Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: http://www.tandfonline.com/loi/gscs20

A sliding window-based multi-stage clustering and probabilistic forecasting approach for large multivariate time series data

Lei Ren, Yongchang Wei, Jin Cui & Yi Du

To cite this article: Lei Ren, Yongchang Wei, Jin Cui & Yi Du (2017): A sliding window-based multi-stage clustering and probabilistic forecasting approach for large multivariate time series data, Journal of Statistical Computation and Simulation, DOI: 10.1080/00949655.2017.1299151

To link to this article: http://dx.doi.org/10.1080/00949655.2017.1299151

Published online: 05 Mar 2017.

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2017
http://dx.doi.org/10.1080/00949655.2017.1299151

Taylor & Francis
Taylor & Francis Group

A sliding window-based multi-stage clustering and probabilistic forecasting approach for large multivariate time series data

Lei Ren[a,b], Yongchang Wei[a], Jin Cui[a] and Yi Du[c]

[a]School of Automation Science and Electrical Engineering, Beihang University, Beijing, People's Republic of China; [b]Engineering Research Center of Complex Product Advanced Manufacturing System, Ministry of Education, Beijing, People's Republic of China; [c]Department of Big Data Technology and Application Development, Computer Network Information Center, Chinese Academy of Sciences, Beijing, People's Republic of China

ABSTRACT

Time series data analysis, such as temporal pattern recognition and trend forecasting, plays an increasingly significant part in temporal data statistics and analytics. Yet challenges still exist in the efficiency of pattern extracting and trend prediction for large multivariate time series. The paper proposes a multi-stage clustering approach towards multivariate time series by using dynamic sliding time windows. The segmented multivariate time series are clustered separately in each time window to product first-stage clustering centres, and which are used to generate second-stage clustering results involving all time windows. The method can simplify large scale multivariate time series mining problems through multi-stage clustering on multiple sliding time windows thus achieve improved efficiency. Based on the clustering outcomes, a correlation rules mining method is given to discover frequent patterns in the time series and generate self-correlation rules. Then, the paper presents a probabilistic forecasting model that leverages the extracted rules to make short-term predictions. Finally, experiments are presented to show the efficiency and effectiveness of the proposed approach.

ARTICLE HISTORY

Received 27 September 2016
Accepted 21 February 2017

KEYWORDS

Time series; statistics; multi-stage clustering; dynamic sliding time window; time series forecasting model

1. Introduction

A time series is defined as a sequence of data points in successive time order. Most commonly, time series are represented by plots in curves or lines chart. Time series statistics and analysis target at extracting meaningful statistics characteristics to identify the underlying patterns as well as predict future events, which are widely used in signal processing, mathematical finance, earthquake prediction, smart control, and transport trajectory forecasting. With the rising of Big Data era, time series mining and forecasting are playing an increasingly significant role in data statistics and analytics [1,2].

Time series statistics and analysis mainly concern two goals. One is discovering the nature of the phenomenon represented by the temporal data, and the other is making predictions of future values by building forecasting models. Time series patterns identification plays an essential part in achieving these goals. Most methods of time series mining involve similarity measurement, time series clustering, categorical time series analysis [3], association rule mining, and anomaly detection method [4]. Time series patterns mining methods usually vary with different problems that involve specific temporal data with characteristics. Romani et al. [5] proposed a time series mining method for identify association rules between meteorological data and remote sensing information data. The method can

CONTACT Lei Ren renlei@buaa.edu.cn

extract the characteristic patterns of all kinds of data and generate association rules by training. Li and Faloutsos[6] introduced an algorithm for mining feature patterns, which can identify the co-evolutional sequences containing missing values or not by exploiting smoothness and correlation. Yao and Kong [7] developed a mode dependent stock mining method that can be used for patterns discovery of stock data and trends predicting. In the method, a time search model was used to find frequent patterns in stock data by converting the continuous time series data into discrete space. And the model can make short-term trend forecasting of stock prices by using the patterns. Novak and Perfilieva [8] applied soft computing methods to time series data mining, such as fuzzy logic and fuzzy transformation, and used these methods to extract a variety of information, and finally integrated them together by natural language. Regarding the research on time series clustering, Xie et al. [9] presented a clustering method based on Singular Value Decomposition (SVD), which can be used to settle the problem of special structure in time series clustering. This method leverages SVD to obtain memory features of time series and cluster these features by canopy and K-means. Tamura and Miyamoto [10] proposed a two-stage clustering method. In the first stage, the one-pass k-median++ method is used to gain the initial clustering centres, but not iterating. And clustering is carried out in the second stage. Xu et al. [11] made a reduction of the dimensions of feature space for ICU data, and segmented the time series by time segmentation technology. This method was applied to prediction of mortality. Wang et al. [12] intercepted the large scale time series into sub ones, and each sub sequence was represented as a group of coarse granularity fuzzy information. As a result, the original problem can be converted from numeric level to granular level and achieve more effective clustering.

Regarding the time series forecasting models and methods, Ching-Hsue Cheng and Shiu [13] built a time series model based on indicators selection through applying multivariate adaptive regression splines and stepwise regression. Further, a forecasting model was established by Support Vector Regression (SVR) and optimized by GA. Ufuk Yolcu et al. [14] introduced a fuzzy artificial intelligence approach to forecast the trend of time series, and applied it to stock market predicting. Ching-Hsue Cheng and Yang [15] also proposed a rough set rule induction method, which was used to predict the stock price. Winita Sulandari and Yudhanto [16] presented a hybrid time series model based on moving average and weighted fuzzy. The model was used to enhance the accuracy of the trend forecasting of time series. Atilla Aslanargun et al. [17] put forward a series of models consisted of different existing models through various combination ways to forecast time series. Bindu Garg et al. [18] also established a predicting model based on fuzzy time series. The model focused on the influence of trend and seasonal components, and was used to forecast the trend in fuzzy environment. Zhibo Zhu et al. [19] introduced a stock trend analysis method based on feature simplification. The method firstly utilized the local least square method to arrange the feature, and then made the selection of features. And this method can be used to trend predicting of both micro stock data and macro stock market.

Most above methods lay emphasis on extracting time sequence features and transforming continuous time series data to a combination of extracted features. This often leads to data loss in calculation process, resulting in lower effectiveness in describing the nature as well as lower efficiency in forecasting, especially for large multivariate time series. In addition, although clustering is widely used, the pattern computing efficiency still faces challenges in dealing with large multivariate time series. The paper proposes a sliding window-based multi-stage clustering approach towards multivariate time series by using dynamic sliding time windows (DSTW). The segmented multivariate time series are clustered separately in each time window to product first-stage clustering centres, and which are used to generate second-stage clustering results involving all time windows. The method can simplify large scale multivariate time series mining problems by multi-stage clustering on multiple sliding time windows, thus achieve improved efficiency. Based on the clustering outcomes, a correlation rules mining method is given to discover frequent patterns in the time series and generate association rules. Then, the paper presents a probabilistic forecasting model that leverages the extracted rules to make short-term predictions. Finally, the experiments are presented to show the efficiency and effectiveness of the proposed approach.

The rest of the paper is organized as follows. Section 2 presents the multi-stage clustering method using DSTW. Section 3 discusses the forecasting model based on rules by clustering. Section 4 gives experiments that verify the proposed approach. Section 5 concludes the work.

2. Multi-stage clustering method using DSTW for multivariate time series

In this section, we present the multi-stage clustering method for multivariate time series data based on DSTW.

2.1. Fuzzy centres selection method for multivariate time series

We improve the clustering algorithm based on K-means [20] to adapt it to multivariate time series.

The steps of K-means algorithm are described as follows.

(1) Select k elements from data source as initial centres.

$$\mu = \{\mu 1, \mu 2, \mu 3, \ldots, \mu k\}. \tag{1}$$

(2) Calculate the distances between each elements and centres, then divide the elements to corresponding clusters according to the minimum distances.

$$c^{(i)} := \arg\min_{j} \|x^{(i)} - \mu j\|^2 \quad [21]. \tag{2}$$

(3) Recalculate the mean value of each clusters.

$$\mu_j := \frac{\sum_{i=1}^{m} 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)} = j\}} \quad [22]. \tag{3}$$

(4) Calculate the standard measuring function, and the algorithm would be terminated once certain conditions were met such as converging to one point, otherwise the process should return to step (2).

As the clustering objects focus on multivariate time series typically represented by a set of curves rather than points, we improve the algorithm by incorporating features of the multiple temporal curves to adapt it to multivariate time series mining. The improvements mainly include two aspects as follows.

(1) Initial centres selection

Initial centres selection has crucial impact on the results of algorithm. Random selection of initial centres is a commonly used method, but it usually leads to unsatisfied results of clustering. In addition, improper selection of initial centres typically causes superfluous iterations that would cost too much computing time. An optimization strategy is running multiple times and choosing a group of different random centres at each run, then selecting the clusters with minimum quadratic sum of variance. This strategy is simple and the results depend on the number of clusters as well as the size of data set. Another strategy is to select successive initial centre that is farthest from the previous initial selected centres. This method can guarantee that the selected initial centres are randomly scattered, but some of the outliers may be chosen.

We integrate two optimization strategies above and make further improvement. Figure 1 shows the improved process of initial centres selection. The first centre is selected randomly from source data set and added to a special collection used for storing centres. Then in the process of successive

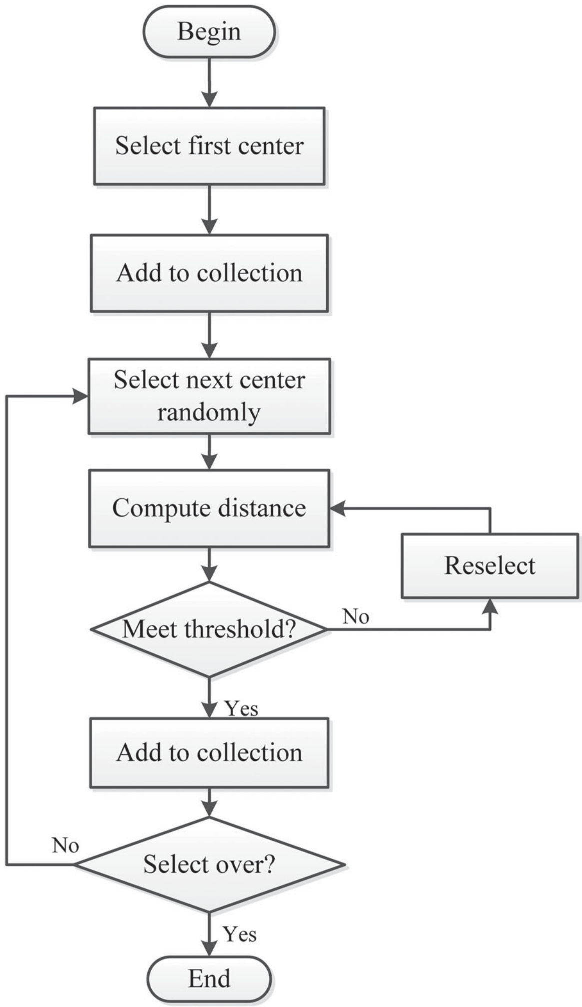

Figure 1. The process of initial centres selection method.

centres selection, once an element is selected randomly, the distance between the element and each existing centres will be calculated. The element could be confirmed as an initial centre and added to the collection if the distance met predefined threshold. Otherwise, the successive centre selection should continue until the collection was filled up. The above process would be running multiple times, and the cluster with minimum quadratic sum of variance should be selected.

(2) Reselection of centres

Since common mathematical methods such as average calculation are not suitable for time series data elements clustering, we present centres reselection method for time series. As shown in Figure 2 is the method which assumes that the initial centres should be selected appropriately with minor deviation in clustering process, and the elements in the cluster could be relatively close to current centre. The method traverses twice the clusters that need reselection of centres, computing the distances between each element and other ones. Once over 50% of the distances are within the scope of preset threshold, the element is chosen as a new centre. If no elements could meet the threshold conditions, the initial centre of this cluster should be regarded as an improper choice, and a new element would be selected randomly as a new centre.

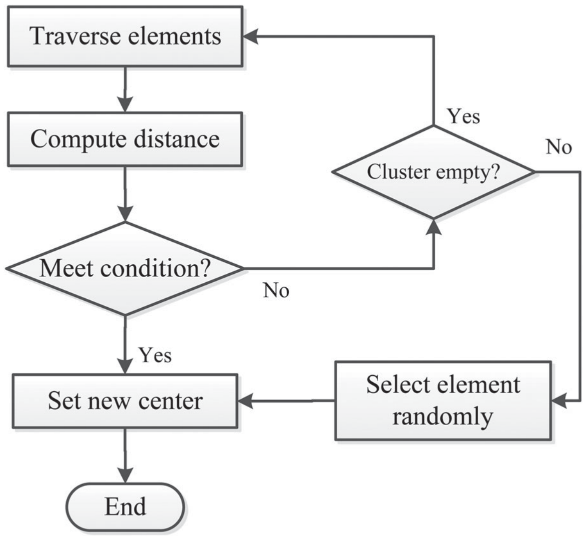

Figure 2. New centre selection method.

2.2. First-stage clustering within time windows

Multivariate time series should be divided into segments before performing clustering in time windows. Based on sliding time windows [23,24], DSTW enable dynamic varying of time windows by configuring the parameters during the segmentation process. Figure 3 shows the principle of DSTW. The key variables involve the size as well as sliding step-length of time windows for data partitioning.

The size configuration of time windows has an essential impact on computational efficiency. Because we use the dynamic time warping (DTW) distance [25,26] to compute the distances [27] between multivariate time series, it is not necessary to make segments with equal length. Actually the

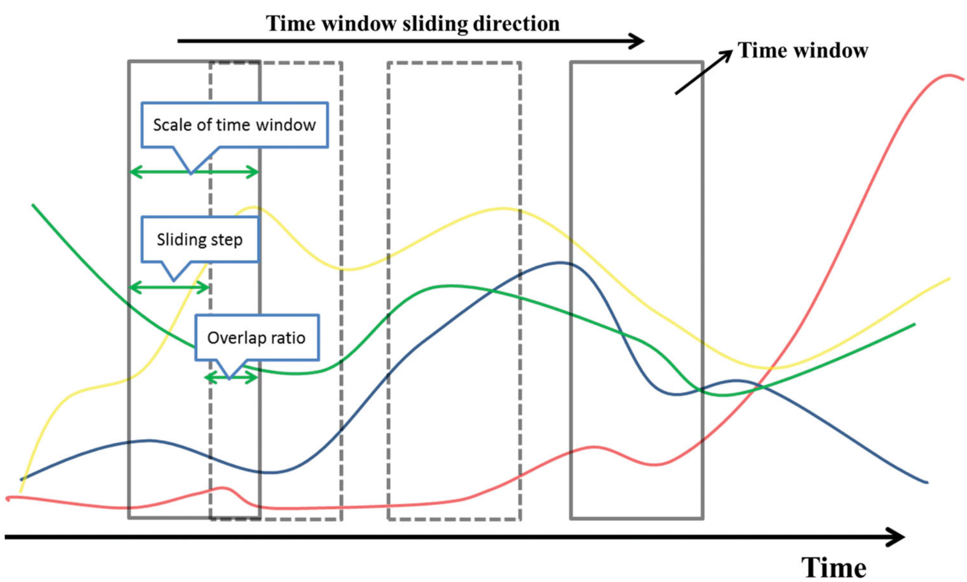

Figure 3. Dynamic sliding time window.

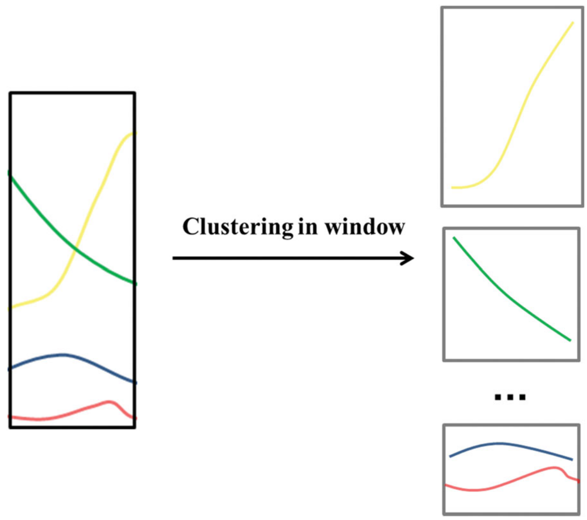

Figure 4. Subsequence clustering in the same period.

overall trend of a segment is more important. Instead, the size of time windows can be set as a variable range [28]. One important thing to note is that too big distance between segments should be avoided because it could not make sense. In our method, the fluctuation range is set to 1–1.3 times of the initial window size. In addition, the sliding step also has critical influence on clustering results. Too small steps could lead to redundant repetitive computation because of overlapping of cross-window time series data sets.

As the parameters are configured, the first-stage inner window clustering will be running. As shown in Figure 4, firstly the inner window data of segmented multivariate time series will be extracted and stored. Then for each collection of multivariate time series data, all segmented time series within the same time window will perform clustering, and each process of clustering within the time window will produce several clusters as shown in Figure 4. As a result, the method will generate a number of independent inner window clusters that cover all time periods.

2.3. Second-stage clustering based on centres

In the second stage, the results of small segmented clusters from the first stage will be aggregated to larger clusters. Firstly, the method gathers the clustering centres of the results in the first stage as the original data set of the second stage. Then second clustering will be running on the data set to generate the clusters of the centres. Further, the corresponding elements of each new cluster centres that appear in the first-stage results will be copied to the corresponding clusters in the second stage to obtain the final clustering results.

Figure 5 shows the overall process of the multi-stage clustering method. The method takes data from the Data Source Pool which contains sub sequences of all time periods. Then these data are organized by same periods separately, and then the first-stage cluster is executed with these data sets by invoking the improved K-means algorithm from the Algorithm Pool. The results are temporarily saved in the Intermediate Result Storage Area which provides service to the second-stage cluster. The centre clustering is carried out with centres extracted from the clusters of the first stage to generate the centre clusters. Finally, the results of the first-stage clustering in the Intermediate Result Storage Area are integrated to the corresponding clusters according to the centres. And the final clusters are stored in the Cluster Result Storage Area.

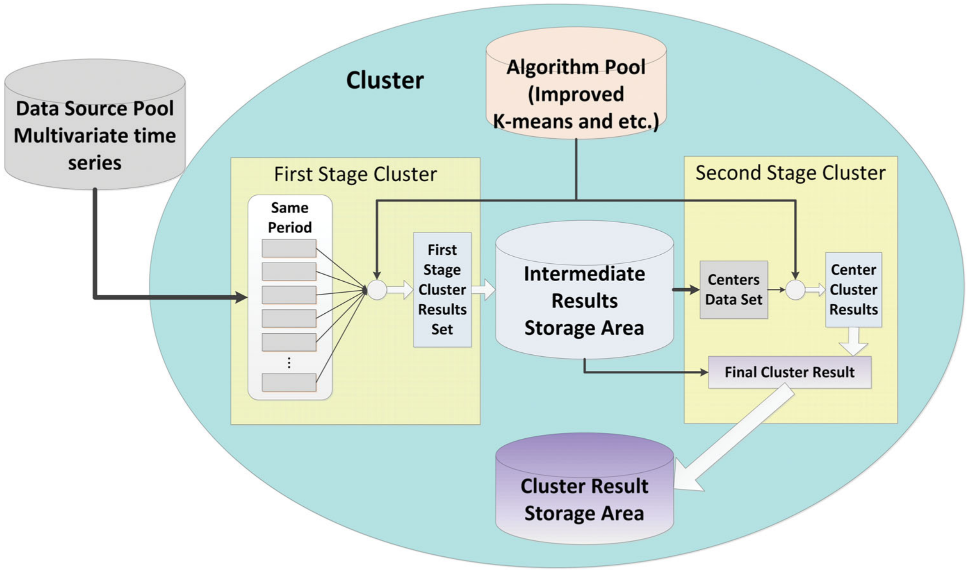

Figure 5. Multi-stage clustering method based on DSTW.

3. Forecasting model based on rules from multi-stage clustering

3.1. Rule base establishment based on clustering results

The self-correlation association rule mining for multivariate time series data is based on rules extraction approach derived from the clustering results. Firstly, the approach gets the first result set *FRset* by pruning the clustering results. Then the approach finds the subsequent elements of the ones in the first result set from the source data, and the acquired data set will be clustered to generate the second result set *SRset*. Finally, the approach computes and discovers the rules, frequently occurred in *FRset* and *SRset*, that represent deductive relationships between *FRset* and *SRset* with confidence. For an item *CenterTS* in *FRset* and an item *SuccessiveTS* in *SRset*, a *Rule* with a confidence is represented as follows:

$$Rule\colon \langle CenterTS \rightarrow SuccessiveTS,\ Confidence\rangle, \tag{4}$$

$$Confidence = Occur(CenterTS,\ SuccessiveTS)/Occur(CenterTS), \tag{5}$$

where $Occur(CenterTS, SuccessiveTS)$ is the frequency of co-occurrence of *CenterTS and SuccessiveTS*, and $Occur(CenterTS)$ is the occurrence frequency of *CenterTS*.

The self-correlation rule mining and rule base establishment process is shown in Figure 6.

3.2. Probabilistic forecasting model of time series based on rule base

Based on the rules pool, we can establish a probabilistic forecasting model to infer probable successive time series by a given time series segment. The forecasting model takes a segment of time series as input, and products several segments of probable time series with different probability as output.

The model is formalized as follows:

$$PredictTSset = RuleMap(CenterTS_i,\ CenterTS_i \in FRset\ and\ Min(DTWdist(InputTS,\ CenterTS_i))). \tag{6}$$

In the model, *InputTS* is the input segment of time series.

PredictTSset is the forecasting time series set. In the set the element is $\langle PredictTS_j, Probability_j\rangle$ where $PredictTS_j$ is a segment of predicted time series with probability $Probability_j$.

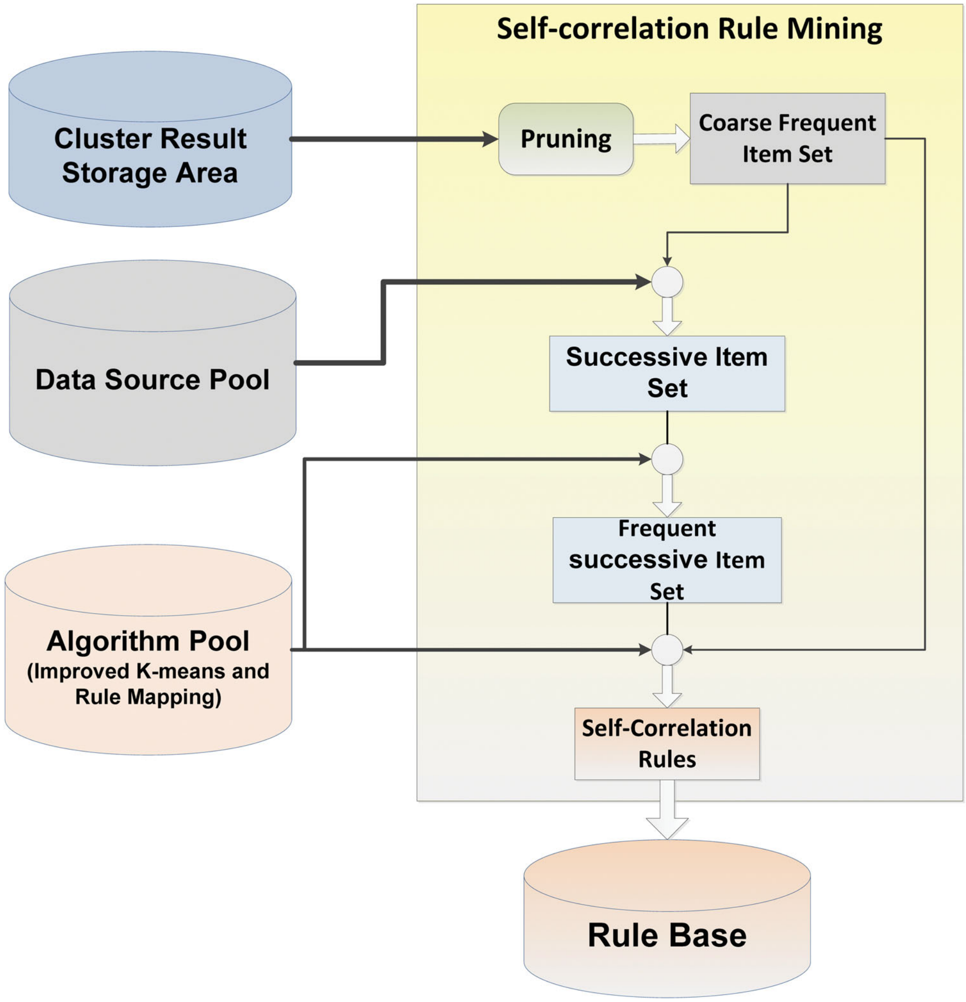

Figure 6. Self-correlation rule mining and pool establishment process.

For each item $CenterTS_i$ in $FRset$, $DTWdist(InputTS, CenterTS_i)$ is used to compute the DTW distances between $InputTS$ and $CenterTS_i$. The $CenterTS_i$ that has the minimum DTW distance to $InputTS$ will be regarded as the $CenterTS$ for a $Rule$. $RuleMap(CenterTS_i)$ is responsible for finding the set of $SuccessiveTS$ according to time series rule base thus generating $PredictTSset$.

The running mechanism of the model is shown in Figure 7.

4. Experiments

4.1. Efficiency of multi-stage clustering method

To verify the multi-stage method, we perform a series of experiments with NASDAQ index data of which the volume is around 400 thousands records of multiple time series involving around 150 stocks spanning 25 years. Figure 8 shows the results of the clustering centres of the time series in which the value k of initial clustering centres is six. Because the multiple time series differ in baseline values and fluctuation ranges, we map all time series data into a normalized number from 0 to 1 by using normalization method. To distinguish the clustering centres more clearly, Figure 9 shows the separated

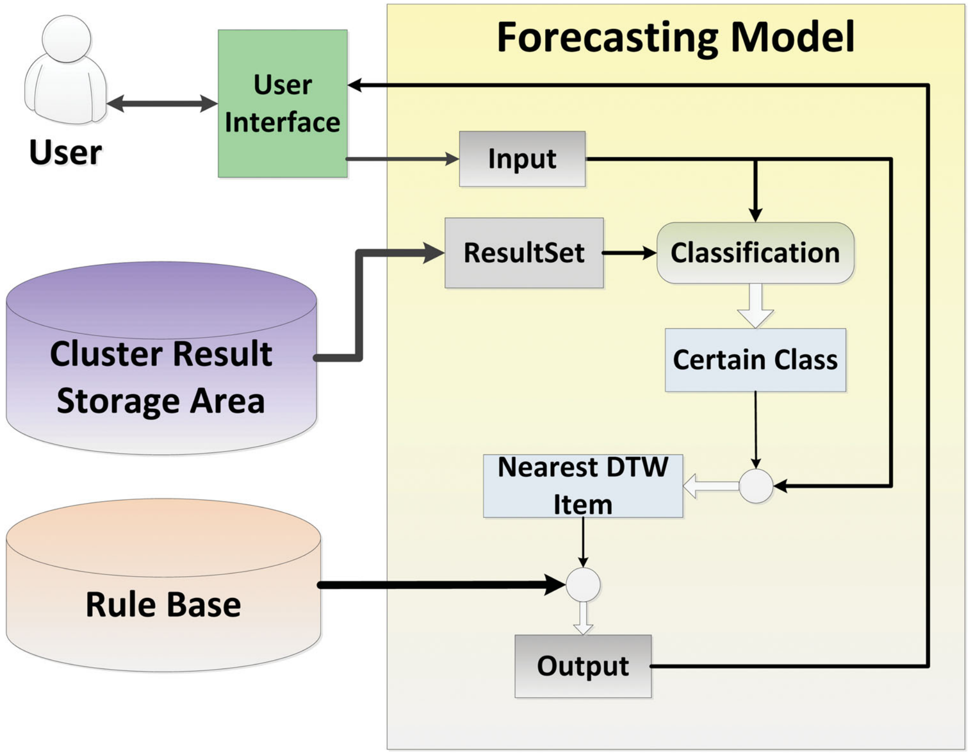

Figure 7. Probabilistic forecasting model of time series based on rule base.

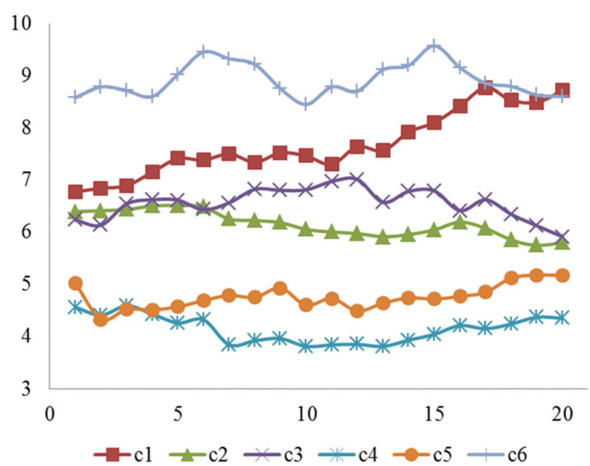

Figure 8. Clustering centres using normalized values ($k = 6$).

clustering centres, and from which we can see that the results conform to the principle of mutual exclusion.

To verify the efficiency of the multi-stage method, we compare the computing time of the proposed method with direct clustering method using the same data set from the NASDAQ data. Since the initial centres selection has an apparent influence on results and computing time, we employ random selection method and perform the experiments several times to calculate the average time as

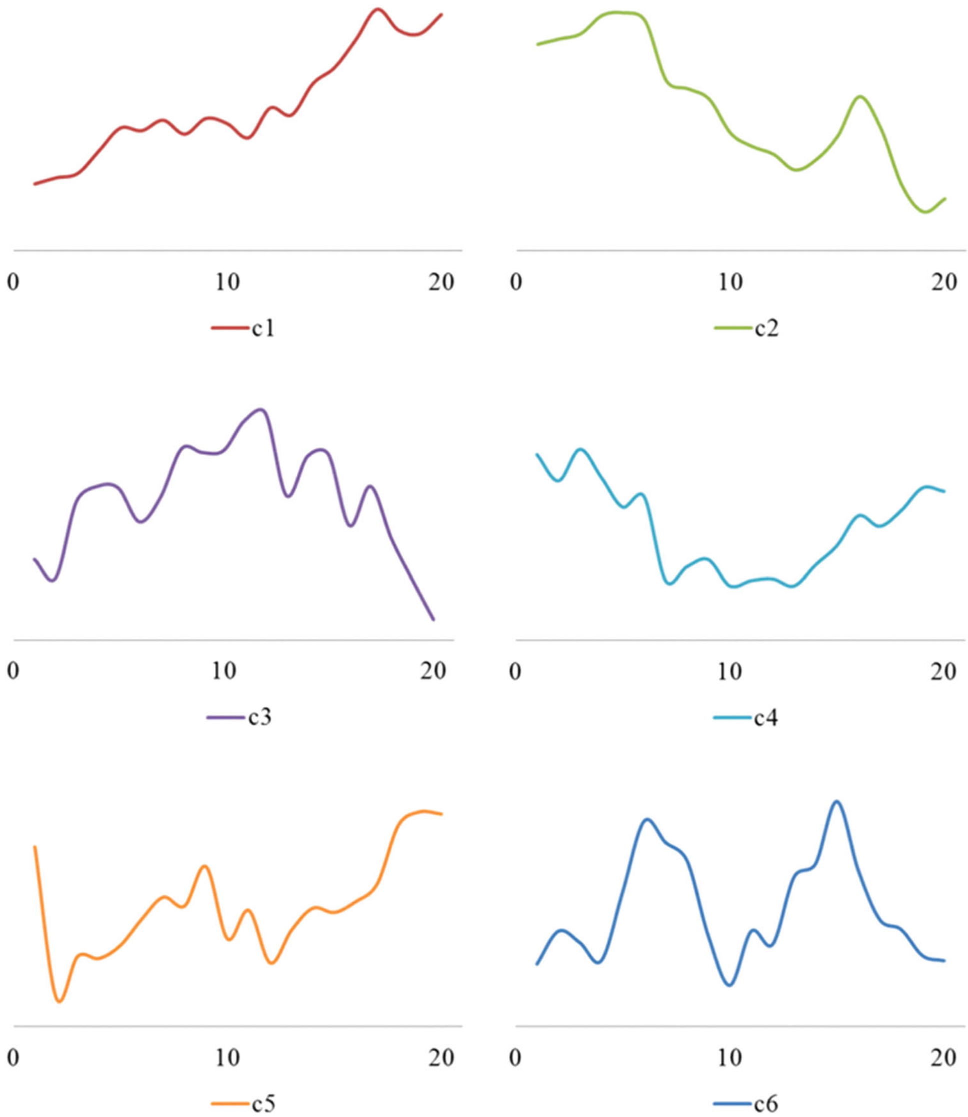

Figure 9. Separated clustering centres.

Table 1. Computing time comparison of two clustering methods.

	1	2	3	4	5	Avg
Direct method	5179	7348	7610	6821	7089	6809
Multi-stage method	656	619	578	630	608	618

algorithm computing time. As Table 1 shown, the average time of traditional direct clustering is 6809 s, while the multi-stage method costs 618 s. Therefore the multi-stage method can greatly improve the efficiency by breaking up the whole into parts.

Since the multi-stage method employs the approach of block computing, the direct relationships between the elements of each blocks might be broken in the first clustering stage and transformed to the indirect relationships of centres computing in the second stage. This may expand the computing errors and lead to low effectiveness for further forecasting. To verify the accuracy especially the trend

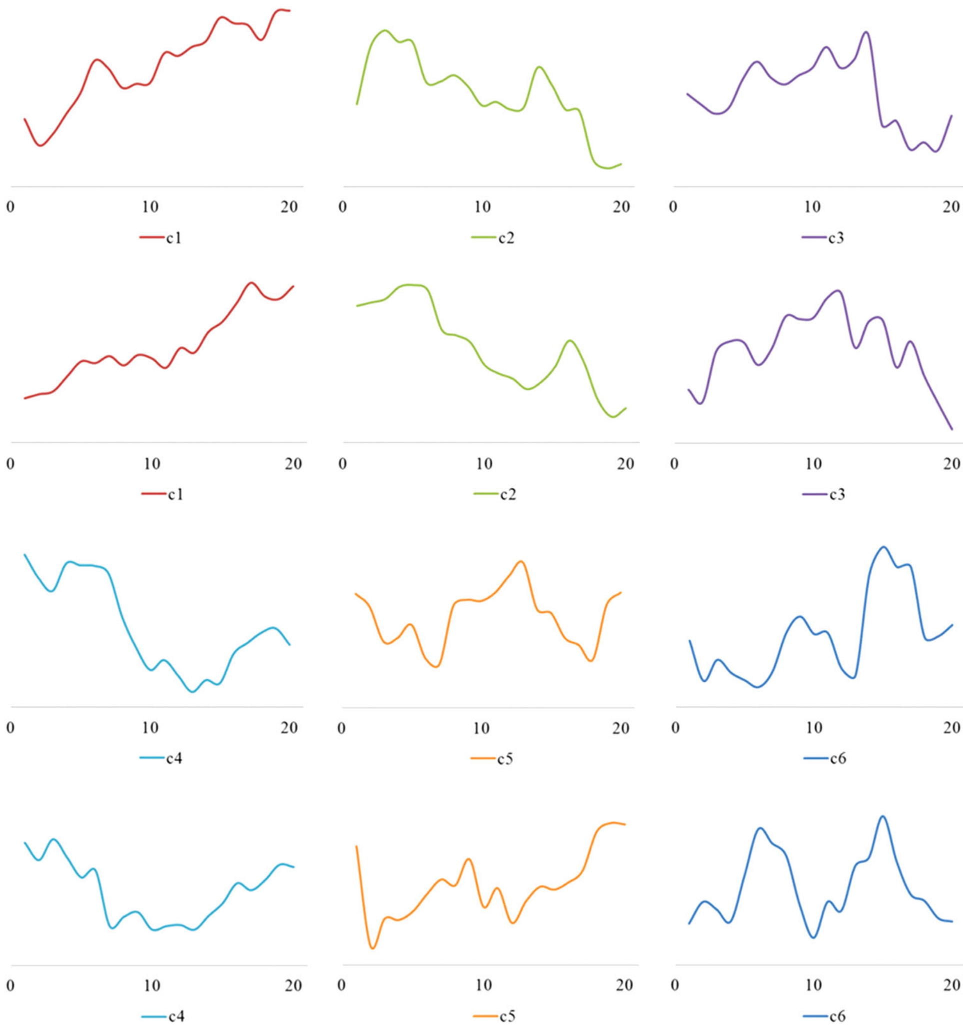

Figure 10. Centres comparison of multi-stage clustering and direct clustering.

consistence of the multi-stage method, we compare the clustering results of the multi-stage method with direct clustering approach.

Figure 10 shows the comparison of clustering centres calculated by the multi-stage method and traditional method. Although there are slight differences between the centres, but from the perspective of the overall trend, the results are consistent and can provide support for further trend forecasting.

4.2. Probabilistic forecasting model experiments

To verify the effectiveness of the proposed forecasting model, we perform experiments using NASDAQ time series data. In the experiment by taking a segment of time series as model input, we verify the trend prediction results by the output of the model.

Figure 11 shows the time series input. As Figure 12 shown, the forecasting model outputs several trend prediction results with different probability as well as different linear trends representing future directions, and among which the output with the highest probability 0.638 might represent the most likely future trend prediction.

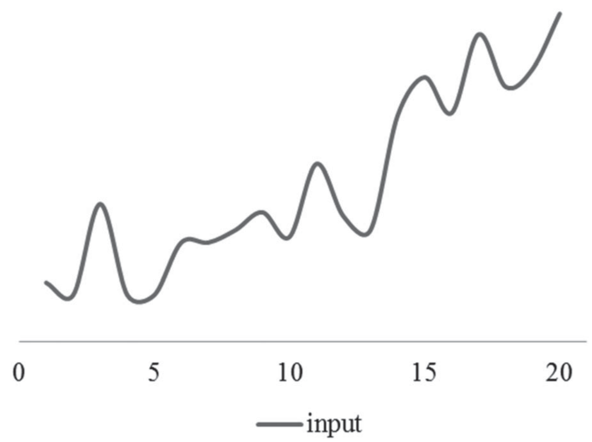

Figure 11. A segment of time series as model input.

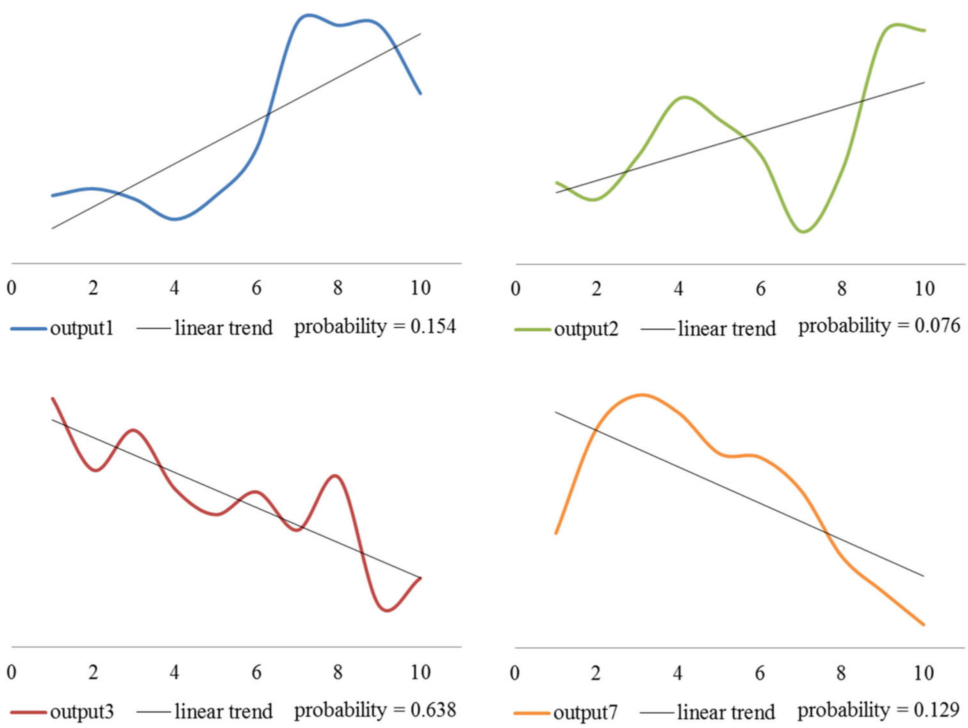

Figure 12. Time series trend prediction outputs with different probability.

Figure 13 shows the actual trend related to the input above, and the linear trend means that the future direction is an overall downtrend. As Figure 12 shown, the model outputs *output3* and *output4* both have downward trends, but among which *output3* has the most similar slope to the actual trend from the perspective of linear trend. In addition, *output3* has the most similar fluctuations to the actual trend. Therefore, the output that has the highest probability 0.638 can be considered as an effective prediction.

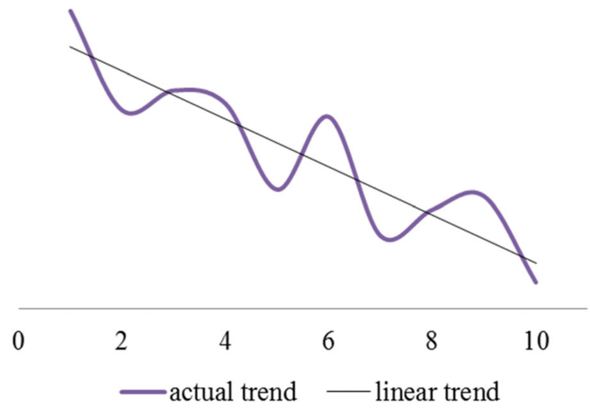

Figure 13. Actual trend related to the input.

5. Conclusion

Multivariate time series pattern mining and trend forecasting is considered as an increasingly crucial area in temporal data statistics and analysis. As frequently used essential approach, large multivariate time series clustering-based methods still face challenges in efficiency. This paper proposes a sliding window-based multi-stage clustering and probabilistic forecasting approach for large multivariate time series. Multivariate time series can be clustered in segment to by using dynamic sliding windows generate the first-stage clustering centres, and based on which we can obtain the second-stage clustering results covering all time windows. The method can simplify large scale multivariate time series mining problems through multi-stage clustering on multiple sliding time windows thus achieve improved efficiency. Based on the clustering outcomes, the paper presents a self-correlation rules mining method used for identifying frequent patterns in the time series and generating association rules. Then, the paper presents a probabilistic forecasting model that leverages the extracted rules to make short-term predictions. The experiment results show that the proposed approach can achieve improved efficiency and can generate effective forecasting results for short-term trend prediction. In future work we are planning to establish a knowledge base used for managing temporal association rules. As the rules increasing, we will improve the accuracy of the time series forecasting model by integrating machine learning methods.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research is supported by the NSFC (National Natural Science Foundation of China) Projects (No. 61572057) in China, the National High-Tech Research and Development Plan of China under [grant no. 2015AA042101].

References

[1] Garima RS. Review on time series databases and recent research trends in Time Series Mining. Paper presented at: Proceedings of the 5th confluence the next generation information technology summit; 2014 Sep 25–26; Noida, India. IEEE; 2014. p. 109–115.
[2] Lee T, Park C. Tests for serial correlation in mean and variance of a sequence of time series objects. J Stat Comput Simul. 2017;87:478–492.

[3] Weiß CH. Empirical measures of signed serial dependence in categorical time series. J Stat Comput Simul. 2011;81(4):411–429.
[4] da Silva PHD, da-Silva CQ. Dynamic changepoint detection in count time series: a particle filter approach. J Stat Comput Simul. 2017;87:42–68.
[5] Romani LAS, De Avila AMH, Chino DYT, et al. A new time series mining approach applied to multitemporal remote sensing imagery. IEEE Trans Geosci Remote Sens. 2013;51(1):140–150.
[6] Li L, Faloutsos C. Fast algorithms for time series mining. Paper presented at: Workshops proceedings of the international conference on data engineering; 2010 Mar 1–6; Long Beach, California, USA. IEEE; 2010. p. 341–344.
[7] Yao J, Kong S. The application of stream data time-series pattern reliance mining in stock market analysis. Paper presented at: Proceedings of IEEE international conference on service operations and logistics and informatics; 2008 Oct 12–15; Beijing, China. IEEE; 2009. p. 159–163.
[8] Novak V, Perfilieva I. Time series mining by fuzzy natural logic and F-transform. Paper presented at: Hawaii international conference on system sciences; 2015 Jan 5–8; Kauai, Hi, USA. IEEE; 2015. p. 1493–1502.
[9] Xie Y, Wulamu A, Wang Y, et al. Implementation of time series data clustering based on SVD for stock data analysis on Hadoop platform. Paper presented at: 9th IEEE conference on industrial electronics and applications; 2014 Jun 9–11; Hangzhou, China. IEEE; 2014. p. 2007–2010.
[10] Tamura Y, Miyamoto S. A method of two stage clustering using agglomerative hierarchical algorithms with one-pass k-means++ or k-median++. Paper presented at: 10th IEEE international conference on granular computing (GrC); 2014 Oct 22–24; Noboribetsu, Hokkaido, Japan. IEEE Computer Society; 2014. p. 281–285.
[11] Xu J, Li D, Zhang Y, et al. CinC challenge: cluster analysis of multi-granular time-series data for mortality rate prediction. Paper presented at: Computing in cardiology (CinC); 2012 Sep 9–12; Krakow, Poland. IEEE; 2012. p. 497–500.
[12] Wang X, Yu F, Zhang H. Granulation-based fuzzy clustering of large-scale time series. Paper presented at: 10th International conference on fuzzy systems and knowledge discovery; 2013 Jul 23–25; Shenyang, China. IEEE; 2013. p. 466–471.
[13] Cheng CH, Shiu HY. A novel GA-SVR time series model based on selected indicators method for forecasting stock price. Paper presented at: International conference on information science, electronics and electrical engineering; 2014 Apr 26–28; Sapporo, Japan. IEEE; 2014. p. 395–399.
[14] Yolcu U, Aladag CH, Egrioglu E, et al. Time-series forecasting with a novel fuzzy time-series approach: an example for Istanbul stock market. J Stat Comput Simul. 2013;83(4):599–612.
[15] Cheng CH, Yang JH. Rough-set rule induction to build fuzzy time series model in forecasting stock price. Paper presented at: International conference on fuzzy systems and knowledge discovery; 2015 Aug 15–17; Zhangjiajie, China. IEEE; 2015. p. 278–284.
[16] Sulandari W, Yudhanto Y. Forecasting trend data using a hybrid simple moving average-weighted fuzzy time series model. Paper presented at: International conference on science in information technology; 2015 Oct 27–28; Yogyakarta, Indonesia. IEEE; 2015. p. 303–308.
[17] Aslanargun A, Mammadov M, Yazici B, et al. Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. J Stat Comput Simul. 2007;77(1):29–53.
[18] Garg B, Beg MMS, Ansari AQ. Fuzzy time series model to forecast rice production. Paper presented at: IEEE international conference on fuzzy systems; 2013 Jul 7–10; Hyderabad, India. IEEE; 2013. p. 1–8.
[19] Zhu Z, Peng Q, Sun S, et al. Stock trend analysis based on feature rank by partial least squares. Paper presented at: IEEE international conference on information and automation; 2015 Aug 8–10; Lijiang, China. IEEE; 2015. p. 3003–3008.
[20] Qiao J, Zhang Y. Study on K-means method based on data-mining. Paper presented at: Chinese automation congress; 2015 Nov 27–29. IEEE; 2015. p. 51–54.
[21] Chen Q, Mo J. Optimizing the ant clustering model based on K-means algorithm. Paper presented at: CSIE 2009, 2009 WRI world congress on computer science and information engineering; 2009 Mar 31–Apr 2; Los Angeles, California, USA. IEEE; 2009. p. 699–702.
[22] Kalam R, Manikandan K. Enhancing K-means algorithm for image segmentation. Paper presented at: International conference on process automation, control and computing (PACC); 2011 Jul 20–22; Coimbatore, India. IEEE; 2011. p. 1–4.
[23] Wang J, Wang L, Wang Y. Computation of coherence cube with variable time window length. Paper presented at: IEEE 11th international conference on signal processing (ICSP); 2012 Oct 21–25; Beijing, China. IEEE; 2012. p. 1397–1400.
[24] Fan J. Dynamic vehicle scheduling problem with changeable time window. Paper presented at: 6th international conference on fuzzy systems and knowledge discovery; 2009 Aug 14–16; Tianjin, China. IEEE; 2009. p. 276–280.
[25] Li G, Wang Y, Li M, et al. Similarity match in time series streams under dynamic time warping distance. Paper presented at: International conference on computer science and software engineering; 2008 Dec 12–14; Hubei, China. IEEE Computer Society; 2008. p. 399–402.

[26] Giao BC, Anh DT. Similarity search in multiple high speed time series streams under dynamic time warping. Paper presented at: Information and computer science; 2015 Sep 16–18; Ho Chi Minh City, Vietnam. IEEE; 2015. p. 82–87.
[27] Jin L. Comparing autocorrelation structures of multiple time series via the maximum distance between two groups of time series. J Stat Comput Simul. 2014;85(17):3535–3548.
[28] Zhang J, Chu Y. The use of sliding time window in the real-time solution of ground-based GPS data. Paper presented at: International conference on electronics and optoelectronics; 2011 Jul 29–31; Dalian, China. IEEE; 2011. p. V4-421–V4-423.