

Research Article

Multivariate Time Series Data Clustering Method Based on Dynamic Time Warping and Affinity Propagation

Xiaoji Wan ^{1,2}, Hailin Li ¹, Liping Zhang,¹ and Yenchun Jim Wu ^{3,4}

¹School of Business Administration, Huaqiao University, Quanzhou 362021, Fujian

²Oriental Enterprise Management Research Center, Huaqiao University, Quanzhou 362021, Fujian

³Graduate Institute of Global Business and Strategy, National Taiwan Normal University, Taipei 106, Taiwan

⁴Leisure & Recreation Administration Department, Ming Chuan University, Taipei 111, Taiwan

Correspondence should be addressed to Yenchun Jim Wu; wuyenchun@gmail.com

Received 5 April 2021; Accepted 17 May 2021; Published 28 June 2021

Academic Editor: Marta Cimitile

Copyright © 2021 Xiaoji Wan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In view of the importance of various components and asynchronous shapes of multivariate time series, a clustering method based on dynamic time warping and affinity propagation is proposed. From the two perspectives of the global and local properties information of multivariate time series, the relationship between the data objects is described. It uses dynamic time warping to measure the similarity between original time series data and obtain the similarity between the corresponding components. Moreover, it also uses the affinity propagation to cluster based on the similarity matrices and, respectively, establishes the correlation matrices for various components and the whole information of multivariate time series. In addition, we further put forward the synthetical correlation matrix to better reflect the relationship between multivariate time series data. Again the affinity propagation algorithm is applied to clustering the synthetical correlation matrix, which realizes the clustering analysis of the original multivariate time series data. Numerical experimental results demonstrate that the efficiency of the proposed method is superior to the traditional ones.

1. Introduction

Time series is a kind of time-dependent and high-dimensional data type, which can be divided into univariate time series (UTS) and multivariate time series (MTS) in terms of the variable type. With the rapid development of society and economy, such time-axis-sorted data widely exist in finance [1], economy [2], engineering [3], wireless communication [4], and other fields. Owing to the high dimensionality of variable attributes, MTS brings extremely difficulties to clustering, classification, prediction, pattern discovery, visualization, and other mining tasks. Therefore, the research on MTS mining has gained growing interests from scholars. In particular, MTS clustering is the most concerned issue in the field of data mining, from which the feature pattern or data classification of MTS are easily found [5].

In fact, the current academic research on the clustering of time series mainly focuses on the UTS data. Generally, they concentrate on the clustering for the whole time series class,

subseries, time points, etc. [6–8]. The traditional time series clustering methods involve model-based clustering [9, 10], feature-based clustering [11, 12], segmentation-based clustering [13], and distance-based clustering [14, 15]. However, since MTS data have the characteristics of high dimensionality, uncertainty, and dynamics [16], the above methods cannot be effectively adapted to MTS clustering. To reduce the high dimensionality of MTS attributes, both principal component analysis (PCA) [17] and wavelet analysis [18] are commonly adopted to conduct the clustering analysis for MTS. However, most clustering algorithms [18–21] cannot solve nonspherical distribution of MTS data well. In the process of clustering, due to the lengths of most MTSs are different, it is necessary to use the dynamic time warping (DTW) [22–26] to measure the similarity between their corresponding component attributes. But the importance of component attributes has not yet received due attention. To further analyze the features of component attributes, Li et al. [27] apply the affinity propagation (AP) algorithm

[28, 29] to adaptively cluster the component attributes which addressed the shortcomings of other traditional methods in a certain extent. However, their calculation process for the comprehensive relationship matrix is still complex. The quality and efficiency of MTS clustering need to be further improved.

To better solve the asynchronous correlation of different time points in MTS clustering, this paper proposes a simple and high-quality clustering method DTW_AP based on the DTW and AP. DTW is used to measure the similarity of both the component attributes and the overall MTS data. After that, from the perspective of local characteristics and global information, we quickly construct the comprehensive correlation matrix of MTS data. Based on this matrix, we further use the AP algorithm to achieve MTS clustering. Numerical experimental results indicate that the proposed method can achieve better clustering results compared with other traditional methods.

The proposed method possesses some advantages that the current MTS clustering research does not have. These advantages are as follows: (1) The similarity among the components of the MTS data is considered during the clustering process. The accurate of MTS clustering is ensured. (2) With a thorough consideration of the complicated relationships between the local component attributes and the overall information, it retains the multipath nature of MTS information.

The rest of this paper is organized as follows. ‘‘Preliminaries’’ introduces some related theoretical methods, such as the DTW and AP algorithms. In ‘‘Multivariate Time Series Clustering,’’ the detailed construction process of the novel method is described. In ‘‘Numerical Experiment,’’ the comparison analysis among the clustering methods is fulfilled in the experimental evaluation. Finally, the conclusions and future work are drawn in ‘‘Conclusions.’’

2. Preliminaries

2.1. Dynamic Time Warping. Euclidean distance is a common and efficient distance measurement method of time series which can be denoted as

$$d(x, y) = \sqrt{\sum_{i=1}^L \|x_i - y_i\|^2}, \quad (1)$$

where $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_m\}$ are two MTSs, and the dimension of x_i is p , $x_i = (x_i^1, x_i^2, \dots, x_i^p)$, $n = m = L$.

From (1), we know that the Euclidean distance usually needs to be measured by two time series with equal length at the same time point. That is, we regard the square difference between x_i and y_i at the same time point as the contribution of the Euclidean distance. Obviously, its characteristics with equal sequence length, equal time matching, and being sensitive to abnormal data make it impossible to obtain satisfactory results when measuring the similarity between time series data.

Compared with the Euclidean distance, dynamic time warping (DTW) [22–26] is a distance measurement method, in which it firstly calculates the distance matrix D between

time series data and then uses the dynamic programming method to construct the cost matrix R . Finally, based on the matrix R , an optimal bending path W is found so that

$$\text{DTW}(X, Y) = \min_W \sum_{k=1}^K \|w_k\|^2, \quad (2)$$

where $w_k \in W$, K is the number of elements in the optimal bending path W , and $\max(n, m) \leq K \leq m + n - 1$. In addition, $w_k = (i_k, j_k)$ indicates that the path element consists of the i^{th} data point in X and the j^{th} data point in Y .

According to the dynamic programming method, the cumulative cost matrix R can be figured out as follows:

$$R(i, j) = D(i, j) + \min \{R(i, j-1), R(i-1, j), R(i-1, j-1)\}, \quad (3)$$

where $R(0, 0) = 0$, $R(i, 0) = \infty$, and $R(0, j) = \infty$.

Obviously, the cumulative distance is $R(n, m)$, that is, $\text{DTW}(X, Y) = R(n, m)$. Meanwhile, a corresponding dynamic bending path W is also found.

Using DTW to measure the distance of time series data not only matches the data points with the same shape but also measures the similarity between time series data with unequal length. In addition, it can also better solve certain sensitive problems caused by the abnormal data points in the Euclidean distance [22, 23]. However, since the time and space complexity of DTW is higher than that of the Euclidean distance, it will result in poor efficiency when directly applying DTW to measure the similarity between high-dimensional time series data. To solve this problem, many improved methods are proposed by scholars. Salvador and Chan [30] presented a fastDTW based on the search range limitation and data abstraction. On the basis of the penalty coefficient, Li and Du [31] proposed an improved DTW algorithm (γ -DTW) to effectively measure the similarity among an asynchronous approximation of the morphology. To promote the quality of DTW, Li [32] designed a novel DTW with time weight in measuring the distance between time series.

In this paper, to reduce the impact of high dimensionality of variable attributes on the clustering, we will utilize the traditional DTW method to measure the similarity among each component of MTS.

2.2. Affinity Propagation. The classical K means algorithm randomly chooses the initial center points, and certain local optimal solutions are easily obtained. To avoid this problem, Frey proposed the affinity propagation (AP) clustering method [33] based on the graph theory. Compared with other algorithms, AP clustering does not limit the number of clusters and does not choose the initial points randomly. It takes each data point as the potential representative point of the candidate class, which effectively reduces the influence of the selection of original clustering centers on the clustering results. In addition, it also has no symmetry requirement for the similarity matrix, and the speed of dealing with multiclass data is also faster. Therefore, AP clustering algorithm can solve non-Euclidean space problems caused by asymmetry

or trigonometric inequalities or large-scale sparse matrix computing issues [28, 34]. Owing to the important functions, AP algorithm is employed to the various research fields, and its performance and efficiency are also improved [35–37]. For instance, Li et al. [35] proposed a new AP algorithm, the adjustable preference affinity propagation (APAP) algorithm, in which the initial value of each element preference $p(k)$ is independently determined on the basis of the data distribution and is automatically adjusted during the iteration process. Experiments verify its better performance comparing with the standard AP algorithm.

As a type of clustering algorithm depending on the information exchange between data points, AP algorithm can locate the optimal set of representative points when the cluster error function is minimized. Each representative point is from a certain point in original data set, and the sum of cumulative information volume between it and other data points reaches the maximum. In this algorithm, there exist two kinds of information between any two data points i and k . They are called the responsibility degree $r(i, k)$ and the availability degree $a(i, k)$, respectively. $r(i, k)$ refers to the information exchanged from the data point i to the candidate representative data point k , which reflects the representative degree of data point k to data point i . And $a(i, k)$ denotes the effectiveness degree in which data point i chooses data point k as its cluster center, which indicates the information transmitted from the candidate cluster center k to data point i .

Given a data set $X = \{x_1, x_2, \dots, x_n\}$, the detailed process of AP clustering algorithm is described as follows.

During the AP clustering process, the purpose of setting the matrices r and a to zero matrix in S1 is to choose each point in data set as the original cluster center. After initializing, each element in the matrices r and a is needed to be updated from S2 to S4. Specifically, to avoid the possible vibration, a damping coefficient λ is introduced to S4. S5 is the condition for the end of algorithm. The corresponding cluster center of each object and the final cluster set can be obtained in S6 and S7, respectively.

In the process of calculation, it is worth noting that the number of iterative cycles is affected by the damping coefficient λ . Generally, they change in opposite directions with each other. When the damping coefficient λ is large, the number of iterative cycles is small, otherwise the opposite. Gui and Yang [38] found that when the range of λ is [0.5, 1], the stability of AP algorithm can be extremely improved. To reduce the probability of the appearance of fluctuations, we take the value of λ to be 0.9 as suggested by Frey and Dueck [33].

3. Multivariate Time Series Clustering

Compared with univariate time series (UTS), multivariate time series (MTS) has high-dimensional attribute variables, which makes it necessary to consider their approximation in clustering. However, to achieve the MTS clustering, some traditional clustering methods usually regard MTS data as the complete data object and select certain appropriate distance measurement methods to calculate their similarity.

Due to the high dimensionality of attribute variables in MTS, the distance measurement method will affect the clustering quality of MTS. To this end, this paper proposes a MTS clustering method based on DTW and AP from the view of univariate attribute series of MTS data.

Given a MTS data set $X = \{X_1, X_2, \dots, X_N\}$, where $X_i = \{X_i^1, X_i^2, \dots, X_i^m\}$ is the i^{th} MTS, X_i^p refers to the p^{th} attribute sequence of the i^{th} MTS, and $p = 1, \dots, m$. DTW is used to measure the following similarity distance between two MTSs X_i and X_j :

$$S_{ij}^0 = -\text{DTW}(X_i, X_j). \quad (4)$$

Obviously, the bending path $W = \{\omega_1, \omega_2, \dots, \omega_K\}$ measured by the minimum bending distance S_{ij}^0 will be obtained, where $\omega_k = (i_k, j_k)$.

Since each dimension of the MTS X_i can be viewed as the UTS, we can compute the similarity between any two MTSs from the data bending path W provided by (4). That is, under the same attribute of different MTS, we can calculate the following similarity of UTS by using the DTW.

$$\begin{aligned} S_{ij}^p &= S(X_i^p, X_j^p) \\ &= -\sum_{k=1}^K (X_i^p(i_k) - X_j^p(j_k))^2. \end{aligned} \quad (5)$$

From (5), we can construct the similarity matrix S^p of the UTS data set under each attribute component.

This paper employs the AP algorithm to cluster the similarity matrices S^0 and S^p constructed by formulas (4) and (5). The clustering results C^j can reflect the relationship R^j among MTSs under the overall information environment and local component. Furthermore, we obtain the relation matrices R^j by the transformation formula $R^j = C2R(C^j)$. Notice that $C^p = AP(S^p)$ and $C^0 = AP(S^0)$ and the element of the relation matrix R^j only include 0 or 1. If two data objects in the final clustering result point to the same representative object, then their relationship is denoted to be 1, otherwise 0.

Based on the above relationship R^j , we design the following comprehensive relationship matrix R .

$$R = R^0 * \sum_{j=1}^m R^j. \quad (6)$$

Regarding the matrix as the similarity matrix, we will further obtain the clustering results by means of AP algorithm. A new clustering algorithm $C = \text{DTW_AP}(X)$ constructed by DTW and AP is described as follows.

Compared with the traditional clustering methods with Euclidean distance, $\text{DTW_AP}(X)$ has better clustering effects. In fact, DTW can measure the similarity of time series with different lengths while not being sensitive to sudden change or outlier of time series. In addition, DTW also can fulfill the asynchronous similarity comparison. Meanwhile,

AP clustering $C = AP(S)$

Input: similarity matrix S between data points, the maximum iteration number cot , damping coefficient λ .

Output: cluster set C .

S1: initialize the responsibility matrix r and the availability matrix a to be $n * n$ zero matrix.

S2: Update the responsibility matrix r .

$$r(i, k) = \begin{cases} s(i, k) - \max_{k' \neq k} \{a(i, k') + s(i, k')\}, & i \neq k \\ s(k, k) - \max_{k' \neq k} \{a(k, k') + s(k, k')\}, & i = k. \end{cases}$$

S3: update the availability matrix a .

$$a(i, k) = \begin{cases} \min \{0, r(k, k) + \sum_{i' \notin \{i, k\}} \max[0, r(i', k)]\}, & i \neq k \\ \sum_{i' \neq k} \max[0, r(i', k)], & i = k. \end{cases}$$

S4: introduce the damping coefficient $\lambda (0 < \lambda < 1)$ to reduce the possible vibration when updating information. Apply the weighted sum of λ times of the last iteration update value and $1 - \lambda$ times of the current information update value to assign each information.

That is

$$\begin{cases} r_{t+1}(i, k) = \lambda \times r_t(i, k) - (1 - \lambda) \times r_{t+1}(i, k) \\ a_{t+1}(i, k) = \lambda \times a_t(i, k) - (1 - \lambda) \times a_{t+1}(i, k). \end{cases}$$

S5: repeat steps S2 to S4. Stop the algorithm when the clustering results tend to be stable or achieve the preset iterative number cot .

S6: construct the matrix g by adding the matrices r and a and further find the corresponding column k where the value of each row i is the maximum. k is the cluster center of object i .

S7: assign the objects with the same cluster center to the same cluster and obtain the finally cluster set C .

ALGORITHM 1

Input: MTS data set $X = \{X_1, X_2, \dots, X_N\}$, where each time series has m -dimensional attribute.

Output: representative object set C of MTS.

S1: obtain the corresponding similarity matrix S_{ij}^0 and optimal bending path W according to the formulas (2) and (4).

S2: based on the optimal bending path W and formula (5), calculate the similarity S_{ij}^p between two MTSs X_i and X_j under the different component attribute.

S3: repeat steps S1 to S2, calculate the similarity between the corresponding component sequences of all MTSs under the different dimensional condition, and further achieve the corresponding similarity matrices S^1, S^2, \dots, S^m . Herein, m is the attribute dimension of MTS.

S4: use the AP algorithm to cluster each similarity matrix and then achieve the clustering results C^p and C^0 under the different component attribute and the perspective of overall MTS, respectively.

S5: transform the clustering results C^p and C^0 to the relation matrix, that is, $R^j = C2R(C^p)$ and $R^0 = C2R(C^0)$. Meanwhile, calculate the synthetic relation matrix R from (6).

S6: employ the AP algorithm to cluster the comprehensive relation matrix R and finally obtain the clustering result C of MTS.

ALGORITHM 2

this paper applies the AP algorithm to cluster the component attribute of MTS. The influence of each component on the clustering is considered in a comprehensive manner. The application of the AP algorithm not only avoids the limitation of the clustering results achieved by the representative point of the initial class but also has no requirement for the symmetry of similarity matrix. When processing the multi-class data, once the operation speed is faster, its performance will be better.

With the help of the effectiveness evaluation, we find that the algorithm proposed by this paper has significantly higher clustering accuracy than other algorithms. Meanwhile, it also achieves the purpose of high accuracy for the high-dimensional data, which proves its component steps to be effective and meaningful. In this paper, the AP clustering algorithm is used twice. The first time is to cluster each com-

ponent and the overall of MTS. The last time is to cluster the comprehensive relationship matrix R . Thus, the achieved clustering results will be more accurate. In addition, it is more universal for MTS with missing data or different length to use DTW as the distance measurement, which can be widely adopted in practical application.

In the method DTW_AP, the AP algorithm solves the problem that the traditional methods are only suitable for data clustering with spherical distribution. Moreover, the application of DTW in the clustering process enhances the clustering effect of asynchronous morphological similarity. In addition, this paper mainly focuses on the effect of each component attribute sequence on the clustering result and achieves MTS clustering under the local component attribute sequence and the overall information environment.

4. Numerical Experiment

To verify the performance of the proposed method, we choose an effective evaluation index and compare it to several other MTS clustering methods when processing various MTS data sets. In the experiments, the computer is Lenovo T420. The CPU is Intel Core i7-2640M, the computer main frequency is 2.8 GHz, and the memory is 4.0 GB. The operation system is Windows 7 (64 bit). The programming environment is MATLAB R2012b.

4.1. Effectiveness Evaluation Index. Due to the existence of different application requirement, the quality of clustering has no uniform standards for a long time. Therefore, it must rely on certain indicators to evaluate the effectiveness of algorithm. The evaluation methods of clustering validity are divided into external effectiveness and internal effectiveness. The former is to judge the validity of clustering results based on the known correct results. By comparing the known correct clustering results with the experimental clustering results, we can get the degree of similarity, which is the accuracy. The latter is a completely opposite method to the former, which usually uses the self-characteristics of original data set to evaluate the results based on the unknown standard clustering results. Since the measurement is labeled by the known clustering results, this paper chooses the former as the valuation index.

Suppose Pre is the external index of measuring the clustering effect and algorithm quality. $Label$ is the correct result, while idx is the experimental result. n is the total number of points in the data set. The evaluation index Pre of clustering validity between $Label$ and idx is defined as

$$Pre = \frac{\sum Label(idx) == Label}{n}. \quad (7)$$

Given that the class label clustered by the AP algorithm is the coordinate of clustering center points in Pre , we only need to figure out the class label value of experimental results. For example, if the class label of data 1 is 5, and the value of data 5 in the real result is also 5, then data 1 should be clustered into the cluster of data 5. Obviously, this clustering proves to be true and records as 1. Similarly, the correct matching is marked by 1, otherwise 0. Obviously, the cumulative sum divided by the value of Pre is between 0 and 1. The closer the value is to 1, the more accurate the clustering result is.

4.2. Clustering Analysis. To demonstrate the performance of the proposed method, ten MTS data sets from the UCR and UCI are selected. They mainly come from the fields of industrial engineering, medical and health, gestures behavior, language pronunciation, and so on. The detailed information of these data sets are shown in Table 1.

As shown in Table 1, the former five MTS data sets have different lengths. For instance, the time dimension of data set AD is from 4 to 93. The last five MTS data sets with the same industry have 6 variables, and their time dimension is 15. However, they contain different numbers of categories and data volume.

TABLE 1: Experimental data sets of MTS.

ID	Name	Variables	Length	Classes	Volume
1	ASL	22	[45-136]	95	1425
2	AD	13	[4-93]	10	2200
3	CMUS16	62	[127-580]	2	29
4	ECG	2	[39-152]	2	100
5	JV	12	[7-29]	9	370
6	LP1	6	15	4	50
7	LP2	6	15	5	30
8	LP3	6	15	4	30
9	LP4	6	15	3	75
10	LP5	6	15	5	100

The traditional clustering methods of time series usually use the feature representation and similarity measurement to process data and further achieve time series clustering based on the designed specific distance measurement method and the classical clustering method. To demonstrate the clustering performance of the method DTW_AP, PCA_AP [19] based on PCA and AP, Kmeans based on the distance measurement, and OAP based on AP clustering for the original data [33] are selected to be compared with it. In addition, our experiment also compares DTW_AP with the permutation distribution clustering (PDC) [20] and the component attributes based affinity propagation clustering for multivariate time series data (cACM) [27]. Different methods are adopted to carry out the clustering analysis on ten different data sets. The experimental results are displayed in Table 2, where the best experimental result in PCA_AP is from the former $P/2$ main ingredient, P represents the attribute dimension of the MTS data set. The bias parameter Pr affecting the number of clustering and the damping factor λ affecting the algorithm convergence in AP clustering algorithm are set to be the median of the similarity matrix and 0.9, respectively [33].

As shown in Table 2, we know that the mean value obtained by DTW_AP is higher than all other methods on ten different data sets. Meanwhile, the standard deviation (SD) value obtained by DTW_AP is also lower than that by OAP, PCA_AP, Kmeans, and PDC. Note that the SD value computed by cACM is lower than that by DTW_AP. However, its stability advantage is not extremely clear. In a word, compared with other MTS clustering methods, DTW_AP is more effective and feasible. Clearly, it is more superior to most traditional AP clustering algorithms based on the whole original data information.

We take the novel method as the benchmark and compare the clustering effect between this method and other traditional methods. The scatter diagrams of clustering results are shown in Figure 1. Obviously, if the data point in the figure is below the gray line, it will indicate that the novel method is better than the traditional method, otherwise the opposite. From the visual comparative analysis of clustering results in Figure 1, we observe that the method DTW_AP achieves better clustering results, compared with other traditional methods on most data sets. In fact, it is easy to find that

TABLE 2: Experimental clustering effectiveness results.

Method	1	2	3	4	5	6	7	8	9	10	Mean	SD
DTW_AP	0.76	1.00	0.93	0.72	0.95	0.80	0.73	0.83	0.89	0.64	0.83	0.12
cACM	0.76	0.99	0.79	0.73	0.95	0.84	0.73	0.83	0.88	0.64	0.81	0.11
OAP	0.51	0.98	0.93	0.81	0.92	0.74	0.73	0.67	0.88	0.57	0.77	0.16
PCA_AP	0.10	0.86	0.72	0.83	0.16	0.64	0.70	0.67	0.76	0.66	0.61	0.26
Kmeans	0.31	0.48	0.59	0.67	0.67	0.52	0.67	0.57	0.77	0.41	0.56	0.14
PDC	0.08	0.10	0.59	0.68	0.31	0.50	0.60	0.67	0.68	0.38	0.46	0.23

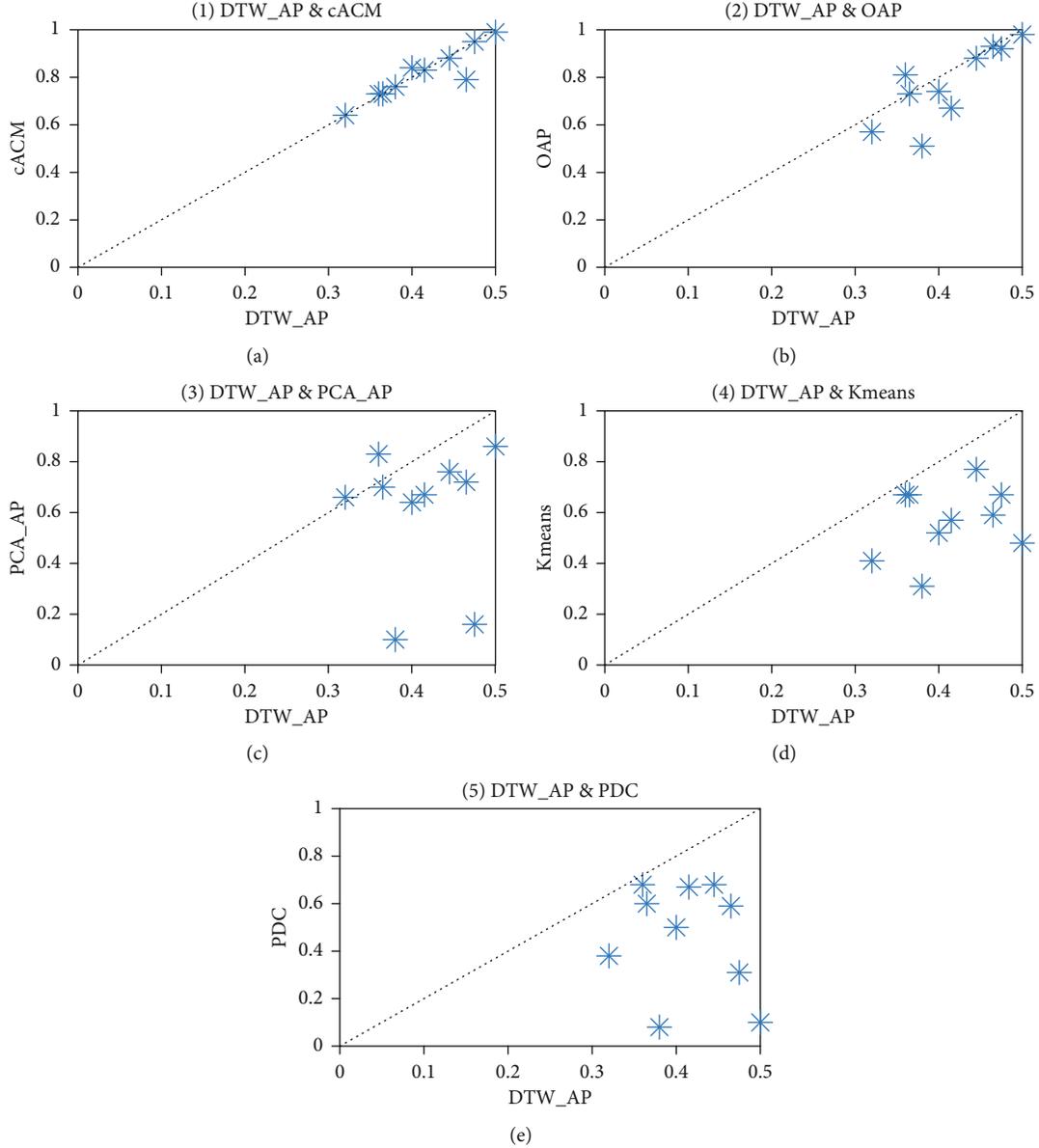


FIGURE 1: Visual comparison of clustering experimental results of different data.

most discrete points in Figure 1 are below the gray line except Figure 1(a), which proves that the novel method has stronger advantages. However, Figure 1(a) generally fluctuates near the gray line, and there is one discrete point deviating from the gray line downward. To a certain extent, it also shows

that the clustering quality of the novel method is better. In particular, from the visual comparison analysis of different methods on all data sets in Figure 2, we find that the ranking of clustering quality is roughly $DTW_AP > cACM > OAP > PCA_AP > Kmeans > PDC$. Meanwhile, based on

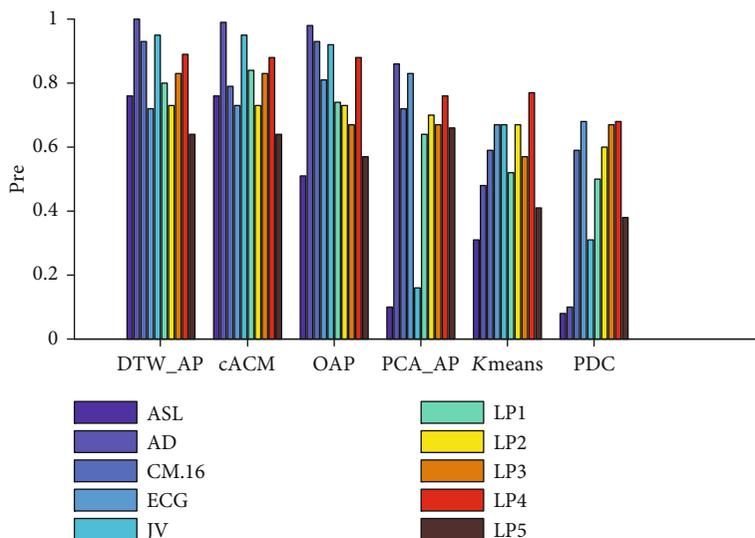


FIGURE 2: Visualization of clustering results of all data sets by different methods.

the average effect and the variance stability of clustering results in Table 2, it also indicates that the clustering performance of DTW_AP is better than other traditional methods.

5. Conclusions

In this paper, a new MTS clustering method is proposed on the basis of DTW and AP. DTW is used to calculate the similarity between the whole information of MTS and parse out the similarity between the component attributes. Meanwhile, the AP algorithm is utilized to construct the clustering relationship between each component attribute and the whole sequence. After that, certain MTS relationship matrices are further presented based on the component attributes and global information. The numerical experimental results indicate that the novel method can achieve better clustering performance when comparing with other traditional clustering methods of time series. However, during the clustering process, the operation using the AP algorithm to cluster the UTS of each component attribute is time-consuming. Therefore, further work should consider improving the time efficiency of the proposed method.

Data Availability

The datasets analyzed in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interests.

Acknowledgments

This study was sponsored by the Huaqiao University's Academic Project Supported by the Fundamental Research Funds for the Central Universities (14SKGC-QG17), Huaqiao University's High Level Talent Research Start-up

Funding Project (14SKBS205), Social Science Planning Project of Fujian Province (FJ2020B088), National Natural Science Foundation (71771094), and Ministry of Science & Technology, Taiwan (MOST 109-2511-H-003-049-MY3).

References

- [1] R. Chandra and S. Chand, "Evaluation of co-evolutionary neural network architectures for time series prediction with mobile application in finance," *Applied Soft Computing*, vol. 49, pp. 462–473, 2016.
- [2] R. A. Davis and L. Song, "Noncausal vector AR processes with application to economic time series," *Journal of Econometrics*, vol. 216, no. 1, pp. 246–267, 2020.
- [3] D. Yang, Z. Dong, L. Lim, and L. Liu, "Analyzing big time series data in solar engineering using features and PCA," *Solar Energy*, vol. 153, pp. 317–328, 2017.
- [4] K. Cao, T. Hu, Z. Li, G. Zhao, and X. Qian, "Deep multi-task learning model for time series prediction in wireless communication," *Physical Communication*, vol. 44, article 101251, 2021.
- [5] Q. Yang and X. Wu, "10 challenging problems in data mining research," *International Journal of Information Technology and Decision Making*, vol. 5, no. 4, pp. 597–604, 2006.
- [6] E. Keogh, J. Lin, and A. Fu, "Hot sax: efficiently finding the most unusual time series subsequence," in *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pp. 226–233, Houston, TX, USA, 2006.
- [7] D. Ye, Y. Wu, and M. Goh, "Hub firm transformation and industry cluster upgrading: innovation network perspective," *Management Decision*, vol. 58, no. 7, pp. 1425–1448, 2020.
- [8] P. Esling and C. Agon, "Time-series data mining," *ACM Computing Surveys*, vol. 45, no. 1, pp. 1–34, 2012.
- [9] C. Matias and V. Miele, "Statistical clustering of temporal networks through a dynamic stochastic block model," *Journal of the Royal Statistical Society*, vol. 60, no. 4, pp. 12–31, 2016.
- [10] A. Cobo, E. R. Rocha, and M. A. Villamizar, "Innovative propensity with a fuzzy multicriteria approach," *Management Decision*, vol. 57, no. 11, pp. 2940–2957, 2019.

- [11] M. S. Yang and Y. Nataliani, "A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 817–835, 2018.
- [12] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 139–156, Munich, Germany, 2018.
- [13] S. Bhati, S. Nayak, K. Murty, and N. Dehak, "Unsupervised acoustic segmentation and clustering using siamese network embeddings," in *Interspeech 2019*, pp. 2668–2672, Graz, Austria, 2019.
- [14] C. Liu, J. Liu, D. Peng, and C. Wu, "A general multiobjective clustering approach based on multiple distance measures," *IEEE Access*, vol. 6, pp. 41706–41719, 2018.
- [15] H. Li, Y. J. Wu, and Y. Chen, "Time is money: dynamic-model-based time series data-mining for correlation analysis of commodity sales," *Journal of Computational and Applied Mathematics*, vol. 370, article 112659, 2020.
- [16] B. T. Pentland, P. Liu, W. Kremser, and T. Haerem, "The dynamics of drift in digitized processes," *MIS Quarterly*, vol. 44, no. 1, pp. 19–47, 2020.
- [17] H. Li, "Accurate and efficient classification based on common principal components analysis for multivariate time series," *Neurocomputing*, vol. 171, pp. 744–753, 2016.
- [18] P. D'Urso and E. A. Maharaj, "Wavelets-based clustering of multivariate time series," *Fuzzy Sets and Systems*, vol. 193, no. 4, pp. 33–61, 2012.
- [19] J. F. Barragan, C. H. Fontes, and M. Embirucu, "A wavelet-based clustering of multivariate time series using a multiscale SPCA approach," *Computers & Industrial Engineering*, vol. 95, no. 5, pp. 144–155, 2016.
- [20] A. M. Brandmaier, "PDC: permutation distribution clustering," *Psychological Methods*, vol. 18, no. 1, pp. 71–86, 2015.
- [21] H. Li, "Multivariate time series clustering based on common principal component analysis," *Neurocomputing*, vol. 349, pp. 239–247, 2019.
- [22] H. Li, "On-line and dynamic time warping for time series data mining," *International Journal of Machine Learning and Cybernetics*, vol. 6, no. 1, pp. 145–153, 2015.
- [23] E. Keogh and M. Pazzani, "Derivative dynamic time warping," in *Proceedings of the 2001 SIAM International Conference on Data Mining (SDM)*, pp. 1–11, Chicago, USA, 2001.
- [24] T. Han, Q. Peng, Z. Zhu, Y. Shen, H. Huang, and N. N. Abid, "A pattern representation of stock time series based on DTW," *Physica A: Statistical Mechanics and its Applications*, vol. 550, article 124161, 2020.
- [25] H. Li and M. Wei, "Fuzzy clustering based on feature weights for multivariate time series," *Knowledge-Based Systems*, vol. 197, article 105907, 2020.
- [26] T. Górecki and M. Łuczak, "Non-isometric transforms in time series classification using DTW," *Knowledge-Based Systems*, vol. 61, pp. 98–108, 2014.
- [27] H. Li, C. Wang, and X. Deng, "Multivariate time series clustering based on affinity propagation of component attributes," *Control and Decision*, vol. 33, no. 4, pp. 649–656, 2018.
- [28] T. Wu, Y. Liang, R. Varela, C. Wu, G. Zhao, and X. Han, "Self-adaptive SVDD integrated with AP clustering for one-class classification," *Pattern Recognition Letters*, vol. 84, pp. 232–238, 2016.
- [29] S. Basloom, N. Akkari, and G. Aldabbagh, "Reducing handoff delay in SDN-based 5g networks using AP clustering," *Procedia Computer Science*, vol. 163, pp. 198–208, 2019.
- [30] S. Salvador and P. Chan, "FastDTW: toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, pp. 561–580, 2007.
- [31] H. Li and T. Du, "Multivariate time-series clustering based on component relationship networks," *Expert Systems with Applications*, vol. 173, article 114649, 2021.
- [32] H. Li, "Time works well: dynamic time warping based on time weighting for time series data mining," *Information Sciences*, vol. 547, pp. 592–608, 2021.
- [33] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.
- [34] Y. Liu, J. Liu, Y. Jin, F. Li, and T. Zheng, "An affinity propagation clustering based particle swarm optimizer for dynamic optimization," *Knowledge-Based Systems*, vol. 195, article 105711, 2020.
- [35] P. Li, H. Ji, B. Wang, Z. Huang, and H. Li, "Adjustable preference affinity propagation clustering," *Pattern Recognition Letters*, vol. 85, pp. 72–78, 2017.
- [36] K. Nabar and G. Kadambi, "Affinity propagation-driven distributed clustering approach to tackle greedy heuristics in mobile ad-hoc networks," *Computers & Electrical Engineering*, vol. 71, pp. 988–1011, 2017.
- [37] Z. Geng, R. Zeng, Y. Han, Y. Zhong, and H. Fu, "Energy efficiency evaluation and energy saving based on DEA integrated affinity propagation clustering: case study of complex petrochemical industries," *Energy*, vol. 179, pp. 863–875, 2019.
- [38] B. Gui and X. Yang, "Research on parameters of affinity propagation clustering," *Advanced Technologies, Embedded and Multimedia for Human-Centric Computing*, vol. 260, pp. 637–644, 2014.