Learning Optimal Warping Window Size of DTW for Time Series Classification

Qian Chen, Guyu Hu, Fanglin Gu, Peng Xiang
PLA University of Science and Technology
Nanjing, China
qqianchen@126.com

Abstract—**The dynamic time warping (DTW) is a classic similarity measure which can handle the time warping issue in similarity computation of time series. And the DTW with constrained warping window is the most common and practical form of DTW. In this paper, the traditional learning method for optimal warping window of DTW is systematically analyzed. Then the time distance to measure the time deviation between two time series is introduced. Finally a new learning method for optimal warping window size based on DTW and time distance is proposed which can improve DTW classification accuracy with little additional computation. Experimental data show that the optimal DTW with best warping window get better classification accuracy when the new learning method is employed. Additionally, the classification accuracy is better than that of ERP and LCSS, and is close to that of TWED.**

Keywords-component; time series; similarity measure; dynamic time warping; warping path, time distance

I. Introduction

Time series are sequences of values or events obtained from repeated measurements in time [1]. Time series data are generated from almost every application field, such as daily fluctuations of stock market, medical and biological experimental observations, signals from industrial monitoring sensors, et al. As a result, similarity queries of time series are practically and widely required, e.g., stock market analysis, discovery of medically abnormal pattern of electrocardiogram, fault alarm monitoring in industry.

Similarity measure is the fundamental problem in similarity queries. Euclidean distance metric is widely used as similarity measurement of time series. One reason is that the computation complexity is linear; another reason is that it satisfies the triangle inequality and supports all kinds of indexing. However, it is very sensitive to distortion in time axis. It is well known that Dynamic Time Warping (DTW) can be used to address the problem of distortion in the time axis [2]. As shown in Figure 1, DTW allows elastic shifting of the time axis between two time series to accommodate sequences that are similar, except the locally out of phase. The main disadvantages of DTW are its quadratic computational complexity, and it does not satisfy the triangle inequality.

Current research on elastic similarity measures that tolerates time warping are focused on the following two aspects:

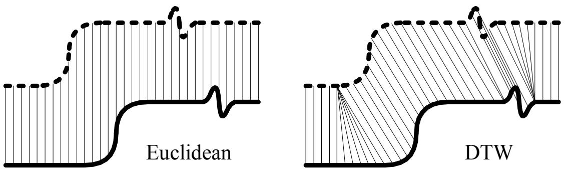


Figure 1. Euclidean distance and DTW distance

(1) To Improve on the DTW. Due to the high computational complexity of DTW, it is indispensable to find fast algorithms to improve the efficiency of DTW calculation. The commonly used methods include: (a) Early abandon technology [3-4]. The main idea is terminating the calculation when the total cost exceeds a threshold and taking them as not similar; (b) Set the search scope using the global constraints [5]. The main idea is confining the warping path in the warping window. Since the DTW does not support indexing, a lower bound function with linear computation complexity, which also satisfying the triangle inequality, is used to construct exact indexing [6-9].

(2) Other elastic measures. Such as Edit distance with Real Penalty (ERP), Longest Common Sub-Sequence (LCSS), Time Warp Edit Distance (TWED), *et al.* ERP attempts to combine the merits of DTW and edit distance, by using a constant reference point for computing the distance between gaps of two time series [10]. LCSS utilizes the longest common subsequence model, to adapt the concept of matching characters in the settings of time series [11,12,13]. TWED combines L_p-norms with the edit distance, and adds a time shift penalty coefficient called stiffness value and a gap constant penalty [14].

In sum, the value of DTW distance is taken as standard for similarity measure in most of the existing literatures. Additionally, the global constraints are often used to deal with the distortion in the time axis. The impact of the time deviation on the similarity has recently become a major concern in some literatures, such as the TWED that use the stiffness value to punish the time deviation. In our opinion, the warping path has certain influence on the similarity, so the DTW would get a better result if the warping path is considered.

The rest of the paper is organized as follows. Section 2 presents the necessary background knowledge about DTW and its global constraints. Section 3 explains the definition of time distance and proposes a new learning method for optimal warping window of ODTW based on DTW and time distance

978-1-4673-0382-8/12/$31.00 ©2012 IEEE

in detail. Section 4 describes a classification experiment that shows the empirical effectiveness of ODTW when the new learning method is employed. Section 5 concludes the paper and presents the directions for further work.

II. Background

A. DTW and Warping Path

Consider two time series $X = \langle x_1, x_2, \cdots, x_n \rangle$, $Y = \langle y_1, y_2, \cdots, y_m \rangle$ with the length n and m. The element (i, j) of a n-by-m distance matrix represents the distance $d(x_i, y_j)$. Find a warping path $W = \langle w_1, w_2, \cdots, w_K \rangle$, $\max(n, m) \le K \le n + m - 1$ from the distance matrix, where $w_k = (i, j)$. The W is usually subjected to several constraints [6]:

- Boundary conditions: $w_1 = (1,1)$, $w_K = (n, m)$.
- Continuity: $w_k = (a_k, b_k)$, $w_{k+1} = (a_{k+1}, b_{k+1})$, then $a_{k+1} - a_k \le 1$, $b_{k+1} - b_k \le 1$.
- Monotonicity: $w_k = (a_k, b_k)$, $w_{k+1} = (a_{k+1}, b_{k+1})$, then $a_{k+1} \ge a_k$, $b_{k+1} \ge b_k$.

As illustrated in Figure 2. DTW path has the minimal cumulative distance in all possible warping paths.

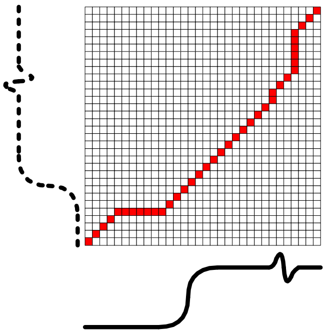


Figure 2. Warping path of DTW

Let X_i be the sub-series of X with discrete time index varying between 1 and i, DTW distance is defined as

$$DTW(X_n, Y_m) = d(x_n, y_m) + \min \begin{cases} DTW(X_{n-1}, Y_m) \\ DTW(X_{n-1}, Y_{m-1}) \\ DTW(X_n, Y_{m-1}) \end{cases} \quad (1)$$

And the computational complexity is $O(n \times m)$.

B. Global Constraints of DTW

When the warping path is far away from the diagonal, it may converge to the undesired path. For example, in Figure 3, DTW will find its optimal mapping between the two time series, and the DTW distance is 0. However, the two time series fall into different classes.

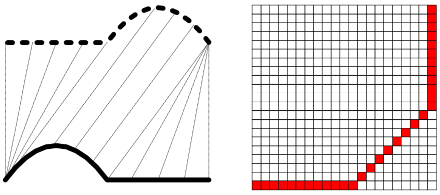
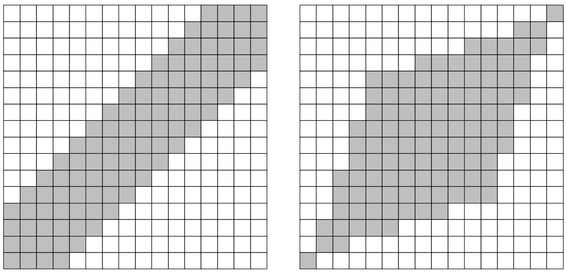


Figure 3. Undesired warping in DTW

The global constraint is used to prevent pathological warping in the originator of DTW algorithm proposed by Sakoe and Chiba [14], in which the global constraint limits the permissible warping paths. The available subset of the matrix is called Warping Window. Figure 4 illustrates two of the most frequently used global constraints. Literature [15] point out that wider bands do not always result in increased accuracy. The global constraints can slightly speed up the DTW distance calculation. What is more, they can improve the classification accuracy under the condition that the correct warping window is employed [16].



(a)Sakoe-Chuba band (b) Itakura Parallelogram

Figure 4. Two most common global constraints

DTW with Sakoe-Chiba Band (uniform warping window) is the most popular practical form of DTW. Compared with full DTW, it has exact indexing and fast algorithm for speeding up sequential searches by using LB_Keogh lower bound. The Optimized DTW with best warping window (OTDW), have been compared with other similarity measures in many literatures [5,14,16].

III. Learning Optimal Warping Window of DTW

A. Traditional Learning Method

To learn the best size of warping window, all the possible classification error rates will be iterated in traditional methods, and then select the best value which minimizes the classification errors estimated for the training data. A leave-one-out method is usually employed in such a procedure. The main idea consists of iteratively selecting one time series from the training set and considering it as a test against the remaining time series within the training set itself. If different values lead to the minimal error rate estimated for the training data, then the lowest value is selected [16].

This method makes the best use of DTW distance information between the time series of training set, and the value is the best in terms of the DTW distance. However, if the information coming from the warping path of DTW is considered, then we may get a better size of warping window by using this information.

B. Time Distance between the Time Series

The concept of time distance is proposed in this paper to make best use of the information coming from the warping path of DTW. Definitions of time distance are based on the warping path of DTW.

Definition 1. The i^{th} path segment of *X*: The segment of warping path which include the i^{th} point of *X*, marked by W_X^i

$$W_X^i = \langle w_k = (a_k, b_k) | a_k = i, w_k \in W \rangle \quad (2)$$

According to the Definition 1, there are n path segments of X in the warping path.

Definition 2. Maximal time deviation of the i^{th} path segment of X: The maximal deviation value of the index of Y from i in the i^{th} path segment of X, marked by $t_X(i)$

$$t_X(i) = \max_{(i,b_k)\in W_X^i} \{|b_k - i|\} \quad (3)$$

Definition 3. Time Distance: The sum of the maximal time deviation of all the path segments of X in warping path, marked by $d_W(X,Y)$

$$d_W(X,Y) = \sum_{i=1}^{n} t_X(i) \quad (4)$$

The physical meaning of time distance is the area between DTW warping path and the diagonal path $W_{best} = \langle w_k = (a_k, b_k) | a_k = b_k \rangle$. As shown in Figure 5, for two time series with the same length, the best warping path should be the diagonal path (gray path in the figure) consisting of points at where $a_k = b_k$. In such case, the two time series match each other point by point, and there is no deviation in time axis. Practically, there is always deviation between DTW warping path (red path in the figure) and the diagonal path. Hence, the deviation of two series in time axis can be presented by the area between DTW warping path and the diagonal path .

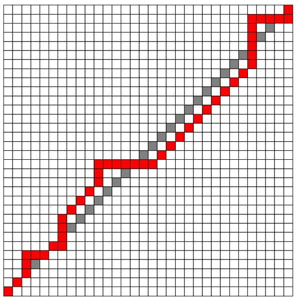


Figure 5. Warping path of DTW and the best warping path

The definition of time distance can also be used for measurement of two time series with unequal length. According to the definition, the distance between two time series with unequal length must be more than zero, the time distance will become larger as the difference between the lengths of time series grows, which is in accordance with our daily experience. The time distance is actually only a measure rather than a metric because the triangle inequality is not satisfied.

Note that the warping path of DTW is an appurtenant of the DTW computation. We construct a orientation matrix G with the source orientations of every DTW value we recorded during DTW computation, and the warping path of DTW will be get from $g(n,m)$ down to $g(1,1)$. Then, we can calculate the time distance according to the definitions. In such a way, it will not add much computation complexity. However, this computation doubles the space complexity because the source orientation matrix needs to be saved.

C. New Learning Method based on both Distances

We introduce our new learning method based on DTW distance and time distance with an example using actual dataset.

First get the classification error rates with all the possible size, according to the leave-one-out method with DTW distance and time distance. As shown in Table I, the error count instead of error rates is listed. E_v (the results of DTW distance) show the influence of value deviation on the error rate, and E_t (the results of time distance) show the influence of time deviation on the error rate. E_{sum} is the sum of E_v and E_t.

TABLE I. CLASSIFICATION ERROR COUNTS WITH DIFFERENT WARPING WINDOW SIZE

r	*Ev*	*Et*	*Esum*
0	14	50	64
1	16	36	52
2	16	25	41
3	16	19	35
4	16	15	31
5	16	18	34
6	17	16	33
…		…	

Since the DTW distance is taken to classify the test set, the size of warping window should come from the values that minimize the classification errors. However, there is possible tiny difference between training sets and test sets, hence, a excursion is allowed. Therefore, we choose the size from

$$E_v(r) \le \min(E_v) + \Delta \quad (5)$$

Δ is error count excursion. To different training set, Δ should be a tiny rate of the size of training set, e.g. $r_n = 1\%$. So $\Delta = \lfloor N_{train} \times r_n \rfloor$, where N_{train} means the size of training set, and $\lfloor\cdot\rfloor$ is a function that rounds to the nearest integers towards minus infinity. On the other hand, the actual classification effect should be taken into consideration, since the classification accuracy of DTW varies with different data set. Error count excursion should be a rate of the maximal error count, e.g. $r_e = 10\%$. Δ is set to be the maximum of two values in consideration of two above factors.

$$\Delta = \max\left(\lfloor N_{train} \times r_n \rfloor, \lfloor \max(E_v) \times r_e \rfloor\right) \quad (6)$$

As shown in Figure 6, the marked points in curve are the candidates. At the point r=0, the DTW distance gets the minimal classification error count, while the error count of time distance might be big, because the warping path is not exact under current warping window. So we should increase the size of warping window to gain a better warping path. On the other hand, the error count of DTW distance might increase while increasing the size of warping window. Therefore, we should select the point that minimizes the sum of E_v and E_t. In this example, the size of warping window should be 4.

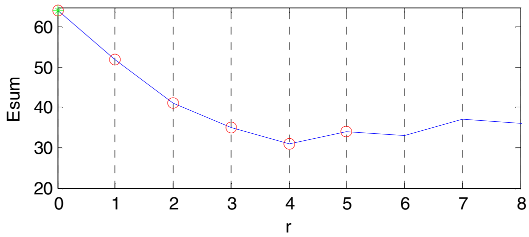


Figure 6. Curve of Esum

According to the above analysis, the rules for selection are:

1) Setting the candidates, select the segment that contains minimum of E_v if candidates consist of several segments.
2) If E_v and E_t. changes with all candidates, select the size leading to the first relative minimal point of E_{sum}.
3) If E_v is a constant with all candidates, select the size leading to the minimum of E_t .
4) If different values lead to the first relative minimal, select the median.

The size of warping window can be learned using (1) and (2) in most of the datasets, (3) and (4) are two special cases. (3) means the DTW distance makes no difference to the classification accuracy with all candidates, then select the size leading to the minimum of E_t to insure the best warping path. (4) represents the median is the best choice for minimizing the classification error when there are multiple extreme values.

Since the warping path is an appurtenant of the DTW computation, the new method improves DTW classification accuracy with little additional computation. At the same time, the work have been researched on DTW can also be used in the new algorithm, e.g. the exact indexing and fast algorithm for speeding up sequential searches by using LB_Keogh lower bound.

IV. EXPERIMENT

In this section, classification experiments are implemented to evaluate the effectiveness of our new learning method. In order to distinguish our new learning method from the traditional one, we name the ODTW using the new learning method as NODTW (New ODTW), whereas the ODTW using the traditional learning method is named as TODTW (Traditional ODTW). The 20 datasets in our experiment come from the UCR repository [17], which is also used in literature [14,15,16]. The classification is based on the simple nearest neighbor decision rule discussed in literature [14,16]. We compared the classification error rates using ED, DTW, ERP, LCSS, TWED, TODTW and NODTW as similarity measure. For ERP, TWED, NODTW, TODTW we used the L_1-norm, while the L_2-norm has been implemented in DTW and ODTW as reported in [5]. L_p-norms in DTW, Daniel Lemire proposed DTW_1 is a good choice to classify time series whereas DTW_2 is a close second through experiments [8]. Train the optimal warping window size using the new learning method, where $r_n = 1\%$, $r_e = 10\%$.

Detailed results of the experiment are presented in Table III. To provide a more intuitive illustration of the performance of the similarity measures compared in Table III, we use scatter plots to conduct pair-wise comparisons. In a scatter plot, the x and y axes show the error rates for the two compared distance (D_x, D_y). The line in the plot has a slope of 1.0 and dots correspond to the error rates for the compared distance and test datasets. A dot above the line indicates that the error rate for D_x is lower than that for D_y. So if the dots above the line are more than the dots below the line, it indicates that the distance D_x is better than the distance D_y. As shown in Figure 7, ODTW with L_1-norm is slightly better than ODTW with L_2-norm, which is consistent with the conclusion in literature [8].

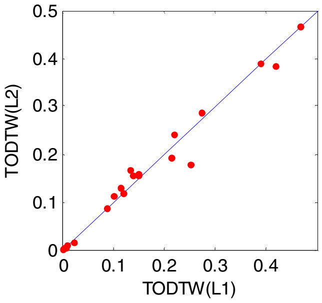


Figure 7. ODTW(L1) VS ODTW(L2)

Figure 8 shows the comparison of distance pairs. It is clear that NODTW is obviously better than TODTW, ED, DTW, ERP, and closed to LCSS, but is slightly worse than TWED. In fact, for the 20 datasets, 9 of them for NODTW are better than that for LCSS, 8 worse and 3 equal. And the average classification error rate for NODTW is much less than that for LCSS, so NODTW is superior over LCSS.

Then we discuss the performance of NODTW against that of TWED. For the 20 datasets, 9 of them for NODTW are better that for TWED, 9 worse and 2 equal. The average classification error rate of NODTW is slightly less than that of TWED, so NODTW and TWED have a very close performance.

On the other hand, TWED have two parameters: stiffness value γ and gap constant penalty λ, while ODTW only have only one parameter: warping window size r. TWED and DTW both search the entire matrix, and have the same computational complexity of $O(n^2)$. ODTW restrict the search range, the less size of warping window result in the less computation complexity. So in the phase of training, the computational complexity of NODTW is less than that of TWED. In the phase of classifying, NODTW have the fast algorithm to speed up sequential searches using LB_Keogh lower bound, the computing speed is much higher than that of TWED. Table II shows the training and classification time of TWED and NODTW in our experiment. For TWED, the stiffness value γ is selected from {10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1}, and λ is selected from {0, 0.25, 0.5, 0.75, 1}, 30 pairs of (γ, λ). For NODTW, the size of warping window r ranges from 0 to 30 in training, and the LB_Keogh lower bound fast searching algorithm is used in classification. The experimental results show that the running time of NODTW is less than that of TWED. So the improved ODTW is clearly the better choice in regarding of practicability.

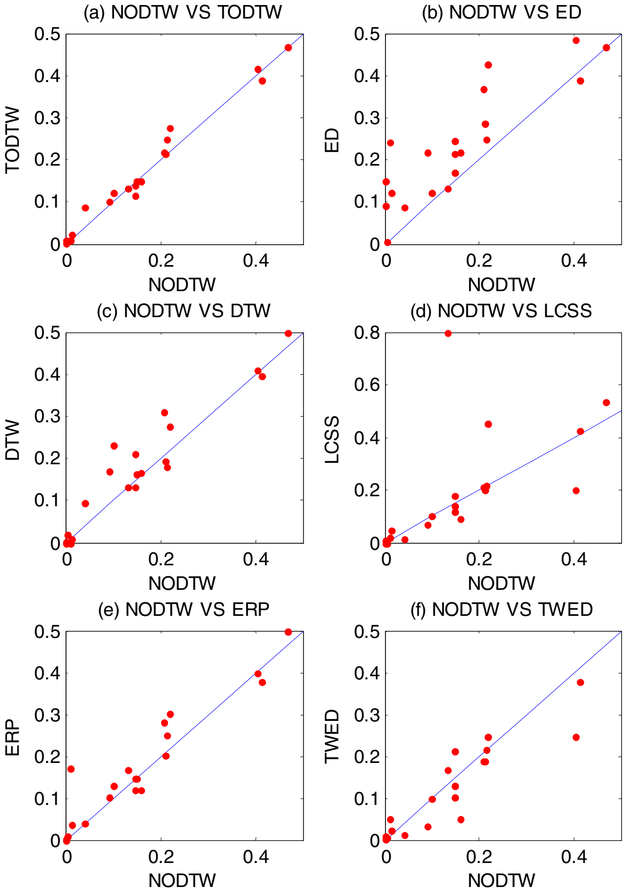


Figure 8. Comparison of distance pairs

TABLE II. RUNNING TIME OF TWED AND NODTW (SECONDS)

Dataset	Training Time		Classification Time	
	TWED	*NODTW*	*TWED*	*NODTW*
Gun-Point	3129.8	813.2	636.6	18.6
CBF	784.5	236.5	1669.5	565.1
Face(four)	4368.1	1141.8	1277.2	221.1
ECG200	5049.8	1776.1	340.6	18.4

V. CONCLUSION

In this paper, we have proposed a new learning method for optimal warping window Size of DTW, our contribution is basically three folded:

1) Introduced the time distance to measure the time deviation between two time series;

2) Proposed a new learning method for optimal warping window size of DTW based on DTW and time distance;

3) Proved that the classification accuracy of improved ODTW is close to TWED via an experiment.

Further research will focus on the reduction of the computational complexity of training with LB_Keogh lower bound to improve the practicability of our new learning method.

ACKNOWLEDGMENT

This work is supported in part by Natural Science Foundation of China under Grant 61001106 and National Program on Key Basic Research Project of China under Grant 2009CB320400.

REFERENCES

[1] J. Han and M. Kamber. Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, CA, 2005.

[2] D.J. Berndt, J. Clifford. "Using Dynamic Time Warping to Find Patterns in Time Series" in Proceedings of KDD-94: AAAI Workshop on Knowledge Discovery in Databases. Menlo Park , CA :AAAI, 1994, pp. 359-370.

[3] W. Li, E. Keogh, H. Herle, A. Mafra-Neto. “Atomic wedgie: Efficient Query Filtering for Streaming Time Series” in Proceedings of the 5th IEEE International Conference on Data Mining. Washington DC: IEEE Computer Society, 2005, pp. 490-497.

[4] E. Keogh, W. LI, X. Xi, S. Lee, M. Vlachos. “LB_Keogh Supports Exact Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures” in Proceedings of the 32nd Very Large Database Conference. Seoul: VLDB Endowment, 2006, pp. 882-893.

[5] C. A. Ratanamahatana, E. Keogh. “Making Time-series Classification More Accurate Using Learned Constraints”, in Proceedings of SIAM International Conference on Data Mining (SDM '04), Lake Buena Vista, Florida, 2004. pp. 11-22.

[6] E. Keogh . “Exact indexing of dynmic time warping” in Proceedings of 28th International Conference on Very Large Databases Conference. Hong Kong, VLDB Endowment, 2002, pp. 406-417.

[7] S. Kim, S. Park, W. Chu. “An Index-Based Approach for Similarity Search Supporting Time Warping in Large Sequence Databases” in Proceedings of the 17th International Conference on Data Engineering. Washington, DC: IEEE Computer Society. 2001, pp. 607-614.

[8] L. Daniel. “Faster Retrieval with a Two-Pass Dynamic-Time-Warping Lower Bound”. Pattern Recognition. 2009, vol. 42, no. 9: 2169-2180.

[9] V. Niennattrakul, P. Ruengronghirunya and C.A. Ratanamahatana. “Exact Indexing for Massive Time Series Databases under Time Warping Distance”. Data Mining and Knowledge Discovery. 2009. vol. 21, no. 3, pp. 509-541.

[10] L. Chen, R. Ng. “On the marriage of Lp-norm and edit distance” in Proceedings of the 30th International Conference on Very Large Data Bases, Toronto Canada, 2004, pp. 792–801.

[11] G. Das, D. Gunopulos, and H. Mannila. “Finding Similar Time Series” in Proceedings of the Conference on Principles of Knowledge Discovery and Data Mining, Springer-Verlag London UK, 1997, pp. 88-100.

[12] W.K.Loh, S.W.Kim, K.Y.Whang. “Index Interpolation: An Aproach for Subsequence Matching Supporting Normalization Transform in Time-Series Database” in Proceedings of the 9th Internation Conference on Information and Knowledge Management. New York, ACM Press. 2000, pp. 314-325.

[13] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, E. Keogh. “Indexing Multi-Dimensional Time-Series with Support for Multiple Distance Measures” in the 9th ACM SIGKDD, Washington, DC, USA. August 24 - 27, 2003, pp. 216-225.

[14] P.F. Marteau. “Time Warp Edit Distance with Stiffness Adjustment for Time Series Matching”. IEEE Transactions on Pattern Analysis and Machine Intelligence Archive. February 2009. Vol.31 Issue 2, pp. 306-318.

[15] H. Sakoe, S. Chiba. “Dynamic Programming Algorithm Optimization for Spoken Word Recognition”. IEEE Trans Acoustics Speech Signal Process , ASSP, 1978, pp:43-49.

[16] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang and E. Keogh. “Querying and Mining of Time Series Data: Experimental Comparison

of Representations and Distance Measures" in Proceedings of the VLDB Endowment. Kyoto Japan, August 2008, Vol. 1, Issue: 2, pp. 1542-1552.

[17] 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

APPENDIX

TABLE III. ERROR RATIO OF DIFFERENT SIMILARITY MEASURES ON 1NN CLASSIFIER

Dataset	Number of classes| Size of training| testing set| Time series Length	ED	LCSS	ERP	TWED	DTW	TODTW L2 (r)	TODTW L1 (r)	NODTW L1 (r)
Synthetic Control	6|300|300|60	0.12	0.047	0.036	0.023	**0.007**	0.017 (6)	0.023 (3)	0.013 (24)
Gun-Point	2|50|150|150	0.087	**0.013**	0.04	**0.013**	0.093	0.087 (0)	0.087 (0)	0.04 (7)
CBF	3|30|900|128	0.148	0.009	0.003	0.007	0.003	0.004 (11)	0.008 (10)	**0 (21)**
Face (all)	14|560|1690|131	0.286	0.201	0.202	**0.189**	0.192	0.192 (3)	0.215 (6)	0.211 (5)
OSU Leaf	6|200|242|427	0.483	**0.202**	0.397	0.248	0.409	0.384 (7)	0.417 (13)	0.405 (16)
Swedish Leaf	15|500|625|128	0.213	0.117	0.12	**0.102**	0.210	0.157 (2)	0.138 (2)	0.147 (4)
50Words	50|450|455|270	0.369	0.213	0.281	**0.187**	0.310	0.242 (6)	0.218 (23)	0.209 (18)
Trace	4|100|100|275	0.24	0.02	0.17	0.05	**0.0**	0.01 (3)	0.01 (12)	0.01 (18)
Two Patterns	4|1000|4000|128	0.09	**0.0**	**0.0**	0.001	**0.0**	0.0015 (4)	0.0013 (5)	**0 (7)**
Wafer	2|1000|6174|152	0.005	**0.0**	0.009	0.004	0.020	0.005 (1)	0.0045 (5)	0.0034 (1)
Face (four)	4|24|88|350	0.216	0.068	0.102	**0.034**	0.170	0.114 (2)	0.102 (3)	0.091 (8)
Lightning-2	2|60|61|637	0.246	0.18	0.148	0.213	0.131	0.131 (6)	**0.115 (8)**	0.148 (6)
Lightning-7	7|70|73|319	0.425	0.452	0.301	0.247	0.274	0.288 (5)	0.274 (3)	**0.219 (9)**
ECG200	2|100|100|96	0.12	**0.10**	0.13	**0.10**	0.23	0.12 (0)	0.12 (0)	**0.1 (4)**
Adiac	37|390|391|176	0.389	0.425	0.378	**0.376**	0.396	0.391 (3)	0.389 (0)	0.414 (2)
Yoga	2|300|3000|426	0.170	0.137	0.147	**0.130**	0.164	0.155 (2)	0.149 (11)	0.149 (11)
Fish	7|175|175|463	0.217	0.091	0.120	**0.051**	0.167	0.160 (4)	0.149 (8)	0.16 (9)
Beef	5|30|30|470	**0.467**	0.533	0.5	0.533	0.5	**0.467 (0)**	**0.467 (0)**	**0.467 (0)**
Coffee	2|28|28|286	0.25	0.214	0.25	0.214	**0.179**	**0.179 (3)**	0.25 (3)	0.214 (7)
Olive Oil	4|30|30|570	**0.133**	0.8	0.167	0.167	**0.133**	0.167 (1)	**0.133 (0)**	**0.133 (0)**
MEAN		0.234	0.191	0.175	**0.145**	0.179	0.164	0.164	0.157