International Journal on
Advanced Science
Engineering
Information Technology

Vol.9 (2019) No. 5
ISSN: 2088-5334

Modified Dynamic Time Warping for Hierarchical Clustering

Mahmoud Sammour[#], Zulaiha Ali Othman[#], Amalia Mabrina Masbar Rus, Rosmayati Mohamed[+]

[#]*Center for Artificial Intelligence Technology, Faculty Information Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor Darul Ehsan, Malaysia*
E-mail: #[1]mahmoud.samour@gmail.com; [#2]zao@ukm.edu.my

[*]*Universitas Syiah Kuala, Darussalam, Banda Aceh 23114, Indonesia*
E-mail: E-mail: amaliammr@gmail.com

[+]*Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia*
E-mail: rosmayati@umt.edu.my

***Abstract*—Time series clustering is the process of grouping sequential correspondences in similar clusters. The key feature behind clustering time series data lies on the similarity/distance function used to identify the sequential matches. Dynamic Time Warping (DTW) is one of the common distance measures that have demonstrated competitive results compared to other functions. DTW aims to find the shortest path in the process of identifying sequential matches. DTW relies on dynamic programming to obtain the shortest path where the smaller distance is being computed. However, in the case of equivalent distances, DTW is selecting the path randomly. Hence, the selection could be misguided in such randomization process, which significantly affects the matching quality. This is due to randomization may lead to the longer path which drifts from obtaining the optimum path. This paper proposes a modified DTW that aims to enhance the dynamic selection of the shortest path when handling equivalent distances. Experiments were conducted using twenty UCR benchmark datasets. Also, the proposed modified DTW result has been compared with the state of the art competitive distance measures which is based on precision, recall and f-measure including the original DTW, Minkowski distance measure and Euclidean distance measure. The results showed that the proposed modified DTW reveal superior results in compared to the standard DTW, either using Minkowski or Euclidean. This can demonstrate the effectiveness of the proposed modification in which optimizing the shortest path has enhanced the performance of clustering. The proposed modified DTW can be used for having good clustering method for any time series data.**

***Keywords*—hierarchical clustering; dynamic time warping; distance measures.**

I. INTRODUCTION

Nowadays, time series has been emerged due to technological advances in various domain [1], such as the climate change domain that used a multi-sensor device to observe the changes in climate condition hourly. Therefore, time series can be defined as a collection of chronological observations that recorded variables over time [2], [3]. Furthermore, the suitable data mining method is needed to extract the exciting pattern from time-series data. One of the techniques that can be used is clustering.

In data mining, clustering time series data aims to identify sequential correspondences among two-time sequences [4], [5]. Clustering can be performed using two main approaches; partitioning (e.g., k-means clustering) and hierarchical clustering (e.g., agglomerative clustering) [6]. Each clustering technique is integrated with a particular similarity (distance) measure that could identify similarity among the objects [7]. Furthermore, finding an appropriate similarity measure that is suitable for the clustering technique is a challenging task [8]. Several similarity functions have been proposed to integrate with clustering techniques such as Minkowski, Euclidean and Dynamic Time Warping (DTW). However, DTW has been widely used for time series data due to its ability to identify sequential correspondences.

DTW is one of the popular functions that have demonstrated competitive results compared to other similarity measures. DTW aims to find the shortest path in the process of identifying sequential matches. DTW relies on dynamic programming to obtain the shortest path where the smaller distance is being computed. However, in the case of equivalent distances, DTW is selecting the path randomly. Hence, the selection could be misguided in such randomization process, which significantly affects the matching quality. This is due to randomization may lead to the longer path which drifts from obtaining the optimum path. Several studies have addressed this problem by proposing specific modifications for the DTW, such as computing a global averaging [9], [10] or computing the shifting estimation [11]. However, these modifications may not fit multiple time series with different lengths [12]. Therefore, this paper proposes a modified DTW that aims to enhance the dynamic selection of the shortest path when handling equivalent distances. To demonstrate the effectiveness of the proposed modification, multiple time series datasets with different lengths, have been used in the experiment. Also, the proposed modified DTW has been compared with three similarity measures, including standard DTW, Euclidean, and Mikowski distance measures.

II. MATERIAL AND METHOD

Many research efforts have proposed enhancement and modification for the Dynamic Time Warping (DTW) to improve its functionality regarding specific domains. Such enhancements have intended to reduce the time consumed, which leads to improving the efficiency, as well as acquiring better accuracy, which leads to improving the effectiveness. The problem of curve matching is considered as a drawback of DTW in which the curves are being sampled as sequences to identify the similarity among two time-series [13]. The authors suggested that such sampling of the curve is directly affecting the quality of the results. In this manner, the authors have proposed a continuous DTW for calculating similarity among curves using exact and approximate matching algorithms. The proposed continuous DTW has been evaluated using signature verification dataset and compared with the traditional DTW. The proposed continuous DTW has outperformed the conventional (i.e., discrete) DTW.

Also, a study proposed a global averaging mechanism for DTW regarding clustering time-series sequences [10]. The authors have concentrated on the drawback of using DTW with pairwise sequence similarity in which two individual sequences are being compared with each other. The main drawback behind the pairwise matching lies in the sensitivity of sequence order. The influence on sequence order has a significant impact on the equality of results. To evaluate the proposed method used twenty UCR benchmark datasets [10]. The results showed superior performance for the proposed method compared to the pairwise matching.

In the context of gene expression, it has been figured out that the gene cannot use traditional DTW because of its significant differences in different time series [11]. Such differences may indicate significant factor rather than a simple time change. Therefore, time-shift estimation appears to be insufficient in terms of identifying similarity among two genes expressions. Therefore, a modified version of DTW has been proposed to overcome both time-shifting estimation and significant indicators detection.

The problem of time-consumption is when using DTW for large-scale time series clustering [14], [29]. A novel approximation for DTW in which the DTW distances can be bound between LB-Keogh (LB) and Euclidian Distance (ED) functions. The key characteristic behind the proposed method lies in the accurate approximation gained by the two bounds LB and ED together. This can be performed by analyzing the best ‘mixing weight’ of the lower and upper bound by sampling a tiny fraction of the true DTW distances. Then, the evaluation process has been performed using UCR benchmark datasets. The results showed a significant reduction regarding consumption when using the proposed approximation of DTW.

A pruning strategy to accelerate the performance of DTW aims to utilize both LB and ED to prune off a large fraction of the expensive distance calculations [15]. Using different domains of time series, the authors have demonstrated the efficiency of the proposed pruning strategy. The problem arises when combining the nearest neighbor classification with DTW for time series data [9]. Such a problem can be represented as the inaccurate centroid resembles in terms of the actual instances. In this vein, the authors have utilized the global averaging mechanism [10] to generate efficient nearest centroid classifiers.

A clustering approach for time series detection community in a social network is proposed [17]. The work has been compared various distance measures, including ED, Infinite Norm, DTW, Person Correlation, Wavelet Transform, and Integrated Periodogram. As a result, DTW performed better than other similarity measures. Finally, DTW in terms of discarding the prefixes and suffixes of two-time series has difficulties [18]. Such a problem can be represented by the warping invariance that is being considered to the body of time series but not for the head and tails. In this vein, the authors have considered more points in the head and tails to be involved in the computation of the DTW to provide more accurate resemble of the time series.

A. *Agglomerative Hierarchical Clustering (AHC)*

AHC can be defined as a bottom-up approach in hierarchical clustering. It started by a single object in a set and merge with another object that has a similar characteristic [5], [6]. This process would be continued until the termination condition achieved and produce a set of the cluster as the endpoint [20], [27]. The result also can be seen as a tree, which can be plotted, as a dendrogram. At each iteration, the two clusters would be measured by a distance. The cluster with the shortest distance would be merge together. In AHC, the use of distance function would be differentiating the different between AHC. Therefore, this paper would be compared to the performance of four-distance function, including Minkowski, Euclidean, standard DTW, and modified DTW as similarity measures.

B. *Euclidean and Minkowski Distance Measures*

Let x_i and v_j be a P-dimensional vector, then the Euclidean distance can be measured [7], [21], [22]:

$$d_E = \sqrt{\sum_{k=1}^{P} \left(x_{ik} - v_{jk}\right)^2} \tag{1}$$

Then, Minkowski distance was introduced as generalization result of Euclidean distance. Minkowski distance can be computed as follows [21]:

$$d_M = \sqrt[q]{\sum_{k=1}^{P}(x_{ik}-v_{jk})^q} \qquad (2)$$

Where q is a positive integer.

C. Dynamic Time Warping (DTW)

DTW has become a popular distance function in time series analysis [21]. The warping mechanism inside DTW aims to align two sequences of time series to identify the correspondences. This two sequence can be arranged in a matrix and start with the bottom left of the matrix. The distance measured would be placed inside each cell to allow the algorithm in comparing the corresponding elements of the two sequences. The algorithm needs to find the shortest distance that including all possible routes in the matrix. The minimum total distances between shows the shortest path. While the total distance can be calculated as a summation of the distances between the single cells on the path, divided by the sum of the weighting function. The weighting function is used to normalize the path length.

For two-time series sequences, Let $S=\{s_1,s_2,\dots,s_i,\dots,s_n\}$ and $T=\{t_1,t_2,\dots,t_j,\dots,t_n\}$. The DTW would representing a matrix for sequence S and T in $n\times m$ form in order to minimize the differences between these series. Then, Euclidean distance would calculate the distance/similarity between s_i and t_j. While, a warping path $P=\{p_1,p_2,\dots,p_k,\dots,p_K\}$ where $\max(m,n)\le K\le m+n-1$ would be one of the elements in DTW function from the matrix that meet three constraints including boundary condition, continuity and monotonicity.

In these cases, the boundary condition constraint needs the warping path to start and finish in diagonally opposite corner cells of the matrix. Where $p_1=(1,1)$ and $p_K=(m,n)$. Then, the continuity constraint would restrict the allowable steps to adjacent cells and finally, the monotonicity constraint forces the points in the warping path to be monotonically spaced in time. The result of minimum distance/similarity between the two series from wrapping the path is on interest. Hence, the DTW can be computed as follows:

$$d_{DTW} = min\frac{\sum_{k=1}^{K}p_K}{K} \qquad (3)$$

D. Modified Dynamic Time Warping (M-DTW)

Let A and B are two-time series that consist of data points as A = $\{PA_1,PA_2,PA_3,PA_4,PA_5,PA_6\}$, and B = $\{PB_1,PB_2,PB_3,PB_4,PB_5,PB_6,PB_7\}$. DTW would establish the similarity matrix between the two time series by computing the Euclidean distance between each corresponding pair of the data points as showed in Table 1.

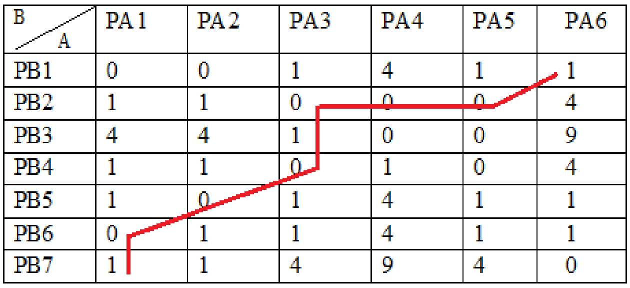
The objective of this matrix is to find the shortest path by starting from the top most right corner $P(PA_6,PB_1)$ and using dynamic programming to reach the point of origin $P(PA_1,PB_7)$ accumulating the sum of the similarity between the went thru points (accumulated path).

TABLE I

SIMILARITY MATRIX

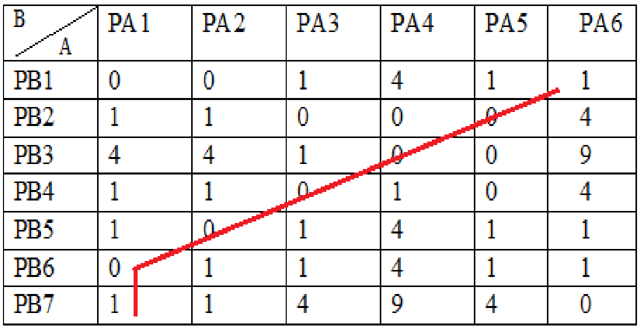
B A	PA1	PA2	PA3	PA4	PA5	PA6
PB1	0	0	1	4	1	1
PB2	1	1	0	0	0	4
PB3	4	4	1	0	0	9
PB4	1	1	0	1	0	4
PB5	1	0	1	4	1	1
PB6	0	1	1	4	1	1
PB7	1	1	4	9	4	0

On $P(PA_6,PB_1)$, DTW would pick the minimum between three points $P(i-1,j)=P(PA_5,PB_1)$, $P(i,j-1)=P(PA_6,PB_2)$ or $P(i-1,j-1)=P(PA_5,PB_2)$. Since $P(PA_5,PB_2)=0$ which is the minimum value thus, it was selected and the accumulated cost was increment by its value. And so on until the point P(PA_1,PB_7) is reached.



B A	PA1	PA2	PA3	PA4	PA5	PA6
PB1	0	0	1	4	1	1
PB2	1	1	0	0	0	4
PB3	4	4	1	0	0	9
PB4	1	1	0	1	0	4
PB5	1	0	1	4	1	1
PB6	0	1	1	4	1	1
PB7	1	1	4	9	4	0

(a) Standard DTW



B A	PA1	PA2	PA3	PA4	PA5	PA6
PB1	0	0	1	4	1	1
PB2	1	1	0	0	0	4
PB3	4	4	1	0	0	9
PB4	1	1	0	1	0	4
PB5	1	0	1	4	1	1
PB6	0	1	1	4	1	1
PB7	1	1	4	9	4	0

(b) Modified DTW

Fig. 1 Standard DTW vs. Modified DTW

On point $P(PA_5,PB_2)$, all the neighboring cells have equivalent values, here the randomization is taking a place in which the movement would be selected randomly as showed in Fig. 1 (a). Similarly, in the points $P(PA_3,PB_2)$ the selection has been performed randomly. In this manner, when handling cells with equivalent values such as in (PA_5,PB_2) and (PA_4,PB_2) the modification would take a place by taking shift using the movement of (PA_i-1,PB_j-1) instead of a random move. This has been depicted in Fig. 1 (b).

As showed in Fig. 1, the summation of the path for the standard DTW can be expressed as$Sum=\{1+0+0+0+1+0+0+0+1\}=3$, whereas for the modified DTW

$Sum = \{1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 1\} = 2$. Apparently, the results of both applications showed variant values. Such difference may have a potential impact on the outcome of the time series clustering. Fig. 2 depicts the pseudo code of the modified DTW.

III. RESULT AND DISCUSSION

This experiment is using a 20-time series benchmark datasets collected from UCR time series data repository [23]. Such a repository contains several time series datasets that are related to multiple domains with different lengths. The evaluation has been conducted based on the distribution of members within the cluster in which the majority would be considered as the class label. The performance evaluation metric such as precision, recall, and f-measure would be used to evaluate the clusters. Such metrics can be computed as follows:

$$Precision = \frac{|TP|}{|TP| + |FP|} \tag{4}$$

$$Recall = \frac{|TP|}{|TP| + |FN|} \tag{5}$$

$$F - measure = 2 \times \frac{P \times R}{P + R} \tag{6}$$

TP is the instance that has been affiliated within a cluster and related to this cluster. FP is the instance that has been affiliated within a cluster but not related, TN is the instance that is not affiliated within a cluster but it is related to it, and finally FN is the instance that is not affiliated within a cluster and it is not related to it.

Algorithm 1. Modified Dynamic Time Warping

```
Input:
$S$ // is a 2-dimensional matrix with $m \times n$
$V_1 = \{s_1, s_2, \ldots., s_n\}$ //is the first vector with $n$ time points
$V_2 = \{s_1, s_2, \ldots., s_m\}$ // is the second vector with $m$ time points
$i, j$ // are loop index, cost is an integer

Steps:
  Initialize the matrix
     S[0,0] = 0;
     FOR i = 1 to m DO LOOP
       S[0 , i ] = ∞
     END
     FOR i = 1 to n DO LOOP
       S[i , 0 ] = ∞
     END
     Populating the similarity matrix
     FOR i = 1 to n DO LOOP
       FOR j = 1 to m DO LOOP
       Cost = d(V1[i], V2[j])//Euclidean distance
       S[i , j] = cost + MINI
       MINI = IF S[I-1, J-1] equals S[i, j-1]
                  MINI = S[i-1, j-1]
             OR if S[i-1, j-1] equals S[i-1, j]
                  MINI = S[i-1, j-1]
             ELSE MINI = Minimum ( S[i-1 ,  j] ,   //increment
                                   S[i ,  j-1] ,   //decrement
                                   S[i-1, j-1] )   //match
       END
     END
  Return S[n , m]
```

Fig. 2 Modified DTW pseudo code

TABLE II
RESULTS OF THE FOUR DISTANCE MEASURES BASED ON PRECISION, RECALL, AND F-MEASURE FOR ALL DATASETS

Dataset		Euclidean			Minkowski			DTW			Modified DTW		
		P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
1	TwoLeadECG	0	0	0	+1	+1	+1	-3	+6	+3	-5	+12	+6
2	OSU Leaf	0	0	0	-4	-2	-3	14	+7	+10	+4	+1	+4
3	Phoneme	0	0	0	0	0	0	-9	+3	+11	-3	+8	+11
4	ArrowHead	0	0	0	0	0	+2	-7	0	0	+16	+1	+8
5	DistalPhalanxTW	0	0	0	+9	+7	+8	10	+1	+3	+5	+12	+10
6	FordA	0	0	0	+2	0	+1	2	+1	+1	+1	+4	+2
7	FISH	0	0	0	+5	-3	0	20	+5	+10	+19	-1	+12
8	50Words	0	0	0	0	-1	-2	-1	+8	+7	+4	+14	+10
9	Beef	0	0	0	-19	+19	+1	2	+4	-1	+1	+19	+7
10	Swedish Leaf	0	0	0	-5	+1	0	6	+7	+8	+12	+10	+12
11	Face (all)	0	0	0	+4	+6	+5	30	+23	+25	+34	+10	+27
12	Computers	0	0	0	+21	0	+8	3	+1	+2	+11	+12	+12
13	Lightning-7	0	0	0	+11	-6	0	1	+9	+9	+10	+18	+17
14	Gun-Point	0	0	0	-4	-2	-3	+12	+8	+10	+14	+4	+6
15	MiddlePhalanxOutlineAgeGroup	0	0	0	+1	+3	-1	+7	17	+9	+6	+7	+7
16	ProximalPhalanxOutlineCorrect	0	0	0	+11	0	+3	+10	30	+20	+10	+21	+16
17	DistalPhalanxOutlineAgeGroup	0	0	0	-10	-20	-19	+20	0	+3	+9	+8	+6
18	BeetleFly	0	0	0	+35	+25	+30	+25	19	+21	+25	+21	+24
19	ProximalPhalanxOutlineAgeGroup	0	0	0	+30	+17	+23	+33	36	+34	+36	+39	+37
20	CBF	0	0	0	-30	-10	-19	-7	6	+1	+11	+31	+22

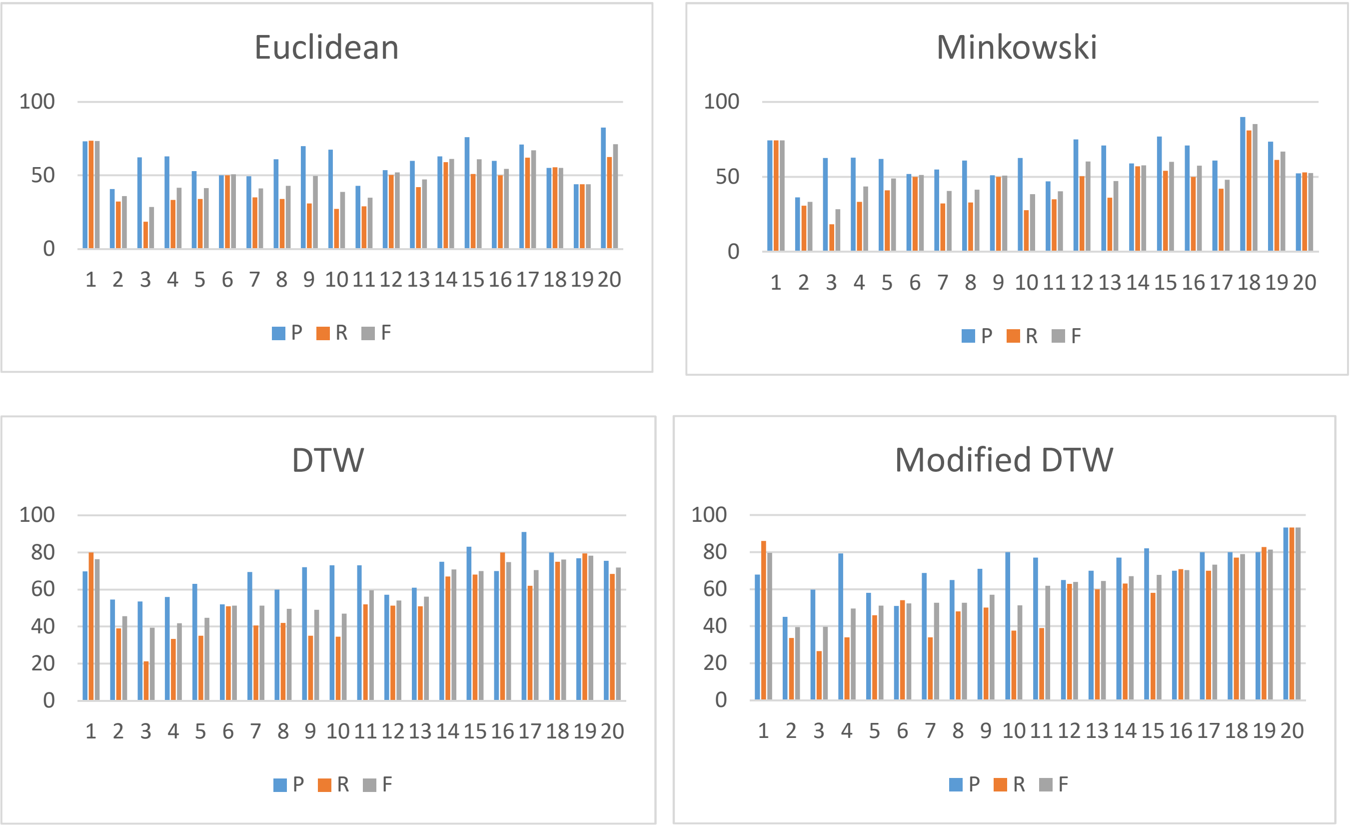


Fig. 3 Results of precision, recall, and f-measure for the four distance measures

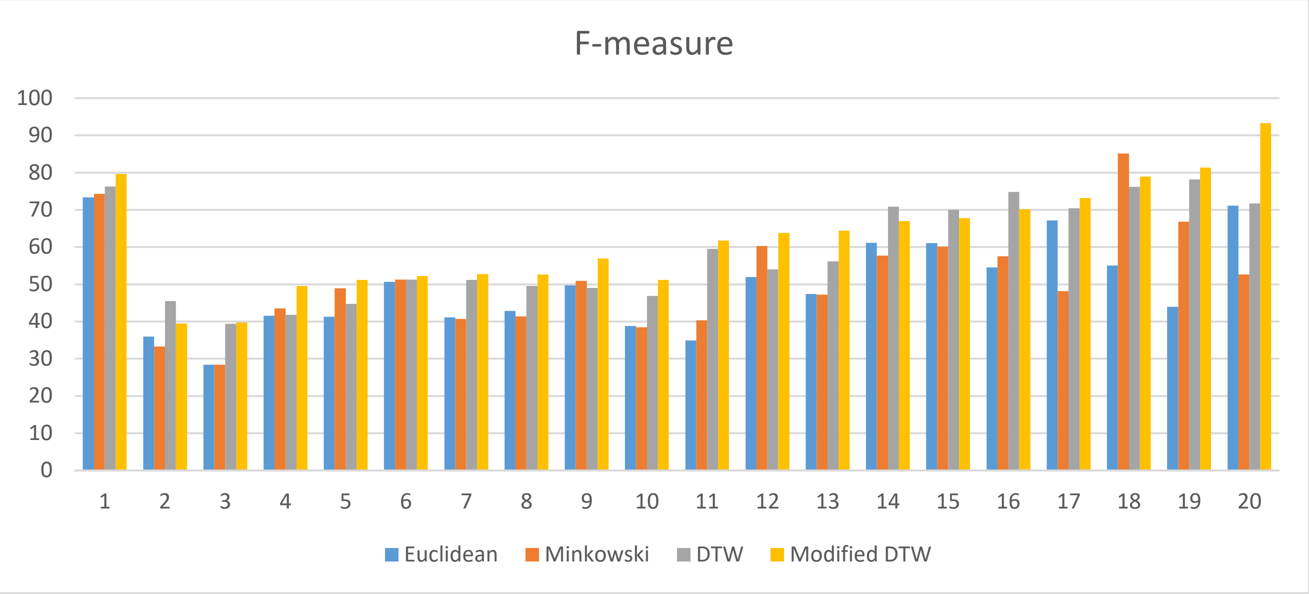


Fig. 4. Comparison of the four distance measures based on f-measure

Based on the mentioned evaluation metrics, Table 2 depicts the results of Minkowski, Euclidean, DTW, and modified DTW for all datasets. As showed in Table 2, for Minkowski, the results were relatively similar to Euclidean except for datasets 18 and 19, where Minkowski has showed relatively higher performance. Vice versa, for datasets 17 and 20 Minkowski showed poor performance compared to Euclidean.

For DTW, the results were relatively higher than Euclidean's results, especially for datasets 11, 16, 18, and 19, where the superiority of DTW was remarkable.

For the modified DTW, the results of all the datasets were better than Euclidean's results, especially for datasets 11, 13, 17, 18, 19, and 20, where the performance has significantly increased. Fig. 3 depicts the performances for each distance measure, respectively, while Fig. 4 shows the comparison among the distance measures based on f-measure. Both figures show the results of precision, recall, and f-measure, which have been computed using Equation (4), Equation (5) and Equation (6).

As showed in Fig 4, The result shows that DTW has outperformed both of Minkowski and Euclidean distance measures in terms of f-measure for 14 out of 20 data sets including '1, 2, 3, 7, 8, 10, 11, 13, 14, 15, 16, 17, 19, 20'. In contrast, Minkowski performs better than other distance functions in terms of f-measure for 6 datasets including '4, 5, 6, 9, 12, 18'. Euclidean distance function has obtained the lowest f-measure values for all datasets.

In contrast between DTW and the modified DTW, DTW has outperformed the modified DTW in terms of F-measure for five datasets including '2, 10, 14, 15, 16'. This is due to the limitation of AHC in which the clusters merging process cannot be undone [24]. In this manner, the modification of DTW would lead to merge irrelevant clusters. Similar to incorrect rejection of true in which the error may lead to better results, the randomization of acquiring the shortest path using DTW would avoid such limitation of AHC.

On the other hand, the modified DTW has outperformed DTW for the remaining 15 datasets. This proves that avoiding the random movement when handling equivalent cells performed by the modified DTW has led to precise results of clustering. Figure 3 shows both performances of DTW and modified DTW. The superiority of DTW compared to Euclidean and Minkowski distance measures was expected from the literature where several clustering approaches have been proposed for time series using DTW for various domains [10], [11], [25], [26].

Multiple comparative analysis studies have been presented to compare different distance measures with clustering technique. A clustering approach for time series community detection in the social network used various distance measures including ED, Infinite Norm, DTW, Person Correlation, Wavelet Transform, and Integrated Periodogram [17]. Among these distance functions, DTW has superior performance compared to other measures.

However, several researchers attempted to modify DTW for instance, a modified DTW in terms of time-consuming [14], and make DTW suitable for specific domain, specifically Gene expressions [11]. Furthermore, a modified DTW is proposed to be able to identify similarity via global averaging mechanism rather than the pair-wise sequence matching [10]. Similar to the mentioned modifications, this study has proposed a modified DTW by optimizing the acquired shortest path. Results showed superior performance compared to the standard DTW.

IV. CONCLUSIONS

This study has proposed a modification of Dynamic Time Warping (DTW) in terms of acquiring the shortest path for hierarchical time series clustering applications using twenty UCR benchmark datasets. The clustering has been performed using Agglomerative Hierarchical Clustering (AHC) using multiple distance measures including Euclidean distance, Minkowski distance, standard DTW and the proposed modified DTW. Results showed that the proposed DTW has outperformed the other distance measures. This can demonstrate the impact of acquiring the shortest path on the clustering results. However, this study has not addressed the impact of the modified DTW on the efficiency thus; future researches could be examine the time and memory consuming using the proposed DTW.

ACKNOWLEDGMENT

This research was funded by Ministry of Higher Education grant (FRGS/1/2016/ICT02/UKM/02/8).

REFERENCES

[1] F. Nasution, N. E. N. Bazin, and A. Zulfikar, "Big Data's Tools for Internet Data Analytics: Modelling of System Dynamics," International Journal on Advanced Science, Engineering and Information Technology, vol. 7, pp. 745-753, 2017.

[2] Y. Lin, K. Ma, R. Sun, and A. Abraham, "Toward a MapReduce-Based K-Means Method for Multi-dimensional Time Serial Data Clustering," in International Conference on Intelligent Systems Design and Applications, 2017, pp. 816-825.doi.

[3] A. Wismüller, O. Lange, D. R. Dersch, G. L. Leinsinger, K. Hahn, B. Pütz, and D. Auer, "Cluster analysis of biomedical image time-series," International Journal of Computer Vision, vol. 46, pp. 103-128, 2002.

[4] N. Alanee and M. A. A. Murad, "A hybrid method of feature extraction and Naive Bayes classification for splitting identifiers," Journal of Theoretical and Applied Information Technology, vol. 95, 2017.

[5] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," Knowledge and information systems, vol. 51, pp. 339-367, 2017.

[6] H. S. Badr, B. F. Zaitchik, and A. K. Dezfuli, "A tool for hierarchical climate regionalization," Earth Science Informatics, pp. 1-10, 2015.

[7] B. Alshaikhdeeb and K. Ahmad, "Integrating correlation clustering and agglomerative hierarchical clustering for holistic schema matching," Journal of Computer Science, vol. 11, pp. 484-489, 2015.

[8] H. Zaifoglu, B. Akintug, and A. Yanmaz, "Regional Frequency Analysis of Precipitation Using Time Series Clustering Approaches," Journal of Hydrologic Engineering, vol. 23, 2018.

[9] F. Petitjean, G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen, and E. Keogh, "Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm," Knowledge and Information Systems, vol. 47, pp. 1-26, 2016.

[10] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," Pattern Recognition, vol. 44, pp. 678-693, 2011.

[11] Y. Yuan, Y.-P. P. Chen, S. Ni, A. G. Xu, L. Tang, M. Vingron, M. Somel, and P. Khaitovich, "Development and application of a modified dynamic time warping algorithm (DTW-S) to analyses of primate brain expression time series," BMC bioinformatics, vol. 12, p. 347, 2011.

[12] B. Alshaikhdeeb and K. Ahmad, "Biomedical Named Entity Recognition: A Review," International Journal on Advanced Science, Engineering and Information Technology, vol. 6, pp. 889-895, 2016.

[13] A. Efrat, Q. Fan, and S. Venkatasubramanian, "Curve matching, time warping, and light fields: New algorithms for computing similarity between curves," Journal of Mathematical Imaging and Vision, vol. 27, pp. 203-216, 2007.

[14] M. Sammour and Z. Othman, "An Agglomerative Hierarchical Clustering with Various Distance Measurements for Ground Level Ozone Clustering in Putrajaya, Malaysia," International Journal on Advanced Science, Engineering and Information Technology, vol. 6, pp. 1127-1133, 2016.

[15] N. Begum, L. Ulanova, J. Wang, and E. Keogh, "Accelerating dynamic time warping clustering with a novel admissible pruning strategy," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 49-58.doi.

[16] Q. Zhu, G. E. Batista, T. Rakthanmanon, and E. J. Keogh, "A Novel Approximation to Dynamic Time Warping allows Anytime Clustering of Massive Time Series Datasets," in SDM, 2012, pp. 999-1010.doi.

[17] L. N. Ferreira and L. Zhao, "Time series clustering via community detection in networks," Information Sciences, vol. 326, pp. 227-242, 2016.

[18] D. F. Silva, G. E. Batista, and E. Keogh, "On the Effect of Endpoints on Dynamic Time Warping," 2016.

[19] L. Zheng, Y. Qu, X. Qian, and G. Cheng, "A hierarchical co-clustering approach for entity exploration over Linked Data," Knowledge-Based Systems, vol. 141, pp. 200-210, 2018.

[20] S. Rani and G. Sikka, "Recent techniques of clustering of time series data: A Survey," Int. J. Comput. Appl, vol. 52, pp. 1-9, 2012.

[21] B. D. Fulcher, "Feature-based time-series analysis," in Feature Engineering for Machine Learning and Data Analytics, ed: CRC Press, 2018, pp. 87-116.

[22] T. W. Liao, "Clustering of time series data—a survey," Pattern recognition, vol. 38, pp. 1857-1874, 2005.

[23] C. Yanping, K. Eamonn, H. Bing, B. Nurjahan, B. Anthony, M. Abdullah, and B. Gustavo, "The UCR Time Series Classification Archive," ed, 2015.

[24] A. R. Bahari, A. Musa, M. Z. Nuawi, Z. I. Rizman, and S. M. Saad, "Novel statistical clustering method for accurate characterization of word pronunciation," International Journal on Advanced Science, Engineering and Information Technology, vol. 7, pp. 1172-1177, 2017.

[25] G. Al-Naymat, S. Chawla, and J. Taheri, "SparseDTW: a novel approach to speed up dynamic time warping," in Proceedings of the Eighth Australasian Data Mining Conference-Volume 101, 2009, pp. 117-127.doi.

[26] V. Niennattrakul and C. A. Ratanamahatana, "On clustering multimedia time series data using k-means and dynamic time warping," in 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07), 2007, pp. 733-738.

[27] Salleh, Zulaiha Ali Othman and Mohamad Salleh,. Characteristics of Agent-based Hierarchical Diff-EDF Schedulability over Heterogeneous Real-Time Packet Networks, European Journal of Scientific Research Sci publication, 2009, Vol 14(3), pp. 431-243.

[28] B. Alshaikhdeeb and K. Ahmad, Integrating Correlation Clustering and Agglomerative Hierarchical Clustering for Holistic Schema Matching, Journal of Computer Science, 2015, Vol 11 (3), pp. 484-489.

[29] M., Islam, M.A. Hannan, H. Basri , A, Hussain and M.Arebey, A Solid waste bin detection and classification using Dynamic Time Warping and MLP classifier, Waste Management , 2014, Vol. 34. pp 281–290.