Engineering Applications of Artificial Intelligence 41 (2015) 17–24

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Time series long-term forecasting model based on information granules and fuzzy clustering

Weina Wang [a,b], Witold Pedrycz [c], Xiaodong Liu [a,*]

[a] Research Center of Information and Control, Dalian University of Technology, Dalian City, 116024, PR China
[b] Department of Mathematics, Jilin Institute of Chemical Technology, Jilin 132022, China
[c] Department of Electrical & Computer Engineering, University of Alberta, Edmonton Canada

ARTICLE INFO

Article history:
Received 15 June 2014
Received in revised form
18 September 2014
Accepted 12 January 2015
Available online 18 February 2015

Keywords:
Information granules
Granular time series
Forecasting
Long-term forecasting
Dynamic time warping

ABSTRACT

In spite of the impressive diversity of models of time series, there is still an acute need to develop constructs that are both accurate and transparent. Meanwhile, long-term time series prediction is challenging and of great interest to both practitioners and research community. The role of information granulation is to organize detailed numerical data into some meaningful, semantically sound entities. With this regard, the design of time series forecasting models used the information granulation is interpretable and easily comprehended by humans. In order to cluster information granules, a modified fuzzy c-means which does not require that data have the same dimensionality is proposed. Then, we develop forecasting model combining the modified fuzzy c-means and information granulation for solving the problem of time series long-term prediction. Synthetic time series, chaotic Mackey–Glass time series, power demand, daily temperatures, stock index, and wind speed are used in a series of experiments. The experimental results show that the proposed model produces better forecasting results than several existing models.

1. Introduction

The analysis of temporal data and the prediction of future values of time series are among the most important problems that data analysts have been facing in many fields, ranging from finance and economics (Bodyanskiy and Popov, 2006; Kang, 2003; Chen and Wang, 2010; Chen and Chen, 2011), to production operations management or telecommunications (Lendasse et al., 2002; Mastorocostas and Hilas, 2012). Different time series models have been proposed, including traditional and fuzzy methods. Traditional time series forecasting, such as statistics and neural networks, is usually highly dependent on historical data, which can be incomplete, imprecise and ambiguous. These uncertainties are likely to be widespread in real-world data and hinder forecasting accuracy, thus limiting the applicability of these models. Unlike traditional time series forecasting approaches, the fuzzy approach is capable of dealing with vague and incomplete time series data under uncertain circumstances. Castillo and Melin (2002) proposed the definition of fuzzy fractal dimension and developed hybrid intelligent systems combining neural networks, fuzzy logic, and the fractal dimension, for problem of time series prediction. Chen and Chang (2010) presented a method for multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques. Song and Chissom (1993a, 1993b, 1994) pioneered the study of fuzzy time series(FTS), in which temporal data are represented as linguistic values rather than numeric ones. Since its emergence, numerous studies have been devoted to improving forecasting performance and efficiency following the steps of Song and Chissom's forecasting framework, and these have resulted insignificant achievements. Egrioglu et al. (2011) used an optimization technique with a single-variable constraint to determine an optimal interval length in high order fuzzy time series models. Chen and Chen (2011) proposed a new method to forecast the TAIEX based on fuzzy time series and fuzzy variation groups. Wang et al. (2013, 2014) used clustering and the concept of information granules to determine temporal intervals of unequal length in fuzzy time series model. Lu et al. (2014) proposed the modeling approach to realize interval prediction, in which the idea of information granules and granular computing is integrated with the classical Chen (1996)'s method.

Most of the above studies involve one-step-ahead forecasting models (single point prediction). Nevertheless, there is an increasing need for long-term forecasting going many time steps in advance, which is difficult to achieve because information is unavailable for the unknown future time steps (Simon et al., 2005). Meanwhile, in the plethora of currently available models

* Correspondence to: Research Center of Information and Control, Dalian University of Technology, No.2, Linggong Road, Ganjingzi District, Dalian City, Liaoning Province, 116024, PR China. Tel.: +86 41184709380; fax: +86 41184707579.
E-mail addresses: wangweina406@sina.com (W. Wang), wpedrycz@ualberta.ca (W. Pedrycz), xdliuros@dlut.edu.cn (X. Liu).

http://dx.doi.org/10.1016/j.engappai.2015.01.006
0952-1976/

of time series, their accuracy has been a holy grail of the overall modeling for a long time. With the emergence of more visible and well-justified need for interpretable models that are easily comprehended by humans, arose an important need to develop models that are not only accurate but transparent as well.

In order to satisfy the above-mentioned needs, a time series long-term forecasting model based on information granules and fuzzy clustering is proposed. The role of information granulation is in the organization of detailed numerical data into some meaningful and operationally viable abstract knowledge, which makes the interpretation of data easier and more transparent as well as becomes helpful navigate through various levels of abstraction/specificity by adjusting sizes of information granules used in the description (Pedrycz and Vukovich, 2001). Therefore, the design of time series forecasting models that used the information granulation is interpretable and easily comprehended by humans. The modified fuzzy c-means based on dynamic time warping is proposed to cluster the granules and extract the fuzzy logical rules. Then, we determine the weight of each fuzzy rule with respect to the input observation and use such weights to determine the predicted output based on the multiple fuzzy rules interpolation scheme. The advantages of the proposed model can be summarized as follows:

- Information granulation is used to design the time series forecasting model, which makes the model interpretable and easily comprehended by humans.
- Many well-known distance-based clustering algorithms (e.g., k-means, fuzzy c-means) require data of the same dimensionality. In order to avoid this drawback, a modified fuzzy c-means based on dynamic time warping is proposed.
- The predicted multiple values can be done in one step instead of iteratively predicting each value separately. The proposed model can simplify the forecasting problem and reduce computational overhead of modeling.

An illustrative example for forecasting a synthetic time series is used to verify the effectiveness of the proposed model. Moreover, we conduct the experiments on the chaotic Mackey–Glass time series, power demand, daily temperatures, stock index, and wind speed. The experimental results show that the proposed model produces better forecasting results than those provided by several existing models.

The paper is organized as follows: Section 2 introduces the principle of justifiable granularity and presents a formation of information granules for given time series. Section 3 proposes a new fuzzy clustering algorithm by combining fuzzy c-means clustering and dynamic time warping. Section 4 proposes a time series long-term forecasting model based on information granules and fuzzy clustering. In Section 5 we compare the forecasting results of the proposed model with the results of the existing models. The conclusions are covered in Section 6.

2. Granular time series

The concept of information granule is first proposed by Zadeh (1979). Pedrycz and Vukovich (2001) introduced a model of generalization and specialization of information granules. Information granulation can split the problem into several manageable subproblems for which we are in a position to produce effective solutions. The formation of the information granules is realized as a compromise between two intuitively compelling requirements–*Justifiable granularity* and *Semantic meaningfulness*. For numeric data, the requirement of *Justifiable granularity* is quantified by counting the number of data falling within the bounds of the granule, and the requirement of *Semantic meaningfulness* is quantified by the length of the granule.

Information granulation play a pivotal role and give rise to granular models of time series or granular time series. Granular time series models offer a new, highly user-centric perspective at description of temporal data. Information granules are tangible and easily interpretable entities, which help perceive, quantify and interpret the data.

Consider a time series $\{x_1, x_2, \ldots, x_N\}$ shown in Fig. 1. A convincing description is realized as a sequence of information granules of magnitude of the time series where each information granule spreads over some time interval (time slice) $T_1, T_2, \ldots, T_p$ where "p" denotes a number of time slices predefined in advance. For all samples of the time series falling within the temporal window T_i, an information granule of the magnitudes of the time series is formed by invoking the principle of justifiable granularity, see Fig. 2. More formally, the principle of the justifiable granularity G is applied to the finite set of data $\{x_1, x_2, \ldots, x_N\}$ and the information granule $\Omega_\alpha^i = G(\{x_1, x_2, \ldots, x_N\}, T_i, \alpha)$ is realized for a certain predefined value of α.

For each temporal window T_i, the volume of the associated information granule $\mathrm{Vol}(\Omega_i)$ is computed by taking a product of the integral of Ω_i (its sigma count) and the length of the temporal window, that is

$$\mathrm{Vol}(\Omega_i) = T_i \int_{x_{min}}^{x_{max}} \Omega_i(z)\, dz, \tag{1}$$

where the bounds of the integration x_{min} and x_{max} are, respectively, minimal and maximal values of the amplitudes of the time series recorded in the temporal window T_i.

The intent is to arrive at information granules $\Omega_1, \Omega_2, \ldots, \Omega_p$ that are the most "*informative*" (compact) so that they carry a clearly articulated semantics. This gives rise to the optimization problem of

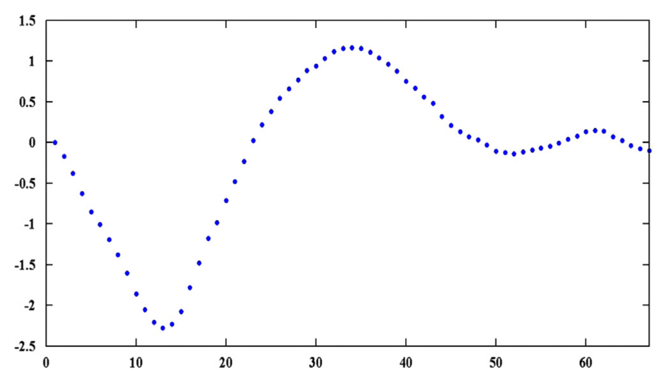

Fig. 1. A discrete time series.

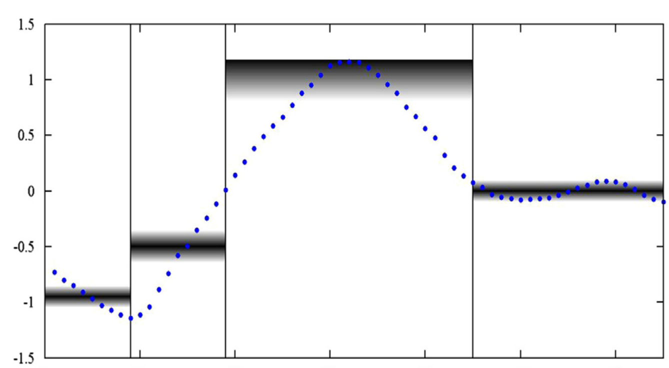

Fig. 2. The description of the time series as a sequence of information granules of magnitude formed over temporal windows.

the form

$$\min_{T_1,T_2,\ldots,T_p} \sum_{i=1}^{p} \mathrm{Vol}(\Omega_i). \tag{2}$$

The time windows are subject to optimization and in this way the resulting granular time series helps capture the very essence of temporal nature of data.

3. A modified fuzzy c-means based on dynamic time warping

3.1. Dynamic time warping

In what follows, the definition of the Euclidean distance and of the DTW similarity measure is recalled.

Euclidean distance: Let $A = \langle a_1, a_2, \ldots, a_T \rangle$ and $B = \langle b_1, b_2, \ldots, b_T \rangle$ be two time series sequences, and let δ be a distance between elements (or coordinates) of sequences. This distance is commonly accepted as the simplest distance between sequences. The distance between A and B is defined by

$$D_{Euclidean}(A,B) = \sqrt{\delta(a_1,b_1)^2 + \delta(a_2,b_2)^2 + \ldots + \delta(a_T,b_T)^2} \tag{3}$$

Euclidean distance is a popular method to define similarity and index time series, but it is very brittle in computing similarity between time series with different time phases.

Dynamic time warping: DTW is based on the Levenshtein distance (also called the edit distance) and was introduced in Sakoe and Chiba (1971, 1978), with applications in speech recognition. It uses a dynamic programming technique to find an optimal warping path between time series sequence, and overcome the weakness of Euclidean distance in computing similarity between time series with different time phases. To calculate the distance, one first forms a distance matrix, where each element in the matrix is a cumulative distance of a minimum of three surrounding neighbors. Let $C = \langle c_1, c_2, \ldots, c_M \rangle$ and $E = \langle e_1, e_2, \ldots, e_N \rangle$ be two time series sequences. First, an M-by-N matrix is created, and then each element, $D_{i,j}$, of the matrix is defined as

$$D_{i,j} = \delta(c_i, e_j) + \min\{D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}\}; \tag{4}$$

where $D_{i,j}$ is the summation of $\delta(c_i, e_j)$ and a minimum cumulative distance of three elements surrounding the element. When all elements in the matrix have been completed, the DTW distance is determined starting from the last element of the matrix, i.e.,

$$\mathrm{DTW}(C,E) = D_{M,N}. \tag{5}$$

The optimal warping path is a path through distance matrix from $D_{1,1}$ to $D_{M,N}$ consisting of those $D_{i,j}$ that have formed the $D_{M,N}$. Suppose the optimal warping path $P = \langle p_1, \ldots, p_k, \ldots, p_T \rangle$, where p_k is the kth coordinate (i_k, j_k) in the optimal path of sequences C and E, where i_k and j_k are indices of data points in sequences C and E, respectively.

The difference of Euclidean distance and DTW is displayed in Fig. 3. By showing this plot, we underline that DTW is able to correctly realign one sequence with the other, and Euclidean distance is unable to capture.

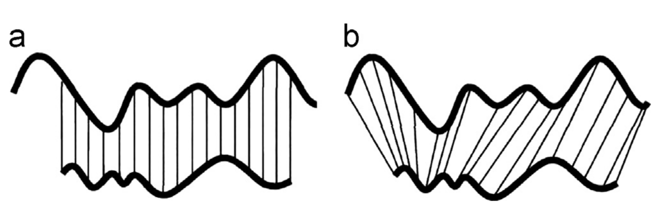

Fig. 3. The difference of Euclidean distance and DTW distance. (a) Euclidean distance and (b) DTW distance.

3.2. Fuzzy c-means based on dynamic time warping

Fuzzy c-means (FCM) is an unsupervised clustering algorithm that has been applied successfully to a number of problems involving feature analysis, clustering and classifier design (Marcelloni, 2003; Pedrycz and Rai, 2008; Tsai and Lin, 2011; Baraldi et al., 2011). DTW is able to find optimal global alignment between time series with same or different time phases and is probably the most commonly used measure to quantify the dissimilarity (Aach and Church, 2001; Bar-Joseph et al., 2002; Gavrila and Davis, 1995; Rath and Manmatha, 2003). Therefore, a modified fuzzy c-means based on dynamic time warping (DTW-FCM) is proposed to cluster time series.

Let $X = \{x_1, x_2, \ldots, x_n\}$ denote a data set to be partitioned into c clusters. The proposed algorithm minimizes the following objective function of FCM to

$$J_m = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \, \mathrm{DTW}^2(x_j, v_i) \tag{6}$$

with the following constraints:

$$\sum_{i=1}^{c} u_{ij} = 1, \ \forall j; \quad 0 \le u_{ij} \le 1, \ \forall i,j; \quad \sum_{j=1}^{n} u_{ij} > 0, \forall i \tag{7}$$

where u_{ij} represents the membership of data x_j in the ith cluster, v_i is the ith cluster center, $m > 1$ is a weighting exponent on each fuzzy membership. The parameter m controls the fuzziness of the resulting partition. Similar to the standard FCM algorithm, the membership functions and cluster centers are updated by the following expressions:

$$u_{ij} = \left[\sum_{k=1}^{c} \left(\frac{\mathrm{DTW}^2(x_j, v_i)}{\mathrm{DTW}^2(x_j, v_k)} \right)^{1/(m-1)} \right]^{-1}, \quad 1 \le i \le c, \ 1 \le j \le n. \tag{8}$$

$$v_i = \frac{\sum_{j=1}^{n} (u_{ij})^m x_j'}{\sum_{j=1}^{n} (u_{ij})^m}, \quad 1 \le i \le c, \tag{9}$$

where $x_j' = \{x_j(i_1'), \ldots, x_j(i_k'), \ldots, x_j(i_T')\}$, and $x_j(i_k')$ is i_k' th feature of x_j. $P' = \langle p_1', \ldots, p_k', \ldots, p_T' \rangle$ is the optimal warping path between x_j and v_i, and (i_k', j_k') is the kth coordinate of p_k'. The DTW-FCM algorithm is executed in the following steps:

Step 1: Given a preselected number of clusters c and fuzzy factor m ($m > 1$), initialize fuzzy cluster center v_i for $i = 1, 2, \ldots, c$.
Step 2: Calculate the fuzzy partition matrix $U = [u_{ij}]$ using Eq. (8).
Step 3: Use Eq. (9) to update fuzzy cluster center v_i for $i = 1, 2, \ldots, c$.
Step 4: If the improvement in J_m is less than a certain threshold (ε), then terminate the algorithm; otherwise go to step 2.

4. The proposed forecasting model

The proposed model is divided into four main phases. In the first phase, information granules of the magnitudes of the time series are formed through the use of the principle of justifiable granularity. Second, DTW-FCM is applied to cluster the time series data corresponding to the obtained information granules. Third, the fuzzy rules are extracted based on the labeled time series data, and the predicted variation is calculated using multiple fuzzy rules interpolation. Finally, the forecasted vector can be obtained by integrating the latest observation with the predicted variation. Figs. 4 and 5 show the basic idea of the proposed model.

4.1. Form information granules of the magnitudes of the time series

Information granularity is inherently present in the perception and interpretation of time series. Describing time series through a vocabulary of information granules makes the interpretation of data easier and more transparent. A way of translating time series into meaningful information granules is realized through the principle of justifiable granularity (Pedrycz and Gomide, 2007). The algorithm of the granular computing proceeds is detailed as follows.

4.1.1. Determine the information granule

Partition the time series $z_1, z_2, \ldots, z_n$ into "p" subseries by temporal windows $T_1, T_2, \ldots, T_p$. The corresponding data sets are $D_1, D_2, \ldots, D_p$. In each temporal window T_i, for a given $\alpha \in [0,1]$, the optimized information granule $\Omega_\alpha^i = [a_i, b_i]$ such that a_i and b_i maximize the index $V(a_i), V(b_i)$ (1,2), that is

$$a_i = \underset{a \le med(D_i)}{\operatorname{argmax}} \{\mathrm{card}\{x_k \in D_i, a \le x_k \le \mathrm{med}(D_i)\} \times \exp(-\alpha|\mathrm{med}(D_i) - a|)\}, \tag{10}$$

$$b_i = \underset{b \ge med(D_i)}{\operatorname{argmax}} \{\mathrm{card}\{x_k \in D_i, \mathrm{med}(D_i) \le x_k \le b\} \times \exp(-\alpha|b - \mathrm{med}(D_i)|)\}. \tag{11}$$

4.1.2. Calculate the volume of each information granule

In each temporal window T_i, computing the volume of the associated information granule $\mathrm{Vol}(\Omega_i)$ by

$$\mathrm{Vol}(\Omega_i) = T_i \int_{x_{min}}^{x_{max}} \Omega_i(z)\, dz, \tag{12}$$

Where the bounds of the integration x_{min} and x_{max} are, respectively, the minimal and maximal values of the amplitudes of the time series recorded in the temporal window T_i.

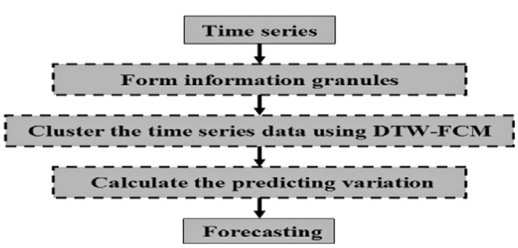

Fig. 4. Illustration of the proposed model.

4.1.3. Optimize the length of the temporal windows

Applying Particle Swarm Optimization (PSO) we minimize the following index:

$$\min_{T_1, T_2, \ldots, T_p} \sum_{i=1}^{p} \mathrm{Vol}(\Omega_i), \tag{13}$$

the optimized temporal windows $T_1, T_2, \ldots, T_p$ are obtained. The time series is segmented into "p" granules to give more efficient interpretation of the data.

4.2. Cluster the time series data using DTW-FCM

4.2.1. Normalize the data sets

The task to be completed is the normalization of the data sets $D_1, D_2, \ldots, D_p$ corresponding to temporal windows $T_1, T_2, \ldots, T_p$ that make the data sets have same original level. The following transformation is considered

$$z'_{i_j} \leftarrow z_{i_j} - z_{i_1}, \tag{14}$$

where z_{i_j} the jth data of D_i. The normalized data is denoted by D'_i.

4.2.2. Cluster the normalized data using DTW-FCM

When the number of granules "p" is small, the detailed information of granules may be lost. When the number of granules "p" is large, then the granules may be minute, thus drastically reducing their interpretability. In order to overcome the above problems, let the number of granules "p" be large, then a modified fuzzy c-means based on dynamic time warping (DTW-FCM) is applied to cluster the granules such that the similar granules belong to a cluster. The clustering results obtained in this way come with a better interpretability.

Apply DTW-FCM to partition the data sets which consist of temporal windows $D'_1, D'_2, \ldots, D'_p$ into clusters $C_1, C_2, \ldots, C_c$, where the membership grade u_{ij} of temporal windows D'_j belonging to C_i are calculated by means of (14), $1 \le i \le c$ and $1 \le j \le p$.

4.3. Calculate the predicted variation

4.3.1. Extract the fuzzy rules

Consider the three order fuzzy logical relationships (Chen and Wang, 2010) meaning that the next state D'_k depends upon a three-step history $(D'_{k-3}, D'_{k-2}, D'_{k-1})$. Assume $D'_{k-3}, D'_{k-2}, D'_{k-1}, D'_k$ belonging to clusters $C_{h_1}, C_{h_2}, C_{h_3}, C_j$, respectively. In general, this form of logical relationship can be schematically expressed in the form

$$C_{h_1}, C_{h_2}, C_{h_3} \rightarrow C_j, \tag{15}$$

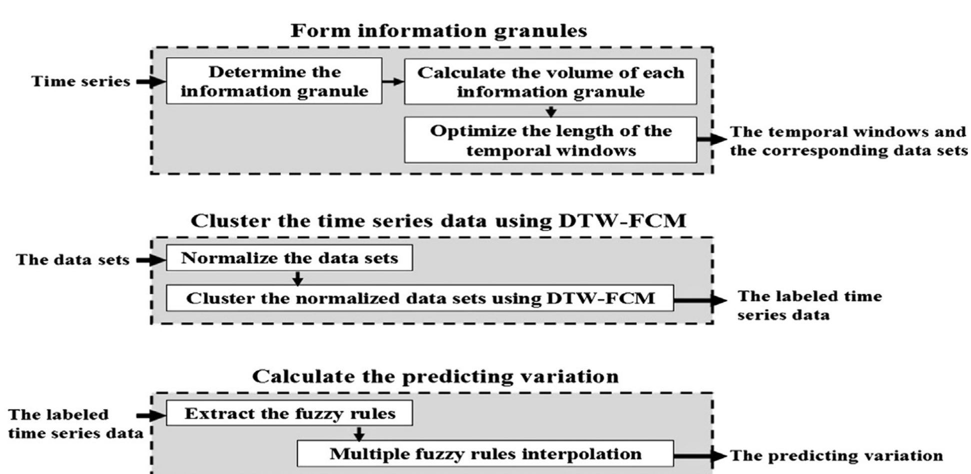

Fig. 5. The detailed phases of form information granules, cluster the time series data using DTW-FCM and calculation of the predicted variation.

where $C_{h_1}, C_{h_2}, C_{h_2}$ is the current state, C_j is the following (forecasted) state.

Fuzzy logical rules can be easily derived based on the fuzzy logical relationships shown as follows:

Rule R: **If** D'_{t-3} is C_{h_1}, D'_{t-2} is C_{h_2}, D'_{t-1} is C_{h_3}, **then** D'_t is C_j.

4.3.2. Multiple fuzzy rules interpolation

The multiple fuzzy rules interpolation scheme comes in the form Chang et al. (2008):

Rule R_1: **If** D'_{t-3} is $C_{h_{1,1}}$, D'_{t-2} is $C_{h_{1,2}}$, D'_{t-1} is $C_{h_{1,3}}$, **then** D'_t is C_{j_1}.

Rule R_2: **If** D'_{t-3} is $C_{h_{2,1}}$, D'_{t-2} is $C_{h_{2,2}}$, D'_{t-1} is $C_{h_{2,3}}$, **then** D'_t is C_{j_2}.

⋮

Rule R_r: **If** D'_{t-3} is $C_{h_{r,1}}$, D'_{t-2} is $C_{h_{r,2}}$, D'_{t-1} is $C_{h_{r,3}}$, **then** D'_t is C_{j_r}.

Observations: D'_{t-3} is C_{s_1}, D'_{t-2} is C_{s_2}, D'_{t-1} is C_{s_3}.

Conclusion: $D'_t = O^*$.

The output O^*, the predicted variation, is computed as follows:

$$O^* = \sum_{i=1}^{r} W_i \times L(C_{j_i}) \tag{16}$$

where $L(C_{j_i})$ is the last granular belonging to the cluster C_{j_i}. For instance, if $D'_{i_1}, D'_{i_2}, \ldots, D'_{i_n}$ belong to the cluster C_{j_i}, then $L(C_{j_i})$ is D'_{i_n}.

The weight matrix can be defined as below:

$$W = [W_1, W_2, \ldots, W_r] = \left[\frac{W'_1}{\sum_{i=1}^{r} W'_i}, \frac{W'_2}{\sum_{i=1}^{r} W'_i}, \ldots, \frac{W'_r}{\sum_{i=1}^{r} W'_i}\right] \tag{17}$$

where $W'_i = m_i \times f_i$. $m_i = u_{h_{i,1},t-3} \times u_{h_{i,2},t-2} \times u_{h_{i,3},t-1}$ is the matching degree of the observations belonging to the antecedent fuzzy sets of *Rule R_i*, and u_{ij} is the membership degree of the data D'_i to the cluster C_j obtained by the DTW-FCM. f_i is the frequency of *Rule R_i*.

The interpretation of predicted variation is obtained as follows:

If D'_{t-3} is C_{s_1}, D'_{t-2} is C_{s_2}, D'_{t-1} is C_{s_3}, then the membership value of predicting variation belonging to $C_{j_i} (1 \le i \le r)$ is W_i.

4.4. Forecasting

Assume $z(n)$ to be the data point at time n, the predicted vector F_{fv} can be determined as follows:

$$F_{fv} = [z_n, z_n, \ldots, z_n] + O^*, \tag{18}$$

where the dimensionality of $[z_n, z_n, \ldots, z_n]$ is the same as that O^*.

5. Experimental results

The proposed granular forecasting model is applied to forecast a synthetic time series, the chaotic Mackey–Glass time series (Oh et al., 2002), power demand time series (Keogh and Folias, 2002), and daily temperatures. Empirical analysis is conducted to validate the performance of the proposed model by comparing the forecasted results with one provided by the standard autoregression (AR) model, granular neural networks, Dong and Pedrycz (2008)'s granular model, Majhi et al. (2009)'s models, and Guo et al. (2012)'s models.

In order to assess the forecasting performance of the proposed model, three performance measures are considered:

- The root mean-square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (F_{forecstedvalue}(t) - F_{actualvalue}(t))^2}{n}} \tag{19}$$

- The mean absolute percentage error (MAPE):

$$\mathrm{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \frac{|F_{forecstedvalue}(t) - F_{actualvalue}(t)|}{F_{actualvalue}(t)} \times 100 \tag{20}$$

- The mean absolute error (MAE).

$$\mathrm{MAE} = \frac{1}{n} \sum_{t=1}^{n} |F_{forecstedvalue}(t) - F_{actualvalue}(t)| \tag{21}$$

where n denotes the number of date, $F_{actualvalue}(t)$ is the actual value of the time series, and $F_{forecstedvalue}(t)$ is the predicted value provided by the forecasting model.

5.1. Synthetic time series

In order to visualize the essence model, a synthetic time series (shown as Fig. 6) is considered.

5.1.1. Formation of information granules of the magnitudes of the time series

The number of temporal windows p is set to 8. The temporal windows obtained by the PSO segmentation and uniform partition are shown in Fig. 7. Fig. 7 visualizes that PSO segmentation based on the principle of justifiable granularity leads to more meaningful and easily interpretable entities.

5.1.2. Clustering the time series data using DTW-FCM

The number of clusters c is set to 3. FCM cannot cluster the time series data with different lengths. Therefore, a modified fuzzy c-means based on dynamic time warping (DTW-FCM) is proposed to cluster time series data. The result of DTW-FCM for the normalized time series data is shown in Fig. 8.

5.1.3. Calculation of the predicted variation

Based on the result produced by the DTW-FCM, the third-order fuzzy logical relationships is constructed as follows:

$C_1, C_2, C_3 \to C_2$,
$C_2, C_3, C_2 \to C_1$,
$C_3, C_2, C_1 \to C_2$,
$C_2, C_1, C_2 \to C_3$,

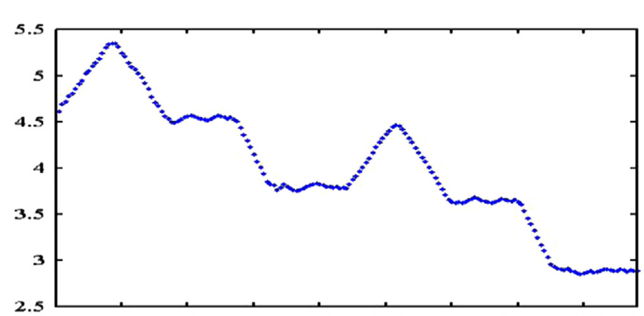

Fig. 6. Synthetic time series.

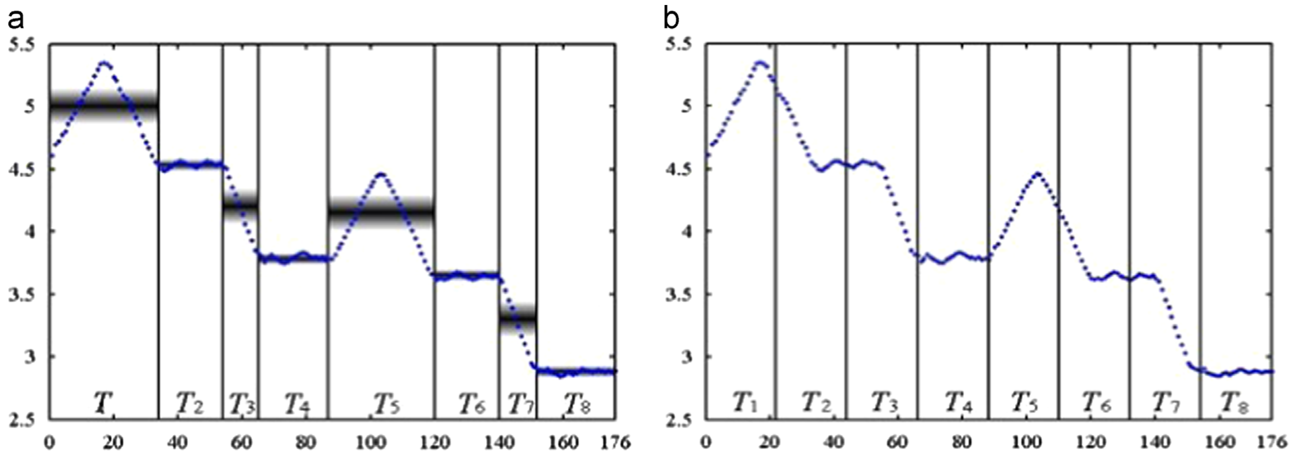

Fig. 7. The temporal windows obtained by the PSO segmentation and uniform partition. (a) PSO segmentation and (b) Uniform partition.

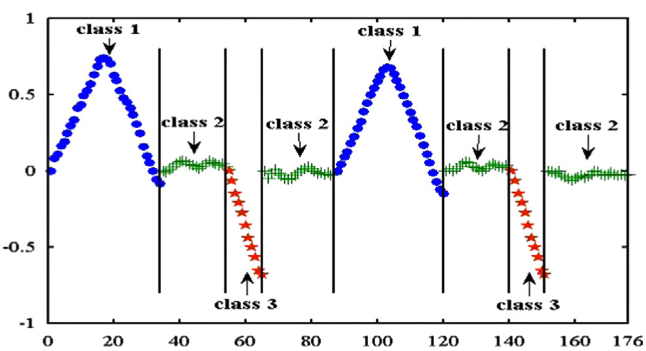

Fig. 8. The result of DTW-FCM obtained for the normalized time series data.

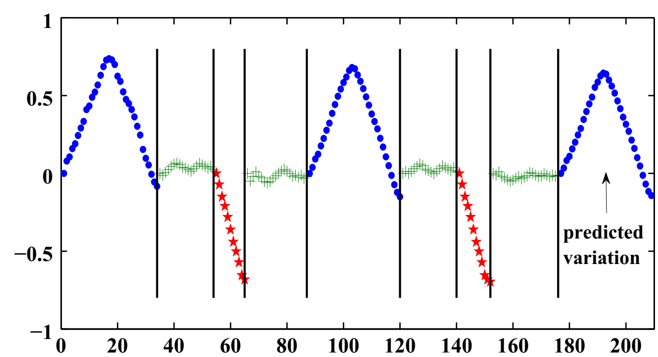

Fig. 9. The predicted variation.

Fuzzy logical rules can be derived based on the fuzzy logical relationships shown as follows:

Rule R_1: If D'_{i1} is class 1, D'_{i2} is class 2, D'_{i3} is class 3, then D'_{i4} is class 2.
Rule R_2: If D'_{i1} is class 2, D'_{i2} is class 3, D'_{i3} is class 2, then D'_{i4} is class 1.
Rule R_3: If D'_{i1} is class 3, D'_{i2} is class 2, D'_{i3} is class 1, then D'_{i4} is class 2.
Rule R_4: If D'_{i1} is class 2, D'_{i2} is class 1, D'_{i3} is class 2, then D'_{i4} is class 3.
Observations: D'_{i1} is class 2, D'_{i2} is class 3, D'_{i3} is class 2.

The matching degrees come in the form $W' = [0.0037, 0.2811, 0.0019, 0.0098]$. The weight matrix is obtained in the following form:

$W = [0.0125, 0.9481, 0.0064, 0.0331]$

The output $O^* = 0.0125 \times D'_8 + 0.9481 \times D'_5 + 0.0064 \times D'_8 + 0.0331 \times D'_7$ is shown as Fig. 9.

The interpretation of predicted variation is obtained as follows:

If D'_{i1} is class 2, D'_{i2} is class 3, D'_{i3} is class 2, then the membership values of predicted variation of membership to class 1, class 2, and class 3 are 0.9481, 0.0188, and 0.0331, respectively.

5.1.4. Forecasting

The forecasted vector F_{fv} is realized in the form:

$$F_{fv} = [z_{176}, z_{176}, \ldots, z_{176}]_{1\times 33} + O^*. \tag{22}$$

The vector is shown in Fig. 10.

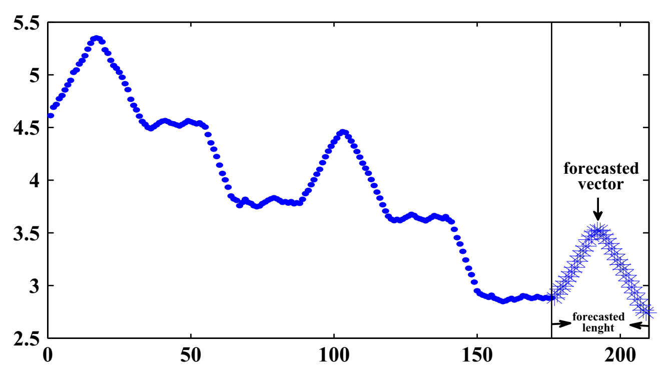

Fig. 10. The forecasted vector.

5.2. Chaotic Mackey–Glass time series

A chaotic Mackey–Glass time series (Oh et al., 2002) is generated by the chaotic Mackey–Glass differential delay equation comes in the form, see Fig. 11:

$$\dot{x}(t) = \frac{0.2{*}x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t). \tag{23}$$

In order to determine the noise sensitivity of the proposed model, random noise are added to the Mackey–Glass time series (see Fig. 12). The original Mackey–Glass time series and the Mackey–Glass time series with added noise are used for model validation. The parameters set in the experiment for proposed model are follows: the number of temporal windows $p = 50$, the number of clusters $c = 2$.

Table 1 summarizes the performance of the models of different forecasting length L for the Mackey–Glass time series. It becomes evident that the proposed model can tolerate noise.

5.3. Demand time series in power systems

The power demand of the Italy Milan Municipal Electric Pow Corporation from the UCR time series data mining archive (Keogh and Folias, 2002) is used for model validation. The following values of the parameters in the experiment were selected: the number of temporal windows $p=35$, the number of clusters $c=3$. Table 2 summarizes the performance of the models of different forecasting length L for the demand time series in power systems.

5.4. Daily temperature time series

The minimum daily temperature of Cowichan Lake Forestry of British Columbia from April 1, 1979 to May 30, 1996 is used for the experiment. The following parameters are chosen: the number of temporal windows $p=35$, the number of clusters $c=2$. Table 3 summarizes the performance of the models of different forecasting length L for the daily temperature time series.

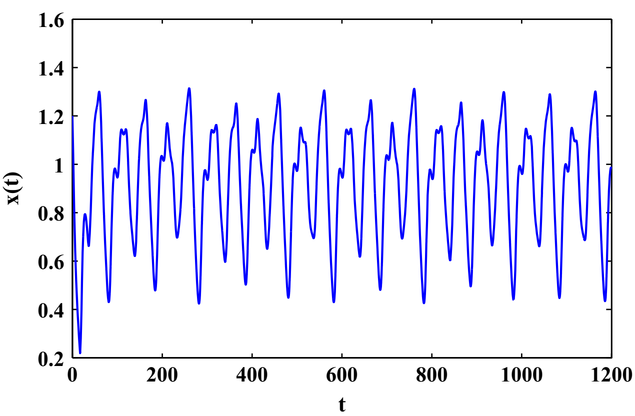

Fig. 11. Plot of data generated by the Mackey–Glass time series.

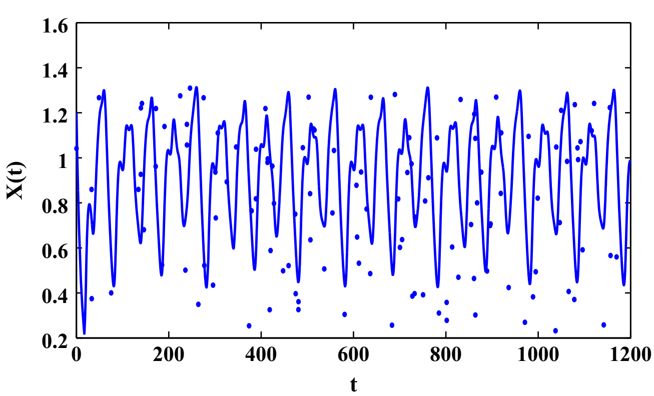

Fig. 12. Plot of data generated by the Mackey–Glass time series with noise.

5.5. Stock index time series

Stock Market prediction is one of the hottest fields of research due to its commercial applications and the attractive benefits it offers. The time series for the stock market prediction experiments have been collected for Standard's and Poor's 500 (S&P 500), USA, and the total number of data points for the stock indices is 3228 trading days, from January 3, 1994 to October 23, 2006. The following parameters are chosen: the number of temporal windows $p=35$, the number of clusters $c=3$. Table 4 summarizes the

Table 2
Comparison of the RMSE for demand obtained from different models.

Models	Forecasting length		
	1	2	24
AR	13.439	18.991	18.161
Dong and Pedrycz (2008)			
Granular neural networks	–	–	–
Granular forecasting model	12.666	18.673	20.045
The proposed model	2.043	4.054	20.895

Table 3
Comparison of the RMSE for daily temperature obtained from different models.

Models	Forecasting length			
	1	6	19	30
AR	2.398	3.100	3.845	5.081
Dong and Pedrycz (2008)				
Granular neural networks	–	–	3.972	3.584
Granular forecasting model	2.443	3.227	3.494	3.473
The proposed model	0.315	3.259	2.808	3.469

Table 4
Comparison of the MAPE for stock index obtained from different models.

Models	Forecasting length		
	1	7	30
Majhi et al. (2009)			
ABFO based model	0.675	1.387	2.409
BFO based model	0.811	1.391	2.531
PSO based model	0.807	1.398	2.385
GA based model	1.847	1.528	2.617
The proposed model	0.539	1.211	2.437

Table 1
Comparison of RMSE for Mackey–Glass obtained from different models.

Models	Forecasting length				
	1	6	16	20	30
AR	0.003	0.715	7.602	4.318	0.289
Dong and Pedrycz (2008)					
Granular neural networks	0.088	0.121	0.201	0.140	0.189
Granular forecasting model	0.043	0.065	0.099	0.034	0.084
The proposed model	0.003	0.011	0.029	0.047	0.054
The proposed model for the time series with noise	0.010	0.019	0.056	0.069	0.081

Table 5
Comparison of the MAE for daily wind speed obtained from different models.

Models	Forecasting length 10
Guo et al. (2012)	
FNN model	0.307
EMD-FNN model	0.344
Modified EMD-FNN model	0.289
The proposed model	0.265

performance of the models of different forecasting length L for the stock index time series.

5.6. Wind speed time series

The mean daily wind speed data of Zhangye of China from May to August in 2006 is used for the experiment. The following parameters are chosen: the number of temporal windows $p=35$, the number of clusters $c=3$. Table 5 summarizes the performance of the models of different forecasting length L for the mean daily wind speed time series.

6. Conclusions

In this paper, we have proposed a time series long-term forecasting model based on information granules and fuzzy clustering. The raw time series are transformed into some meaningful and operationally sound information granules. Then the modified fuzzy c-means based on dynamic time warping is proposed to cluster information granules and extract the fuzzy logical rules. Finally, we determine the weight of each fuzzy rule with respect to the input observation and use such weights to determine the predicted output based on the multiple fuzzy rules interpolation scheme. Experiments have been carried out for 5 publicly available time series to examine the improvements in forecasting performance of the suggested model when compared with the “standard” autoregression (AR) model, different “conventional” granular forecasting models (Dong and Pedrycz, 2008), Majhi et al. (2009)'s models, and Guo et al. (2012)'s models.

Future works will aim at generalizing the model to deal with more complicated real-world problems. A question on how to determine the optimal the number of temporal windows p and the number of clusters c are worth studying in the future.

Acknowledgments

This work is supported by the Natural Science Foundation of China under Grant 61175041. We would like to express thanks to the anonymous reviewers for their invaluable comments and suggestions.

References

Aach, J., Church, G., 2001. Aligning gene expression time series with time warping algorithms. Bioinformatics 17, 495–508.
Bar-Joseph, Z., Gerber, G., Gifford, D., Jaakkola, T., Simon, I., 2002. A new approach to analyzing gene expression time series data. In: Proceedings of the Sixth Annual International Conference on Computational Biology. pp. 39-48.
Baraldi, P., Razavi-Far, R., Zio, E., 2011. Bagged ensemble of fuzzy c-means classifiers for nuclear transient identification. Ann. Nucl. Energy 38, 1161–1171.
Bodyanskiy, Y., Popov, S., 2006. Neural network approach to forecasting of quasiperiodic financial time series. Eur. J. Oper. Res. 175, 1357–1366.
Castillo, O., Melin, P., 2002. Hybrid intelligent systems for time series prediction using neural networks, fuzzy logic, and fractal theory. IEEE Trans. Neural Netw. 13, 1395–1408.
Chang, Y., Chen, S., Liau, C., 2008. Fuzzy interpolative reasoning for sparse fuzzy rule-based systems based on the areas of fuzzy sets. IEEE Trans. Fuzzy Syst. 16, 1285–1301.
Chen, S., 1996. Forecasting enrollments based on fuzzy time series. Fuzzy Sets Syst. 81, 311–319.
Chen, S., Chang, Y., 2010. Multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques. Inf. Sci. 180, 4772–4783.
Chen, S., Chen, C., 2011. Taiex forecasting based on fuzzy time series and fuzzy variation groups. IEEE Trans. Fuzzy Syst. 19, 1–12.
Chen, S., Wang, N., 2010. Fuzzy forecasting based on fuzzy-trend logical relationship groups. IEEE Trans. Syst. Man Cybern.-Part B: Cybern. 40, 1343–1358.
Dong, R., Pedrycz, W., 2008. A granular time series approach to long-term forecasting and trend forecasting. Physica A 387, 3253–3270.
Egrioglu, E., Aladag, C.H., Basaran, M.A., Yolcu, U., Uslu, V.R., 2011. A new approach based on the optimization of the length of intervals in fuzzy time series. J. Intell. Fuzzy Syst. 22, 15–19.
Gavrila, D., Davis, L., 1995. Towards 3-d model-based tracking and recognition of human movement: a multi-view approach. In: IEEE International Work-shop on Automatic Face- and Gesture-Recognition. pp. 272–277.
Guo, Z., Zhao, W., Lu, H., Wang, J., 2012. Multi-step forecasting for wind speed using a modified emd-based artificial neural network model. Renew. Energy 37, 241–249.
Kang, I., 2003. Multi-period forecasting using different models for different horizons: an application to U.S. economic time series data. Int. J. Forecast. 19, 387–400.
Keogh, E., Folias, T., 2002. The UCR Time Series Data Mining Archive. University of California, Irvine.
Lendasse, A., Lee, J., Wertz, V., Verleysen, M., 2002. Forecasting electricity consumption using nonlinear projection and self-organizing maps. Neurocomputing 48, 299–311.
Lu, W., Pedrycz, W., Liu, X., Yang, J., Li, P., 2014. The modeling of time series based on fuzzy information granules. Expert Syst. Appl. 41, 3799–3808.
Majhi, R., Panda, G., Majhi, B., Sahoo, G., 2009. Efficient prediction of stock market indices using adaptive bacterial foraging optimization (abfo) and bfo based techniques. Expert Syst. Appl. 36, 10097–10104.
Marcelloni, F., 2003. Feature selection based on a modified fuzzy c-means algorithm with supervision. Inf. Sci. 151, 201–226.
Mastorocostas, P., Hilas, C., 2012. A computational intelligence-based forecasting system for telecommunications time series. Eng. Appl. Artif. Intell. 25, 200–206.
Oh, S., Pedrycz, W., Ahn, T., 2002. Self-organizing neural networks with fuzzy polynomial neurons. Appl. Softw. Comput. 2, 1–10.
Pedrycz, W., Gomide, F., 2007. Fuzzy Systems Engineering: Toward Human-Centric Computing. John Wiley, Hoboken, NJ.
Pedrycz, W., Rai, P., 2008. Collaborative clustering with the use of fuzzy c-means and its quantification. Fuzzy Sets Syst. 159, 2399–2427.
Pedrycz, W., Vukovich, G., 2001. Abstraction and specialization of information granules. IEEE Trans. Syst. Man Cybern.-Part B: Cybern. 31, 106–111.
Rath, T., Manmatha, R., 2003. Word image matching using dynamic time warping. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 521–527.
Sakoe, H., Chiba, S., 1971. A dynamic programming approach to continuous speech recognition. In: Proceedings of the Seventh International Congress on Acoustics. pp. 65–69.
Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Signal 26, 43–49.
Simon, G., Lendasse, A., Cottrell, M., Fort, J., Verleysen, M., 2005. Time series forecasting long term trends with self-organizing maps. Pattern Recognit. Lett. 26, 1795–1808.
Song, Q., Chissom, B., 1993a. Forecasting enrollments with fuzzy time series—Part I. Fuzzy Sets Syst. 54, 1–9.
Song, Q., Chissom, B., 1993b. Fuzzy time series and its models. Fuzzy Sets Syst. 54, 269–277.
Song, Q., Chissom, B., 1994. Forecasting enrollments with fuzzy time series—Part II. Fuzzy Sets Syst. 62, 1–8.
Tsai, D., Lin, C., 2011. Fuzzy c-means based clustering for linearly and nonlinearly separable data. Pattern Recognit. 44, 1750–1760.
Wang, L., Liu, X., Pedrycz, W., 2013. Effective intervals determined by information granules to improve forecasting in fuzzy time series. Expert Syst. Appl. 40, 5673–5679.
Wang, L., Liu, X., Pedrycz, W., Shao, Y., 2014. Determination of temporal information granules to improve forecasting in fuzzy time series. Expert Syst. Appl. 41, 3134–3142.
Zadeh, L., 1979. Fuzzy sets and information granularity. In: Advances in Fuzzy Set Theory and Application, vol. 1. pp. 3–18.