Procedia Computer Science
Volume 51, 2015, Pages 2838–2842
ICCS 2015 International Conference On Computational Science

Flexible Dynamic Time Warping for Time Series Classification*

Che-Jui Hsu[1], Kuo-Si Huang[2], Chang-Biau Yang[1]†, and Yi-Pu Guo[1]

[1] Department of Computer Science and Engineering
National Sun Yat-sen University, Kaohsiung, Taiwan
cbyang@cse.nsysu.edu.tw
[2] Department of Information Management
National Kaohsiung Marine University, Kaohsiung, Taiwan

Abstract
Measuring the similarity or distance between two time series sequences is critical for the classification of a set of time series sequences. Given two time series sequences, X and Y, the dynamic time warping (DTW) algorithm can calculate the distance between X and Y. But the DTW algorithm may align some neighboring points in X to the corresponding points which are far apart in Y. It may get the alignment with higher score, but with less representative information. This paper proposes the flexible dynamic time wrapping (FDTW) method for measuring the similarity of two time series sequences. The FDTW algorithm adds an additional score as the reward for the contiguously long one-to-one fragment. As the experimental results show, the DTW and DDTW and FDTW methods outperforms each other in some testing sets. By combining the FDTW, DTW and DDTW methods to form a classifier ensemble with the voting scheme, it has less average error rate than that of each individual method.

Keywords: dynamic time warping, time series, classification, longest common subsequence

1 Introduction

The time series classification problem is an important topic in data mining. It can be applied to many areas, such as finance, pharmacy, biometrics, chemistry, astronomy, robotics, networking, industry, etc. [8]. Given an input time series sequence s, the goal of this problem is to classify s in a prepared dataset D, which is composed of several classified clusters and each cluster has several time series sequences. To solve this problem, one of the methods is to calculate the distance or similarity between s and each time series sequence in D, and then choose the cluster of the highest frequency based on the k *nearest neighbors* (k-NN) algorithm [7, 10].

*This research work was partially supported by the National Science Council of Taiwan under contract NSC102-2221-E-110-033-MY2.
†Corresponding author.

 Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2015

doi:10.1016/j.procs.2015.05.444

For the time series classification problem, to measure the distance between two sequences is a critical issue. One of the commonly used methods for measuring the distance is the *dynamic time warping* (DTW) algorithm [4]. The DTW algorithm may consider the one-to-many, many-to-one and one-to-one mappings between two sequences by using the dynamic programming approach. While applying the DTW algorithm, the special case that most points in one sequence are mapped to only a few points of the other sequence may occur. To avoid such an extreme situation, one may apply the *warping window* scheme to the DTW algorithm [10].

In this paper, we propose a new algorithm, the *flexible dynamic time warping* (FDTW) method, which is based on the flexible longest common subsequence algorithm [3]. In the proposed algorithm, we give additional score to reward the longer one-to-one fragments. That is, we try to make the lengths of contiguous one-to-one fragments as long as possible. Though our new algorithm does not apply the warping window, our algorithm can still have the property that the points in one sequence are not aligned to the points far away in the other sequence.

The organization of this paper is as follows. In Section 2, the background knowledge of the FDTW method is introduced. In Section 3, we propose a new algorithm for determining the similarity of two time series sequences. Section 4 presents the materials, results and discussions of the experiments. Finally, in Section 5, we give the conclusion of this paper.

2 Preliminaries

Let $A = a_1a_2\cdots a_m$ and $B = b_1b_2\cdots b_n$ be two input sequences (or strings). The *subsequence* of A is obtained by removing zero or more symbols from A. The *common subsequence* of A and B is a subsequence of A and B simultaneously. The *longest common subsequence* (LCS) of A and B is the common subsequence with the maximum length among all possible common subsequences of A and B. The similarity of two sequences can be measured by their LCS length, and then the corresponding alignment can also be constructed accordingly [5, 12]. Several LCS variants have also been studied [1, 6, 11].

The *flexible longest common subsequence* (FLCS) problem [3], a variant from the LCS problem, considers the preference of longer continuously matched fragments (substrings) in A and B. If a contiguously matched fragment is longer, its *flexible alignment score* will be higher. Let $FLCS(i, j)$ denote the maximum of the flexible alignment score of $A_{1,i}$ and $B_{1,j}$, where $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$. Note that $A_{k,i}$ denotes the substring $a_ka_{k+1}\cdots a_i$ of A. Let c and γ denote the length of the matched fragment and the parameter to control the magnitude of emphasized score, respectively. The formula for calculating $FLCS(i, j)$ is given in Equation (1) [3].

$$FLCS(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0; \\ \max \begin{cases} FLCS(i-1,j) \\ FLCS(i,j-1) \\ \{FLCS(i-c,j-c) + c^{\gamma} \mid c > 0 \\ \ \text{and } A_{(i-c+1),i} = B_{(j-c+1),j}\} \end{cases} & \text{otherwise.} \end{cases} \tag{1}$$

The *dynamic time warping* (DTW) algorithm is applied to measuring the distance between two sequences of time series [4]. Similar to LCS algorithm, the DTW algorithm is also based on the dynamic programming approach and it can be solved in $O(mn)$ time.

Let $X = \{x_1, x_2, \cdots, x_m\}$ and $Y = \{y_1, y_2, \cdots, y_n\}$ be two time series sequences. The distance between $X_{1,i}$ and $Y_{1,j}$ is denoted as $DTW(i, j)$. The formula for calculating $DTW(i, j)$ is given in Equation (2). In Equation (2), the function $d(x_i, y_j)$ indicates the distance of x_i and

y_j, which may be any function for measuring the distance between x_i and y_j.

$$DTW(i,j) = \begin{cases} d(x_i, y_j) + \min \begin{cases} DTW(i-1,j) \\ DTW(i,j-1) \\ DTW(i-1,j-1) \end{cases} & \text{if } i \neq 0 \text{ and } j \neq 0; \\ 0 & \text{if } i = 0 \text{ and } j = 0; \\ \infty & \text{otherwise.} \end{cases} \tag{2}$$

3 The Flexible Dynamic Time Warping Algorithm

We propose a new algorithm, the *flexible dynamic time warping* (FDTW), for solving the time series classification problem. The traditional DTW algorithm measures the distance between two given sequences. However, our FDTW calculates the similarity between two sequences of time series, based on the concept of the FLCS algorithm [3]. The principle of our algorithm is to increase the score of one-to-one fragments. With this property, we may avoid that the neighboring points in one sequence are aligned to the points whose coordinates are far away in the other sequence.

In the FDTW algorithm, we first have to define the threshold t for determining whether a point pair are matched or not. If the distance of two points is no more than t, we say that these two points are matched; otherwise, the two points are viewed as mismatched. That is, we transform the distance of two points into the state of only match or mismatch, and also transform the distance measurement into the similarity measurement.

Let c, γ and t denote the length of a one-to-one fragment, the parameter to control the magnitude of emphasized score and the predefined distance threshold, respectively. The *maximum flexible similarity* of $X_{1,i}$ and $Y_{1,j}$ in the FDTW algorithm, denoted as $FDTW(i,j)$, is given in Equation (3), where $|X_{(i-c+1)..i} - Y_{(j-c+1)..j}| \leqslant t$ means $\{|x_{i-c+1} - y_{j-c+1}| \leqslant t, |x_{i-c+2} - y_{j-c+2}| \leqslant t, \cdots, |x_i - y_j| \leqslant t\}$.

$$FDTW(i,j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0; \\ \max \begin{cases} FDTW(i-1,j) \\ FDTW(i,j-1) \\ \{FDTW(i-c,j-c) + c^{\gamma} \mid c > 0 \\ \;\text{and } |X_{(i-c+1)..i} - Y_{(j-c+1)..j}| \leqslant t\} \end{cases} & \text{otherwise.} \end{cases} \tag{3}$$

The FDTW algorithm tries to find several contiguously long fragments which are one-to-one mapping, and then to construct the alignment based on these fragments. For example, suppose $X = \{0.3, 0.6, 0.2, 0.7, 0.5, 0.4\}$, $Y = \{0.2, 0.4, 0.8, 0.1, 0.4\}$, $t = 0.2$, and $\gamma = 2$. One possible alignment is shown in Figure 1. The flexible similarity score of this alignment is $2^2+1^2+1^2 = 6$.

X	0.3	0.6	-	0.2	0.7	0.5	0.4
Y	0.2	0.4	0.8	0.1	-	-	0.4

Figure 1: A possible alignment of $X = \{0.3, 0.6, 0.2, 0.7, 0.5, 0.4\}$, $Y = \{0.2, 0.4, 0.8, 0.1, 0.4\}$, with $t = 0.2$ and $\gamma = 2$.

4 Experimental Results

In our experiments, 15 datasets were selected for performance comparison with the DTW [4] and DDTW [10] methods from the UCR time series classification/clustering homepage [9]. The UCR datasets contain a lot of real data of time series and these datasets have been adopted in several previous studies [2, 7]. In the UCR datasets, each dataset has been separated into a training set and a testing set. The experiments were performed on a computer with Microsoft Windows 7 Professional OS, Intel Core i5-4570 CPU and 4GB RAM.

For a dataset, after computing the score of the target time series in the testing set and each time series in the training set, we need to classify the target into its class. Here, we apply the k-NN classification method. In the method, the k sequences of the training set with the best measurement (shortest distance or highest similarity score) are examined. Then, the majority vote is used to decide the class of the target sequence.

In the experiments, the arithmetic average (AVG) is the sum of error rates divided by the number of testing sets, and weighted average (wAVG) is the average with considering the size of each testing set. The formulas are shown in Equation (4), where E_i and $|T_i|$, $1 \leqslant i \leqslant 15$, denote the error rate and the size of each testing set, respectively.

$$\text{AVG} = \sum_{i=1}^{15} E_i/15, \quad \text{wAVG} = \sum_{i=1}^{15}(E_i \times |T_i|)/\sum_{i=1}^{15}|T_i|, \tag{4}$$

The experimental results are summarized in Table 1, where t' is the controlling parameter to decide the threshold t in the FDTW algorithm. The error rates of FDTW with $t' = 0.05$ are better in Beef, Coffee and Motestrain datasets. In Face (four), ECGFiveDays and Italy-PowerDemand datasets, FDTW with $t' = 0.1$ has better performance. And in ECG dataset, FDTW is better with $t' = 0.2$. DTW and FDTW (with $t' = 0.1$) have better performance for the arithmetic and the weighted averages of all datasets, respectively.

For improving the error rates of classification, we classify the data with the voting scheme. That is, we first aggregate the results of DTW, DDTW and FDTW and then decide the class of the target time series according to the majority vote. For example, if DTW regards the target as class C_1, DDTW regards the target as class C_2 and FDTW regards the target as class C_1. Then the target time series is determined to belong to class C_1. The decision of FDTW is also obtained from the majority vote of FDTW with $t' \in \{0.05, 0.1, 0.2\}$. The voting ensemble improves performance in three datasets, and it gets improvement in both the arithmetic and weighted averages.

5 Conclusion

In this paper, we propose the flexible dynamic time wrapping (FDTW) algorithm to compute the similarity of two time series sequences. With the FDTW algorithm, we may avoid that the neighboring points in one sequence are aligned to the points separated far away in the other sequence. Hence, the long contiguous one-to-one fragments can be found to construct the alignment. Afterwards the FDTW algorithm is applied for classification of time series.

In the experimental results, we find that the performance of the FDTW algorithm is comparable to the DTW and the DDTW algorithms in some datasets. And we use the voting scheme to improve the accuracy of classification. By the comparison of error rates, the voting scheme has the better average performance than each individual method. In the future, it is worthy to analyze the characteristics of datasets to find which properties are especially suitable for the FDTW algorithm.

Dataset \ Method	DTW	DDTW	FDTW $t' = 0.05$	FDTW $t' = 0.1$	FDTW $t' = 0.2$	Vote
Synthetic Control	0.013 (1)	0.3567 (3)	0.1233 (9)	0.08 (11)	0.1 (7)	0.04
Gun-Point	0.12 (1)	0 (3)	0.0333 (3)	0.2067 (3)	0.2667 (1)	0.0267
CBF	0 (1)	0 (3)	0.0589 (7)	0.0889 (9)	0.1044 (5)	0.0856
Face (four)	0.1591 (1)	0.375 (1)	0.125 (1)	0.0341 (1)	0.0455 (1)	0.1477
Lightning-7	0.2192 (3)	0.3836 (3)	0.3151 (7)	0.4247 (5)	0.4384 (11)	0.2466
ECG	0.17 (5)	0.14 (3)	0.18 (3)	0.15 (3)	0.07 (3)	0.14
Beef	0.4333 (1)	0.3333 (1)	0.3 (1)	0.5667 (7)	0.6667 (1)	0.3667
Coffee	0 (11)	0 (13)	0 (5)	0.2143 (15)	0.4643 (1)	0.0357
OliveOil	0.1 (3)	0.1333 (1)	0.7 (11)	0.7 (11)	0.7 (11)	0.1667
ECGFiveDays	0.2253 (1)	0.3182 (1)	0.1301 (1)	0.0221 (1)	0.0348 (1)	0.1336
ItalyPowerDemand	0.0535 (7)	0.0845 (1)	0.0962 (5)	0.0535 (9)	0.207 (1)	0.0476*
MoteStrain	0.1094 (1)	0.2819 (1)	0.1046 (3)	0.1086 (13)	0.1446 (13)	0.103*
SonyAIBORobot SurfaceII	0.1574 (1)	0.1259 (1)	0.2193 (9)	0.1532 (1)	0.1511 (3)	0.1112*
SonyAIBORobot Surface	0.2879 (1)	0.2612 (1)	0.3927 (1)	0.3328 (1)	0.3428 (1)	0.3128
TwoLeadECG	0.0674 (1)	0.0053 (1)	0.0386 (1)	0.1414 (13)	0.4539 (7)	0.0334
Arithmetic average (AVG)	0.1411	0.1866	0.1878	0.2185	0.2793	0.1332*
Weighted average (wAVG)	0.1345	0.1693	0.1452	0.1267	0.2273	0.1076*

Table 1: The error rates of various methods, where each number marked by a red underline and blue star indicate the least error rate in each dataset by an individual method and by the voting classifier ensemble, respectively.

References

[1] H.-Y. Ann, C.-B. Yang, and C.-T. Tseng. Efficient polynomial-time algorithms for the constrained lcs problem with strings exclusion. *Journal of Combinatorial Optimization*, 28:800–813, 2014.

[2] T. Górecki. Using derivatives in a longest common subsequence dissimilarity measure for time series classification. *Pattern Recognition Letters*, 45(0):99–105, 2014.

[3] Y.-P. Guo, Y.-H. Peng, and C.-B. Yang. Efficient algorithms for the flexible longest common subsequence problem. In *Proceedings of the 31st Workshop on Combinatorial Mathematics and Computation Theory*, pages 1–8, Taipei, Taiwan, 2014.

[4] L. Gupta, D. L. Molfese, R. Tammana, and P. G. Simos. Nonlinear alignment and averaging for estimating the evoked potential. *IEEE Transactions on Biomedical Engineering*, 43(4):348–356, April 1996.

[5] D. S. Hirschberg. A linear space algorithm for computing maximal common subsequences. *Communications of the ACM*, 18:341–343, 1975.

[6] K.-S. Huang, C.-B. Yang, K.-T. Tseng, H.-Y. Ann, and Y.-H. Peng. Efficient algorithms for finding interleaving relationship between sequences. *Information Processing Letters*, 105(5):188–193, 2008.

[7] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu. Weighted dynamic time warping for time series classification. *Pattern Recognition*, 44(9):2231–2240, 2011.

[8] E. Keogh and S. Kasetty. On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Mining and Knowledge Discovery*, 7(4):349–371, Oct. 2003.

[9] E. Keogh, Q. Zhu, B. Hu, Y. Hao, X. Xi, L. Wei, and C. A. Ratanamahatana. The UCR time series classification/clustering homepage. `http://www.cs.ucr.edu/~eamonn/time_series_data/`, 2011.

[10] E. J. Keogh and M. J. Pazzani. Derivative dynamic time warping. In *Proceedings of the First SIAM International Conference on Data Mining*, 2001.

[11] C.-T. Tseng, C.-B. Yang, and H.-Y. Ann. Efficient algorithms for the longest common subsequence problem with sequential substring constraints. *Journal of Complexity*, 29:44–52, 2013.

[12] C.-B. Yang and R. C. T. Lee. Systolic algorithms for the longest common subsequence problem. *Journal of the Chinese Institute of Engineers*, 10(6):691–699, 1987.