2014 IEEE International Conference on Data Mining Workshop

Dynamic time warping based on cubic spline interpolation for time series data mining

Hailin Li
College of Business Administration
Huaqiao University
QuanZhou, China 362021
Email: hailin@mail.dlut.edu.cn

Xiaoji Wan, Ye Liang and Shile Gao
College of Business Administration
Huaqiao University
QuanZhou, China 362021
Email:WanXiaoji@hqu.edu.cn

***Abstract*—Dynamic time warping (DTW) and derivative dynamic time warping (DDTW) are two robust distance measures for time series, which allows similar shapes to match even if they are out of phase in the time axis. In this paper, we propose a novel dynamic time warping based on cubic spline interpolation (SIDTW) to improve the performance. The derivative of every point of time series is calculated by cubic spline interpolation and is used to replace the estimated derivatives in DDTW. After interpolation we use derivative-based sequences to represent the original time series, which is better to describe the trend of the original time series and more reasonable to warp. Meanwhile, we empirically point out that the quality of similarity measure for the three warping methods is nothing to do with the amount of warping. We experimentally perform the proposed method and compare with the existing ones, which demonstrates that in most cases our approach not only can produce much less singularities and obtain the best warping path with shorter length but also is an alternative version of DTW when time series datasets are not suitable for DTW to be measured.**

***Keywords*-dynamic time warping, time series data mining, cubic spline interpolation, similarity measure**

I. Introduction

Time series is a type of common data existing in our daily life. Valuable information and knowledge are hiding in large time series database, including bioinformation, engineering, financial market, medicine, etc. Recently more and more attention has been paid on time series mining, many models and algorithms have been applied to this field, such as clustering [2], [23], classification [9], [21], motifs finding [13] and indexing [8].

In most cases, we should compare one time series to another and obtain similarity before time series clustering, classification and other time series data mining tasks. Euclidean distance measure is a common way to calculate the similarity between two time series, which is widely used in many cases [11], [6]. However, Euclidean distance measure is very brittle for time series [20], [5], [8] measure. Moreover, time series approximately have the same overall component shapes which often do not line up in the time axis. Therefore, the work is looking for a way to align the time axis so that these shapes can line up in time axis.

The available solution to address the above issue is dynamic time warping (DTW) [18], [4], [7], [2], which is often used to measure the similarity of signals or time series with equal length [16]. It is also used to measure the similarity between two time series by warping the time axis of one sequence (or two sequences simultaneously). In the past decade, DTW was successfully used in different fields, such as DNA expression data [1], time series patterns discovery [4] and articulated motion recognition [15]. Although it is more effective than Euclidean distance and also applied widely to various domains, it has some defects. The most obvious one is the abnormal results it produced, which is found by Keogh and Pazzani [7]. They stated that DTW trying to explain the variability in the Y-axis by warping time axis (X-axis) will cause non-intuitive alignments where a single point on one time series maps into a large subsection of another time series. To overcome the drawbacks, they modified it and provided derivative dynamic time warping (DDTW). It is also used in many fields. For examples, Muscillo et al. used it to classify accelerometer data [14], Zhou and Wong used it for time scaling searching [24].

DDTW [7] produced less “singularities” (less warping) to measure the time series than DTW. However, it still produces many “singularities” and needs a large amount of warping in some cases. In this paper, for the corresponding cases we propose a novel method called spline-interpolation-based dynamic time warping (SIDTW) and denote it as a template for an improvement of DTW. We also suggest that anyone much more efficient and effective than cubic spline interpolation can be used. The authors [15] also combines cubic spline interpolation with DTW to match the normalized sequences, but it focuses on the values of the re-sampled sequences rather than the derivatives of interpolated points. SIDTW uses the cubic spline interpolation to calculate the much more accurate derivative of every point of time series. Moreover, the derivatives are used to replace those of DDTW which are computed by the roughly estimated method.

The main motivation of our approach is obtaining more accurate derivatives to reflect the trend of time series and improve the effectiveness of the similarity measure for the

978-1-4799-4274-9/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDMW.2014.21

time series mining. It decreases unnecessary warping and obtain intrinsic similarity of two time series to improve the performance of the techniques used in time series mining. Furthermore, It is an alternative version of DTW when the similarity measure in some time series datasets is not suitable to be computed by DTW. To our best knowledge, we first and empirically point out that the quality of similarity measure between two time series using the dynamic warping methods is nothing to do with the amount of warping, which is one of the most important contributions in our work. The experimental results demonstrate that our approach not only can produce much less number of warping than DTW and DDTW but also is efficiently and effectively used to measure the similarity when a dataset is not suitable to be measured by DTW.

The rest of the paper is organized as follows. In section 2 we give the background and related work. Section 3 simply introduces cubic spline interpolation. Section 4 contains a detailed description of the new algorithm (SIDTW) and section 5 illustrates some experimental evaluation. Finally we give conclusions and discuss the future work in section 6.

II. Background and related work

Dynamic time warping (DTW) is not only a method of similarity measure between two time series with equal (or unequal) length but also a warping approach to align the same shapes better. The proposed method proposed by [4] predefines a pattern to be a template for similarity measure for time series.

In DTW, to align two time series, $Q = \{q_1, q_2, \cdots, q_n\}$ and $C = \{c_1, c_2, \cdots, c_m\}$, a n by m matrix D is constructed, whose element's value is the square distance between q_i and c_j, that is, $d(i,j) = (q_i - c_j)^2$.

One warping path p is a contiguous set of the elements of D, which means that there is a mapping between Q and C. The lth element of p is defined as $p_l = (i,j)_l$, so we have $p = \{p_1, p_2, \cdots, p_l, \cdots, p_k\}$, where $k \in [\max(m,n), m+n-1)$, especially $k \in [m, 2m-1)$, when $m = n$.

The warping path must be typically subject to several constraints, such as boundary conditions, continuity and monotonicity [7]. There are many warping paths P satisfying those constraints, but the best one p is required, $p \in P$. It can be picked out according to the minimum warping cost given by

$$DTW(Q,C) = \min_{p \in P} \left\{ \frac{1}{k}\sqrt{\sum_{l=1}^{k} p_l} \right\}. \quad (1)$$

Since there are various warping paths with different length, the k is often used to average the sum cost so that different warping path has qualification to compare with another one. Generally, the best path can be found by using dynamic programming which defines the cumulative distance $r(i,j)$ as the distance $d(i,j)$ adding the minimum of the cumulative distance of the three adjacent elements, i.e.

$$r(i,j) = d(i,j) + \min \begin{cases} r(i, j-1) \\ r(i-1, j-1) \\ r(i-1, j) \end{cases}. \quad (2)$$

However, in DDTW, the steps are the same as DTW besides the square of difference between q_i and c_j. The estimated derivative $d'(i,j)$ in DDTW is used to replace the $d(i,j)$ in DTW, i.e.

$$d'(i,j) = (d_i(q) - d_j(c))^2, \quad (3)$$

where

$$d_l(x) = \frac{(x_l - x_{l-1}) + (x_{l+1} - x_{l-1})/2}{2}. \quad (4)$$

Paper [7] indicates that DTW tries to explain the variability in the Y-axis by warping the X-axis, which causes no good alignment where a single point on time series maps to a large subsection of another one. They call it " singularity". Moreover, if two points q_i and c_j are respectively on the rising trend in one subsequence and on the falling trend in the other subsequence, it is unreasonable for DTW to map one to the other directly because of the obvious feature. To deal with this problem, DDTW modifies it by considering higher level feature of shapes for better alignment. Some authors [17] proposed to learn arbitrary constraints on the warping path. In the same year, Regression time warping (RTW) [12] was proposed to address the challenges of scaling, shifting, complexity and robustness.

Besides the improvement of accuracy of DTW, some methods are proposed to improve the efficiency of DTW used to time series mining. [24] presented a segment-wise time warping (STW) for time scaling search with high efficiency. FastMap method [22] was for an approximate indexing using DTW to filter the non-qualifying time series. Another exact indexing approach [8] based on PAA representation of time series for DTW similarity measure was further proposed. Fast similarity search under the time warping distance (FTW) was proposed by [19] for efficiently pruning a number of search candidates. Moreover, it is significantly faster than the best existing method, approximately up to 222 times.

Although DDTW is an improved version of DTW and it can produce less number of warping when time series are warped, the estimated method to calculate the derivatives of the points in time series seems to be out of work in some cases and affects the last result of time series mining. Therefore, a fast technique to calculate much more accurate derivatives of the points are required to improve the quality of time warping measure. Perhaps there were many such techniques, but in our work we select the cubic spline interpolation and provide a template. Anyone more effective and efficient for the derivative computation than cubic spline interpolation can be suggested to improve the performance of the warping measure.

III. Cubic spline interpolation

Interpolations, such as Lagrange interpolation, Newton interpolation and Hermite interpolation, can create one or more functions to fit the discrete data points. These functions are used to draw the corresponding lines or curves, and also can obtain derivatives of different data points by executing derivative function. In our method cubic spline interpolation [10] is used to fit the data points so as to obtain the derivatives of the points on a smoother curve with less error between fit functions and actual functions.

We assume that a table of values has been given as $(x_i, y_i), i = 0, 1, 2, \cdots, n$, $x_0 < x_1 < x_{n-1} < x_n$, and the actual function is $y_i = f(x_i)$. Cubic spline interpolation can create a piecewise continuous curve passing through each value in the former table. At the same time, we can construct a cubic interpolation function $s_i(x)$ for each subinterval $[x_i, x_{i+1}]$ of the X-axis. The overall cubic spline interpolation function is

$$S(x) = \begin{cases} s_0(x) & x \in [x_0, x_1] \\ s_1(x) & x \in [x_1, x_2] \\ \vdots & \vdots \\ s_{n-1}(x) & x \in [x_{n-1}, x_n] \end{cases} . \tag{5}$$

According to cubic spline interpolation function, the cubic polynomial $s_i(x)$ should be

$$s_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i. \tag{6}$$

There are four unknown coefficients in (6). Therefore, there are $4n$ coefficients need to be defined for the overall function. We can get $2n$ constraints according to

$$\begin{cases} s_i(x_i) = y_i \\ s_i(x_{i+1}) = y_{i+1} \end{cases} \tag{7}$$

and another $2(n-1)$ ones by

$$\begin{cases} s'_{i-1}(x_i) = s'_i(x_i) \\ s''_{i-1}(x_i) = s''_i(x_i) \end{cases} . \tag{8}$$

We totally have $4n - 2$ conditions. We need to find two more conditions to completely fix the overall spline interpolation. The two conditions are always given by users. For example, the overall curve has two known values of derivatives of the lower-bound x_0 and the upper-bound x_n, i.e. $s'(x_0) = m_0$, $s'(x_n) = m_n$. Thus we can solve a set of n equations and define all the unknown coefficients by the pursuit method. People interesting in the detailed information about the cubic spline interpolation algorithm are suggested to refer to paper [10].

IV. Cubic-spline-interpolation-based dynamic time warping

[7] employed an index to indicate the amount of warping implied by the corresponding algorithm, i.e.,

$$W = \frac{K - m}{m}, \qquad W \in [0, 1]. \tag{9}$$

where K is the number of warping and m is the length of time series.

From the above formula (9), it is easy to know that W is in direct proportion to K because of the fixed length m. It also means that the small W is, the less warping it will produce. In order to avoid the unnecessary warping happened in DTW, we direct complete the algorithm of DTW before the cubic spline interpolation intervenes. After the construction of cumulative distance matrix r, the algorithm should backward find out the best path beginning at $r(n, m)$. The original way to carry out this step is picking out the minimum one from the three adjacent elements in r as the next member of the best warping path, which neglects some case where the three adjacent elements are equal. So we change the original strategy in DTW (or DDTW) to another, i.e.,

$$p_l = \begin{cases} (i^*, j^*)_l & if\ r(i^*, j^*) \leq \min(r(i^*, j), r(i, j^*)) \\ (i^*, j)_l & else\ if\ r(i^*, j) < \min(r(i^*, j^*), r(i, j^*)) \\ (i, j^*)_l & else\ r(i, j^*) < \min(r(i^*, j^*), r(i^*, j)) \end{cases} , \tag{10}$$

where $1 < i \leq n$ and $1 < j \leq m$, n and m are respectively the length of the two time series.

The formula (10) tells us that we should select the element $(i-1, j-1)$ in cumulative distance matrix r when the three elements are equal, that is $r(i-1, j-1) = r(i-1, j) = r(i, j-1)$, which is able to decrease the number of warping when some sections with equal values are respectively in the two series. We denote this completed method as CDTW. As shown in Figure 1, the amount of warping in CDTW is much less than in DTW.

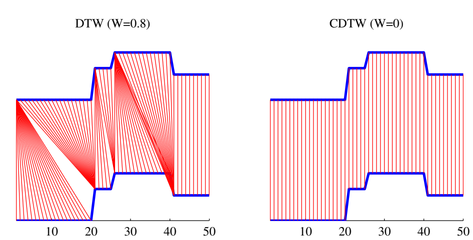


Fig. 1. The number of warping is 40 in DTW when the two identical time series of length 50 is used, but no warping happens in CDTW.

More detailed information about the case will be given at the experimental section (see subsection 5.1), which further testifies that such case exists in the real dataset and the competed method CDTW can avoid the unnecessary warping.

[7] indicates that DDTW makes less singularities than DTW when warping time series. In this paper, we objectively point out that the amount of warping in time series is dependent on dataset. The similarity of some time series datasets is suitable to be measured by DTW because the values of Y axis benefit the nature alignment. However, DDTW does well in dealing with the time series dataset with the remarkable shapes. DDTW uses a higher level to do it according to the derivatives of the points despite of their identical value of the Y-axis. To calculate the derivative of the point of time series, DDTW uses the formula (4) to " estimate" the value $d_l(x)$ in time series x for simplicity and so-called objectivity.

Estimating the derivative not only creates the inaccurate value of $d_l(x)$ but also affects the final distance measure. When one point is at the peak of the curves, its derivative should be equal to 0. However, the estimated result of DDTW is not equal to 0. It is often greater or smaller. The peak is a notable shape of time series and is very important for the similarity computation. If we deal with it roughly by the inaccurate estimation, these notable features of the data points in time series mining will be lost. Thereby, it seems to be unreasonable to estimate the derivative value roughly. Instead, we had to find a better way to obtain the accurate value and let them be close to the objective value. In our work we select the cubic spline interpolation to calculate the derivatives. We also state that any method reflecting the shape of time series more accurate can be suggested to compute the derivatives.

Cubic spline interpolation (CSI) can transform the data points of time series into the ones on a smooth curve. The derivatives of the points at the peak of piecewise lines are close to 0, which means that the derivatives produced by cubic spline interpolation can reflect the points at the peaks. As shown in Figure 2, (a) shows the curve produced by the interpolation to approximate the original time series, which illustrates that some peaks of the original time series are at the ramp and the others are close to the maximum (or the minimum)of the local section. It also shows that the curve can approximate the time series well. (b) shows the derivative-based series respectively produced by DDTW and CSI to approximate the original time series. These two series basically have the same shapes, but some points of the derivative series of CSI are close to 0, which reflects the corresponding shapes. For instance, the $51th$ point of the original time series is at the valley, whose shape can be reflected by the derivative of 0. Its derivative in DDTW is -0.5947, which can't reflect the derivative change (from the negative to the positive). However, the one in CSI is 0.0980 (be close to 0), which reflects the change. Therefore, the derivatives produced by CSI are more accurate to reflect the original time series.

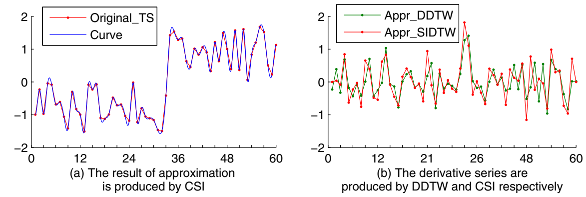


Fig. 2. (a) shows the curve produced by CSI to approximate the time series. (b) shows the two derivative series produced by DDTW and CSI.

After the above analysis, we present the cubic-spline-interpolation-based dynamic time warping (SIDTW) as follows:

Setp 1. Input two time series data, Q and C respectively, of length n and m. Note that x_i means the time of X-axis and y_j is the corresponding point value of the time series.

Step 2. Let boundary derivatives of Q and C be $M_{q1} = q_2 - q_1$, $M_{q1} = q_n - q_{n-1}$, $M_{c1} = c_2 - c_1$ and $M_{cm} = q_m - q_{m-1}$ respectively.

Step 3. Bring these parameters of the two sequences respectively into the cubic spline interpolation algorithm. Finally we can get two functions, $QS(y)$ and $CS(y)$.

Step 4. Compute the derivative functions, $QS^{'}(y)$ and $CS^{'}(y)$. We can get the derivatives of every points of two time series, $QS^{'}(y_i)$ and $CS^{'}(y_j)$.

Step 5. Replace the values of $d_i(q)$ and $d_j(c)$ according to the formula (4) with $QS^{'}(y_i)$ and $CS^{'}(y_j)$,

$$\begin{cases} d_i(q) = QS^{'}(y_i) \\ d_j(c) = CS^{'}(y_j) \end{cases} \tag{11}$$

Step 6. Calculate the distance matrix according to the formula (3) and use the dynamic programming to find the minimum warping cost. Meanwhile, construct cumulative distance matrix r and execute CDTW to obtain the number of warping K and the best path p.

Through the above algorithm we can get more accurate derivatives of the points. We let them replace the original points. for this way, besides the correct warping, SIDTW at least has three advantages:

(1) The points with positive derivative in one sequence will align to the points which also have the positive derivative in the other sequence. It means that the points on the different time series will align to each other on the same trend.

(2) In most cases, the length of the warping path will be shorter than the DDTW and DTW. Moreover, the number of singularities produced by the SIDTW also will be least in all.

(3) SIDTW is an alternative version of DTW. If some database are not suitable for DTW to be measured, then they can be measured by SIDTW well, and vice-versa.

It is obvious that SIDTW has the first advantage because more accurate derivatives to reflect the shapes of time series are produced by SIDTW than by DDTW. However, some conditions constrain the second advantage. Since most of the techniques in time series mining are dependent on dataset, the amount of warping in SIDTW is not least in some cases in which the similarity measure of the dataset is not suitable to be computed by SIDTW but may be suitable to calculated by DDTW or DTW. But the shape based comparisons exist in most of time series datasets, much more accurate derivatives of the points in time series reflect the shapes and benefit the shape based comparisons. Therefore, the SIDTW is suitable to warp the time series with less singularities and measure the similarity of most of time series datasets, which is testified in section 5.2. We also objectively suggest that SIDTW is an alternative if DTW could not measure the similarity well for some time series dataset, which can be confirmed by the classification experiment in the section 5.3.

In this work, we point out that in the field of time series mining the performance of the warping methods including DTW, DDTW and SIDTW is nothing to do with the amount of warping. It means that one method produced less warping may be rough to measure the similarity of time series and affects the effectiveness of data mining. Similarly, one method produces much more warping than the others may be robust to measure the similarity. Therefore, we can choose the reasonable method in the three to measure the similarity according to the data and the practice.

For the time complexity, SIDTW is actually close to the two existing algorithms, DTW and DDTW, They all cost the complexity of $O(nm)$. The additional time consumption of DDTW and SIDTW is used to compute the derivatives of the elements of the time series. For each series of length n, the time complexity of SIDTW used to compute the points' derivatives is $O(n)$. We use the pursuit method to solve the equations in process of the cubic spline interpolation and its time complexity is linear to the number of points. Therefore, though additional time consumption should be taken to interpolate, time complexity of SIDTW is close to the DDTW and DTW.

V. Experimental Evaluation

In this section there are four subsections to demonstrate the quality of our approach (SIDTW). In the first subsection, we further testify that CDTW used to construct SIDTW can avoid the unnecessary warping. The next subsection tells us that the amount of warping produced by SIDTW in most of time series datasets is much less than that produced by DTW and DDTW, which demonstrates that SIDTW can measure the similarity without much singularities. In the third subsection we testify whether the amount of warping affects the result of classification in time series data mining. An illustration of time consumption is given in the last subsection.

A. Unnecessary warping

Before the intervention of cubic spline interpolation, the original DTW produces much unnecessary warping. The proposed method CDTW can filter those warping. Now there exists two questions to be answered. (a) are there the unnecessary warping existing in DTW and DDTW. (b) whether CDTW can discover those unnecessary warping and correct them. In this subsection we use the real stock dataset to testify that whether CDTW can avoid the unnecessary warping.

We intentionally pick out 3 subsequences (whose length are respectively 51, 81 and 41) from the stock time series. There are some sections with equal values existing in these three subsequences, which is a common phenomenon in time series and especially be widely existing in financial time series dataset. We make a copy of each subsequence Q and obtain the copy Q'. Meanwhile, we use the three methods to compare subsequence Q with their copy Q'. As shown in Figure 3, we find that the unnecessary warping really exists in DTW and DDTW because none should have existed in the two identical series. We also find that CDTW recognizes the case and avoids the unnecessary warping.

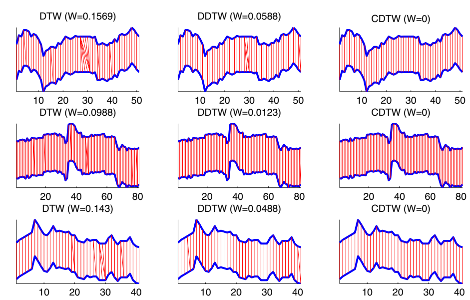


Fig. 3. No warping is produced by the proposed CDTW. The two existing methods (DTW and DDTW) produce the unnecessary warping

As shown in Figure 3, in DTW and DDTW, W corresponding to the amount of warping is not equal to 0. However, in CDTW, W is always equal to 0. The reason is that when the three adjacent elements (r(i-1,j-1), r(i,j-1), r(i-1,j)) are equal, CDTW regards the element $r(i-1, j-1)$ as a member of the best path, which avoids the unnecessary warping. Therefore, CDTW can recognize such cases and does not produce the unnecessary warping.

B. Warping quality comparison

In this subsection we test the quality of the warping methods (DTW, DDTW and SIDTW) so as to know whether in most cases our approach can produce the less singularities than the others. Meanwhile, we also test that whether the three methods mistakenly find the warping where none exists, which is the same as the spurious warping test in [7].

As the previous work had done, the pairs of time series used to measure similarity by the three methods should be highly correlated but no identical. To test the quality of the warping methods, we utilize 12 UCR time series datasets deriving from the website [9]. Those time series datasets are different in some attributes including the length of time series L, the number of classes N_c, the training size S_{tr} and the testing size S_{te}. The detailed information about the datasets are in Table I.

Table I
THE UCR TIME SERIES DATASETS ARE USED TO TEST THE QUALITY OF THE WARPING METHODS.

ID	Name	N_c	S_{tr}	S_{te}	L
1	Adiac	37	390	391	176
2	Beef	5	30	30	470
3	CBF	3	30	900	128
4	Coffee	2	28	28	286
5	ECG	2	100	100	96
6	FISH	7	175	175	463
7	FaceAll	14	560	1690	131
8	Gun_Point	2	50	150	150
9	Lighting7	7	70	73	319
10	OliveOil	4	30	30	570
11	Synthetic_Control	6	300	300	60
12	Two_Patterns	4	1000	4000	128

We also use the formula (9) to denote the amount of the warping in the different methods. In the same UCR time series dataset we merge the two sets, the training and the testing. We compare each time series to every other time series. Furthermore, a pair of the two time series in the same class is correlated but no identical. We compare them by order and average the sum of W for all of the pairs of time series. Take Synthetic_Control dataset for an example, the whole dataset is obtained by merging the 300 training ones and the 300 testing ones. We classify them into 6 groups according to their class labels and obtain 100 time series in each group. We compare every pair of adjacent time series in the same group so that there are 99 pairs of time series used to be compared. It means that there are $99 \times 6 = 594$ pairs of time series used to be compared in the Synthetic_Control datasets. We compute W for every pair of time series and average them at last. The results of W for the 12 UCR datasets listed above are presented in Table II.

In Table II AW_X denotes the averaged value of W for the method X. The numerical values in bold represent the minimum values of the averaged AW in the same time series dataset. Furthermore, the ones in italic type denote the sub-minimum values of the averaged W in the same dataset. The table indicates that, in contrast DTW and DDTW, the amount of warping produced by SIDTW is smaller for most of time series dataset. Only two datasets have minimum AW for DTW and another two have minimum AW for DDTW. However, in the four datasets the values of sub-minimum AW are all in SIDTW. Moreover, the values of sub-minimum AW is close to that in DDTW. Therefore, we

Table II
THE RESULTS OF WARPING QUALITY COMPARISON ARE GIVEN FOR THE 12 URC DATASETS USING THE THREE METHODS.

ID.name	AW_DTW	AW_DDTW	AW_SIDTW
1. Adiac	**0.1004**	0.3443	*0.3262*
2. Beef	0.6185	0.4382	**0.4022**
3. CBF	0.4359	0.2779	**0.2629**
4. Coffee	0.3626	0.2721	**0.2587**
5. ECG	0.4882	0.3188	**0.3093**
6. FISH	**0.2370**	0.5674	*0.5625*
7. FaceAll	0.2467	**0.2330**	*0.2370*
8. Gun_Point	0.5616	0.4524	**0.4334**
9. Lighting7	0.5417	0.3378	**0.3234**
10. OliveOil	0.1046	0.0881	**0.0879**
11. Synthetic_Control	0.3214	**0.2437**	*0.2458*
12. Two_Patterns	0.3675	0.3309	**0.3155**

empirically state that the proposed SIDTW can produce least singularities for the most time series datasets and is much less inclined to make a warping where none exists.

For more detailed information about the amount of warping, we give the distribution graph of W for every pair of time series as shown in Figure 4. It is obvious that most of the values of W for SIDTW is smaller than the other two methods. Meanwhile, it also shows that the amount of warping is dependent on the data. For some time series dataset, such as Adiac and Fish, the minimum value of W is inclined to DTW or DDTW. Fortunately, the minimum values of W for most of time series datasets are inclined to SIDTW.

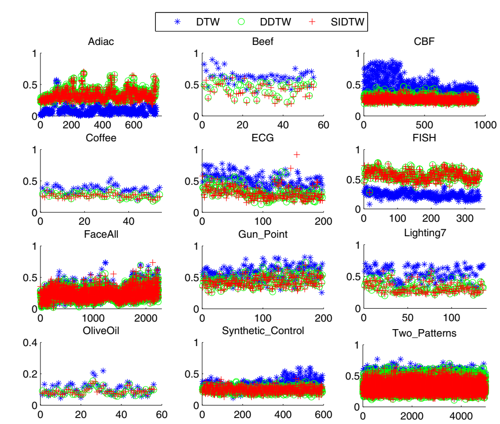


Fig. 4. The distribution graph of W for each pair of time series using the three methods in different UCR dataset.

C. Classification

To testify the performance of our approach SIDTW applied to the field of time series data mining, we let the three methods (DTW, DDTW, SIDTW) on the 12 UCR time series datasets [9] make a classification experiment. We use the 1-nearest-neighbor classification to class the time series. For each datasets, every time series from the testing set is used to find the most similar one in training set. If the label of the retrieve one is the same to that of the testing one, we consider it as a correct classification. Otherwise, it is regard as an incorrect classification. Finally, The ratio of the number of incorrect classifications over the number of time series in the testing set is regarded as the last result of the classification. The classification results produced by the three methods for the different datasets are presented in Table III.

Table III
THE CLASSIFICATION RESULTS PRODUCED BY THE FOUR METHODS FOR DIFFERENT TIME SERIES DATASET.

$Methods$ / $ID.name$	DTW	DDTW	SIDTW
1. Adiac	**0.396**	0.399	0.468
2. Beef	0.500	0.300	**0.300**
3. CBF	**0.003**	0.377	0.501
4. Coffee	0.179	0.071	**0.071**
5. ECG	0.230	0.190	**0.170**
6. FISH	0.167	0.109	**0.103**
7. FaceAll	0.192	0.147	**0.133**
8. Gun_Point	0.093	0.02	**0.007**
9. Lighting7	**0.274**	0.411	0.452
10. OliveOil	**0.133**	0.02	0.167
11. Synthetic_Control	**0.007**	0.293	0.533
12. Two_Patterns	**0.000**	0.002	0.007

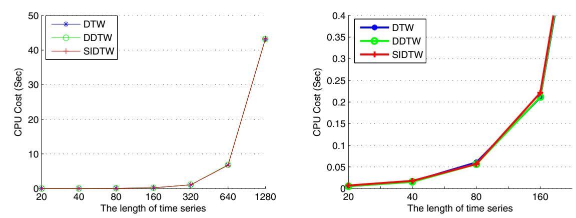
From Table III, we know that DTW and SIDTW have good performance of classification. It also indicates that the dataset is not suitable to be mined by DDTW and SIDTW when the similarity of some dataset is suitable to be measured by DTW, such as Adiac, CBF, Coffee, Lighting7, OliveOil, Synthetic_Control and Two_Patterns. Similarly, the datasets being suitable to be mined by SIDTW have the worst results if DTW is used. Therefore, we can state that SIDTW is an alternative version of DTW and they are dependent on the special dataset.

Combining Table II with Table III, we find that the performance of the three methods in the field of classification is nothing to do with the amount of warping. The method with high performance in classification may has the much unnecessary warping (singularities), such as DTW in Coffee dataset, who has the high performance of classification but produces much singularities. Inversely, less singularities produced by a method are not corresponding to less misclassifications occurred. So we empirically state that the performance of the methods (DTW, DDTW and SIDTW) is nothing to do with the amount of warping. Since they are dependent on dataset, the application of the methods should agree with the demand of the practice.

D. Efficiency comparison

We have analyzed that the time complexity of SIDTW is the same to that of DTW and DDTW, $O(mn)$, where m and n are the length of two time series respectively. In contrast to DTW, additional time needs to be consumed for the derivative computation in DDTW and SIDTW. However, the addition time consumption is linear to the length of time series, which means that we can neglect the additional time series.

In order to testify whether the above analysis is right or not, we arbitrarily divide the stock time series into some subsequences with unequal length. Let the three methods to measure the similarity between two different subsequences who have the same length. Moreover, for each length we execute the methods 50 times for different pairs of time series and average the sum of their time consumption. The final results are shown in Figure 5. Since much time is cost for the time series of length 1280, the changeable lines of the time consumption for the three methods overlap and seem to be one only in subfigure 5(a).



(a) (b)

Fig. 5. (a) shows the results of the time consumption for the three methods. (b) shows the zooming version of the results in (a).

From the subfigure 5(b), we can find that the time consumption of SIDTW is slightly larger than that of DTW and DDTW. The reason is that the additional time used to compute the derivatives in SIDTW is slightly more than that in DDTW. However, since the additional time of the two methods is linear to the length of the time series, the difference between them is very small and is often neglected. Therefore, the three methods have the same efficiency.

VI. CONCLUSIONS

DTW and DDTW are two conventional approaches to measure the similarity of time series. In this paper, we propose an alternative version of dynamic time warping which is based on cubic spline interpolation (SIDTW) for similarity measure of time series. The experiments on time series datasets demonstrate that SIDTW can decrease the number of the singularities and align the points between the two time series more suitably. The reason is that it is impossible that the accurate derivative of the points calculated by the SIDTW makes the point on the rising trend of one time series map to the one on the falling trend of another time series. In other words, the more accurate slope of time series determines the correct warping. Additionally, when there exists many adjacent points with equal values in time series, the proposed CDTW can avoid the unnecessary warping.

We also empirically state that SIDTW is suitable for most of the time series databases to measure the similarity. It can produce less singularities and avoid the unnecessary warping. In particular, the experimental results of classification demonstrate that the performance of the warping methods used to mine time series is nothing to do with the amount of warping. The choice of the three methods used to warp and mine time series is dependent on the dataset and the practice. Furthermore, SIDTW does well in the time series datasets which are not suitable for DTW to be measured. Therefore, SIDTW is an alternative version of DTW and can produce much less singularities.

In addition to the accuracy of the warping methods (DTW, DDTW and SIDTW), since they need the time complexity of $O(nm)$ to measure the time series, their efficiency also should be considered. The future work is that we may extract the features and find the most important elements which can represent the time series well. For this way, the number of the most important elements is lower than the original time series, which could improve the effectiveness of SIDTW.

Acknowledgments

This work has been supported by the National Natural Science Foundation of China (61300139) and Promotion Program for Young and Middle-aged Teacher in Science and Technology Research of Huaqiao University (ZQN-PY220).

References

[1] J. Aach, G. M. Church, Aligning gene expression time series with time warping algorithms, Bioinformatics, 17(6)(2001)495-508.

[2] Z. Bank*ó*, J. Abonyi, Correlation based dynamic time warping of multivariate time series, Expert Systems with Applications, 39(17)(2011)12814-12823.

[3] B. Boecking, S. K. Chalup, D. Seese, Support vector clustering of time series data with alignment kernels, Pattern Recognition Letters, 45(2014)129-135.

[4] D. J. Berndt, J. Clifford, Using dynamic time warping to find patterns in time series, in: Proceedings of AAAI-94 Workshop on Knowledge Discovery in Databases,1994, pp.359-371.

[5] S. Chu, E. Keogh, D. Hart, et al., Iterative deepening dynamic time warping for time series, in:Proceedings of the 2nd SIAM International Conference on Data Mining, 2002, pp.195-212.

[6] C. H. Joh, T. Arentze, H. Timmermans, Pattern recognition in complex activity travel patterns, comparison of Euclidean distance, signal-processing theoretical, and multidimensional sequence alignment methods, Transportation Research Record, 2003, pp.16-22.

[7] E. Keogh, M. J. Pazzani, Derivative dynamic time warping, in:Proceedings of the First SIAM International Conference on Data Mining, 2001, pp.1-11.

[8] E. Keogh, Exact indexing of dynamic time warping, in: Processings of the 28th VLDB Conference, 2005, pp.358-380.

[9] E. Keogh, X. P. Xi, et al., UCR time series classification and clustering page. http://www.cs.ucr.edu/ẽamonn/time_series_data/dataset.zip, 2013.

[10] D. Kincaid, W. Cheney, Numerical analysis: mathematics of scientific computing, Thomson Learning, 2002.

[11] F. Korn, H. V. Jagadish, C. Faloutsos, Efficiently supporting ad hoc queries in large datasets of time sequences, in:Proceedings of SIGMOD '97, 1997, pp.289-300.

[12] H. Lei, V. Govindaraju, Regression time warping for similarity measure of sequence, in:Proceedings of the 4th International Conference on Computer and Information Technology, 2004, pp.826-830.

[13] F. Marinez-*Á*lvarez, A. Troncoso, J.C. Riquelme, et al., Discovery of motifs to forecast outlier occurrence in time series, Pattern Recognition Letters, 32(12)(2011)1652-1665.

[14] R.Muscillo, S. Conforto, S. Schmid, et al., Classification of motor activities through derivative dynamic time warping applied on accelerometer aata, in:Proceedings 29th Annual International Conference of Engineering in Medicine and Biomedicine, 2007, pp.4930-4933.

[15] S. Nayak, S. Sarkar, B. Loeding, Distribution-based dimensionality reduction applied to articulated motion recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(5)(2011)795-810.

[16] L. Rabiner, B. H. Juang, Fundamentals of speech recognition, Prentice Hall, 1993.

[17] C. A. Rantanamahatana, E. Keogh, Making time-series classification more accurate using learned constraints, in:Proceedings of the 4th SIAM International Conference on Data Mining, 2004, pp.11-22.

[18] H. Sakoe, Chiba, S. Dynamic programming algorithm optimization for spoken word recognition, IEEE Transaction of Acoustics Speech and Signal Processing, 2(1978)43-49.

[19] Y. Sakurai, M. Yoshikawa, C. Faloutsos, FTW: fast similarity search uner the time warping distance, in:Proceedings of the 24th ACM SIGACT-SIGMODSIGART Symposium on Principles of Database Systems, 2005, pp. 326-337.

[20] M. Vlachos, D. Gunopoulos, G.Kollios, Discovering similar multidimensional trajectories, in:Proceedings of the 18th International Conference on Data Engineering, 2002, pp.673-684.

[21] B. V. Kini, C. C. Sekhar, Large margin mixture of AR models for time series classification, Applied Soft Computing, 13(1)(2013)361-371.

[22] B. Yi, H. V. Jagadish, C. Faloutsos, Efficient retrival of similar time sequences under time warping, in:Proceedings of the 14th IEEE International Conference on Data Engineering, 1998, pp.13-22.

[23] X. H. Zhang, J. Q. Liu, D. Yu, et al., A novel clustering method on time series data, Expert Systems with Applications, 38(9)(2011)11891-11900.

[24] M. Zhou, M. H. Wong, A segment-wise time warping method for time scaling searching, Information Sciences, 173(2005)227-254.