta) \cdot \tilde{\rho}\left(\boldsymbol{T}_k, \boldsymbol{x}\right) \tag{12}$$

where $\theta \in [0, 1]$ is a positive weight. A larger θ indicates that the similarity measure depends mainly on the amplitude difference between time series. On the contrary, a smaller θ means that the similarity measure depends mainly on the time distortion between time series. Thus, the choice of weight can be determined based on the characteristics of the data. For the datasets with large amplitude differences between different classes, a larger weight is chosen.

5. Experiment

In this section, a number of experiments are conducted to evaluate the proposed similarity measure. First, the time distortion between time series is analyzed using the time distortion coefficient. Then, the 1-NN classifier is utilized to verify the performance of the proposed fusional similarity measure. In this experiment, DTW and four different DTW variants including cDTW, wDTW, dDTW and shapeDTW are selected to evaluate the proposed similarity measure. In addition to the DTW distance, other four metrics, such as average warping distance, warping path length ratio, TAM distance, singularity score are utilized for comparison.

5.1. Experimental datasets and settings

The UCR Time Series Classification Archive database is a widely used open source database for time series analysis [30]. The database contains numerous types of data, such as images, motions, ECG, sensors, devices, spectro and simulated data. The number of classes varies from 2 to 60. These datasets have standard partitions of training and test data, we experiment with these given partitions and report classification accuracies on the test data.

In order to make these experiments reproducible, we provide details of the experimental settings:

Table 1
The comparison of averaged accuracy rate (%) of the 1-NN classifier for DTW and four variants using different similarity measures.

Method	DTW distance	Fusional distance	Num (Avg_rate)	Avg_wdis	W_L	TAM	SS	Overall time distortion
DTW	74.95	78.61	56 (5.55)	74.21	69.78	69.78	59.25	69.82
cDTW	76.31	78.69	44 (4.59)	75.83	51.28	51.28	52.7	51.87
dDTW (Raw)	75.86	79.51	66 (4.69)	74.9	68.89	68.89	64.74	69.06
dDTW (Feature)	72.10	78.60	71 (7.78)					
wDTW (Raw)	75.68	77.41	37 (3.97)	75.48	42.47	42.47	39.93	42.57
wDTW (Feature)	75.58	76.66	30 (3.07)					
shapeDTW-PAA (Raw)	75.66	78.11	49 (4.23)	75.12	45.7	45.7	39.18	45.83
shapeDTW-PAA (Feature)	77.86	79.06	49 (2.08)					
shapeDTW-HOG1D (Raw)	74.25	75.47	30 (3.47)	73.93	42.66	42.66	46.82	42.66
shapeDTW-HOG1D (Feature)	76.98	79.01	59 (2.92)					

- The squared Euclidean distance is utilized as the local cost measure for both DTW and variants.
- For cDTW, the warping constraint parameter w follows the report in [24].
- For wDTW, the parameters of the modified logistic weight function are set as follow: $w_{max} = 1$, $g = 0.5$ [25].
- For shapeDTW, the length of subsequence is set to 30. Two shape descriptors, Piecewise Aggregate Approximation (PAA) and HOG1D, are used.
- As the time series in UCR have a uniform sampling frequency, the down-sampling rate d in the singularity score is set to 1.
- For feature-based DTW variants, such as dDTW and shapeDTW, the alignment is generated based on the feature sequence, the DTW distances are calculated based on the raw sequence and the feature sequence, respectively.
- For the fusional similarity distance, the weight θ varies from 0 to 1 with a fixed step of 0.1.
- The warping path length ratio is calculated as follows:

$$W = \frac{L - \max(m, n)}{\max(m, n)}$$

- The average warping distance is calculated as the ratio of the DTW distance using the raw sequence to the length of the warping path.

All algorithms are programmed in the Matlab language using Matlab 2017b. The experiments are implemented on a computer with Intel Core i5-5700, 3.10 GHz CPU, 4 GB RAM, and Windows 10 64-bit operating system.

5.2. Experimental results and analysis

5.2.1. Analysis of time distortion between time series

We first analyze the time distortion between the time series shown in Fig. 1. The time distortions between time series $\boldsymbol{A}$ and three test time series $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$ are 10, 14 and 30, respectively, which is consistent with the differences of length between them. Compared to the TAM distance, the proposed time distortion measure can provide an accurate and meaningful evaluation of the similarity between time series in the time domain.

Fig. 3 shows the alignment between two time series in the first class of 50Words calculated by DTW and DDTW. Due to the local amplitude difference, the alignment calculated by DTW suffers from pathological problem near the extreme points. Fig. 4 illustrates the corresponding time distortion coefficient, providing a detailed time distortion information at each point. The time distortion calculated by DTW and DDTW are 222 and 196, respectively. From Fig. 4, we can see that DDTW can significantly reduce the pathological alignment compared to DTW.

5.2.2. Comparison of 1-NN classification

The accuracy of the 1-NN classifier directly reflects the performance of the similarity measure and can be used as an evaluation criterion. For DTW and four different DTW variants, Table 1 summarizes the comparison of the averaged accuracy rate of 1-NN classifiers using different similarity measures. The “Raw” and “Feature” in the first column of the Table 1 indicate that the DTW distances are calculated using the raw sequence and the feature sequence, respectively. “shapeDTW-PAA” and “shapeDTW-HOG1D” denote the shapeDTW using the PAA and HOG1D shape descriptors respectively. The fourth column of Table 1 shows the number of datasets where the fusional distance outperforms the DTW distance and the average improvement rate. Among all the combinations of DTW algorithms with distance measures, the dDTW with fusional similarity distance using the raw sequence achieves the highest accuracy. Except for wDTW and shapeDTW-HOG1D (Raw), the proposed fusional similarity distance significantly improves the accuracy compared to the DTW distance on more than half of the datasets. For both DTW and dDTW, the fusional similarity distance yields superior performance, on more than two-thirds of the datasets, with an improvement of more than 5%. For all DTW-based methods, the average warping distance achieves results comparable to the DTW distance. Compared to the DTW distance, the four metrics that calculated based on the warping path including warping path length, TAM distance, singularity score and overall time distortion, always achieve lower classification accuracy. As mentioned above, the DTW-based methods inevitably suffer from the pathological alignment problem. Therefore, these metrics cannot effectively capture the differences between time series. From

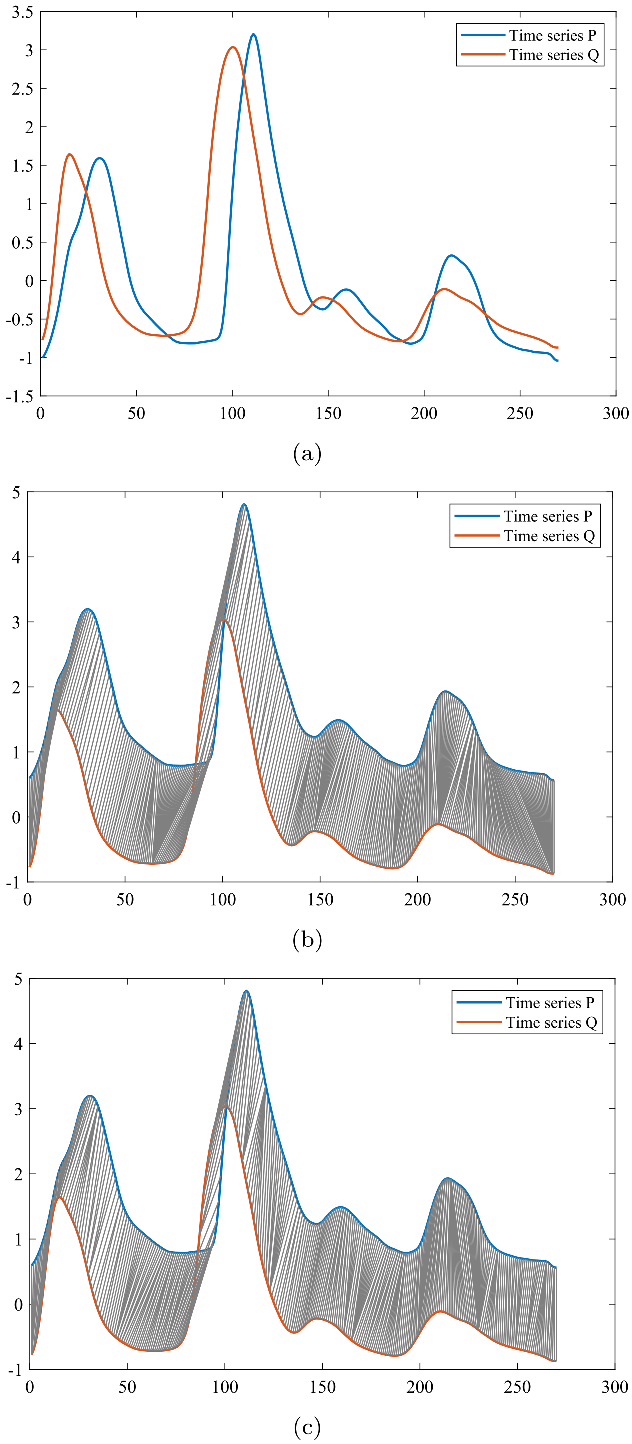


Fig. 3. Time series alignments calculated by DTW and DDTW. (a) Two time series in the first class of 50Words. (b) Time series alignment calculated by DTW. (c) Time series alignment calculated by DDTW.

Table 1, we find that the TAM distance always obtains the same accuracy as the warping path length ratio for all methods. Given any two equal-length time series, the TAM distance is linearly relative to the warping path length ratio. In other words, the TAM distance and the warping path length ratio are equivalent. Each method is then discussed in detail.

DTW: Fig. 5 shows the comparison of the accuracy rate between the fusional similarity distance and the DTW distance for the DTW method. The red dots above the diagonal line indicate that the fusional similarity distance achieves a higher accuracy rate

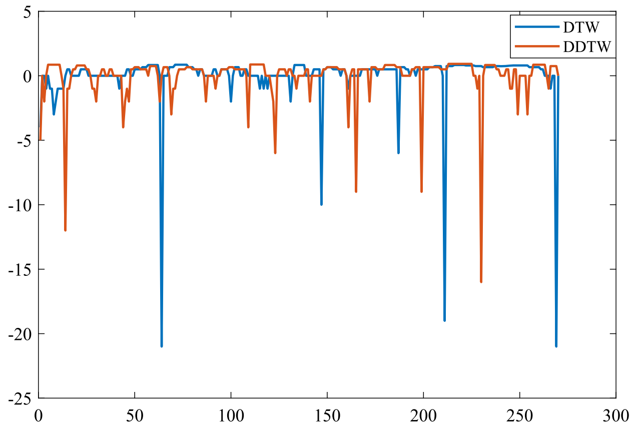


Fig. 4. Time distortion coefficient calculated by DTW and DDTW.

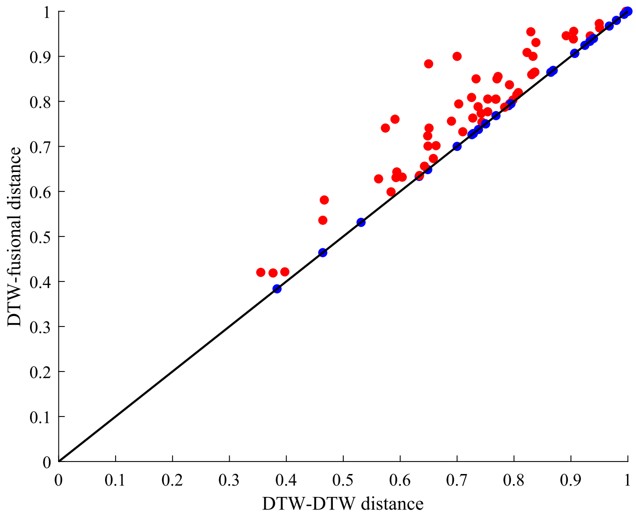


Fig. 5. Graphical comparison of accuracy rates between fusional similarity distance and the DTW distance for the DTW method.

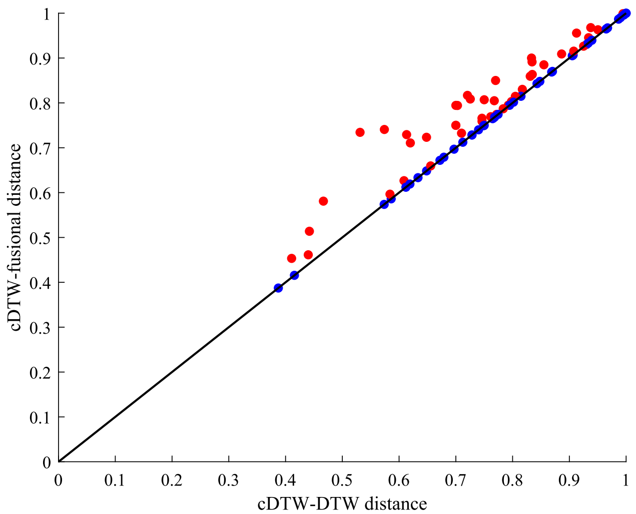
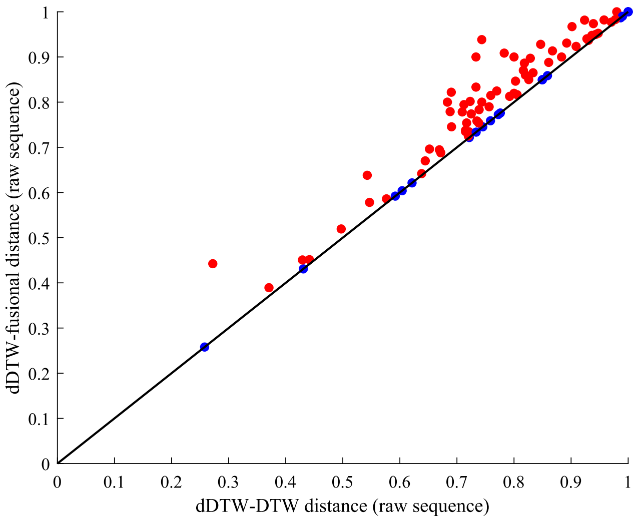
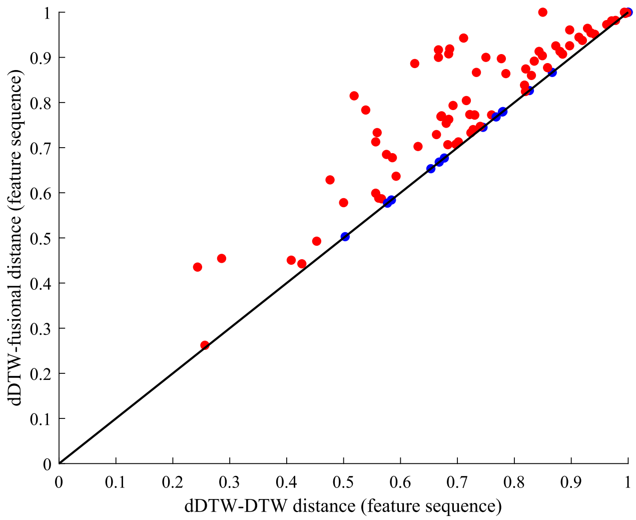


Fig. 6. Graphical comparison of accuracy rates between fusional similarity distance and the DTW distance for the cDTW method.



(a)



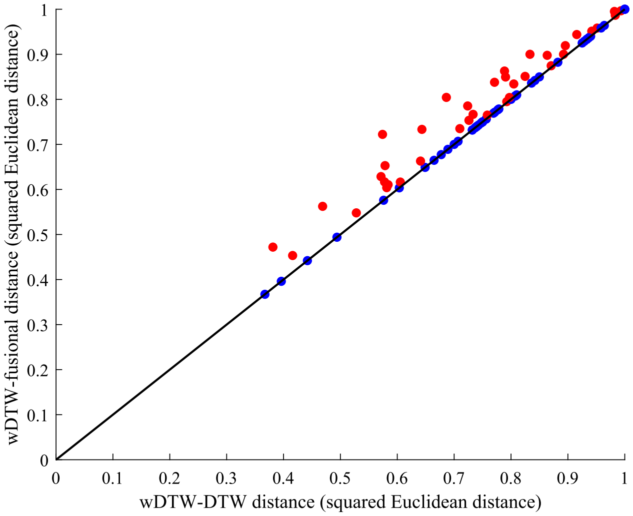
(b)

Fig. 7. Graphical comparison of accuracy rates between fusional similarity distance and the DTW distance for the dDTW method. (a) DTW distance is calculated using the raw sequence. (b) DTW distance is calculated using the derivative sequence.

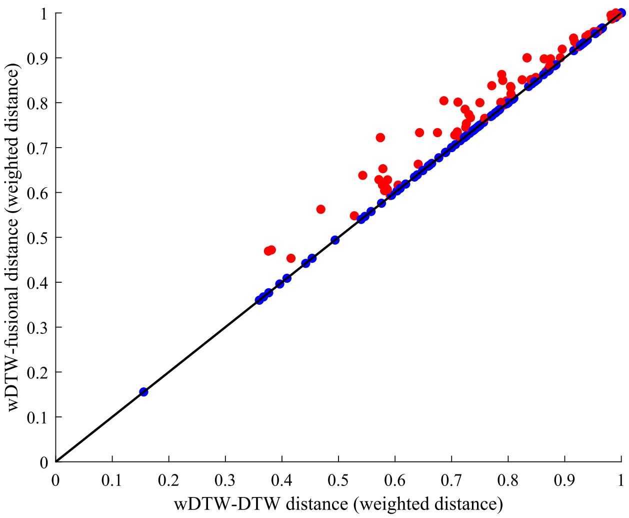
compared to the DTW distance. As shown in Fig. 5, the fusional similarity distance outperforms the DTW distance on 56 out of 85 datasets, with an average improvement of 3.66%. It is important to note that 4 datasets have an accuracy of 1 when using the DTW distance and there is no more room for improvement. For these 56 datasets, the classification accuracy rate improves by an average of 5.55%. There are 25 datasets with an accuracy improvement of more than 5%, 7 datasets with an accuracy improvement of more than 10%. The highest improvement is achieved in the ShapeletSim dataset, with an improvement of 23.33%.

cDTW: Fig. 6 shows the comparison between the fusional similarity distance and the DTW distance for the cDTW method. The fusional similarity distance outperforms the DTW distance for 44 out of 85 datasets, with an average improvement of 2.38%. For these 44 datasets, the classification accuracy rate improves by an average of 4.59%. There are 17 datasets with an accuracy improvement of more than 5%, 4 datasets with an accuracy improvement of more than 10%. The highest improvement is obtained in the ShapeletSim dataset, with an improvement of 20.31%. cDTW reduces the pathological alignment by constraining the global warping path. As a result, cDTW improves the accuracy for both the fusional similarity distance and the DTW distance compared to DTW. However, the global warping path constrain may have the risk of missing the correct warping. Since the pathological alignment always produces a meaningless time distortion measurement, the fusional similarity distance can only improve the accuracy for a small fraction of the datasets.

dDTW: dDTW is a feature based DTW method, the DTW distances are calculated based on the raw sequence and the feature sequence, respectively. Fig. 7 illustrates the comparison of dDTW between the fusional similarity distance and the DTW distance.



(a)

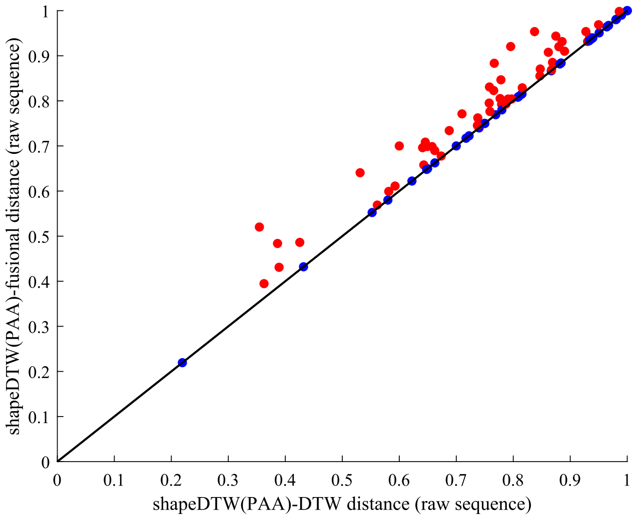


(b)

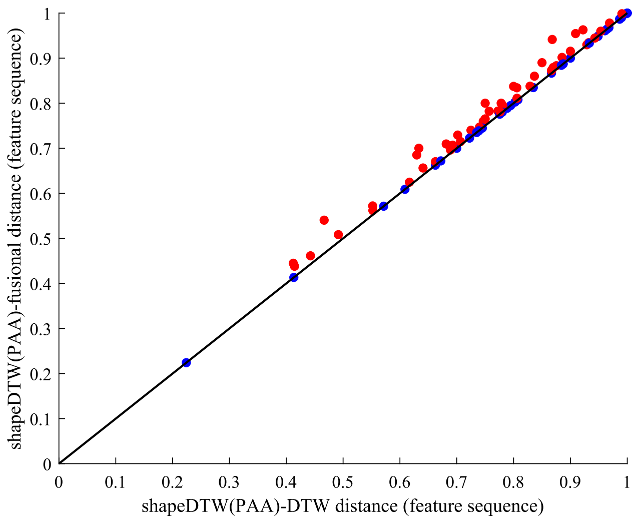
Fig. 8. Graphical comparisons of accuracy rates between fusional similarity distance and the DTW distance for the wDTW method. (a) DTW distance is calculated by taking the sum of the squared Euclidean distances of all aligned elements. (b) DTW distance is calculated by taking the sum of the weighted distances of all aligned elements.

For these two cases, the fusional similarity distance outperforms the DTW distance in 66 and 71 out of 85 datasets, with an average improvement of 3.64% and 6.5%, respectively. There are 25 and 36 datasets with an accuracy improvement of more than 5%, 8 and 19 datasets with an accuracy improvement of more than 10%. When using the first derivative sequence, the fusional similarity distance significantly improves the accuracy compared to the DTW distance. As the first derivative is sensitive to noise and outline, both the fusional similarity distance and the DTW distance calculated using the raw sequence have a higher accuracy than those using the derivative sequence for the majority of the datasets.

wDTW: For wDTW, the sequence alignment is generated using the weighted distance as the local cost measure. The DTW distances are calculated by taking the sum of the squared Euclidean distances and the weighted distances of all aligned elements, respectively. Fig. 8 illustrates the comparisons of wDTW between the fusional similarity distance and the DTW distance for two local cost measures. The fusional similarity distance outperforms the DTW distance in 37 and 30 out of 85 datasets, with an average improvement of 1.73% and 1.08%, respectively. There are 12 and 7 datasets with an accuracy improvement of more than 5%, 2 and 1 datasets with an accuracy improvement of more than 10%. wDTW adjusts the weight of the local cost function according to the phase difference between two points to obtain a better alignment. The choice of parameters for the weighting function has a direct effect on the performance of wDTW. Since the fixed parameters are used in this experiment, wDTW achieves a lower accuracy compared to other methods.



(a)

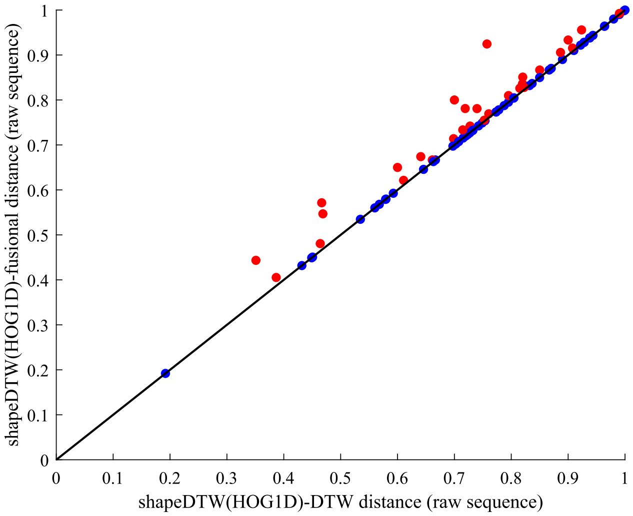


(b)

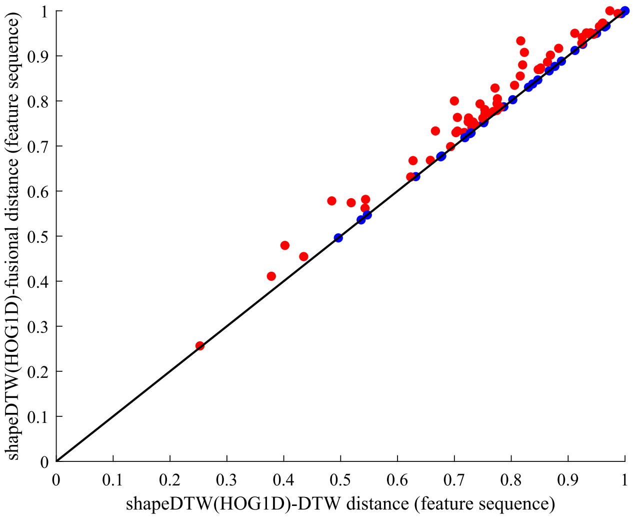
Fig. 9. Graphical comparison of accuracy rates between fusional similarity distance and the DTW distance for the shapeDTW method with PAA. (a) DTW distance is calculated using the raw sequence. (b) DTW distance is calculated using the feature sequence.

shapeDTW: For shapeDTW, two shape descriptors, PAA and HOG1D, are utilized. Similar to dDTW, the DTW distances are calculated based on the raw sequence and the feature sequence, respectively. Fig. 9 and Fig. 10 show the comparison of shapeDTW with two shape descriptors between the fusional similarity distance and the DTW distance, respectively. For shapeDTW with PAA, the fusional similarity distance outperforms the DTW distance in 49 and 49 out of 85 datasets, with an average improvement of 2.12% and 1.2% when using the raw sequence and the feature sequence, respectively. For shapeDTW with HOG1D, the fusional similarity distance outperforms the DTW distance in 30 and 59 out of 85 datasets, with an average improvement of 2.12% and 1.2%, respectively. Table 1 shows that the DTW distances calculated based on the feature sequence always achieve higher accuracy than those calculated based on the raw sequence for both PAA and HOG1D. The computational complexity of shapeDTW is $O(l \cdot m \cdot n)$, where l is the length of the subsequence. Compared to other DTW variants, shapeDTW obtains a small improvement at a very high computational cost.

Effect of weight: The weight of the proposed fusional similarity measure represents the proportion of amplitude difference and time distortion in the similarity measure. The optimal weight should be chosen based on the characteristics of the datasets. Next, the effect of the weight in the fusion similarity measure on the 1-NN classifier is analyzed. For the DTW algorithm, the average weights on the 56 datasets where the fusional similarity measure outperforms the DTW distance is 0.39. Fig. 11 shows the accuracy rates of the 1-NN classifier using the fusional similarity measure with different weights for some datasets. For the ArrowHead dataset, the accuracy first increases and then decreases as the weight increase. The optimal weight is 0.5. This means that the ArrowHead dataset



(a)



(b)

Fig. 10. Graphical comparison of accuracy rates between fusional similarity distance and the DTW distance for the shapeDTW method with HOG1D. (a) DTW distance is calculated using the raw sequence. (b) DTW distance is calculated using the feature sequence.

has differences in both amplitude and temporal domain between time series of inter-class. In this situation, a moderate weight can achieve best results. For the BeetleFly and SonyAIBORobotSurface datasets, the optimal weight is close to 1. A closer look to at these datasets shows that the differences between time series of inter-class are mainly caused by time distortion. For the Meat dataset, the differences between time series of inter-class are mainly caused by amplitude variation. Therefore, the optimal weight is 0. The optimal weight can be determined experimentally. First, the training data is used to calculate the DTW distance and the overall time distortion between the inter-class and intra-class time series, respectively. Then, the classification accuracy is calculated using different weights. The weight with the highest classification accuracy is the optimal weight.

6. Conclusions

In this study, the time distortion coefficient based on the DTW warping path is proposed to quantify the difference of time series in the time domain, which can compensate for the shortcoming that the DTW distance can only evaluate the difference in the amplitude. It can provide a detailed time distortion information, including the type and degree of time distortion at each point. For the 1-NN-based time series classification, a fusional similarity measure that combines the DTW distance and the overall time distortion is proposed, which can evaluate the similarity in both amplitude and time domains. The experimental results demonstrate that the proposed fusional similarity distance can significantly improve the classification accuracy of the 1-NN classifier with only

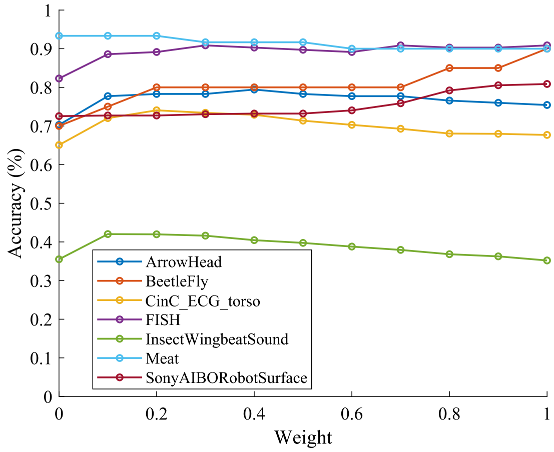


Fig. 11. The accuracy of the 1-NN classifier using fusional similarity measure with different weights.

a small amount of additional computational cost compared to the DTW distance and other metrics. The choice of weights directly affects the performance of the fusional similarity distance in the 1-NN classifier. An appropriate weight should be chosen based on the characteristics of the datasets. Our future work focuses on developing a learning scheme to determine the optimal weight.

CRediT authorship contribution statement

Yutao Liu: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Yong-An Zhang:** Validation, Writing – review & editing. **Ming Zeng:** Writing – review & editing. **Jie Zhao:** Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This research was partially supported by the Joint Construction Project for Medical Science and Technology of Henan Province, China under Grant (LHGJ20210341).

References

[1] V. Maus, G. Câmara, R. Cartaxo, A. Sanchez, F.M. Ramos, G.R. De Queiroz, A time-weighted dynamic time warping method for land-use and land-cover mapping, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (8) (2016) 3729–3739.
[2] M. Belgiu, O. Csillik, Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis, Remote Sens. Environ. 204 (2018) 509–523.
[3] E.E. Berman, D.K. Bolton, N.C. Coops, Z.K. Mityok, G.B. Stenhouse, R.D. Moore, Daily estimates of Landsat fractional snow cover driven by modis and dynamic time-warping, Remote Sens. Environ. 216 (2018) 635–646.
[4] P. D'Urso, L. De Giovanni, R. Massari, Trimmed fuzzy clustering of financial time series based on dynamic time warping, Ann. Oper. Res. 299 (1) (2021) 1379–1395.
[5] D. Clifford, G. Stone, I. Montoliu, S. Rezzi, F.-P. Martin, P. Guy, S. Bruce, S. Kochhar, Alignment using variable penalty dynamic time warping, Anal. Chem. 81 (3) (2009) 1000–1007.
[6] P. Tormene, T. Giorgino, S. Quaglini, M. Stefanelli, Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation, Artif. Intell. Med. 45 (1) (2009) 11–34.
[7] H. Skutkova, M. Vitek, P. Babula, R. Kizek, I. Provaznik, Classification of genomic signals using dynamic time warping, BMC Bioinform. 14 (10) (2013) 1–7.
[8] Z. Xing, J. Pei, E. Keogh, A brief survey on sequence classification, ACM SIGKDD Explor. Newsl. 12 (1) (2010) 40–48.
[9] A. Abanda, U. Mori, J.A. Lozano, A review on distance based time series classification, Data Min. Knowl. Discov. 33 (2) (2019) 378–412.
[10] H. Kaya, Ş. Gündüz-Öğüdücü, A distance based time series classification framework, Inf. Syst. 51 (2015) 27–42.
[11] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. Acoust. Speech Signal Process. 26 (1) (1978) 43–49.
[12] D.S. Hirschberg, Algorithms for the longest common subsequence problem, J. ACM 24 (4) (1977) 664–675.
[13] L. Chen, M.T. Özsu, V. Oria, Robust and fast similarity search for moving object trajectories, in: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, 2005, pp. 491–502.

[14] L. Chen, R. Ng, On the marriage of lp-norms and edit distance, in: Proceedings of the Thirtieth International Conference on Very Large Data Bases-Volume 30, 2004, pp. 792–803.
[15] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, E. Keogh, Experimental comparison of representation methods and distance measures for time series data, Data Min. Knowl. Discov. 26 (2013) 275–309.
[16] A. Bagnall, J. Lines, An experimental evaluation of nearest neighbour time series classification, arXiv preprint, arXiv:1406.4757.
[17] D. Folgado, M. Barandas, R. Matias, R. Martins, M. Carvalho, H. Gamboa, Time alignment measurement for time series, Pattern Recognit. 81 (2018) 268–279.
[18] H. Li, J. Liu, Z. Yang, R.W. Liu, K. Wu, Y. Wan, Adaptively constrained dynamic time warping for time series classification and clustering, Inf. Sci. 534 (2020) 97–116.
[19] Z. Zhang, R. Tavenard, A. Bailly, X. Tang, P. Tang, T. Corpetti, Dynamic time warping under limited warping path length, Inf. Sci. 393 (2017) 91–107.
[20] H. Kirchhoff, A. Lerch, Evaluation of features for audio-to-audio alignment, J. New Music Res. 40 (1) (2011) 27–41.
[21] J. Zhao, L. Itti, Shapedtw: shape dynamic time warping, Pattern Recognit. 74 (2018) 171–184.
[22] Y. Jiang, Y. Qi, W.K. Wang, B. Bent, R. Avram, J. Olgin, J. Dunn, Eventdtw: an improved dynamic time warping algorithm for aligning biomedical signals of nonuniform sampling frequencies, Sensors 20 (9) (2020) 2700.
[23] E.J. Keogh, M.J. Pazzani, Derivative dynamic time warping, in: Proceedings of the 2001 SIAM International Conference on Data Mining, SIAM, 2001, pp. 1–11.
[24] H.A. Dau, D.F. Silva, F. Petitjean, G. Forestier, A. Bagnall, A. Mueen, E. Keogh, Optimizing dynamic time warping's window width for time series data mining applications, Data Min. Knowl. Discov. 32 (2018) 1074–1120.
[25] Y. Jeong, M.K. Jeong, O.A. Omitaomu, Weighted dynamic time warping for time series classification, Pattern Recognit. 44 (9) (2011) 2231–2240.
[26] Z. Zhang, P. Tang, R. Duan, Dynamic time warping under pointwise shape context, Inf. Sci. 315 (2015) 88–101.
[27] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 509–522.
[28] A. Bagnall, J. Lines, A. Bostrom, J. Large, E. Keogh, The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances, Data Min. Knowl. Discov. 31 (3) (2017) 606–660.
[29] R.J. Kate, Using dynamic time warping distances as features for improved time series classification, Data Min. Knowl. Discov. 30 (2) (2016) 283–312.
[30] Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, G. Batista, The UCR time series classification archive, www.cs.ucr.edu/~eamonn/time_series_data/, July 2015.