International Journal of Innovative Computing, Information and Control
Volume **14**, Number **5**, October **2018**

ICIC International ©2018 ISSN 1349-4198
pp. 1583–1602

THE DATA-DRIVEN FUZZY COGNITIVE MAP MODEL AND ITS APPLICATION TO PREDICTION OF TIME SERIES

Dan Shan[1], Wei Lu[2] and Jianhua Yang[2]

[1]Department of Electronic Engineering
Dalian Neusoft University of Information
No. 8, Software Park Rd., Dalian 116023, P. R. China
shandan@neusoft.edu.cn

[2]School of Control Science and Engineering
Dalian University of Technology
No. 2, Linggong Road, Ganjingzi District, Dalian 116024, P. R. China
luwei@dlut.edu.cn

Received December 2017; revised April 2018

Abstract. *As a soft computing method, fuzzy cognitive map (FCM) has become a hot research topic in recent years. However, the traditional method of constructing FCM depends on expert knowledge which has obvious drawbacks of subjectivity and limitation especially for long-term or complex time series. In this paper, a new method for constructing FCM is proposed which extracts knowledge from data by exploiting clustering, and FCM's weights can be automatically obtained by particle swarm optimization algorithm (PSO) according to historical data of time series. Further the proposed method is applied to predicting, in which FCM is used to represent, storing fuzzy logical relationships of time series and realizing prediction by iterations. Three benchmark time series: the enrollments time series, the TAIEX time series and the Wolf's sunspot time series are applied to verifying prediction performance of the proposed method, whose results show that the proposed numerical prediction method of time series is effective and can obtain better prediction accuracy than traditional methods, and the potential advantage of the proposed method is capable of processing the prediction problem of long-term or complex time series. In addition the influence of parameters of the method is analyzed individually.*

Keywords: Fuzzy cognitive map, Fuzzy c-means clustering, Time series

1. **Introduction.** As an effective soft computing tool, fuzzy cognitive map (FCM) [1] is put forward by Kosko as an extension of cognitive maps, which has become a hot issue for researchers in numerous scientific fields for modeling, prediction [2, 3, 4, 5, 6], decision making [7, 8], and pattern recognition [9], etc.

FCM has the advantages as follows. (1) FCM is convenient to represent knowledge, describe dynamic behavior of systems, quantify relationships between concepts (variables) in systems and perform reasoning with interpretable results. (2) FCM has direct causal representation, user-friendly approach, easiness of use, practicality and low time consumption. In the virtues of these, FCM can become an available alternative for modeling and prediction of time series.

To construct FCM, there are two crucial issues that are encountered – one is how to construct architecture of FCM based on historical data, and the other is how to learn weights of the constructed FCM architecture based on historical data. At present, many researches associated with FCM mainly focus on the second issue. Papageorgiou et al.

DOI: 10.24507/ijicic.14.05.1583

[10] proposed a representative method of nonlinear Hebbian learning (NHL) algorithm, illustrated in medical domain and industrial process control domain. Stach et al. [11] proposed an evolutionary-based learning method of genetic algorithms to learn fuzzy cognitive maps (FCMs) directly from data, performed on both synthetic and real-life data. Also Papageorgiou et al. [12] used particle swarm optimization (PSO) algorithm as evolutionary-based learning method, applied on an industrial process control problem. Recently, Zou and Liu [13] proposed a mutual information based two-phase memetic algorithm for large-scale fuzzy cognitive map learning, applied for the gene regulatory network (GRN) reconstruction problem. Ahmadi et al. [14] used imperialist competitive algorithm as FCMs learning method, compared with other well-known FCM learning algorithms. Chi and Liu [15] proposed a multi-objective evolutionary algorithm for FCM learning, validated on both synthetic and real data with varying sizes and densities. However, they rarely researched with respect to the first issue of architecture of FCM. In general, the architectures of FCMs are created by expert experience, whereas expert experience may be subjective and not always reliable for FCM, and cannot describe dynamic behaviour of systems very well. Especially for complex system or the architecture with more concepts, constructing FCM model becomes quite difficult for experts.

Fortunately, clustering technology, especially fuzzy c-means algorithm, can discover the knowledge implying in data. Therefore, it can be considered to generate architecture of FCM. Inspired by this idea, in this study, we propose a novel constructing method of FCM to avoid the shortcomings of traditional method of constructing FCM mentioned above, in which we first extract knowledge from historical data by exploiting clustering and form concepts (nodes) of FCM, the architecture of FCM, and then FCM's weights can be learnt by some optimization algorithms on the basis of historical data. Consequently the ensuing FCM based on historical data is constructed. The proposed method is used to build FCM model for three real-time series. The corresponding experimental results show that the proposed method of constructing FCM is feasible and effective.

The remainder of this paper is organized as follows. Section 2 briefly describes some relevant preliminaries. Section 3 provides details of the proposed method of constructing FCM. In Section 4, three benchmark time series data sets are used to validate feasibility and effectiveness of the proposed method. Finally, Section 5 provides conclusions.

2. Preliminaries.

2.1. Using improved fuzzy c-means clustering algorithm to extract architecture of FCM. Fuzzy c-means clustering [16] is widely applied in the field of pattern recognition, which introduces clusters for representing groups of observations that are "close" or "similar" in the sense of some predefined metric from numerical data. In this paper, we focus on this feature to cluster nodes of FCM, i.e., concepts. How to construct framework of FCM according to numeric data is important to realize FCM-based prediction of time series. Several approaches [3, 5, 6] have been proposed. One of them is by the usual fuzzy c-means clustering algorithm to obtain nodes (concepts) of FCM. An advantage of this method is that it can build framework of FCM automatically depending on numeric data.

However, as an unsupervised machine learning algorithm, the usual fuzzy c-means clustering algorithm is sensitive to isolated data points, i.e., does not take outliers into consideration which are likely caused by noise, affect the derivation of FCM nodes, and influence the accuracy of prediction further. In order to overcome this shortcoming, we propose an improved fuzzy c-means clustering algorithm for constructing framework of FCM, which involves optimizing membership matrix with a membership optimization function to reduce sensitivity to isolated data point. The improved fuzzy c-means algorithm is detailed as follows: first take a brief review of traditional version, and then introduce the improved one.

Assuming the sample data set is $X = \{x_1, x_2, \ldots, x_n\}$ which is divided into c groups, and clustering centers v_i $(i = 1, \ldots, c)$ are derived to minimum objective function (1):

$$\min : Q = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \|x_j - v_i\|^2 \tag{1}$$

$$\sum_{i=1}^{c} u_{ij} = 1, \quad \forall j = 1, \ldots, n \tag{2}$$

where x_j $(j = 1, \ldots, n)$ is the jth measured data, v_i is the center of the ith cluster (prototype), c is the number of clusters, n is the number of data points, and u_{ij} is membership of x_j to the ith cluster center satisfying Equation (2). In Equation (3) and Equation (4), $m \in [1, \infty)$ is any real number greater than 1, and $\|\cdot\|$ is any norm implying the similarity between any measured data and the centers.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_j - v_i\|}{\|x_j - v_k\|} \right)^{\frac{2}{m-1}}} \tag{3}$$

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m} \tag{4}$$

Fuzzy partitioning is carried out by iterative optimization of objective function Equation (1) with update of membership and cluster centers according to Equation (3) and Equation (4), and this iteration will stop when $\max_{ij}\{|u(s+1) - u(s)|\} \leq \varepsilon$, where ε is threshold between 0 and 1, and s is the iteration step.

Owing to the advantage of perfect theory and deep mathematical foundation, fuzzy c-means clustering has become a popular method in the process of fuzzy clustering, and relatively factory results can be obtained.

In the usual fuzzy c-means algorithm, the causal relationships are significant in two folds. One is that the membership of a data point is determined by the distance between measured data and the centers individually. That is to say each membership indicates the extent to which the data belongs to the cluster, the greater the membership is, the smaller the uncertainty is, likewise the smaller the membership is, the greater the uncertainty is. The other fold is that during update of the clustering centers by iteration, the memberships also show contributions of each data point to the new clustering centers. The higher the membership is, the greater the impact of corresponding data point to the new clustering center is, vice versa. Because the membership is relative, they are likely unsuitable for typical applications. The new clustering centers obtained by current memberships are likely not to be expected positions. Just imagine if an isolated data point is encountered, the membership is $1/c$ for all the cluster centers, where c is the number of the clustering centers. Consequently it may cause undesirable clustering result. In order to take this situation into consideration, we introduce a correcting function Equation (5) as membership optimization function to reduce the influence of outliers on the clustering center, so as to optimize the result of clustering analysis. After correcting, the lower the membership is, the lower influence of corresponding data point on the new cluster center position will be, vice versa.

$$N_{ij} = u_{ij} \cdot a^{(u_{ij} - 1)} \tag{5}$$

$$C_{ij} = \frac{N'_{ij}}{\sum_{i=1}^{c} N'_{ij}} \tag{6}$$

where u_{ij} is original membership, N_{ij} in Equation (5) is the one after correcting, C_{ij} is the new membership after normalization by Equation (6), and $a \in [1, \infty)$ first named as steepness index is any real number greater than 1. Obviously, when membership u_{ij} is 1 in usual fuzzy c-means, new membership C_{ij} is also 1, and when it is 0, the new one is also 0. By optimization of membership matrix, it can be seen in $[0, 1]$ interval, membership decreases in comparison with the value obtained from usual version. Moreover, the smaller the original membership is, the more apparent the new one decreases relatively. When it is applied to the clustering center formula Equation (4), it implies that the impact of those small membership data points on updating new clustering center is reduced. Therefore, to a certain extent we reduce the influence of isolated point that may be caused by noises. The optimized clustering centers are calculated according to Equation (7).

$$P_i = \frac{\sum_{j=1}^{n} (N_{ij})^m x_j}{\sum_{j=1}^{n} (N_{ij})^m} \tag{7}$$

In the membership optimization function, parameter a is associated with the extent of influence of each data point on new center position during iteration. The role of this parameter is to provide some additional calibration of new center. Figure 1 shows the impacts between the sensitivity of N_{ij} and steepness parameter a. Higher value of a increases the steepness of the curve and makes it more sensitive to isolated point.

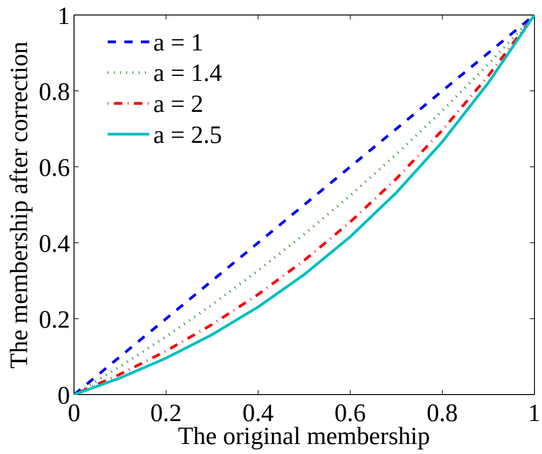
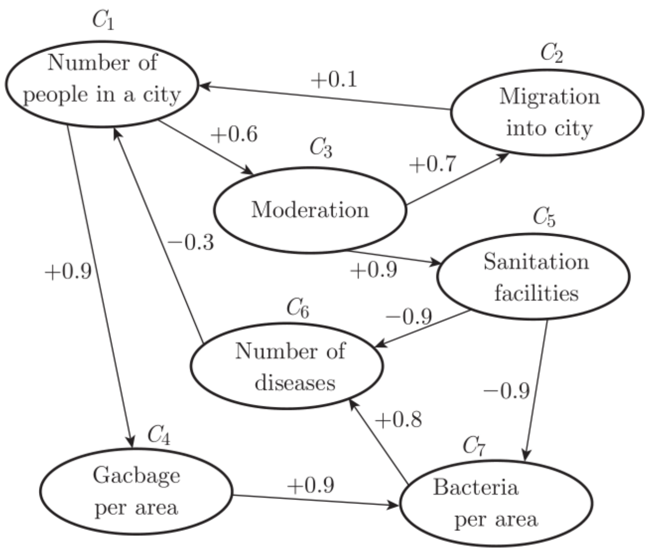


FIGURE 1. The shape of membership optimization functions with different a

In the improved fuzzy c-means clustering algorithm, we can take clustering result, i.e., prototypes containing certain fuzzy semantics as nodes of FCM, so the framework of FCM is generated. Weights among all nodes of FCM can be obtained by PSO algorithm on the basis of these numerical data (refer to Section 2.3).

2.2. **Fuzzy cognitive map (FCM).** Fuzzy cognitive map is a simple and powerful tool for representing human knowledge and performing reasoning. FCM as a modeling methodology describes given system by means of concepts and mutual relationships among them, which plays a critical role of time series modeling and predicting in this paper. Within the framework of FCM, concepts stand for variables, events, goals, actions, terms, value, etc. of given system, which are of interest to researchers. The interaction between concepts can be qualitatively described as three kinds of relationships: positive, negative, and neutral. Each relationship is directed, strength of which can be quantitatively expressed as a real number in $[-1, 1]$ interval, where -1 stands for the strongest negative relationship, 0 for the neutral, and 1 for the strongest positive relationship between corresponding concepts. Each type of relationship expresses different causality between two concepts. Positive relationship indicates that an increase of the value of one concept leads to an increase of the value of the other concept linked with it (and vice versa). Negative relationship indicates that an increase of the value of one concept leads to a decrease of the value of the other concept linked with it (and vice versa). Neutral relationship means no relationship between two concepts linked with each other.

FCM can be shown conveniently in the form of a graph which is composed of a collection of nodes and directed edges with weights between the nodes. The nodes, directed edges and values of edge weights (weights, for short) are used to represent concepts, relationships between corresponding concepts, and values of strength of relationships between concepts, respectively. Specially, the directed edge between concepts is removed from the graph if the relationship between them is neutral, i.e., the value of weight between them is zero or near-zero. Equivalently, FCM can also be represented as a square matrix (it is also called relationship matrix), which stores values of weights among all nodes. Figure 2 shows an example of FCM model for public city health issues [21] and its relationship matrix.



	C_1	C_2	C_3	C_4	C_5	C_6	C_7
C_1	0	0	0.6	0.9	0	0	0
C_2	0.1	0	0	0	0	0	0
C_3	0	0.7	0	0	0.9	0	0
C_4	0	0	0	0	0	0	0.9
C_5	0	0	0	0	0	−0.9	−0.9
C_6	−0.3	0	0	0	0	0	0
C_7	0	0	0	0	0	0.8	0

Figure 2. FCM model and its relationship matrix

The FCM in Figure 2 reflects the main behavior characteristics of the public city health issue system. There are 7 concepts, C_1, C_2, C_3, C_4, C_5, C_6, C_7, with different semantics. The edges reflect how the nodes affect one another and the weight on each edge or alternatively shown in the relationship matrix quantifies the strength of affection into the interval $[-1, 1]$. For example, when the weight is greater than 0 like from node C_7 to node C_6, an increase of value of C_7 leads to an increase of value of C_6; when the weight is less than 0 like from node C_5 to node C_7, an increase of value of C_5 leads to an decrease of value of C_7; when the weight is equal to 0, no relationship between nodes and the edge is removed.

FCM can also be represented mathematically as following Equation (8):

$$C_i(t+1) = f\left(\sum_{j=1}^{c} w_{ji} C_j(t) + w_{0i}\right) \tag{8}$$

where $C_i(t)$ is the activation level of the ith node at t moment, i.e., the value of the ith node at t moment, $w_{ji} \in [-1, 1]$ is the value of weight from the jth node to the ith node, $w_{0i} \in [0, 1]$ is the bias associated with the ith node, t is the time point, c is the number of nodes, $C_i(t + 1)$ is the predicting value of the ith node at $t + 1$ moment and f is the transformation function, which is generally selected to be sigmoid function with steepness parameter σ version $f(u) = 1/(1 + \exp(-\sigma u))$ $(u \in R,\ \sigma > 0)$, where the steepness parameter σ is associated with individual node of the FCM. The role of this parameter is to provide some additional calibration of the value of the node. Higher values of σ increase the steepness of the function f and make it more sensitive to the changes of u. Actually it is remarkable that values of nodes of FCM at $t+1$ moment are determined by values of all the nodes which exert influence mutually through their weights at t moment, and also at moments before, for example $t-1$ moment, etc. Therefore, as simple and clear as possible we introduce second-order FCM to perform prediction as following Equation (9).

$$C_i(t+1) = f\left(\sum_{j=1}^{c} w1_{ji} C_j(t) + \sum_{j=1}^{c} w2_{ji} C_j(t-1) + w_{0i}\right) \tag{9}$$

where $w1_{ji} \in [-1, 1]$ is the value of weight from the jth node to the ith node at t moment, and similarly $w2_{ji} \in [-1, 1]$ is the value of weight from the jth node to the ith node at $t - 1$ moment, w_{0i} is the same as mentioned before.

Once FCM is constructed, it starts with an initial state to perform successive iteration according to Equation (9) until it goes into stable situation. In other words, FCM can reach to an equilibrium point within finite iterations.

2.3. **Using PSO algorithm to learn weights of FCM.** As previously mentioned, how to reasonably determine weights of FCM is critical to realize FCM-based prediction of time series. Reasonable weights enable FCM to describe the dynamic behavior of system accurately. Several approaches for automated learning weights of FCM models from numeric data have been proposed. One of them is by using particle swarm optimization algorithm (PSO) [12], which provides capabilities of global optimization based on population yet not bring about a very heavy computational overload.

PSO algorithm involves a population of particles whose dynamic characteristics are guided by the mechanisms of social interactions and individual experience. Each particle has two important attributes: one is that it can memorize and follow its previous direction, and the other is that it can move and gather towards the best position (solution) searched by individual particle and the entire population in the solution space. The details of usual PSO and many improved versions are shown in [23, 24]. The search strategy of PSO exhibits better biological and societal background, which can be described as following Equation (10) and Equation (11):

$$w(t) = w(t-1) + v(t) \tag{10}$$

$$v(t) = \xi v(t-1) + \Phi_1 r_1 (p - w(t-1)) + \Phi_2 r_2 (P_{total} - w(t-1)) \tag{11}$$

where v is velocity of corresponding particle, w is the current position, p is the best position for individual particle and P_{total} is the best one of all best positions for all particles in the whole population. The parameters Φ_1 and Φ_2 are acceleration constants, r_1 and r_2 are random numbers defined over the $[0, 1]$ interval under normal distribution, and ξ is inertial weight smaller than 1. Each particle explores the solution space at velocity of $v(t)$ from current position $w(t)$ which includes three components: the first one is inertia component which reveals moving habit of particle, the second one is cognition component which reflects memory function of particle for past and the third one is social component which shows collaboration and knowledge sharing between the particles.

The objective here is to develop candidate FCM. In essence, this is an optimization problem, for second-order prediction which requires to establish $c(2c+2)$ parameters which include all weights of relationships among c nodes of FCM, biases w_{0i} $(i = 1, 2, \ldots, c)$ associated with the ith node and a vector of steepness parameters σ_i $(i = 1, 2, \ldots, c)$ associated with the ith node. Consequently, the particles structure is defined as Equation (12):

$$\begin{aligned} W = \{ & w1_{11}, w1_{12}, \ldots, w1_{1c}, w1_{21}, w1_{22}, \ldots, w1_{2c}, \ldots, w1_{c1}, w1_{c2}, \ldots, w1_{cc}, \\ & w2_{11}, w2_{12}, \ldots, w2_{1c}, w2_{21}, w2_{22}, \ldots, w2_{2c}, \ldots, w2_{c1}, w2_{c2}, \ldots, w2_{cc}, \\ & w_{01}, w_{02}, \ldots, w_{0c}, \sigma_1, \sigma_2, \ldots, \sigma_c\}^T \end{aligned} \tag{12}$$

where $w1_{ji}$ and $w2_{ji} \in [-1, 1]$ $(i, j = 1, 2, \ldots, c)$ are weights from the jth to the ith node at t moment and $t-1$ moment respectively, $w_{0i} \in [0, 1]$ $(i = 1, 2, \ldots, c)$ is bias associated with the ith node, and σ_i $(i = 1, 2, \ldots, c)$ is steepness parameters greater than 0 for the ith node. W can be regarded as a single particle in the $c(2c+2)$ dimension solution space.

Objective function (13) defined by exploiting an inherent property of FCM execution model is used to evaluate particles quality in population, where $C_i(t)$ is the real response associated with the ith node, $\hat{C}_i(t)$ is the candidate FCM response predicted, n is the number of input data points (observations), and c is the number of concepts.

$$\min : f = \frac{1}{(n-2)c} \sum_{t=2}^{n-1} \sum_{i=1}^{c} \left\| \hat{C}_i(t) - C_i(t) \right\|^2 \tag{13}$$

3. **The FCM-Based Time Series Prediction Method.** Considering a time series $X = \{x(1), x(2), \ldots, x(n)\}$, block diagram of the proposed prediction method is illustrated in Figure 3, which includes five function modules: fuzzy c-means clustering module, fuzzifying original data module, learning FCM weights module, defuzzifying data module and calculating accuracy module. In what follows, function of each module in our proposed method is detailed respectively.

Fuzzy c-means clustering module is to construct framework of FCM and fuzzifying original data module is to transform original time series into fuzzy time series [17]. All observations in X are firstly clustered by improved fuzzy c-means clustering algorithm to obtain optimized prototypes. Subsequently, a prototype vector $P = (P_1, P_2, \ldots, P_c)$

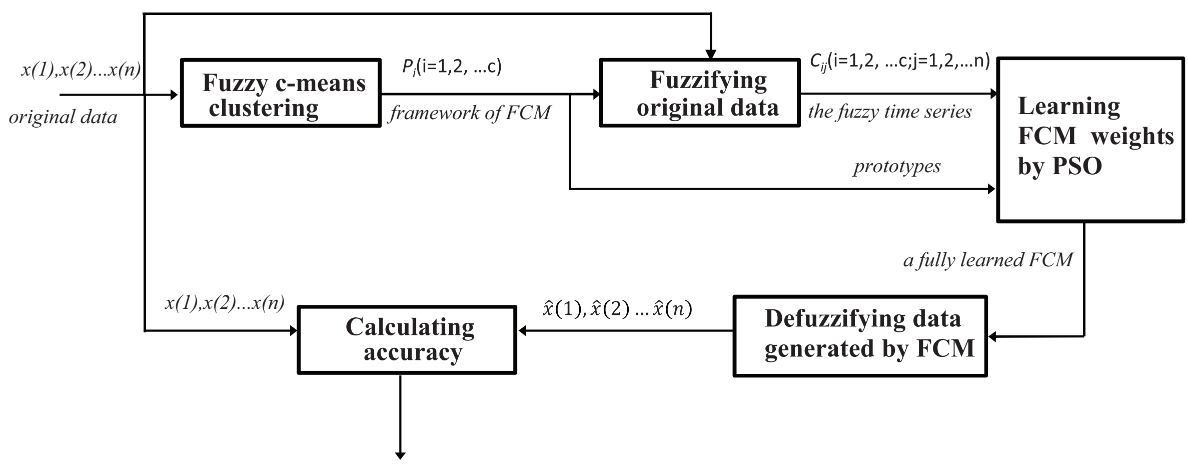


FIGURE 3. The framework of the proposed prediction method for time series

is generated. Each prototype P_i $(i = 1, 2, \ldots, c)$ is assigned into a certain semantic g_i $(i = 1, 2, \ldots, c)$. Next, each observation $x(t)$ $(t = 1, 2, \ldots, n)$ is transformed into their membership values corresponding to each prototype P_i, which means originally numerical time series is converted to fuzzy time series according to Equation (14).

$$C_i(j) = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x(j)-P_i\|}{\|x(j)-P_k\|} \right)^{\frac{2}{m-1}}} \tag{14}$$

$$\sum_{i=1}^{c} C_i(j) = 1, \quad \forall j = 1, \ldots, n \tag{15}$$

where $x(j)$ $(j = 1, \ldots, n)$ is the jth measured data, P_i is the center of the ith cluster (prototype), c is the number of clusters, n is the number of data points, and $C_i(j)$ is membership of $x(j)$ to the ith cluster center satisfying Equation (15). In Equation (14), $m \in [1, \infty)$ is any real number greater than 1, and $\|\cdot\|$ is any norm implying the similarity between any measured data and the centers.

The fuzzy time series becomes available consisting of c fuzzy subsequence $F_i = [C_{i1}, C_{i2}, C_{i3}, \ldots, C_{in}]$ $(i = 1, 2, 3, \ldots, c)$, which is represented in form of following:

$$F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_c \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(n) \end{bmatrix}^T = \begin{bmatrix} C_1(1) & C_1(2) & \cdots & C_1(n) \\ C_2(1) & C_2(2) & \cdots & C_2(n) \\ \vdots & \vdots & \vdots & \vdots \\ C_c(1) & C_c(2) & \cdots & C_c(n) \end{bmatrix} \tag{16}$$

Note that each row of F in Equation (16), say $F_1, F_2, \ldots, F_c$, expresses the level that the given time series can be characterized by corresponding fuzzy semantics P_i $(i = 1, 2, \ldots, c)$, whereas each column of F, say $R(1), R(2), \ldots, R(n)$, expresses the level that an observation of time series $x(t)$ at t moment $(t = 1, 2, \ldots, n)$ can be characterized by all fuzzy semantics. Actually, each element of each column of F in Equation (16) can also be regarded as the activation level of the corresponding node of FCM. The learning FCM weights module is used to construct a fully learned FCM by PSO algorithm for all parameters W (see Equation (12)). The defuzzifying data module and calculating accuracy module exploit the fully learned FCM to validate accuracy of the FCM constructed by iteration. For example, take observations x_1 and x_2 as initial state vector transformed into $R(1)$ and $R(2)$ which indicate the activation level of corresponding nodes in FCM according to Equation (12), and then FCM executes iteration step by step to obtain its response vector $\hat{R}(k) = \left[\hat{C}_1(k), \hat{C}_2(k), \ldots, \hat{C}_c(k)\right]$ at k moment. The numerical result can be obtained by using Equation (17) according to the activation value of all nodes in FCM and a prototype vector P formed by the first module.

$$\hat{x}(k) = \frac{\sum_{i=1}^{c} \hat{C}_i(k) P_i}{\sum_{j=1}^{c} \hat{C}_i(k)} \tag{17}$$

where $\hat{C}_i(k)$ is the ith node response of FCM for $C_i(k)$, P_i is corresponding the ith prototype and $\hat{x}(k)$ is the numerical value at t moment. The accuracy between the data generated from FCM $X = \{\hat{x}(3), \hat{x}(4), \ldots, \hat{x}(n)\}$ and the real observations $X = \{x(1), x(2), \ldots, x(n)\}$ is calculated in calculating accuracy module. The proposed prediction algorithm is detailed as follows.

Step 1. Generating prototypes P_i $(i = 1, 2, \ldots, c)$ and framework of FCM by the improved fuzzy c-means clustering algorithm. So far the concepts of FCM are generated.

Step 2. Transforming original time series into fuzzy time series $C_i(j)$ $(i = 1, 2, \ldots, c;\ j = 1, 2, \ldots, n)$ according to Equation (14).

Step 3. Carrying out PSO algorithm for learning weights W of FCM based on the whole subset. So far the model of FCM for fuzzy times series is established.

Step 4. Carrying out activation level calculating by iteration from the fully learned FCM built by Step 4 according to Equation (7), and then defuzzifying result can be obtained according to Equation (13) in which the activation value of all nodes in FCM and a prototype vector P are used to reconstruct numeric value.

Step 5. Calculating accuracy between the data generated by FCM and the real observations to validate the algorithm.

4. Experimental Study.

In this section, three time series, the enrollments of University of Alabama time series [17, 18, 19, 20, 21, 22, 23, 24], the daily value of Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) [17, 21, 22] and the Wolf's sunspot time series [25, 26, 27, 28, 29], are used to carry out experiment, respectively.

4.1. The enrollments time series.

The yearly data of enrollments of University of Alabama from 1971 to 1992 are commonly used to validate the prediction method of fuzzy time series. In this paper, the same data set is also used to validate the constructed FCM. Based on the implementation steps described in Section 3, the validation is accomplished with multiple experiments to illustrate the influence of diverse parameters in improved fuzzy c-means clustering algorithm on the prediction result. Table 1 records the specific experimental results of the root mean square error (RMSE) under some typical enumeration of clustering number c, fuzzification coefficient m, and steepness index a.

According to Table 1, the maximum RMSE is 480.0 and the minimum one is 16.6. Therefore, we can achieve better prediction accuracy by adjusting the values of parameters c, a, and m. Furthermore, we summarize the experimental data and draw a variety of curves which illustrate the influence trend of each parameter on the RMSE of prediction.

4.1.1. *The impact of parameters c, m and a on the prediction accuracy.*

Figure 4 shows the impact of some selected parameters c and m on the performance of constructed FCM model in the case of fixing value of a. From Figure 4 we can clearly see that the RMSE values exhibit a downtrend when m moves from low value to high value. However, when going toward higher value of m, the RMSE values begin to show an uptrend. For example, when c is fixed at 6, the RMSE is 243.3 for $m = 1.5$, whereas it is 114.8 for $m = 2.5$ and 169.5 for $m = 2.7$. The explanation for this is that the fuzzification coefficient m impacts the sharing degree between fuzzy clustering centers. When the value of m is too low, fuzzy c-means seems to degenerate to hard c-means, and when m is too high, fuzzy c-means seems to lose division characteristics, so there is a certain value which is not too low or too high, which can reach better prediction accuracy. Further when the value of m is fixed, the RMSE values decrease with increasing of the value of c. For example, when m is fixed at 2.5, the RMSE is 176.6 for $c = 5$, whereas it is 114.8 for $c = 6$ and 63.4 for $c = 8$. There is small difference in number of clusters, but there is large difference in RMSE. The explanation for this is that when the number of clusters is too low, very few clustering centers are generated which are not sufficient to describe the dynamic behavior of time series and give rise to the prediction error. Figure 5 shows the impact of some selected parameters c and a on the performance of constructed FCM model in the case of fixing value of m – when the value of m is fixed, the RMSE values slightly decrease with increasing of the value of a. For example, when c is fixed at 6, the RMSE is 161.8 for $a = 1$, whereas it is 146.1 for $a = 1.4$ and 99.5 for $a = 2$. The explanation for this is that as the increasing of a, the model is less sensitive to outliers, so the prediction accuracy

Table 1. The experimental result of RMSE under typical clustering number c, fuzzification coefficient m, and steepness index a

Clustering number c	Steepness index a	RMSE with different fuzzification coefficient m					
		1.5	1.75	2	2.25	2.5	2.7
3	1	480.0	441.2	422.9	396.2	398.7	422.0
	1.2	476.5	431.7	432.2	396.0	400.9	417.9
	1.4	476.7	431.5	427.4	391.7	388.7	406.3
	1.6	472.7	430.9	428.1	390.9	388.3	410.6
	2	390.0	394.6	406.8	389.1	358.4	413.5
4	1	280.1	278.2	236.2	225.1	265.0	303.1
	1.2	270.9	269.1	230.4	220.2	264.5	281.3
	1.4	270.6	250.7	223.0	205.7	222.1	220.3
	1.6	266.6	248.0	225.0	202.4	220.2	214.4
	2	191.8	195.0	191.2	172.0	134.3	208.5
5	1	258.9	247.4	234.8	216.4	164.4	218.1
	1.2	253.5	242.9	227.4	206.9	220.7	222.5
	1.4	253.5	239.8	212.6	200.4	176.6	220.7
	1.6	244.0	236.1	187.4	185.4	200.1	210.7
	2	222.0	235.4	138.2	157.3	139.9	175.9
6	1	280.4	237.0	190.1	161.8	161.6	175.6
	1.2	268.0	228.7	187.2	155.7	155.7	171.4
	1.4	243.3	221.2	176.8	146.1	114.8	169.5
	1.6	353.9	220.0	175.8	130.8	104.6	167.2
	2	243.6	215.7	168.6	99.5	91.9	150.0
7	1	255.6	213.2	156.0	146.4	155.5	152.0
	1.2	256.6	218.7	151.7	126.3	133.6	151.6
	1.4	251.4	215.4	140.2	120.6	112.4	150.3
	1.6	213.3	214.8	135.0	106.3	104.8	121.3
	2	139.4	118.4	96.3	84.2	78.2	111.8
8	1	240.9	189.9	135.6	88.6	92.5	125.6
	1.2	242.6	188.6	127.9	86.2	87.2	124.2
	1.4	230.8	174.9	123.8	82.7	63.4	109.5
	1.6	230.3	169.9	121.8	82.4	40.1	104.7
	2	210.9	146.5	95.0	81.0	17.4	98.6
9	1	212.5	167.6	139.2	98.4	89.6	96.3
	1.2	204.7	141.1	112.4	88.1	78.9	92.4
	1.4	205.6	140.5	111.5	86.4	78.9	84.3
	1.6	203.8	114.6	110.3	58.5	70.6	81.3
	2	202.2	109.8	107.9	52.6	63.9	79.6
10	1	208.9	204.1	79.0	85.5	87.3	89.0
	1.2	204.2	187.6	48.0	84.5	85.5	87.9
	1.4	189.7	182.3	41.8	71.7	83.3	84.1
	1.6	205.1	180.0	36.8	16.6	34.4	80.1
	2	202.4	140.1	18.5	16.9	38.7	74.2

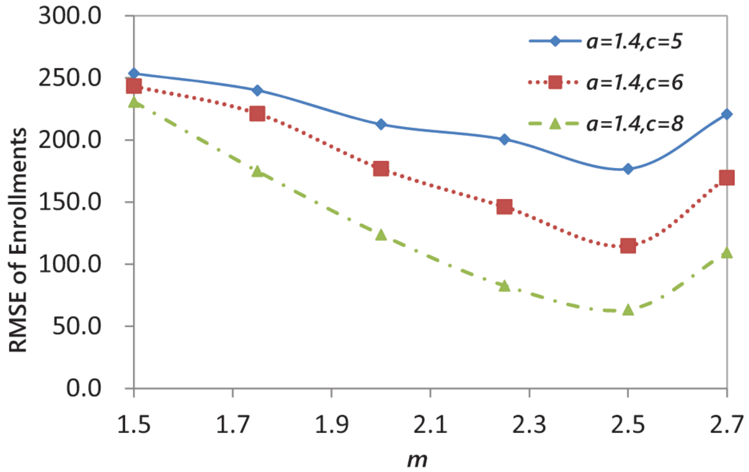


FIGURE 4. The influence of fuzzification coefficient m on average RMSE

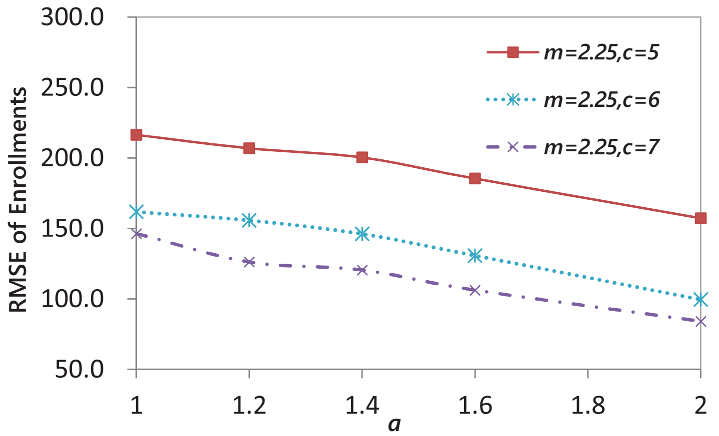


FIGURE 5. The influence of fuzzification coefficient a on smallest RMSE

may increase to some extent. Also we can see that when the value of a is fixed, the RMSE values decrease with increasing of the value of c.

4.1.2. *Prediction example.* Specially taking $c = 6$, $m = 2.5$, $a = 1.4$ as an example, the process of realizing the method is described in detail as follows. Here for comparison with other prediction methods proposed before, we take the whole enrollment time series data set into account directly to construct a fully learned FCM model and carry out prediction because the size is small relatively which is not suitable for division of two subsets separately. First of all, the prototypes of enrollment data set are extracted in the form of numerical value, where the improved version of fuzzy c-means clustering algorithm is applied to cluster. The framework of FCM is shown in Figure 6. Moreover, each prototype can be described in the form of a certain fuzzy semantics:

G_1 – the value of enrollments is lower, where locate around 13502;
G_2 – the value of enrollments is low, where locate around 15237;
G_3 – the value of enrollments is medium, where locate around 15838;
G_4 – the value of enrollments is high, where locate around 16389;
G_5 – the value of enrollments is higher, where locate around 16873;
G_6 – the value of enrollments is highest, where locate around 19109.

And then, Equation (14) is used to fuzzify each observation the original time series into the fuzzy time series $\{C(1), C(2), \ldots, C(21)\}$ in the form of Equation (6), where

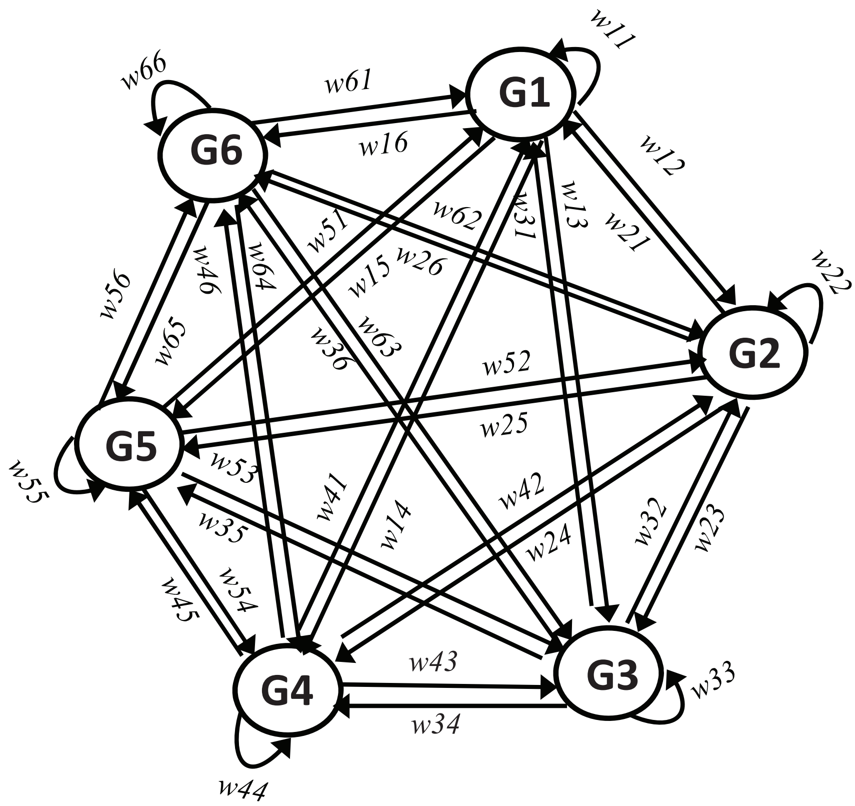


Figure 6. FCM model of enrollment series

fuzzification coefficient m is assigned to 2.25 and a is 1.4. After fuzzification, the original numeric value of time series transforms to membership matrix which expresses the fuzzy relationships between each measured data to concepts of FCM individually. Next, the standard version of PSO algorithm is used to learn all parameters W of FCM according to the fuzzy time series data $C(t)$ $(t = 1, 2, \ldots, 21)$. In this step, we define the particle structure as Equation (18).

$$W = \{w_1, w_2, w_{0i}, \sigma\}^T \tag{18}$$

where w_1 and w_2 are weights matrix for the second-order prediction, w_{0i} is biases vector, and σ is steepness vector. w_1, w_2, w_{0i}, and σ are detailed in Figure 6 respectively.

Finally, based on the former four steps, second-order prediction is accomplished as Equation (9) with 2 initial data points $C(t)$ $(t = 1, 2)$ for 19 data points, and then defuzzifying the predicted $\hat{C}(t)$ $(t = 3, \ldots, 21)$ to numeric value by Equation (17). Further RMSE is calculated to validate the accuracy of the method proposed in this paper.

Figure 7 shows the original time series and the predicted results at $c = 6$, $m = 2.5$ and $a = 1.4$. Table 2 reports the comparison results of the proposed prediction algorithm with other classic methods [17, 19, 20, 21, 22, 23, 24] based on fuzzy sets theory. The RMSE of the proposed methods can achieve 115, while the best result Yu's method [22] scored 295 for the time-invariant method and Hwang et al.'s method [19] scored 567 for the time-variant method. The proposed method can not only get higher prediction accuracy, but also automatically extract fuzzy features from original time series by using improved fuzzy c-means clustering, obtain the relationships among these fuzzy features by using PSO algorithm and carry out reasoning by using FCM without human intervention. The parameters of PSO algorithm are shown in Table 3. We also tried other parameter combinations, but no obvious improvement appeared, so we determined the parameters

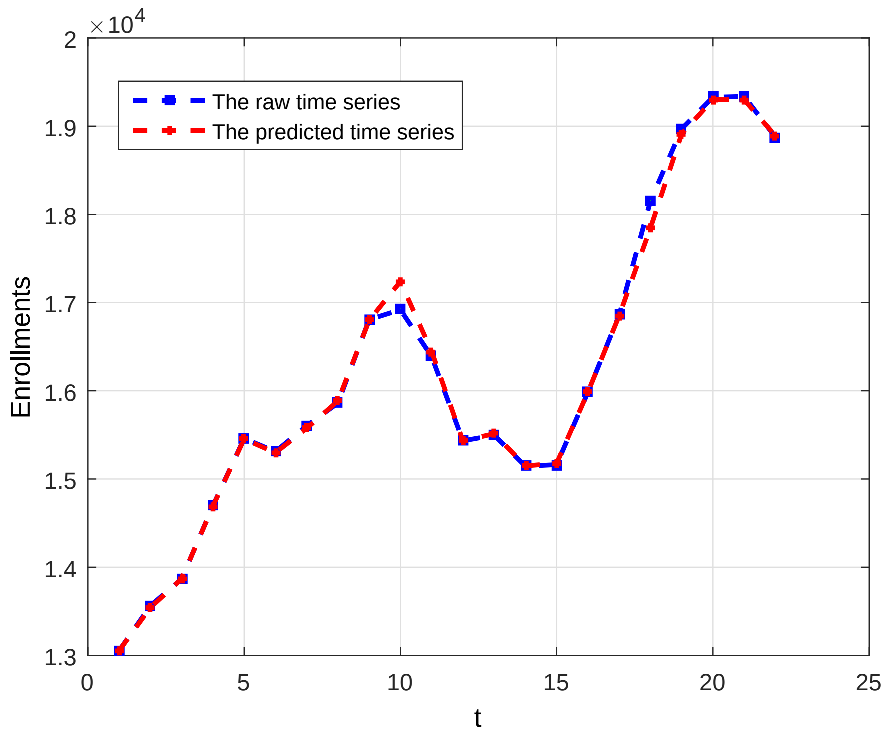


FIGURE 7. The original data and predicted data of enrollment time series ($c = 6$, $m = 2.5$, $a = 1.4$)

TABLE 2. The comparision results for enrollment data set

Methods	RMSE
Song and Chissom's method [17]	677
Chen's method [20]	663
Huarng's method [21]	489
Yu's method [22]	295
Sullivan and Woodall's method [23]	638
Singh's method [24]	1020
Hwang et al.'s method [19]	567
Proposed method ($c = 6$, $m = 2.5$, $a = 1.4$)	115

TABLE 3. The parameters of PSO algorithm for all experiment

Description	Value
Population size	50
Acceleration constant Φ_1	2
Acceleration constant Φ_2	2
Inertial weight	0.9
Initial positions	Random number
maximum number of iterations	1000
minimum objection function value	105

in Table 3 to get relatively good results.

$$w_1 = \begin{bmatrix} -0.80 & -0.40 & -0.70 & -0.30 & -0.90 & 1.00 \\ -1.00 & 0.80 & -0.80 & 0.20 & -0.90 & 0.00 \\ -1.00 & -0.90 & -0.30 & 0.10 & -0.50 & -0.30 \\ -1.00 & 0.60 & -1.00 & -0.10 & 0.10 & 0.00 \\ -0.90 & 0.20 & 1.00 & -0.60 & -1.00 & -0.20 \\ 0.60 & -1.00 & -1.00 & -1.00 & -0.90 & -0.20 \end{bmatrix}$$

$$w_2 = \begin{bmatrix} -0.20 & -0.60 & -0.90 & -1.00 & -0.80 & 1.00 \\ -1.00 & 1.00 & -1.00 & -0.30 & -0.70 & 1.00 \\ -1.00 & -0.90 & -0.50 & -1.00 & -0.90 & 0.60 \\ -1.00 & -0.80 & -0.60 & -0.90 & -0.90 & 0.10 \\ 0.20 & -0.80 & 0.70 & 0.40 & 0.40 & -0.80 \\ -0.20 & -0.30 & 0.30 & -1.00 & -0.80 & -0.30 \end{bmatrix} \tag{19}$$

$$w_{0i} = \begin{bmatrix} 0.00 \\ 0.10 \\ 0.00 \\ 0.00 \\ 0.20 \\ 0.00 \end{bmatrix} \quad \sigma = \begin{bmatrix} 4.0 \\ 4.0 \\ 4.0 \\ 4.0 \\ 4.0 \\ 4.0 \end{bmatrix}$$

4.2. **The TAIEX time series.** The TAIEX data set [17, 21, 22] concerns 242 observations of Taiwan Stock Exchange Capitalization Weighted Stock Index during January 1, 2000 to December 30, 2000, which is used to validate our proposed prediction method.

The same as the enrollments time series, the validation is accomplished with multiple experiments. Table 4 records the specific experimental results of RMSE.

4.2.1. *The impact of parameters c, m and a on the prediction accuracy.* According to Table 4, the maximum RMSE is 308.6 and the minimum one is 96.5. Therefore, the difference of the values of parameters c, a, and m influence the prediction accuracy. And from Figures 8 and 9, we can see that the impact trend of each parameter on the prediction accuracy is in the same way as the enrollments time series in Section 4.1.1. (1) When the value of m moves from low value to high value, the RMSE values show a downtrend at first, and after reaching the lowest point, show an uptrend. (2) The RMSE values decrease as the value of c increases. (3) The RMSE values slightly decrease as the value of a increases.

4.2.2. *Prediction example.* Specially taking $c = 6$, $m = 2.25$, $a = 1.4$ of the experiment as an example, the process of realizing the method is described in detail as follows.

The prototypes of TAIEX data are generated in the form of numerical value, where the improved version of fuzzy c-means clustering algorithm proposed before is applied to clustering entire TAIEX data set. Each prototype can be described in the form of a certain fuzzy semantics as follows:

G_1 – the value of TAIEX is lower, where locate around 5150.5;
G_2 – the value of TAIEX is low, where locate around 5878.8;
G_3 – the value of TAIEX is medium, where locate around 6904;
G_4 – the value of TAIEX is high, where locate around 8069.1;
G_5 – the value of TAIEX is higher, where locate around 8869.5;
G_6 – the value of TAIEX is highest, where locate around 9838.8.

Equation (14) is used to fuzzify each observation from the raw time series, which leads to form the fuzzy time series $\{C(1), C(2), \ldots, C(242)\}$ in the form of Equation (6). In

TABLE 4. The experimental result of RMSE under typical clustering number c, fuzzification coefficient m, and steepness index a

Clustering number c	Steepness index a	RMSE with different fuzzification coefficient m					
		1.5	1.75	2	2.25	2.5	2.7
3	1	308.6	291.2	282.8	276.6	284.3	289.9
	1.2	307.9	285.3	275.7	268.0	273.8	278.1
	1.4	301.0	280.0	268.0	261.6	265.1	267.4
	1.6	300.4	277.0	261.6	255.4	257.3	262.5
	2	294.5	270.2	256.9	248.7	249.2	252.0
4	1	251.1	239.0	223.8	214.0	222.5	222.1
	1.2	248.7	230.7	219.1	210.4	212.8	215.3
	1.4	248.2	227.7	215.0	208.5	192.3	195.9
	1.6	248.2	227.9	206.7	185.4	189.2	193.3
	2	247.3	224.2	187.4	179.2	185.6	189.0
5	1	186.8	168.2	167.1	163.2	164.3	172.3
	1.2	183.3	170.6	166.6	161.5	163.1	170.9
	1.4	182.6	169.0	165.2	160.3	161.5	165.7
	1.6	181.3	167.0	160.2	158.0	158.5	163.9
	2	178.3	165.3	157.9	151.3	150.5	162.5
6	1	183.6	157.9	158.3	151.5	150.8	151.1
	1.2	179.8	153.2	157.2	147.8	149.9	150.5
	1.4	173.3	152.8	151.7	146.9	148.1	150.3
	1.6	168.1	151.9	148.2	141.8	146.0	169.9
	2	167.8	143.9	141.1	131.1	134.5	147.2
7	1	152.7	136.1	134.9	131.6	139.5	149.6
	1.2	144.8	135.1	135.5	130.1	141.0	148.9
	1.4	149.8	137.1	135.0	130.0	139.8	147.4
	1.6	150.3	139.1	136.4	128.5	137.9	147.0
	2	141.8	128.1	121.7	118.9	130.3	140.0
8	1	149.8	137.1	135.5	128.5	140.7	150.9
	1.2	148.8	136.2	135.3	128.1	140.2	148.2
	1.4	148.2	136.1	134.9	126.3	139.8	144.4
	1.6	141.9	129.6	126.4	119.2	133.5	140.0
	2	131.2	129.7	124.6	115.9	132.7	136.5
9	1	161.3	137.1	127.2	122.4	143.6	164.4
	1.2	148.1	124.1	124.4	130.0	137.1	149.4
	1.4	149.0	137.7	134.6	129.5	131.1	138.9
	1.6	126.8	135.1	135.9	128.5	131.4	143.6
	2	115.1	117.5	113.9	112.1	121.2	131.8
10	1	137.1	131.8	126.1	124.0	124.0	145.8
	1.2	136.7	124.9	120.5	123.5	124.0	144.3
	1.4	136.6	124.3	110.0	120.3	126.8	140.8
	1.6	125.5	124.1	111.9	109.3	107.9	126.5
	2	124.1	124.0	111.0	96.5	110.0	125.0

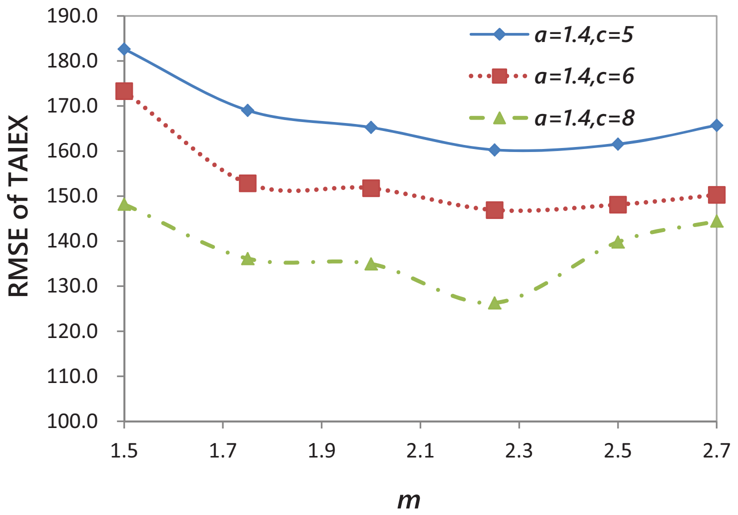


FIGURE 8. The influence of fuzzification coefficient c on average RMSE

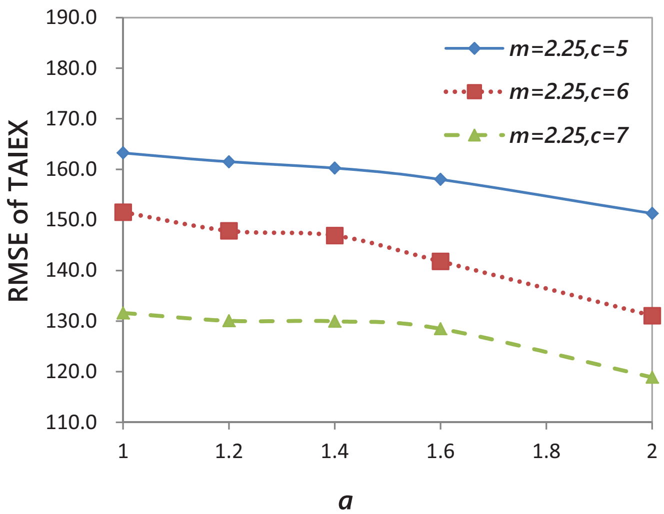


FIGURE 9. The influence of fuzzification coefficient a on smallest RMSE

order to compare with other methods based on fuzzy set theory directly, the formed fuzzy time series data is divided into two subset: one consisting of the first 195 vectors $\{C(1), C(2), C(195)\}$ is regarded as training set which is used to construct a fully learned FCM, and the other consisting of the last 47 vectors $\{C(196), C(197), C(242)\}$ is taken as the testing set which is used to carry out numerical prediction. The standard version of PSO algorithm is used to learn the weights of FCM for constructing a fully learned FCM according to training dataset. The parameters of PSO algorithm and the framework are the same as in enrollment dataset prediction, and corresponding relationship matrix of a fully learned FCM is shown in Equation (20).

$$w_1 = \begin{bmatrix} 0.7515 & -0.2970 & 0.0350 & -0.3448 & 0.9877 & 0.0326 \\ -0.1671 & 0.4599 & 0.0587 & 0.1588 & 0.0359 & 0.6457 \\ -0.9971 & -0.2109 & 0.3485 & -0.0786 & 0.1143 & 0.1445 \\ -0.3468 & -0.6086 & -0.0411 & 0.6138 & -0.2344 & -0.0410 \\ -0.2327 & 0.4762 & -0.8361 & 0.9940 & -0.2933 & -0.1627 \\ 0.0735 & 0.2956 & -0.1793 & -0.2026 & -0.9999 & -0.1770 \end{bmatrix}$$

$$w_2 = \begin{bmatrix} 0.9935 & -0.4816 & -0.2890 & 1.0000 & 1.0000 & -0.3348 \\ -0.6690 & 0.9970 & -0.0544 & -0.2108 & -0.8317 & 0.0382 \\ -0.8902 & -0.0694 & -0.7688 & 0.0387 & -0.2361 & 0.9966 \\ -0.0553 & -1.0000 & -0.0491 & 1.0000 & 0.1414 & -0.0582 \\ -0.0820 & -0.9712 & -0.5484 & -0.2285 & 0.9990 & 0.7220 \\ -0.4356 & -0.6489 & 0.0659 & 0.0648 & 0.1717 & 0.9999 \end{bmatrix}$$

$$w_{0i} = \begin{bmatrix} 0.0882 \\ 0.0446 \\ 0.0865 \\ 0.0681 \\ 0.0448 \\ 0.0768 \end{bmatrix} \qquad \sigma = \begin{bmatrix} 1.1735 \\ 1.3403 \\ 2.3861 \\ 3.9988 \\ 4.0 \\ 1.8175 \end{bmatrix} \tag{20}$$

Figure 10 shows the original TAIEX time series and the predicted results at the value of $c = 6$, $m = 2.25$, $a = 1.4$ for training and test data. Table 5 compares the RMSE of the proposed prediction method with other classic methods based on fuzzy sets theory. The result in Table 5 shows that the proposed prediction method can obtain better accuracy. The RMSE of the proposed methods can achieve 125, while the best result of other method is scored 138 of Huarng's method [21]. Although comparison shows that the proposed method has similar prediction precision with Huarng's method for the TAIEX

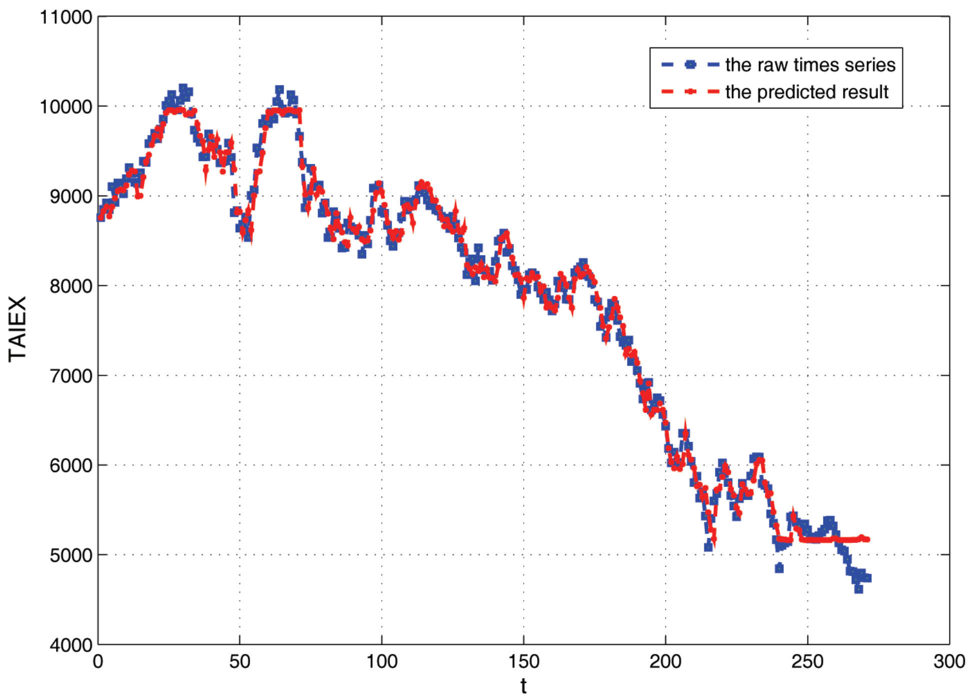


Figure 10. The original data and predicted data of TAIEX time series ($c = 6$, $m = 2.25$, $a = 1.4$)

Table 5. The comparision results for TAIEX data set

Methods	RMSE
Chen's method [20]	176
Huarng's method [21]	128
Yu's method [22]	170
Proposed method ($c = 6$, $m = 2.25$, $a = 1.4$)	125

time series, the potential advantage is the ability of processing the prediction problem of long-term or complex time series, and avoiding tedious and complex process for mining fuzzy logical relationships. Actually, the proposed prediction method exploits improved fuzzy c-means clustering to mine fuzzy logical relationships implied in time series, fuzzy cognitive map to express these fuzzy logical relationships and carry out prediction by its iteration computation. The entire process is automatic and depends on data of time series.

4.3. **The Wolf's sunspot time series.** The Wolf's sunspot time series is concerned with the annual number of sunspots from the year of 1700 to 1987, 288 observations totally. The time series is regarded as nonlinear and non-Gaussian and as such is often used to evaluate the effectiveness of some hybrid models. The hybrid models based on ARIMA, say ARIMA-ANN models [25, 26], ARIMA-SVM models [27, 28] and ARIMA-GP (combining ARIMA with genetic programming) [29], are used to carry out a comparative analysis. For this time series, the first 221 observations (data from the year of 1700 to 1920) are regarded as the training sunset and the last 67 observations (data from the year of 1921 to 1987) are used as the testing subset. According to the previous study findings, we use the method detailed in Section 5 to respectively construct the proposed FCM prediction model with $c = 6$, $m = 2.25$, $a = 1.4$ and $c = 9$, $m = 2.25$, $a = 1.4$ for modeling the time series and performing prediction. The predicted results are compared with those produced by the hybrid models based on ARIMA, refer to Table 6.

Table 6. The comparison with hybrid models base ARIMA for the Wolf's sunspot time series

Methods	MSE
ARIMA-ANN1 [25]	280.2
ARIMA-ANN2 [26]	268.7
ARIMA-SVM [27]	259.3
ARIMA-LSSVM [28]	283.6
ARIMA-GP [29]	265.1
Proposed method ($c = 6$, $m = 2.25$, $a = 1.4$)	181.6
Proposed method ($c = 9$, $m = 2.25$, $a = 1.4$)	122.6

From Table 6, we see that the prediction accuracies of the proposed prediction models with $c = 6$ and $c = 9$ are better than other models based on ARIMA, and also it has the advantage of high interpretability. Figure 11 presents the original sunspot time series and the predicted results obtained by the prediction model with $c = 9$ which is constructed by our approach.

5. **Conclusions.** A new time series prediction method based on fuzzy cognitive map is presented in this paper. Two important components of this prediction method are improved fuzzy c-means clustering algorithm and fuzzy cognitive map. The former extracts the numeric feature from original time series automatically and converts original time series into fuzzy time series and the latter is used to model fuzzy time series by PSO algorithm and realize the prediction. Three benchmark time series: the enrollments time series, the TAIEX time series and Wolf's sunspot time series are used to validate effectiveness and feasibility of the proposed method. The results of experiments show that the proposed prediction method can get better prediction accuracy. Further, by analyzing the details of prediction results of the first two benchmark time series, several interesting conclusions are drawn about the impacts of clustering number c, fuzzification coefficient

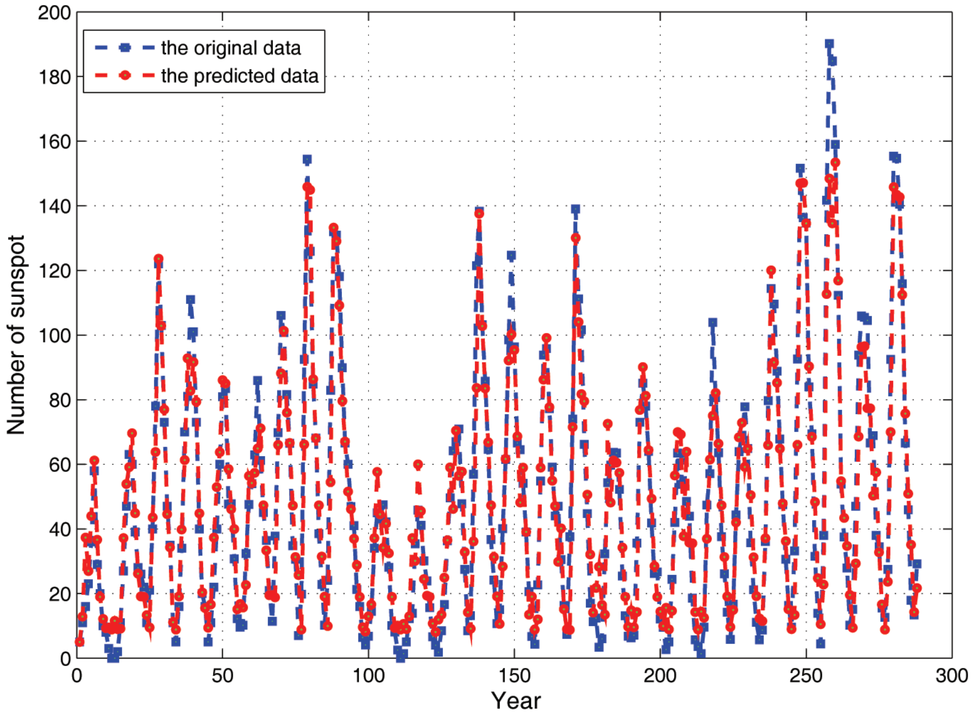


FIGURE 11. The original data and predicted data of sunspot time series ($c = 9$, $m = 2.25$, $a = 1.4$)

m, and steepness index a. Additionally, the potential advantages of the proposed prediction method are the abilities of handling the prediction problem of long-term or complex time series, at the same time avoiding tedious and complex work in the process of mining fuzzy logical relationships.

REFERENCES

[1] B. Kosko, Fuzzy cognitive maps, *International Journal of Man-Machine Studies*, vol.24, no.8, pp.65-75, 1986.

[2] C. D. Stylios and P. P. Groumpos, Modeling complex systems using fuzzy cognitive maps, *IEEE Trans. Systems, Man and Cybernetics, Part A*, vol.34, no.1, pp.155-162, 2004.

[3] W. Pedrycz, A. Jastrzebska and W. Homenda, Design of fuzzy cognitive maps for modeling time series, *IEEE Trans. Fuzzy Systems*, vol.24, no.1, pp.120-130, 2016.

[4] G. Napoles, I. Grau, R. Bello and R. Grau, Two-steps learning of fuzzy cognitive maps for prediction and knowledge discovery on the HIV-1 drug resistance, *Expert Systems with Applications*, vol.41, no.3, pp.821-830, 2014.

[5] W. Froelich and J. L. Salmeron, Evolutionary of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series, *International Journal of Approximate Reasoning*, vol.55, no.6, pp.1319-1335, 2014.

[6] W. Lu, J. Yang, X. Liu and W. Pedrycz, The modeling and prediction of time series based on synergy of high-order fuzzy cognitive map and fuzzy c-means clustering, *Knowledge-Based Systems*, vol.70, pp.242-255, 2014.

[7] S. Mei, Y. Zhu, X. Qiu, X. Zhou, Z. Zu, A. V. Boukhanovsky and P. M. Sloot, Individual decision making can drive epidemics: A fuzzy cognitive map study, *IEEE Trans. Fuzzy Systems*, vol.22, no.2, pp.264-273, 2014.

[8] S. Lee, J. Yang and J. Han, Development of a decision making system for selection of dental implant abutments based on the fuzzy cognitive map, *Expert Systems with Applications*, vol.39, no.14, pp.11564-11575, 2012.

[9] G. Nápoles, E. I. Papageorgiou, R. Bello and K. Vanhoof, Learning and convergence of fuzzy cognitive maps used in pattern recognition, *Neural Processing Letters*, vol.45, no.2, pp.431-444, 2017.

[10] E. I. Papageorgiou, C. Stylios and P. Groumpos, Fuzzy cognitive map learning based on nonlinear Hebbian rule, *Australasian Joint Conference on Artificial Intelligence*, pp.256-268, 2003.

[11] W. Stach, L. Kurgan, W. Pedrycz and M. Reformat, Genetic learning of fuzzy cognitive maps, *Fuzzy Sets and Systems*, vol.153, no.3, pp.371-401, 2005.

[12] E. I. Papageorgiou, K. E. Parsopoulos, C. S. Stylios, P. P. Groumpos and M. N. Vrahatis, Fuzzy cognitive maps learning using particle swarm optimization, *Journal of Intelligent Information Systems*, vol.25, pp.95-121, 2005.

[13] X. Zou and J. Liu, A mutual information based two-phase memetic algorithm for large-scale fuzzy cognitive map learning, *IEEE Trans. Fuzzy Systems*, 2017.

[14] S. Ahmadi, N. Forouzideh, S. Alizadeh and E. I. Papageorgiou, Learning fuzzy cognitive maps using imperialist competitive algorithm, *Neural Computing and Applications*, vol.26, no.6, pp.1333-1354, 2015.

[15] Y. Chi and J. Liu, Learning of fuzzy cognitive maps with varying densities using a multi-objective evolutionary algorithm, *IEEE Trans. Fuzzy Systems*, vol.24, no.1, pp.71-81, 2016.

[16] J. C. Bezdek, R. Ehrlich and W. Full, FCM: The fuzzy C-means clustering algorithm, *Computers & Geosciences*, vol.10, no.2, pp.191-203, 1984.

[17] Q. Song and B. S. Chissom, Forecasting enrollments with fuzzy time series – Part I, *Fuzzy Sets and Systems*, vol.54, no.1, pp.1-9, 1993.

[18] Q. Song and B. S. Chissom, Forecasting enrollments with fuzzy time series – Part II, *Fuzzy Sets and Systems*, vol.62, no.1, pp.1-8, 1994.

[19] J. R. Hwang, S. M. Chen and C. H. Lee, Handling forecasting problems using fuzzy time series, *Fuzzy Sets and Systems*, vol.100, nos.1-3, pp.217-228, 1998.

[20] S. M. Chen, Forecasting enrollments based on fuzzy time series, *Fuzzy Sets and Systems*, vol.81, no.3, pp.311-319, 1996.

[21] K. Huarng, Heuristic models of fuzzy time series for forecasting, *Fuzzy Sets and Systems*, vol.123, no.3, pp.369-386, 2001.

[22] H. K. Yu, A refined fuzzy time-series model for forecasting, *Physica A: Statistical Mechanics and Its Applications*, vol.346, nos.3-4, pp.657-681, 2005.

[23] J. Sullivan and W. H. Woodall, A comparison of fuzzy forecasting and Markov modeling, *Fuzzy Sets and Systems*, vol.64, no.3, pp.279-293, 1994.

[24] S. R. Singh, A simple time variant method for fuzzy time series forecasting, *Cybernetics and Systems*, vol.38, no.3, pp.305-321, 2007.

[25] G. P. Zhang, Time series forecasting using a hybrid arima and neural network model, *Neurocomputing*, vol.50, no.1, pp.159-175, 2003.

[26] O. Valenzuela, I. Rojas, F. Rojas, H. Pomares, L. J. Herrera, A. Guillén, L. Marquez and M. Pasadas, Hybridization of intelligent techniques and ARIMA models for time series prediction, *Fuzzy Sets and Systems*, vol.159, no.7, pp.821-845, 2008.

[27] P. F. Pai and C. S. Lin, A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega*, vol.33, no.6, pp.497-505, 2005.

[28] B. Zhu and Y. Wei, Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology, *Omega*, vol.41, no.3, pp.517-524, 2013.

[29] Y. S. Lee and L. I. Tong, Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming, *Knowledge-Based Systems*, vol.24, no.1, pp.66-72, 2011.