Accurate and Fast Dynamic Time Warping

Hailin Li and Libin Yang

College of Business Administration, Huaqiao University, Quanzhou 362021, China
hailin@mail.dlut.edu.cn, ylib1982@163.com

Abstract. Dynamic time warping (DTW) is widely used to measure similarity between two time series by finding an optimal warping path. However, its quadratic time and space complexity are not suitable for large time series datasets. To overcome the issues, we propose a modified version of dynamic time warping, which not only retains the accuracy of DTW but also finds the optimal warping path faster. In the proposed method, a threshold value used to narrow the warping path scope can be preset automatically, thereby resulting in a new method without any parameters. The optimal warping path is found by a backward strategy with reduced scope which is opposite to the forward strategy of DTW. The experimental results demonstrate that besides the same accuracy, the proposed dynamic time warping is faster than DTW, which shows that our method is an improved version of the original one.

Keywords: Dynamic time warping, Time series, Similarity measure, Data mining, Computational complexity.

1 Introduction

In the field of time series data mining, distance function or similarity measure is often used to describe relationships between two time series, such as Euclidean distance [1, 2], edit distance [3], K-L distance [4] and dynamic time warping (DTW) [5–9]. Euclidean distance is one of the most popular functions used to fast measure time series with same length. Its time and space complexity are linear to the length of time series. It often combines with time series representations [10–12] to improve the results. However, Euclidean distance is sensitive to abnormal points and helpless to measure time series with different lengths.

Dynamic time warping (DTW) [6] is another popular method to measure similarity between two time series. It is not only robust to the abnormal points but also suitable for time series with different length. DTW finds an optimal alignment between two time series to measure the similarity. In this way, the points with same shape in each time series can be mapped. However, since DTW must search the best warping path in a cumulated matrix, it has to cost the quadratic time and space complexity, which causes to be not suitable for the long time series. Thereby, a new DTW with the same accuracy to the original one and less time consumption is required.

In this paper, we propose a modified version of DTW, which we call accurate and fast dynamic time warping (AF_DTW). The framework of AF_DTW consists

H. Motoda et al. (Eds.): ADMA 2013, Part I, LNAI 8346, pp. 133–144, 2013.

of three parts. The first is a backward strategy used to find the optimal warping path, which is opposite to the forward strategy of DTW. The second is the process of reduced scope and the third is a choice method of threshold value which can be preset to narrow the warping path scope. Especially, the backward strategy and the way of reduced scope run concurrently with one another. The contribution of the backward strategy and threshold intervening is to discard a remaining scope when the current cost of the warping path is negative. In this way, AF_DTW finds the optimal warping path in a reduced scope and its accuracy is the same to DTW.

The remainder of the paper is organized as follows. In section 2, we give the background and related work. The framework of the improved version of DTW is presented in section 3. Some experiments are performed on serval time series datasets in section 4. In the last section the conclusions and future work are discussed.

2 Background and Related Work

Given two time series $Q = \{q_1, q_2, \cdots, q_m\}$ and $C = \{c_1, c_2, \cdots, c_n\}$, where m and n respectively represent the length of Q and C, Euclidean distance [1,13,14] is used to measure the similarity between the two time series when $m = n$. Each pair of points with same time-stamp in the two time series is used to calculate the distance. When the shapes of the two time series are considered, DTW is a good choice to measure time series. Especially, when the lengths of the two time series are different, i.e., $m \neq n$, DTW is often used instead of Euclidean distance.

DTW [2,5,6] minimizes the distance between two time series by constructing an optimal warping path P which often makes the points with same shapes map to each other. A warping path can be denoted as $P = \{p_1, p_2, \cdots, p_K\}$ and p_i means the mapping information about the time of two points (q_i and c_j) respectively deriving from the two time series, i.e., $p_k = [i, j]$, where $K \in [max(m, n), m + n - 1]$ represents the length of the path, $i \in [1, m]$ and $j \in [1, n]$. $d(p_k)$ denotes the distance between two points q_i and c_j, i.e., $d(p_k) = d(i, j) = (q_i - c_j)^2$. At the same time, the warping path much satisfy at least three constraints, such as boundary conditions, continuity and monotonicity [5].

There are many such path existed in the mapping sets, but we only need an optimal one with the minimal warping cost, i.e.

$$DTW(Q, C) = \min_{P} \sum_{k=1}^{K} d(p_k). \tag{1}$$

Generally, the best warping 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.

$$R(i,j) = d(i,j) + \min \begin{cases} R(i,j-1) \\ R(i-1,j-1) \\ R(i-1,j) \end{cases}, \tag{2}$$

where $R(0,0) = 0$, $R(i,0) = R(0,j) = \infty$. We often call the cumulative distance R as a cost matrix.

For example, if we have two time series Q and C, i.e., $Q = [-0.6, 4.6, -6.4, -22.3, 0.9, 0.4, 0.3, 9, 12]$ and $C = [0.1, 5.3, -5.8, -22.1, 1.5, 0.7, 0.4, 9.2]$ The best warping path can be calculated by starting from ($i = 1$ and $j = 1$) to ($i = 9$ and $j = 8$). There are 10 path elements constructing the best warping path, that is, $P = \{p_1, p_2, \cdots, p_{10}\}$. Thereby, the minimum distance of DTW to measure the similarity between two time series Q and C is $DTW(Q, C) = R(9, 8) = 9.67$.

DTW is widely applied to the field of time series data mining [15], speech recognition [16] and other disciplines [2], such as medicine, meteorology, gesture recognition and finance. However, the quadratic time and space complexity ($O(nm)$) of DTW constrain its performance. So far, there are some methods [18,19] used to speed up the calculation of DTW. Two of the most commonly used methods is the Sakoe-Chuba Band [17] and the Itakura Parallelogram [16] that limit the number of cells in the cost matrix R and reduce the path scope for searching the suboptimal warping path. However, The performance of DTW using the two methods depends on a constant factor. Moreover, it often cannot retrieve the best warping path which is often out of the path scope. FTW [20] is often applied to similarity search in time series and can faster retrieve the results than the original DTW. However, it uses lower bounding functions to reduce time cost when they are applied to similarity search. So FTW is preferable to similarity search and indexing rather than clustering and classification. Similarly, FastDTW [21] is proven to be a method with linear time and space complexity, but its accuracy depends on a factor which is hard to decide and the returned distance value is a result approximating to the minimum one. At the same time, the larger the factor is, the more the calculation of time and space is cost. Thereby, a technique with less time and space consumption and without any factors is required, which is the main topic and contribution of our paper.

3 Accurate and Fast Dynamic Time Warping

Accurate and fast dynamic time warping (AF_DTW) is an improved version of DTW, which includes three parts. The first one is a backward strategy to construct the main idea of AF_DTW. The second one is the introduction of a way to reduce scope in the cost matrix. The last one is a choice method of a threshold value used to reduce the scope.

3.1 Backward Strategy

DTW starting from $(i, j) = (1, 1)$ to $(i, j) = (m, n)$ is a forward strategy based algorithm. The cost matrix R is constructed cumulatively from $(1, 1)$ to (m, n).

Each element in the cost matrix R is larger than the three bottom left adjacent elements. In contrast to DTW, AF_DTW uses a backward strategy to construct the algorithm. It means that AF_DTW starts from (m, n) to $(1, 1)$ and each element in cost matrix R is less than the three top right adjacent ones.

In DTW a current element $R(i, j)$ in cost matrix is the distance $d(i, j)$ adding the minimum of the cumulative distance of the three bottom left adjacent elements. Contrary, in AF_DTW the element $R'(i, j)$ is the maximum of three top right adjacent elements subtracting the distance $d(i, j)$, i.e.,

$$R'(i,j) = \max \begin{cases} R'(i, j+1) \\ R'(i+1, j+1) \\ R'(i+1, j) \end{cases} - d(i,j), \tag{3}$$

where $i = m, m-1, \cdots, 1$, $j = n, n-1, \cdots, 1$, $R'(m+1, n+1) = 0$, $R'(i, n+1) = R'(m+1, j) = -\infty$.

In this way, a warping path $P' = \{p'_1, p'_2, \cdots, p'_K\}$, which is a contiguous set of distance matrix elements that defines a mapping between Q and C, can be constructed as DTW does. The warping path also must be subject to three constraints including boundary conditions, continuity and monotonicity.

The dynamic programming is also used to find the best warping path with a maximum warping cost, i.e.,

$$BS\_DTW(Q, C) = \underset{P'}{max} \sum_{k=1}^{K} d(p'_k). \tag{4}$$

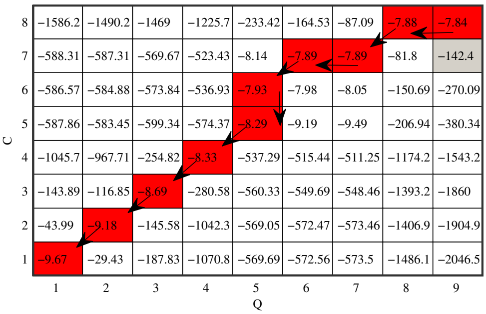

Fig. 1. The cost matrix is constructed by the backward strategy and the best warping path can be found

Using the above example, our new method starts at $(m, n) = (9, 8)$ and ends at $(1, 1)$. The warping cost can be calculated by the backward strategy as shown in Fig. 1. The red cells denote the elements of the best warping path. It is easy to discover that the element is $R'(1, 1) = -9.67$, which is the minimum value of the best warping path in red cells. It also means that $BS\_DTW(Q, C) = R'(1, 1) = -9.67$. The arrows show the direction of our strategy.

Comparing the result of DTW with that of BS_DTW, we found that their absolute values are equal, i.e., $|BS\_DTW(Q,C)| = DTW(Q,C) = 9.67$. BS_DTW has the same accuracy of DTW. Moreover, the time and space complexity of BS_DTW are the same to that of DTW and equal to $O(mn)$. So the next subsection is proposing a method to reduce the time and space consumption when running AF_DTW based on BS_DTW.

3.2 Reduced Scope

In this section, a method is proposed to obtained the reduced scope in which the best warping path exists. It is inspired by the the Sakoe-Chuba Band [17] and the Itakura Parallelogram [16] which limit the cells in the cost matrix.

In BS_DTW, all the elements in the cost matrix R' are negative as shown in Fig. 1. The reason is that BS_DTW depends on the initial value of $R'(m+1, n+1)$. If $R'(m+1, n+1)$ is initially set to be 0, all the elements in R' are negative. If we set $R'(m+1, n+1)$ to be a positive value, then some of the elements are positive. Moreover, these positive elements are adjacent. As shown in Fig. 2, the red cells denote the best warping path which is obtained by BS_DTW according to different initial values of $R'(m+1, n+1)$. The positive cells are in gray color. Since some cells in red color are positive, they are also a part of the ones in gray color.

If we set $R'(m+1, n+1)$ to be equal to 9.5 (i.e., $R'(m+1, n+1) = 9.5$) and run BS_DTW, some elements in R' are positive as shown in Fig. 2(a). Moreover, it is obvious that the best warping path only appears in the scope of the gray cells. As shown in Fig. 2(b), the best warping path is surrounded by the gray cells when $R'(m+1, n+1)$ is set to be 200. In Fig. 2(a), except for the cell $(1, 1)$, the remaining red cells are also surrounded by the positive cells. Thereby, if $R'(m+1, n+1)$ is big enough, all the elements of the best warping path can be surrounded by the positive cells filled in gray color. We regard these positive cells as reduced scope.

For simplicity, we denote $R'(m+1, n+1)$ to be a threshold value θ, i.e., $\theta = R'(m+1, n+1)$. In addition, we know that the minimum distance between the two time series Q and C is equal to 9.67 in subsection 3.1. This same result can also be obtained by our method, i.e.,

$$AF\_DTW(Q, C, \theta) = \theta - R'(1, 1), \tag{5}$$

AF_DTW is the new method based on BS_DTW. Fig. 2 shows that if $\theta = 9.5$, then $AF\_DTW(Q, C, \theta) = 9.5 - (-0.17) = 9.67$. If $\theta = 200$, then $AF\_DTW(Q, C, \theta) = 200 - 190.33 = 9.67$. So AF_DTW retains the same accuracy of DTW.

The above analysis tells us that the best warping path always exists in the reduced scope with regards to a special value of θ. So we only force AF_DTW to find the best warping path in the scope of the positive cells when θ is enough large. In this way, the number of the positive cells is less than that of all cells in the original cost matrix so that the time and space consumption depending on the number of the positive cells can be reduced. Therefore, AF_DTW costs less time and space than DTW and retrieves the same accurate result.

(a)R'(m+1,n+1)=9.5

8	−1576.7	−1480.7	−1459.5	−1216.2	−223.92	−155.03	−77.59	1.62	1.66
7	−578.81	−577.81	−560.17	−513.93	1.36	1.61	1.61	−72.3	−132.9
6	−577.07	−575.38	−564.34	−527.43	1.57	1.52	1.45	−141.19	−260.59
5	−578.36	−573.95	−589.84	−564.87	1.21	0.31	0.01	−197.44	−370.84
4	−1036.2	−958.21	−245.32	1.17	−527.79	−505.94	−501.75	−1164.7	−1533.7
3	−134.39	−107.35	0.81	−271.08	−550.83	−540.19	−538.96	−1383.7	−1850.5
2	−34.49	0.32	−136.08	−1032.8	−559.55	−562.97	−563.96	−1397.4	−1895.4
1	−0.17	−19.93	−178.33	−1061.3	−560.19	−563.06	−564	−1476.6	−2037
	1	2	3	4	5	6	7	8	9

(b)R'(m+1,n+1)=200

8	−1386.2	−1290.2	−1269	−1025.7	−33.42	35.47	112.91	192.12	192.16
7	−388.31	−387.31	−369.67	−323.43	191.86	192.11	192.11	118.2	57.6
6	−386.57	−384.88	−373.84	−336.93	192.07	192.02	191.95	49.31	−70.09
5	−387.86	−383.45	−399.34	−374.37	191.71	190.81	190.51	−6.94	−180.34
4	−845.7	−767.71	−54.82	191.67	−337.29	−315.44	−311.25	−974.15	−1343.2
3	56.11	83.15	191.31	−80.58	−360.33	−349.69	−348.46	−1193.2	−1660
2	156.01	190.82	54.42	−842.34	−369.05	−372.47	−373.46	−1206.9	−1704.9
1	190.33	170.57	12.17	−870.81	−369.69	−372.56	−373.5	−1286.1	−1846.5
	1	2	3	4	5	6	7	8	9

Fig. 2. Some positive elements exist in the cost matrix and the best warping path is surrounded by gray cells

Why should we choose an enough large θ? In order to let the best warping path be surrounded by the positive cells, θ must be larger than the minimum distance value, i.e., $\theta \geq DTW(Q, C)$. Otherwise, the positive cells cannot surround the best warping path as shown in Fig. 2(a). The cell (1,1) is an element of the best warping path, but it is out of the scope of the positive cells. Actually, the value of cell (1,1) is smallest in the red cells. If cell (1,1) is negative, then at least one element of the best warping path is out of the scope of the positive cells. Thereby, to make cell (1,1) be positive, we must let $\theta \geq DTW(Q, C)$, which is also inferred by Eq. (5).

After we choose a suitable θ, we take the backward strategy to construct the reduced scope in the cost matrix. If the value of current cell in the cost matrix is negative, then let it be zero. If the three top right adjacent cells are zero, then the current cell is set to be zero and the current iteration is broken. In this way, AF_DTW goes on calculating other cells whose values are positive. Finally, AF_DTW can obtain the reduced scope and retrieve the minimum distance between time series Q and C.

Different θ produces different number of non-white cells. At the same time, large difference between two values of θ does not crazily influence the changed number of not-white cells. In other words, θ and $\theta + \eta$ (where η may be very large) may retrieve the same number of non-white cells in the cost matrix, which means that AF_DTW using a large θ may be speeded up because of the reduced scope. Thereby, a suitable θ is important for AF_DTW. In next subsection, we will address this problem.

3.3 Choice of Threshold Value

We know that the bigger θ is, the larger the scope will be, and the more time and space are consumed. It means that the smallest scope is obtained when $\theta = DTW(Q, C)$. However, these minimum distance value is unknown and need us to compute. Thereby, we should find some value of θ close to $DTW(Q, C)$.

According to the Sakoe-Chuba Band and the Itakura Parallelogram, they all search the suboptimal warping path by limiting the scope along the opposite diagonal of the cost matrix, which inspires us to initialize θ. In other words, we regard one of the opposite diagonal warping paths as the initial warping path to assign θ a relative small value. The initial value of θ is the sum of elements in the initial warping path, i.e., $\theta = \sum_{k=1}^{K} d(p_k)$.

However, if the length of two time series used to measure the similarity is equal (i.e., $m = n$), then Euclidean distance between the two time series is the initial value of θ. We call this special path the Euclidean path. Actually, when $m = n$, the best warping path in opposite diagonal scope will be generalized into Euclidean path. Thereby, if $m = n$, then

$$\theta = \sum_{k=1}^{L} (q_k - c_k)^2, \qquad L = m = n. \tag{6}$$

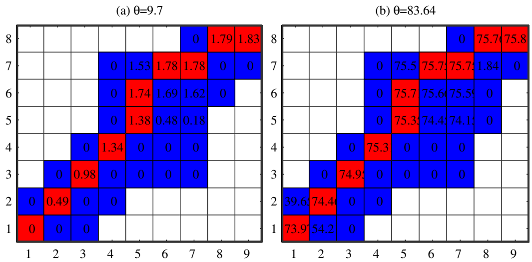

Fig. 3. The reduced scope produced by θ deriving from the initial warping path is the same to that produced by the minimum distance value $\theta = DTW(Q, C)$

In the above case, if $\theta = 83.64$ which is obtained by our method, then the reduced scope is identical to the optimal one which is produced by $\theta = DTW(Q, C) = 9.67$ as shown in Fig. 3. Thereby, our method to choose a suitable value of θ is feasible. At the same time, we must point out that any ways to produce a small θ which is bigger than $DTW(Q, C)$ can be used to reduce the search scope so as to speed up AF_DTW. Of course, the time and space complexity of the ways used to obtain a suitable θ must be linear to the length of time series. In our method, since the opposite diagonal scope is linear to the length of the two time series, the time and space complexity of the proposed method to decide the threshold value θ are linear to the length of time series, which also can be inferred by Euclidean path.

4 Experiments

Three experimental subsections about the performance of the proposed method AF_DTW are given. In the first subsection, we testify that the method about the choice of threshold θ is feasible and is used to improve the efficiency of DTW. In the second subsection, an experiment shows that the difference of time consumed by DTW and AF_DTW is more obvious with regards to the increased length of time series. The last experiments demonstrate that in contrast to DTW, the proposed AF_DTW not only has the same accuracy but also has faster calculation.

4.1 Different Thresholds Based Comparison

It is well know that the efficiency of AF_DTW depends on a reduced scope which further depends on a choice of threshold θ. In other words, the efficiency of AF_DTW indirectly depends on the choice of threshold θ. To testify that the method about the choice of threshold θ is feasible and useful, we take the following experiments on different time series datasets.

We choose the well-known UCR time series datasets [22] whose ID can be obtained in table 1. Each kind of time series dataset is consist of a training set and a testing set. Their length and size are also different from each other. For each kind of time series, we combine the training set with the testing set and regard them as a whole dataset. For a special value of threshold θ, every adjacent time series are used to compute the similarity (distance) between them and we regard the averaged time consumed as the result of the experiment with regards to the θ. Take Adiac dataset for example, for each value of θ, the number of time series in the whole dataset is 781 and the number of the pairs to compute similarity measure is 780. Let AF_DTW and DTW do this experiment. The final result is the averaged time they consumed.

In addition, the method about the choice of threshold θ runs in advanced and returns the initial value of θ for each pair of time series. We denote the initial value as a standard value θ'. At the same time, we also calculate the minimum distance value v between the pair of time series using the original DTW, i.e., $v = DTW(Q, C)$. Let l (step length) be $l = (\theta' - v)/3$. The values of θ chosen to compute the similarity by AF_DTW are $[v, v+l, v+2l, \theta', \theta'+l, \theta'+2l, \theta'+3l]$ or $[\theta'-3l, \theta'-2l, \theta'-l, \theta', \theta'+l, \theta'+2l, \theta'+3l]$ which is also identified as [1,2,3,4,5,6,7] shown in the label of axis X in the Fig. 4 for simplicity.

The results of experiments in Fig. 4 shows that for each θ, the proposed AF_DTW consumes much less time than DTW. Moreover, the result marked by a cycle in each subplot is the averaged time consumed by AF_DTW according to the choice method of threshold as shown in algorithm 3. The result marked by the cycle tells us that the proposed choice method of threshold makes AF_DTW be faster than DTW and the time consumption of AF_DTW with the standard value θ' is close to the optimal one which is consumed by AF_DTW with the minimum distance value v. At the same time, the ascending trend of time consumed by AF_DTW tells us that the bigger θ is, the more time AF_DTW consumes.

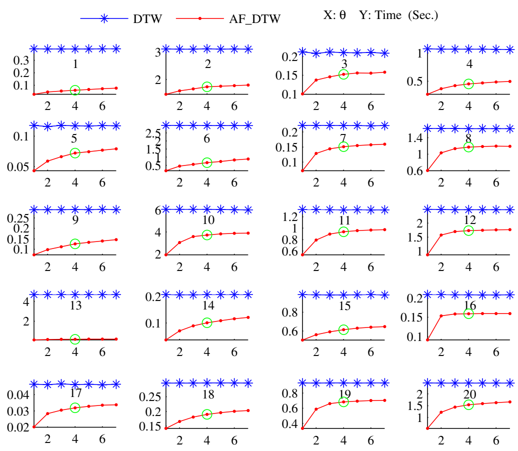

Fig. 4. The time consumption comparisons between AF_DTW and DTW according different θ

4.2 Different Length Based on Comparison

In this subsection, time series with different length are used to compare the efficiencies of AF_DTW and DTW. A long stock time series of length 2119415 is used [23] and we segment it seven groups of subsequences according to the length $L = [32, 64, 128, 256, 512, 1024, 2048]$. Each group has 50 subsequences. In each group we use AF_DTW and DTW to measure the similarity between each adjacent subsequences. In other words, each group has 49 pairs of subsequences used to measure the similarity. For each group, the result of the experiment is the averaged time.

Fig. 5 shows the result of comparison between AF_DTW and DTW according to different length of time series. It is easy to find that the longer the time

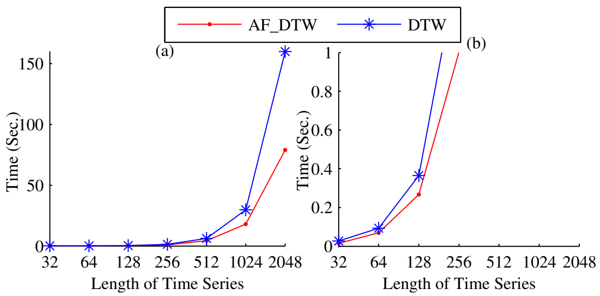

Fig. 5. The time consumption comparison between AF_DTW and DTW according different length of time series. (a) shows the result in the range $[0, 160]$ of axis Y; (b) shows the result in the range [0,1] of axis Y.

series is, the more obvious the difference of time consumption between the two methods will be. It means that when the time series is much long, in contrast to DTW, AF_DTW used to measure the similarity is more efficient.

4.3 Classification

Although the previous work [16, 17, 20, 21] can fast run the variants of dynamic time warping, they retrieve the approximated results instead of the accurate one. It means that in most cases they can find the best warping path as the original DTW does. In order to experimentally shows the accuracy and efficiency of AF_DTW, we do the classification experiments on the UCR time series [22].

We use the nearest neighbor classification to classify the time series. In every time series dataset, each time series in the testing set using the classification is respectively measured by AF_DTW and DTW to find the most similar one in training set. If the label of the most similar object in the training set is not same to that of the one used to classify in the testing set, then we regard it as an error classification. Finally, we record the number of the error classifications and consider the error ratio as the accuracy of the algorithm. At the same time, for each time series dataset, we record the CPU time consumption and denote it as T. The averaged time consumption $\bar{T}$ (sec.) is used to show the efficiency of the algorithm.

Table 1. Experiment results on the UCR datasets

No.	Name	DTW & AF_DTW	DTW($\bar{T}$)	AF_DTW($\bar{T}$)
1	Adiac	0.396	168.57	25.16
2	Beef	0.5	101.39	54.11
3	CBF	0.003	6.92	4.89
4	Coffee	0.179	32.64	13.01
5	ECG200	0.23	12.99	8.08
6	FISH	0.167	563.86	121.44
7	FaceAll	0.192	124.64	94.66
8	FaceFour	0.170	39.20	29.00
9	Gun	0.093	14.65	6.10
10	Lighting2	0.131	365.60	232.28
11	Lighting7	0.274	94.34	69.30
12	OSULeaf	0.409	503.58	352.86
13	OliveOil	0.133	142.70	6.37
14	Swedish Leaf	0.210	105.85	52.83
15	Trace	0	100.56	64.48
16	2Patterns	0	212.27	162.47
17	Control	0.007	13.91	10.36
18	Wafer	0.02	201.71	194.33
19	50Words	0.310	429.12	316.75
20	Yoga	0.164	752.99	459.80

The final results of the experiments are shown in table 1. It is easy to discovery that the accuracy of classification using AF_DTW and DTW is identical, which means that the proposed method AF_DTW can retain the same accuracy as the original DTW does. At the same time, the averaged time consumption by AF_DTW is less than that of DTW, which means that AF_DTW is faster than DTW for time series classification. It also easily know that AF_DTW can be more efficient than DTW. Especially in some datasets, such as No.1, No.6 and No.13, AF_DTW is obviously more efficient than DTW.

5 Conclusions

The proposed method, accurate and fast dynamic time warping (AF_DTW), is an improved version of the original DTW. Comparing to DTW, it uses backward strategy to find the best warping path. At the same time, a new method using the backward strategy reduces the search scope in the cost matrix so that not only the accuracy of AF_DTW is the same to that of DTW but also the computation performance is more efficient. In addition, the choice method of threshold we proposed makes AF_DTW be a new version of DTW without any parameters. Actually, any values of the threshold greater than the minimum distance are able to reduce the search scope, which makes AF_DTW be faster than DTW.

Although the computation performance of AF_DTW is more efficient than DTW, the speeding-up degree indirectly depends on a choice of threshold. The proposed choice method of threshold can retrieve a good result to speed up the computation. However, other better choice methods of threshold to narrow the search scope may be existed, which will let AF_DTW be more faster. So finding a more efficient choice method of threshold is one of the most important tasks in the future.

Acknowledgements. This work has been partly supported by the National Natural Science Foundation of China (61300139) and the Society and Science Planning Projects in Fujian (2013C018).

References

1. Faloutsos, C., Ranganathan, M., Manolopoulos, Y.: Fast subsequence matching in time series databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 419–429 (1994)
2. Chu, S., Keogh, E., Hart, D., Pazzani, M.: Iterative deepening dynamic time warping for time series. In: Proceedings of the Second SIAM International Conference on Data Mining, pp. 195–212 (2002)
3. Chen, L., Ng, R.: On the marriage of Lp-norm and edit distance. In: Proceedings of the 30th International Conference on Very Large Databases, pp. 792–801 (2004)
4. Burnham, K.P., Anderson, D.R.: Kullback-Leibler information as a basis for strong inference in ecological studies. Wildlife Research 28(1), 111–119 (2001)
5. Keogh, E., Pazzani, M.J.: Derivative dynamic time warping. In: Proceedings of The First SIAM International Conference on Data Mining, pp. 1–11 (2001)

6. Keogh, E.: Exact indexing of dynamic time warping. In: Processings of the 28th International Conference on Very large Data Bases, pp. 358–380 (2005)
7. Lemire, D.: Faster retrieval with a two-pass dynamic-time-warping lower bound. Pattern Recog. 42(9), 2169–2180 (2009)
8. Li, H., Guo, C., Yang, L.: Similarity measure based on piecewise linear approximation and derivative dynamic time warping for time series mining. Expert Systems with Applications 38(12), 14732–14743 (2011)
9. Petitjean, F., Ketterlin, A., Gancarski, P.: A global averaging method for dynamic time warping, with applications to clustering. Pattern Recogn. 44(3), 678–693 (2011)
10. Lai, C.P., Chung, P.C., Tseng, V.S.: A novel two-level clustering method for time series data analysis. Expert Systems with Applications 37(9), 6319–6326 (2010)
11. Guo, C., Li, H., Pan, D.: An improved piecewise aggregate approximation based on statistical features for time series mining. In: Bi, Y., Williams, M.-A. (eds.) KSEM 2010. LNCS, vol. 6291, pp. 234–244. Springer, Heidelberg (2010)
12. Li, H., Guo, C.: Piecewise cloud approximation for time series mining. Knowledge-Based Systems 24(4), 492–500 (2011)
13. Vlachos, M., Gunopoulos, D., Kollios, G.: Discovering similar multidimensional trajectories. In: Proceedings of the 18th International Conference on Data Engineering, pp. 673–684 (2005)
14. Lin, J., Keogh, E., Lonardi, S., Chiu, B.: A symbolic representation of time series with implications for streaming algorithms. In: Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, pp. 2–11 (2003)
15. Muscillo, R., Conforto, S., Schmid, S.: Classification of motor activities through derivative dynamic time warping applied on accelerometer aata. In: Proceedings of the 29th Annual International Conference of Engineering in Medicine and Biomedicine, pp. 4930–4933 (2007)
16. Itakura, F.: Minimum prediction residual principle applied to speech recognition. IEEE Transactions on Acoustous, Speech, and Signal Process 23(1), 52–72 (1975)
17. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustous, Speech, and Signal Process 26(1), 43–49 (1978)
18. Keogh, E., Pazzani, M.: Scaling up dynamic time warping for datamining applications. In: Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 285–289 (2000)
19. Zhou, M., Wong, M.H.: A segment-wise time warping method for time scaling searching. Information Sciences 173(1), 227–254 (2005)
20. Sakurai, Y., Yoshikawa, M., Faloutsos, C.: FTW: fast similarity search under the time warping distance. In: Proceedings of the 2005 Principles of Database Systems, pp. 326–337 (2005)
21. Salvador, S., Chan, P.: FastDTW: toward accurate dynamic time warping in linear time and space. Intelligence Data Analysis 11(5), 561–580 (2007)
22. Keogh, E., Xi, X., Li, W.: Welcome to the UCR time series classification/clustering (2007), http://www.cs.ucr.edu/~eamonn/time_series_data/
23. Stock datasets (2005), http://www.cs.ucr.edu/~wli/FilteringData/stock.zip