Indian Journal of Science and Technology, Vol 9(21), DOI: 10.17485/ijst/2016/v9i21/93886, June 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Efficient Dynamic Time Warping for Time Series Classification

Kumar Vasimalla[1]*, Narasimham Challa[2] and Manohar Naik S[3]

[1]Research and Development Center , Bharathiar University, Coimbatore - 641046, Tamil Nadu, India; Kumar.sanvi@gmail.com

[2]Department of CSE, VIIT Visakhapatnam, Gajuwaka - 530046, Andhra Pradesh, India; narasimham_c@yahoo.com

[3]Department of Computer Science, Central University of Kerala, Padannakkad, Nileshwar - 671314, Kerala, India; manoharamen@gmail.com

Abstract

Background/Objective: Dynamic Time Warping (DTW), a similarity measure works in O(N^2) complexity. Cause of this it will be used for small datasets only. **Methods/Statistical Analysis**: In this work, we introduced Efficient DTW (EDTW), which works in linear time. It uses two level approaches. In the first level data reduction is performed, and in the second level warping distance and path are calculated. **Findings:** While calculating the values of distance matrix, values along the warping path only considered and calculated. For time series of length n, maximum n values of distance matrix are calculated. So it works in linear time. **Improvements/Applications:** We applied this distance measure to UCR Time Series archive and calculated error rate of 1NN classification. Most of the cases it is matching, some cases it is better, and some other cases error rate is high.

Keywords: Dynamic Time Warping, Efficient DTW, 1NN Classification, Time Series Classification, UCR Time Series Data Sets

1. Introduction

Dynamic time warping (DTW) finds similarity between sequences, it is one of the similarity methods used in pattern recognition and time series data mining and other fields.

Figure 1 shows example of non-warping and warping between time series. DTW is efficient similarity method for time series datasets, but it has quadratic complexity.

Here we developed an Efficient Dynamic Time Warping (EDTW) algorithm, it takes linear time and works in two level. In first level the time series data is reduced by applying data reduction technique, i.e. the series is reduced to one by third by taking the average of three consecutive values. Second level calculates warping distance and path, while calculating, only limited values around warping path of distance matrix considered. The shape of the wrapping path of reduced data set and original dataset are almost similar as shown in the Figure 5.

The remaining sections are organized as follows. Section 2 outlines the details of standard DTW, relevant prior approaches to accelerate it. An in depth description of EDTW algorithm covers in Section 3. Section 4 is experimental evaluation discusses experimentation on synthetic temperature dataset and explains the experimentation results on the UCR Time Series dataset. Section 5 summarizes the EDTW.

2. Related Work

The DTW distance measure is used to find similar instances in a time series database and to overcome[1] problems of other similarity methods. Finding similar series is base for classification and clustering and other data mining techniques.

Given two time series X and Y, of lengths n and m.

$$X = (x_1, x_2, x_3, \ldots, x_n)$$

*Author for correspondence

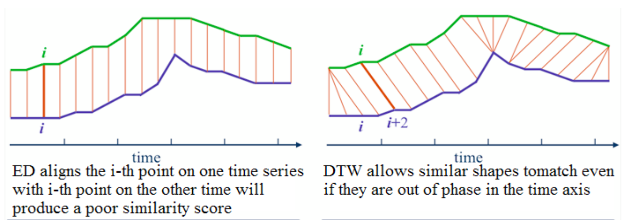

Figure 1: A non-warping and warping between two time series.

$$Y = (y_1, y_2, y_3, \ldots, y_m)$$

Construct a warping path P, P= p_1, p2… p_k
Where k greater than or equal to length of the larger time series and less than or equal to sum of time series.

The optimal warping path P is calculated as follows

$$Dist(P) = \sum_{k=1}^{k} Dist(p_{ki}, p_{kj})$$

DTW distance follows boundary condition, monotonicity condition, step size condition.

Definition: A warping path is a sequence p= (p_1, p_2, p_L) with three properties

(i) Boundary[2] condition: p_1= (1, 1), p_L= (N, M).
(ii) Monotonicity[2] condition: n1 ≤ n2 ≤ …≤ n_L and m_1 ≤ m_2 ≤ … ≤ m_L.
(iii) Step size[2] condition: p_L+ 1 – p_L ∈ {(0, 1), (1, 0), (1, 1)} for L ∈ (1: L-1).

To calculate minimum-distance warping path, dynamic programming approach is used. It follows principle of optimality. Figure 2 shows dynamic time warping path, for given two series. X=(7,8,7,8,9,8,11,14,16,17,15,13), Y=(5,8,9,10,12, 12,13,17,16,16,15,14), it start with *D*(1,1) and ends with *D*(12,12).

The wrap path is w={(1,1),(1,2),(2,3),(2,4),(3,5),(4,5),(4,6),(5,7),(6,7),(7,8),(8,9),(8,10),(9,11),(10,11),(11,11),(12,12)}. Initially Y_1 is warped with $X_{1,2}$ and Y_2 is warped with $X_{3,4}$, X_5 warped with $Y_{3,4}$ and so on. At the end X_{11} is warped with $Y_{9,10,11}$.

Time and Space complexity of the DTW is O(NM). Because we need to fill N*M distance matrix, each value calculation of matrix takes constant time, if size of X and Y are N and M respectively. There are so many attempts in the literature to minimize the space and time complexity, some attempts succeeded but it will be applicable to some applications, may not be suitable to applications requiring full path[3], and merging of time series together[4,5].

2.1 Speedup Dynamic Time Warping

To speedup DTW three methods are available.

1) *Constraints* –It reduce the quantity of values in the cost matrix while calculating distance and warping path.
2) *Data reduction* – Before applying DTW try to reduce the data set size.
3) *Indexing* – It will decrease the DTW calculation in Data Mining functionalities.

The constraints S-C Band[6] and the Itakura Parallelogram[7], which are shown in figure 3.

Applying constraints while calculating DTW will accelerate process by constant factor, still it requires quadratic time. Data reduction[8] speeds up, even it requires O(N^2) time. Figure 4 below show DTW path on reduced dataset.

14												*
15											*	
16											*	
16											*	
17									*	*		
13								*				
12							*					
12							*					
10					*	*						
9					*							
8			*	*								
5	*	*										
Y/X	7	8	7	8	9	8	11	14	16	17	15	13

Figure 2: DTW path shown with stars

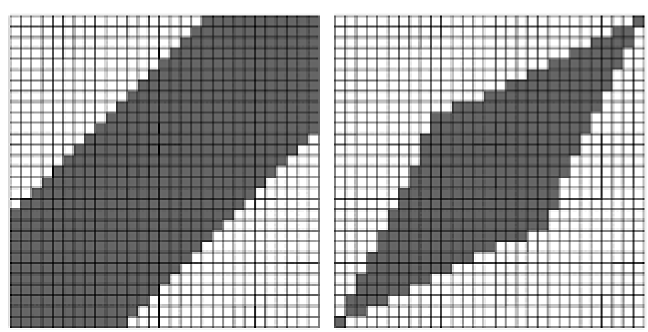

Figure 3: S-C Band, Itukara Parallelogram (right)[8].

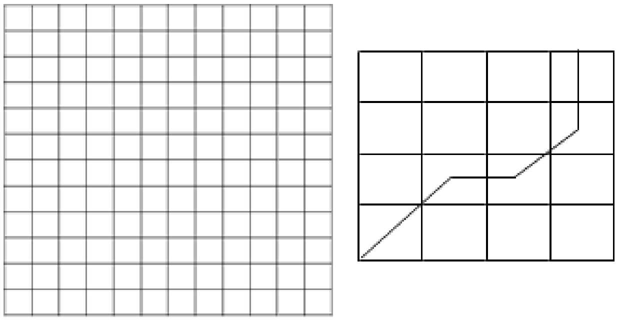

Figure 4: Speeding up DTW by Data Reduction

Indexing can be used in time series data mining to speedup similarity and other techniques[9, 10].

2.2 Fast DTW

This method uses multilevel approach by three operations *Coarsening,Projection,Refinement*[11].

FastDTW[11] is a recursive implementation of DTW. The solution to the problem is available if time series size is less than min Size else it uses the above three operations.

The space complexity and time complexity of Fast DTW is O(N) and O(N) respectively.

Fast DTW is an approximation of DTW, the limitations of Fast DTW is it does not always find an optimal solution. The possible extensions for Fast DTW is trying for different step sizes conditions, Investigate search algorithms to help improve refinement, examining number of cells evaluated while calculating warping path and compare accuracy.

3. Efficient DTW

The EDTW follows two stages. It is motivated by data mining[12] process of finding various patterns through functionalities. Where the dataset is reduced to small size suitable for existing algorithm. Then apply algorithm on reduced dataset to get the patterns.

3.1 Efficient DTW Algorithm

The EDTW works in two stages

1. *Reduction* – Reduce an instance into a small size by using some data reduction techniques.
2. *Constraints–Reduce the number of values in the cost matrix while calculating distance and warping path*

Reduction reduce an instance into a small size by replacing every three values by a single value which is the average of three values. This reduce instance size by 1/3[rd] of the original instance.

Constraints limits the count of values evaluated in distance matrix, just above the warping path one line and below the warping path one line of values will be evaluated, remaining values will not be evaluated.

The EDTW algorithm first uses reduction to create the reduced dataset by taking the average of three consecutive values. The EDTW applied on reduced data, while calculating optimal warping path it follows constraints. That is only the values around the path will be calculated. The dark shaded cells is the warping path and light shaded cells are the extra cells required for calculation. This path will be approximately equivalent to original dataset optimal warping path. Figure 5 shows the optimal warping paths of reduced dataset as well as original dataset.

We will calculate maximum three values corresponding to each cell in the entire warping path using the following formula[13].

$$minDist(i,j) = \begin{cases} (X_1 - Y_1)^2 & i = j = 1 \\ minDist(i',j') + min\begin{cases}(X_i - Y_{j+1})^2 \\ (X_{i+1} - Y_j)^2 \\ (X_{i+1} - Y_{j+1})^2\end{cases} & \end{cases}$$

Where (i',j') is the previous pointon the warping path.

Whenever calculating mindist. We followed Boundary condition, monotonicity condition and step size condition.

The optimal warping paths have similar shape and the number of value of full size matrix is approximately three times the reduced matrix.

EDTW evaluated 10 cells at reduced size 4*4, and 38 cells at the original size (12*12), while DTW evaluates all

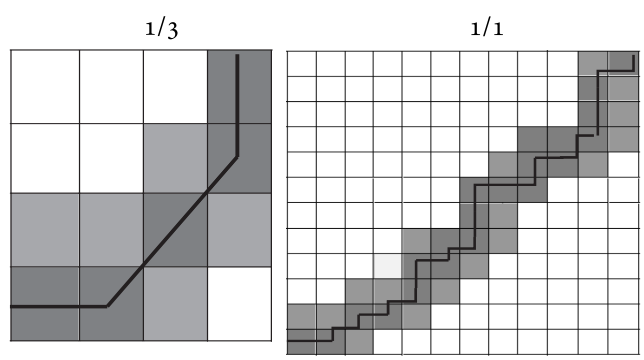

Figure 5. Efficient DTW paths for reduced and originalsizes

16 cells at reduced size and 144 cells at original size. The original size is 12 and it reduced to 1/3(=4), in reduced size we calculating only 10 values. In the worst case also for time series of size N maximum N values only calculated. This increases efficiency significantly. The EDTW algorithm is shown in table 1.

Table 1. Efficient DTW algorithm

Algorithm Efficient DTW (X,Y)
i/p: *X* and Y are Time Series
o/p: P –warping path
minDist- distance of path
{
1. M=|X|/3; N=|Y|/3;
2. //calculate Reduced dataset of size M by N
3. i:=1;j:=1;
4. P*[i,j]=[1,1]*
5. *minDist* :=$(X_1\text{-}Y_1)^2$
6. While (i!=N and j!=M)
{
7. *If (i==N) then*
8. j=j+1;
9 end
10. if (j==M) then
11. i=i+1;
end
12. if (i!=N&&j!=M) then

$$[i,j] = min\begin{cases}(X_i - Y_{j+1})^2\\(X_{i+1} - Y_j)^2\\(X_{i+1} - Y_{j+1})^2\end{cases}$$

end
13. minDist=minDist+$(X_i\text{-}Y_j)$*^2*
14. P[i,j]=[i,j]
}
}

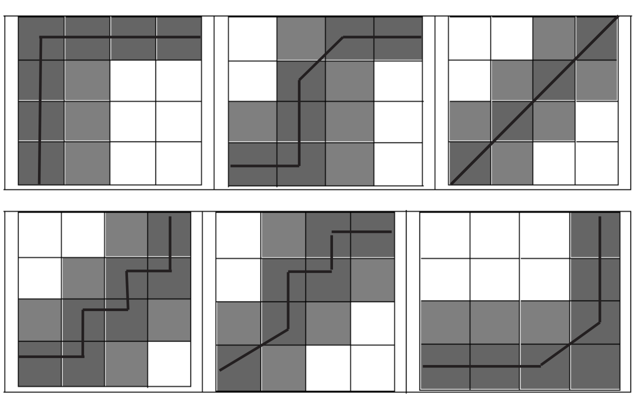

Figure 6. Some of possible warping paths in reduced data.

The number of values needed to calculate are maximum 12, means it is equivalent to original size. Original size N, reduced size N/3, for reduced size EDTW calculates N cell values in worst case. So it's time complexity is linear. The possible wrapping paths of size 4 by 4 distance matrix is shown in figure 6. In all possible cases we computed 10 to 12 values in the distance matrix, so the space complexity of EDTW is O(N).

4. Experimentation

In this section we want to test the new distance measure EDTW against synthetic temperature dataset and standard UCR Time series datasets to measure the accuracy. The following subsections explain the results of EDTW.

4.1 Synthetic Dataset

This section explains experimental evaluation of the EDTW algorithm using synthetic temperature dataset. In data set we took temperature for every two hours between 0 to 24 hours for 10 days. The reduced dataset is generated by taking the average of consecutive three values. The original and reduced data sets are shown in tables2 and 3. On reduced dataset we applied EDTW and calculated neighbors and summarized in the table 4.The neighbors of original dataset using DTW(NODDTW)and reduced dataset using DTW(NRDDTW), reduced dataset using EDTW(NRDEDTW) are mostly similar (table 4).

4.2 UCR Time Series Data Set

EDTW is tested against UCR Time series datasets [14]. The reference[14] contains various datasets names. For

Table 2. Original Dataset

5	8	9	10	12	12	13	17	16	16	15	14
6	9	11	12	12	12	14	16	18	17	16	14
7	8	9	11	11	12	13	14	17	16	12	12
6	7	7	9	9	10	10	14	16	18	14	13
7	7	8	7	7	10	10	12	14	15	13	12
6	7	7	8	8	10	10	14	16	16	12	12
5	6	7	8	8	9	9	11	12	13	13	11
6	6	7	8	8	10	10	12	14	15	15	12
7	8	7	8	9	8	11	14	16	17	15	13
8	8	10	10	11	10	14	15	16	17	13	12

each dataset the following details available. Number of classes(NC), size of training set(STRS), size of test set(STES), time series length(TL), error rates of 1-NN classifier using Euclidean Distance(1NNED), DTW with best warping window size(DTWBWW), and DTW with no warping window(DTWNWW).

On UCR time series dataset we applied1-NN classification with EDTW and calculated classification error rate. The error rate is compared with existing classification methods. Surprisingly its *performance is good on some datasets*. For example if we observe wine dataset, its error rate is 0.000. It's a good improvement compared to rival methods. The list of data sets for which error rate is less (performance improved) is shown in the table 5.

On some data sets its performance is equivalent to existing methods. The list of datasets whose error rate is equal to existing methods is shown in the table 6. The list of datasets where error rate is high compared to rival methods is shown in the table 7. For example car dataset, Cricket_X, Face, Strawberry datasets we may observe more error rate compared to rival methods.

Table 3. Reduced dataset

7	11	15	15
9	12	16	16
8	11	15	13
7	9	13	15
7	8	11	13
7	9	13	13
6	8	11	12
6	9	12	14
7	8	14	15
9	10	15	14

Table 4. Neighbors of original/reduced dataset using DTW/EDTW distance.

Dataset No.	NODDTW	NRDDTW	NRD EDTW
1	2,3,10	2,3,4,10	3,10,2
2	1,3,10	1,3,10	1,4,10
3	10,1,2	10,1,2	10,1,8
4	9,6,5	9,8,6	9,8,6
5	8,6,4,7,9	7,8,6	7,8,6
6	4,5,8,9	4,8,5	8,4,5
7	8,5,6	5,6,8	5,6,3
8	5,6,7	4,5,6	4,6,5
9	4,6,5,8	4,6,8	4,8,6
10	3,1,2	3,1,4	3,1,4

Table 5. Data sets[14] with less error rate

S.No.	Name	NC	STRS	STES	TL	1NNED	DTWBWW	DTWNWW	1-*NN Efficient DTW*
1	Bird Chicken	2	20	20	512	0.450	0.300(6)	0.250	0.020
2	Distal Phalanx TW	6	139	400	80	0.273	0.272 (0)	0.290	0.272
3	Ford A	2	1320	3601	500	0.341	0.341 (0)	0.438	0.404
4	Ham	2	109	105	431	0.400	0.400 (0)	0.533	0.523
5	Hand Outlines	2	370	1000	2709	0.199	0.197 (1)	0.202	0.095
6	Haptics	5	155	308	1092	0.630	0.588 (2)	0.623	0.577
7	Herring	2	64	64	512	0.484	0.469 (5)	0.469	0.468
8	Insect Wing beat Sound	11	220	1980	256	0.438	0.422 (2)	0.645	0.607
9	Large Kitchen Appliances	3	375	375	720	0.507	0.205 (94)	0.205	0.229
10	Middle Phalanx Outline Age Group	3	154	400	80	0.260	0.253 (5)	0.250	0.242
11	Proximal Phalanx Outline Age Group	3	400	205	80	0.215	0.215 (0)	0.195	0.165
12	Refrigeration Devices	3	375	375	720	0.605	0.560 (8)	0.536	0.538
13	Screen Type	3	375	375	720	0.640	0.589 (17)	0.603	0.517
14	Wine	2	57	54	234	0.389	0.389 (0)	0.426	0.000

Table 6. The datasets[14] with similar error rate

S.No.	Name	NC	STRS	STES	TL	1NNED	DTWBWW	DTWNWW	1-*NN Efficient DTW*
1	50Words	50	450	455	270	0.369	0.242 (6)	0.310	0.325
2	Adiac	37	390	391	176	0.389	0.391 (3)	0.396	0.450
3	Distal Phalanx Outline Age Group	3	139	400	80	0.218	0.228 (1)	0.208	0.227
4	ECG 5000	5	500	4500	140	0.075	0.075 (1)	0.076	0.079
5	Ford B	2	810	3636	500	0.442	0.414 (1)	0.406	0.436
6	Medical Images	10	381	760	99	0.316	0.253(20)	0.263	0.271
7	Middle Phalanx Outline Correct	2	291	600	80	0.247	0.318 (1)	0.352	0.371
8	Middle Phalanx TW	6	154	399	80	0.439	0.419 (2)	0.416	0.442
9	Plane	7	105	105	144	0.038	0.000 (6)	0.000	0.047
10	Proximal Phalanx Outline Correct	2	600	291	80	0.192	0.210 (1)	0.216	0.240
11	Trace	4	100	100	275	0.240	0.010 (3)	0.000	0.020
12	Two Patterns	4	1000	4000	128	0.090	0.002 (4)	0.000	0.003
13	Synthetic Control	6	300	300	60	0.120	0.017 (6)	0.007	0.190

Table 7. The datasets[14] with high error rate

S.No.	Name	NC	STRS	STES	TL	1NNED	DTWBWW	DTWNWW	1-*NN Efficient DTW*
1	Arrow Head	3	36	175	251	0.200	0.200 (0)	0.297	0.371
2	Beef	5	30	30	470	0.333	0.333 (0)	0.367	0.400
3	Beetle Fly	2	20	20	512	0.250	0.300 (7)	0.300	0.450
4	Car	4	60	60	577	0.267	0.233 (1)	0.267	0.533
5	CBF	3	30	900	128	0.148	0.004 (11)	0.003	0.183
6	Chlorine Concentration	3	467	3840	166	0.35	0.35 (0)	0.352	0.396
7	Cricket_X	12	390	390	300	0.423	0.228 (10)	0.246	0.497
8	Diatom Size Reduction	4	16	306	345	0.065	0.065 (0)	0.033	0.3
9	Distal Phalanx Outline Correct	2	276	600	80	0.248	0.232 (2)	0.232	0.495
10	Earth quakes	2	139	322	512	0.326	0.258 (22)	0.258	0.413
11	ECG Five Days	2	23	861	136	0.203	0.203 (0)	0.232	0.319
12	Face (four)	4	24	88	350	0.216	0.114 (2)	0.170	0.590
13	Face (all)	14	560	1690	131	0.286	0.192 (3)	0.192	0.384
14	Toe Segmentation 1	2	40	228	277	0.320	0.250 (8)	0.228	0.302
15	Toe Segmentation 2	2	36	130	343	0.192	0.092 (5)	0.162	0.492
16	Strawberry	2	370	613	235	0.062	0.062 (0)	0.060	0.386
17	Word Synonyms	25	267	638	270	0.382	0.252 (8)	0.351	0.426

5. Conclusion and Future Work

In this paper we introduced the EDTW algorithm. It has linear time and space complexity. Efficient DTW uses a two level approach, first level it reduce the dataset and second level calculates limited values in a distance matrix to find distance and wrapping path. It is efficient than DTW, and accelerate data mining functionalities. Technique applied on synthetic dataset and UCR Time series datasets. In future we will apply this technique to existing classification and clustering methods of time series data to test efficiency.

6. References

1. Ratanamahatana CA, Keogh E. Everything you know about dynamic time warping is wrong. International Conference on Knowledge Discovery and Data Mining (KDD); Seattle, WA. 2004 Aug. p. 22–5.
2. Muller M. Information Retrieval for Music and Motion. Springer; 2007.
3. Lia H, Chen X. Unifying time reference f smart card data using dynamic time warping. Procedia Engineering. 2016; 137:513–22.
4. Kumar V. A survey on time series data mining. IJIRCCE. 2014; 2(5):170–9.
5. Gupta L, Molfese D, Tammana R, Simos P. Non-linear alignment and averaging for estimating the evoked potential. IEEE Transactions on Biomedical Engineering. 1996; 43(4):346–56.
6. Keogh E, Pazzani M. Scaling up dynamic time warping for data mining applications. Proceedings of the 6th ACM SIGKDD; Boston, Massachusetts. 2000. p. 285–9.
7. Niennattrakul V, Ratanamahatana C A. Learning DTW Global constraint for time series classification. SIGKDD. 2009 Feb; 1–8.
8. Santosh VC. Spoken digits recognition using weighted MFCC and improved features for dynamic time warping. IJCA. 2012 Feb; 40(3):975–8887.
9. Forman G. An extensive empirical study of feature selection metrics for text classification. J Mach Learn. 2003 Mar; 1289–305.
10. Kim S, Park S, Chu W. An index-based approach for similarity search supporting time warping in large sequence databases. Proceedings of 17th International Conference on Data Engineering; Heidelberg, Germany. 2001. p. 607–14.
11. Stan S, Philip C. Fast DTW: Toward Accurate Dynamic Time Warping in Linear Time and Space, Intelligent Data Analysis archive. 2007 Oct; 11(5).
12. Arun KP. Data Mining Technique. University press; 2001.
13. Muller M. Dynamic Time Warping. Information Retrival for Music and Motion. 2007; 69–84.
14. Yanping C, Keogh E, Bing H, Nurjahan B, Anthony B, Abdullah M, Gustavo B. The UCR Time Series Classification Archive. 2015.