Computers and Mathematics with Applications 56 (2008) 3052–3063

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

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

A FCM-based deterministic forecasting model for fuzzy time series

Sheng-Tun Li [a,b,*], Yi-Chung Cheng [b,c], Su-Yu Lin [a]

[a] *Institute of Information Management, National Cheng Kung University, Taiwan, ROC*

[b] *Department of Industrial and Information Management, National Cheng Kung University, Taiwan, ROC*

[c] *Department of International Business Management, Tainan University of Technology, Taiwan, ROC*

ARTICLE INFO

Article history:
Received 21 December 2007
Received in revised form 25 May 2008
Accepted 10 July 2008

Keywords:
Fuzzy time series
Forecasting
Interval partitioning
Fuzzy logical relationship
Monte Carlo simulation

ABSTRACT

The study of fuzzy time series has increasingly attracted much attention due to its salient capabilities of tackling uncertainty and vagueness inherent in the data collected. A variety of forecasting models including high-order models have been devoted to improving forecasting accuracy. However, the high-order forecasting approach is accompanied by the crucial problem of determining an appropriate order number. Consequently, such a deficiency was recently solved by Li and Cheng [S.-T. Li, Y.-C. Cheng, Deterministic Fuzzy time series model for forecasting enrollments, Computers and Mathematics with Applications 53 (2007) 1904–1920] using a deterministic forecasting method. In this paper, we propose a novel forecasting model to enhance forecasting functionality and allow processing of two-factor forecasting problems. In addition, this model applies fuzzy *c*-means (FCM) clustering to deal with interval partitioning, which takes the nature of data points into account and produces unequal-sized intervals. Furthermore, in order to cope with the randomness of initially assigned membership degrees of FCM clustering, Monte Carlo simulations are used to justify the reliability of the proposed model. The superior accuracy of the proposed model is demonstrated by experiments comparing it to other existing models using real-world empirical data.

1. Introduction

The forecasting problem of time series data, a series of data ordered in time sequence segmented by fixed time intervals [2], is an interesting and important research topic. In various disciplines it has been commonly tackled by using a variety of approaches such as statistics, artificial neural networks, etc. Traditional time series forecasting models are usually extensively dependent on historical data, which can be incomplete, imprecise and ambiguous. If these uncertainties were widespread in real-world data, they could hinder forecasting accuracy, thus limiting the applicability of forecasting models.

Song and Chissom [3–5] first introduced the definition of fuzzy time series, which is capable of dealing with incomplete and vague data under uncertain circumstances by applying the theory of fuzzy logic. Fuzzy time series differs from traditional time series in that it is represented as linguistic values instead of numeric values. Since its emergence, there has been much research devoted to improving forecasting performance, which has resulted in significant achievements. Song and Chissom explored forecasting of fuzzy time series with enrollment data of the University of Alabama and proposed a forecasting framework composed of four steps: (1) determining and partitioning the universe of discourse, (2) defining the fuzzy sets on the universe of discourse and fuzzifying the time series, (3) constructing fuzzy relationships from the fuzzified time series, and (4) forecasting and defuzzifying the forecasting outputs.

* Corresponding address: No. 1, Ta-Hsueh Road, Tainan 701, Taiwan, ROC. Tel: +886 6 2757575x53126; fax: +886 6 2362162.
E-mail address: stli@mail.ncku.edu.tw (S.-T. Li).

0898-1221/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2008.07.033

Since the work of Song and Chissom, numerous studies have been conducted to improve forecasting accuracy or reduce computational overhead. Song and Chissom [3–5] developed a fuzzy time series forecasting model by applying fuzzy relations to represent the temporal relationships of linguistic values in a fuzzy time series. In order to facilitate forecasting, these fuzzy relations are constructed as relational matrices. Sullivan and Woodall [6] later established a first-order time-invariant fuzzy time series forecasting model based on the Markov model. Chen [7] modified Song and Chissom's model [5] and alleviated the computational overload of matrix manipulation by developing first-order fuzzy logic relationship rules. By integrating domain-specific heuristic knowledge used in Chen's model [7], Huarng proposed heuristic time-invariant fuzzy time series forecasting models [8].

Chen [9] extended his previous work [7] to a high-order time-invariant fuzzy time series forecasting model, while Hsu et al. [10] applied fuzzy Markov relation matrix to fuzzy time series forecasting. Later, Own and Yu [11] extended Chen's model [9] as a heuristic high-order fuzzy time series model, which depends strongly on the trend of fuzzy time series. Huarng and Yu [12,13] applied neural networks and type 2 fuzzy set to refine the fuzzy relationships and further build up forecasting accuracy. Yu proposed models of refinement relation [14] and weighting scheme [15] to enhance forecasting accuracy. Lee et al. [16] also extended Chen's [9] one-variable high-order time-invariant fuzzy time series forecasting model to allow two-factor forecasting. Data of daily average temperature and daily cloud density in Taipei from June to September, 1996 were used to test the two-factor model. These high-order models come with the major hurdle of determining an appropriate order number, k, and order redundancy [1]. Such a hurdle was addressed recently by Li and Cheng [1]; however, its applicability is rather limited since it can only deal with one-factor forecasting problems.

All previously mentioned works mainly focused on improving steps (3) and (4) of Song and Chissom's framework. However, forecasting performance can also be significantly affected by interval partitioning [17]. Huarng [17] investigated the impact of interval length on the forecasting results and proposed two heuristic approaches, namely, distribution- and average-based to determine the interval length. Unfortunately, the method of the 'base-mapping table' was not disclosed. Li and Chen [18] used a natural partition-based approach to study the effect of interval partitioning. Chen and Hsu [19] used a two-phase partitioning method based on the statistical distributions of the historical data. However, it usually resulted in numerous intervals, which was unrealistic for real-world applications and could result in fewer fluctuations in the fuzzy time series, as Huarng [17] indicated. Moreover, it complicated the task of defuzzification. Huarng and Yu [20] proposed 'ratio-based' length of intervals to improve forecasting results. Chen and Chung [21] and Lee, Wang and Chen [22] used genetic algorithms (GA) to adjust the length of each interval, for one-factor and two-factor high-order forecasting models, respectively.

In spite of the aforementioned approaches, the most commonly used method for interval partitioning, perhaps, is equal-width partition [1,3,5–7,9,11,16,23], however when the distribution of the continuous values is not uniform, it might not yield good results [24]. In addition, it has been recently shown that using unequal-sized intervals might produce better forecasting accuracy than traditional equal-width partitioning [22].

In this paper, we propose a novel forecasting model to remedy these two deficiencies. First, we apply fuzzy c-means (FCM) clustering, proposed by Bezdek [25], an effective discretization approach in data mining that handles the issue of interval partitioning since it can take the density of data points into account and produce unequal-sized intervals. In contrast to the traditional hard k-means and GA clustering, FCM allows each data point to belong to more than one interval with membership degrees so that the uncertainty inherent in the time series data can be dealt with appropriately. Second, since events in the real world can be affected by many factors, we extend our previous work [1] to allow processing two-factor forecasting problems so that better forecasting results can be obtained [26]. Bivariate fuzzy time series models have proven to yield better forecasting results than univariate models [10,16,27]. In addition, due to the randomness of the initially assigned membership degrees of FCM clustering, the model reliability and the distribution of forecasting values are analyzed by box plot and Monte Carlo simulation, a well-recognized statistical simulation method [28]. It is possible to simulate natural phenomena by iteratively evaluating a deterministic model using sets of random numbers and probability distribution [28,29]. With the availability of Monte Carlo simulation results, the reliability of the forecasting model can be justified by box plot.

There are five sections in this paper. A brief introduction to fuzzy time series is given in Section 2. In Section 3, a FCM-based deterministic forecasting model is proposed and discussed. Section 4 presents performance evaluation by comparing our model with other models and analyzes model reliability through distribution of forecasting values. The last section describes our conclusions and future works.

2. Fuzzy time series

Song and Chissom [4] first introduced the definition of fuzzy time series as follows.

Definition 1. Let $Y(t) \in R^1$ $(t = 0, 1, 2, \ldots)$ be the universe of discourse on which fuzzy sets $f_i(t)$ $(i = 1, 2, \ldots)$ are defined and let $F(t)$ be a collection of $f_i(t)$. Then, $F(t)$ is called a fuzzy time series on $Y(t)$ $(t = 0, 1, 2, \ldots)$.

Song and Chissom [4] defined fuzzy relations among fuzzy time series, which are based on the assumption that the values of fuzzy time series $F(t)$ are fuzzy sets, and the observation of time t is caused by the observations of the previous times.

Table 1
FCM-based deterministic forecasting procedure

Begin
Step 1: *Perform interval partitioning for each factor and fuzzify each time series using FCM* (Section 3.1)
Step 2: *Construct certain transition rules from fuzzy time series using Algorithm CTR* (Section 3.2)
Step 3: *Forecast in accordance with certain transition rules using Algorithm Forecast* (Section 3.3)
Step 4: *Defuzzify the forecasting outputs using* Eq. (11) (Section 3.3)
End

Definition 2. If for any $f_j(t) \in F(t)$, there exist $f_i(t-1) \in F(t-1)$ and a fuzzy relation $R_{ij}(t, t-1)$ such that $f_j(t) = f_i(t-1) \circ R_{ij}(t, t-1)$, where '$\circ$' is the max–min composition, then $F(t)$ is said to be caused by $F(t-1)$ only. Denote this as

$$f_i(t-1) \to f_j(t) \quad \text{or} \quad F(t-1) \to F(t). \tag{1}$$

Definition 3. If for any $f_j(t) \in F(t)$, there exist $f_i(t-1) \in F(t-1)$ and a fuzzy relation $R_{ij}(t, t-1)$ then $f_j(t) = f_i(t-1) \circ R_{ij}(t, t-1)$. Let $R(t, t-1) = \cup_{i,j} R_{ij}(t, t-1)$ where '$\cup$' is the union operator, then $R(t, t-1)$ is called the fuzzy relation between $F(t)$ and $F(t-1)$. Thus, a fuzzy relational equation is defined as $F(t) = F(t-1) \circ R(t, t-1)$.

The first-order and kth-order models are formed by applying the operators 'or' and 'and' of fuzzy relation equations.

Definition 4. Suppose $F(t)$ is caused by $F(t-1)$ only or by $F(t-1)$ or $F(t-2)$ or ... or $F(t-k)$, $k > 0$. This relation is denoted as $F(t-1) \to F(t)$ or $F(t-2) \to F(t)$ or ... or $F(t-k) \to F(t)$ and can be expressed as follows:

$$F(t) = F(t-1) \circ R(t, t-1)$$

or

$$F(t) = (F(t-1) \cup F(t-2) \cup \cdots \cup F(t-k)) \circ R_o(t, t-k) \tag{2}$$

where '$\cup$' is the union operator and '$\circ$' is the composition. $R(t, t-1)$ is called the fuzzy relation between $F(t)$ and $F(t-1)$. $R_o(t, t-k)$ is defined as a fuzzy relationship between $F(t)$ and $F(t-1)$ or $F(t-2)$ or ... or $F(t-k)$. Eq. (2) is called the first-order model of $F(t)$.

Definition 5. Suppose that $F(t)$ is caused by $F(t-1), F(t-2), \ldots$, and $F(t-k) (k > 0)$ simultaneously. Then their relations can be represented as:

$$F(t) = (F(t-1) \times F(t-2) \times \cdots \times F(t-k)) \circ R_a(t, t-k). \tag{3}$$

Eq. (3) is called the kth-order model of $F(t)$, and $R_a(t, t-k)$ is a relation matrix expressing the fuzzy relationship between $F(t-1), F(t-2), \ldots, F(t-k)$, and $F(t)$.

3. FCM-based deterministic forecasting model

We now propose the FCM-based deterministic forecasting model, which is outlined in Table 1 and discussed in detail in the subsequent subsections.

3.1. FCM-based interval partitioning and fuzzifying the time series

Interval partitioning is a process of partitioning continuous values of a variable into a set of intervals. It is related to the discretization and has been an important issue widely studied in the fields of machine learning and data mining [24,30]. The major issues of discretization include determining the number of intervals, allocating an observed value to an interval, and choosing the representative value for each interval [30]. In recent years, a variety of discretization techniques tackling these issues have been developed. Dougherty, Kohavi, and Sahami [30] classified the techniques along three axes: unsupervised vs. supervised, global vs. local, and static vs. dynamic. Unsupervised discretization differs from supervised discretization in that no class information is available, which is where fuzzy time series can be incorporated.

In most of the fuzzy time series literature, the universe of discourse was partitioned into equal-width intervals, which forms the most popular unsupervised discretization method. However, there are two major problems with interval partitioning. First, some of the parameters used for partitioning were selected arbitrarily. For instance, Song and Chissom [3,5], Chen [7,9], and Lee et al. [16] defined the universe of discourse U as $U = [D_{\min} - D_1, D_{\max} + D_2]$, where $D_{\min}$ and $D_{\max}$ were the minimal and maximal values in the historical data, and D_1 and D_2 proper positive numbers. However, how D_1 and D_2 were determined was not explained. The other problem is that equal-width interval partitioning may not produce good results in cases where the distribution of continuous values is not uniform [24]. In this paper, we turn to the unsupervised discretization method, unsupervised clustering [30], and generating clusters where the similarity within each cluster is minimized and the dissimilarity between clusters is maximized. In particular, the FCM partitioning method is used since it takes the distribution of data points and uncertainty into account, and assigns the membership degrees of clusters for each data point belonging to it. The resultant clusters are unequally sized without the overhead of determining any meaningless parameters in advance. The only information required is the appropriate number of clusters that can be usually obtained from the advice of an expert in the domain of interest.

The purpose of FCM is to minimize the following objective function [25]:

$$J_w(\gamma, M) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\gamma_{ij}^{a}) \left\| x_i - m_j \right\|^2, \tag{4}$$

where x_i is the ith element of a data set $\{x_1, x_2, \ldots, x_n\}$, $x_i \in R$. c is the number of clusters, $c \in \{2, 3, \ldots, n-1\}$. w is a weighted constant and $a \in (1, \infty)$. $\gamma = [\gamma_{ij}]$, where γ_{ij} is the degree of membership of x_i belonging to cluster j. $M = \{m_1, m_2, \ldots, m_c\}$, where m_j is the center of cluster j. $\|*\|$ is the similarity measure between x_i and m_j. The objective function is iteratively minimized. For the pth iteration, the membership degree matrix γ and the cluster centroid M are updated as follows:

$$\gamma_{ij}^{(p+1)} = \left[\sum_{k=1}^{c} \left(\frac{\left\| x_i - m_j^{(p)} \right\|}{\left\| x_i - m_k^{(p)} \right\|} \right)^{\frac{2}{a-1}} \right]^{-1} \tag{5}$$

$$m_j^{(p)} = \frac{\sum_{i=1}^{n} (\gamma_{ij}^{(p)})^a x_i}{\sum_{i=1}^{n} (\gamma_{ij}^{(p)})^a}, \quad 1 \le j \le c. \tag{6}$$

The iteration stops until $\left\| J_a^{(p+1)} - J_a^{(p)} \right\| < \delta$, where δ is the minimum amount of improvement.

Upon finding the interval (cluster) centroids, one may proceed to defining fuzzy sets as A_i $(i = 1, 2, \ldots, c)$ as:

$$A_i = \sum_{j=1}^{c} \mu_{ij}/m_j, \tag{7}$$

where μ_{ij} is the membership degree of m_j belonging to A_i, which is defined by

$$\mu_{ij} = \begin{cases} 1 & \text{if } j = i \\ 0.5 & \text{if } j = i-1 \text{ or } i+1 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

To fuzzify the raw time series, for all the elements x_i in X, we calculate the membership degree of x_i belonging to m_j by (9), for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, c$.

$$\gamma_{ij} = \left[\sum_{k=1}^{c} \left(\frac{\|x_i - m_j\|}{\|x_i - m_k\|} \right)^{\frac{2}{a-1}} \right]^{-1}. \tag{9}$$

Then x_i is fuzzified as A_j, where γ_{ij} is the maximum. The second factor is processed in a similar way.

3.2. Constructing certain transition rules

Li and Cheng [1] pioneered the study on the issue of uncertainty existing in a group of fuzzy relationships. They found that forecasting errors could result from uncertainty and proposed a deterministic forecasting model to improve forecasting accuracy. They developed a novel algorithm based on state transitions and backtracking process for constructing a set of certain transition rules of fuzzy relationships. In other words, the next state of any state transition is deterministically decided.

Fig. 1 illustrates the concepts of state transition and backtracking process. It indicates that state $F(t)$ is reached when state $F(t-1)$ moves forward in a one-time step with edge A_j. Each state s may have more than one state transition leaving it; in this case, s is named an uncertain state, otherwise it is a certain state. The state transition resulting in a certain state is called 'certain transition'. To eliminate uncertainties, a backtracking action is invoked (denoted as the negation '-' sign in Fig. 1) to find the previous state of s, i.e. a fuzzy time series beginning at A_i followed by $F(t-1)$.

Li and Cheng [1] developed an effective algorithm for identifying certain transition rules for a given fuzzy time series, however it can only handle one-factor problems. For real-world problems, an event can be affected by a number of factors [16], so considering more factors would mean a better representation of the problem closer to reality and

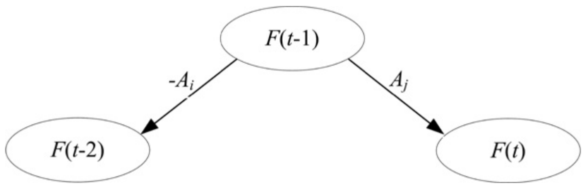


Fig. 1. State transition and backtracking.

Table 2
Certain transition rule algorithm (Algorithm CTR)

Input: A fuzzy time series $F = f_1 f_2 \ldots f_t \ldots f_n$.
Output: The set of certain transition rules, P
Begin
Let C be a subset of subseries of F;
Let S be a subset of symbols that appear in F;
***For** each element f in fuzzy series F*
If** $f \notin C$ **Then
Add f to C;
End If
End For
Add a special symbol, (%, %), at the beginning and end of series F;
***For** each element c in C*
Delete c from C;
Set S to be an empty set;
Search for all the subseries in F matching c exactly;
Add all distinct symbols following these subseries to S;
If** the cardinality of S is equal to 1 **Then
Add a new certain transition rule $\{c \to S\}$ into the rule set P;
***Else For** all these subseries*
Extend each subseries by adding its previous symbol to it and add it to C;
End For
End If
End For
***Return** P*;
End

consequently, better forecasting results. Several studies concern two-factor forecasting problems, such as temperature forecasting [16,22,23], stock index forecasting [10,12], financial prediction problems [31], and disruption prediction in Tokamak reactors [32]. In this study we extend the algorithm of Li and Cheng to allow processing two-factor forecasting. A concise representation of constructing the Certain Transition Rule set (Algorithm CTR) is illustrated in Table 2. The resultant certain transition rule set, P, contains the fuzzy relationship represented as production rules in the form of $c \to S$, where c and S are the cause and effect of the state transition, respectively. Therefore, P is known as a 'cause–effect' set in [1]. Each element in C is a candidate which could cause a state transition in F. One needs to determine if the element causes an uncertain transition, and if so, backtracking is invoked and C is updated by adding the new candidate state, which is the concatenation of the previous state of the element with the element itself. In the case where the element leads to a certain state, P will be added to the certain transition rule and the element will be removed from C. The resultant rule set P will serve as a knowledge base in the forecasting stage.

We demonstrate how Algorithm CTR works using a two-factor real-world example as shown in Table 5, which will be discussed in the next section. The first 10 two-factor fuzzy time series of the average temperature and cloud density are

$$F = (A_1, B_4)\,(A_2, B_2)\,(A_5, B_2)\,(A_8, B_1)\,(A_8, B_1)\,(A_7, B_4)\,(A_7, B_5)\,(A_6, B_4)\,(A_4, B_3)\,(A_6, B_2),$$

where $f_t = \left(A_i, B_j\right)$, A_i and B_j are the fuzzified temperature and cloud density at time t.

First, add a special symbol, %, at the beginning and end of series F, $f_0 = (\%, \%)$ and $f_{11} = (\%, \%)$. The initial candidate set C contains each element f in fuzzy time series F, i.e.

$$C = \{(A_1, B_4)\,,(A_2, B_2)\,,(A_5, B_2)\,,(A_8, B_1)\,,(A_7, B_4)\,,(A_7, B_5)\,,(A_6, B_4)\,,(A_4, B_3)\,,(A_6, B_2)\}.$$

Note that (A_8, B_1) occurs twice in F.

For element (A_1, B_4), there is only one next state, (A_2, B_2), meaning that $(A_1, B_4) \to (A_2, B_2)$ is a certain transition rule and is thus added into P. On the other hand, element (A_8, B_1) can result in two next states: (A_8, B_1) and (A_7, B_4), therefore backtracking is needed and two new candidates $(A_5, B_2)\,(A_8, B_1)$ and $(A_8, B_1)\,(A_8, B_1)$ are added into C. Both candidates will finally produce two certain transitions: $(A_5, B_2)\,(A_8, B_1) \to (A_8, B_1)$ and $(A_8, B_1)\,(A_8, B_1) \to (A_7, B_4)$, respectively. It is worth noting that in this demonstration, the length of c is not fixed as it can take values such as 1, 2, or others, which correspond to orders 1, 2, or others in Chen's high-order model [9]. In other words, it can produce certain transition rules without having to choose in advance a proper order number.

Table 3
Forecasting algorithm (Algorithm forecast)

Input: The maximum length w of the left hand side of certain transition rules in P and the query fuzzy time series $F' = f_1' f_2' \dots f_r'$.
Output: The forecasting result f_{r+1}'.

Begin
  If $r \geq w$ ***Then***
    $(i, k, S) = \mathit{forecasting}(i, k, F'_{r-w+1,w})$;
  Else
    $(i, k, S) = \mathit{forecasting}(i, k, F'_{0,r+1})$;
  End If
  If $S = \{A_e\}$ ***Then Return*** A_e;
  End If
  If $S = \{\%\}$ ***Then Return*** f_r';
  End If
End
Procedure *forecasting*$(i, k, F_{i,k})$;
Begin
  If *a value is bound to the key* $F_{i,k}$ *in* P ***Then***
    Return$(i, k, F_{i,k})$;
  End If
  If $k = 1$ ***Then***
    Return(i, k, ϕ);
  Else
    forecasting$(i + 1, k - 1, F_{i+1,k-1})$;
  End If
End

3.3. Forecasting and defuzzifying

In the forecasting step, we apply the important heuristic rule established by Theorem 3 in [1],

$$\text{If } F_{i,j} \to S \in P \quad \text{Then } F_{i+1,j-1} \to S \notin P, \tag{10}$$

where P is the certain transition rule set created from fuzzy time series F. $F_{i,j}$ is a subsequence of F starting at f_i with length j. The theorem quantifies a maximum length of subsequences in a fuzzy time series that leads to a certain state.

Let w be the maximum length of the left hand side of certain transition rules in P, $F' = f_1' f_2' \dots f_r'$ be the query fuzzy time series, the next state of F', f_{r+1}', be the forecasting result, and r be the length of F'. The forecasting process is determined by the parameters w and r. If $r \geq w$, one only needs to look into the subsequence with length w, which begins at $F'_{r-w+1,w} = f'_{r-w+1} f'_{r-w+2} \dots f_r'$. On the contrary, if $r < w$, the subsequence $F'_{0,r+1} = f_0' f_1' \dots f_r'$ needs to be investigated. Table 3 illustrates the forecasting algorithm based on this heuristic.

The final defuzzification step is to obtain the crisp value of the forecasted result A_j which is deterministically decided. Since A_j is defined by the maximum membership degree occurring at m_j, the forecasting result A_j can be defuzzified by Eq. (11).

$$\text{Defuzzified } A_j = \begin{cases} \dfrac{m_1 + 0.5m_2}{1.5} & \text{if } j = 1 \\ \dfrac{0.5m_{j-1} + m_j + 0.5m_{j+1}}{2} & \text{if } 2 \leq j \leq c-1 \\ \dfrac{0.5m_{c-1} + m_c}{1.5} & \text{if } j = c. \end{cases} \tag{11}$$

It can be observed that the computation time of the Algorithm CTR is

$$\sum_{i=0}^{w-1} (n - i) = wn - \frac{1}{2} w^2$$

and thus its time complexity is $O(wn)(w \leq n)$. For Algorithm Forecast, there are at most n certain transition rules given a fuzzy time series $F = f_1 f_2 \dots f_t \dots f_n$, thus the time complexity is $O(n)$. As a result, the time complexity of the proposed forecasting model is $O(wn) + O(n) \approx O(wn)$. We compare the time complexities of our model with different forecasting models [9] in Table 4, where k denotes the number of fuzzy logical relationships. It demonstrates that the proposed model achieves the lower time complexity.

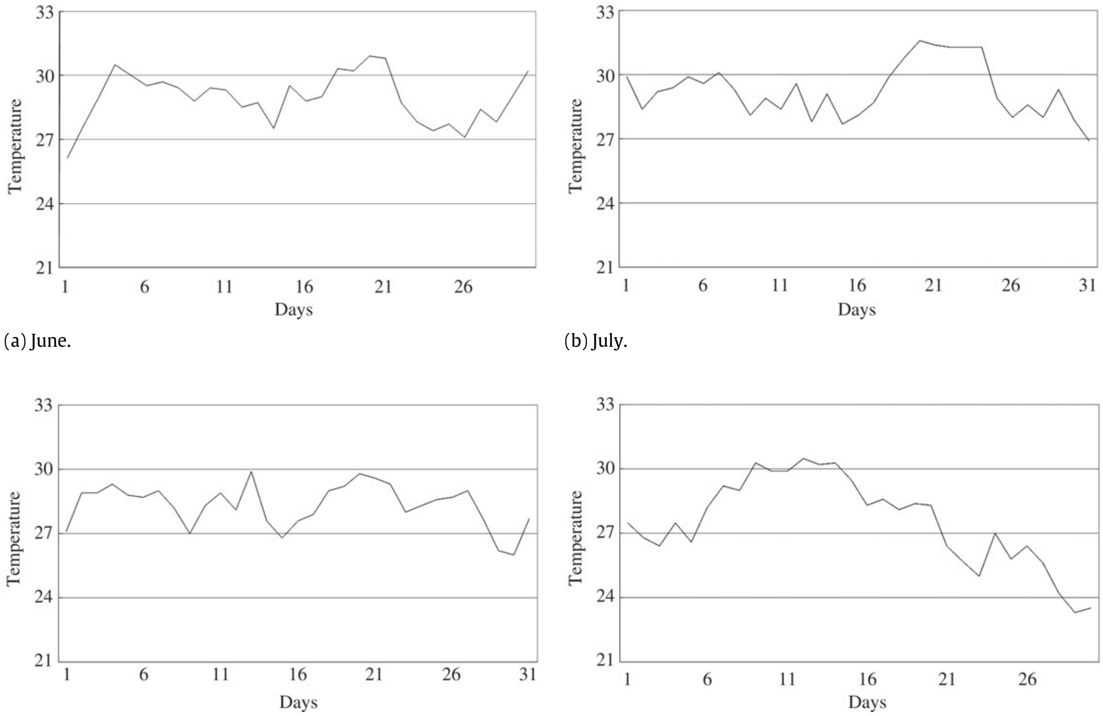
4. Experiments and analysis

4.1. Experimental design and demonstration

In order to validate the effectiveness of the proposed forecasting model, we conducted experiments on a real-world two-factor data set and made comparisons with other models in the literature. The benchmark was the average temperature and

Table 4
Comparison of the time complexity among different forecasting models

Model	Song and Chissom [3]	Song and Chissom [5]	Chen [7]	Chen [9]	Proposed model
Time complexity	$O(kn^2)$	$O(kn^2)$	$O(n^2)$	$O(n^2)$	$O(wn)$



(a) June.

(b) July.

(c) August.

(d) September.

Fig. 2. Average temperature in Taipei from June to September, 1996.

cloud density in Taipei from June to September, 1996, of which the average temperatures are depicted in Fig. 2. The former variable was the main factor to be forecasted whereas the latter the second factor. For illustration purposes, the input time series data of June are represented as $Y(t) = (26.1, 36)(27.6, 23)\ldots(29.0, 29)(30.2, 19)$, as listed in Table 5.

We followed the same number of partitioning intervals ($c_{\text{main factor}} = 9$, $c_{\text{second factor}} = 7$) used in Lee's model [16]. FCM clustering was applied to partition the main and second factors individually and the resulting set of cluster centroids were as follows, respectively:

$$M_{\text{main factor}} = \{27.4, 27.7, 28.4, 28.7, 29.0, 29.4, 29.5, 30.2, 30.8\}$$
$$M_{\text{second factor}} = \{13.8, 22.8, 29.0, 30.0, 44.9, 55.5, 63.1\}.$$

Thus, fuzzy sets $A_i(i = 1, 2, \ldots, 9)$ and $B_j(j = 1, 2, \ldots, 7)$ for both factors were determined by Eqs. (7) and (8), correspondingly.

Next, the historical average temperatures and cloud densities were fuzzified according to where the maximum membership degree occurred. For example, the fuzzy set for temperature 30 °C was A_8 because m_8 had the greatest membership degree. Table 5 gives the fuzzified results of the two-factor time series $Y(t)$, which are represented as $F = (A_1, B_4)\,(A_2, B_2)\ldots(A_8, B_2)$.

In the following, a set of certain transition rules P was constructed by executing Algorithm CTR and the resultant P is given in Table 6. With the availability of P, forecasting could be proceeded by following Algorithm Forecast (Table 3). For example, to forecast the average temperature in Taipei in June 15, 1996, $w = 2$ in P, and the state of the previous two days was $(A_4, B_4)\,(A_1, B_3)$. The forecasting result for June 15 was given to A_7 due to the existence of the certain rule $(A_4, B_4)\,(A_1, B_3) \to A_7$.

The last step of the proposed model was to generate a crisp forecasting output by defuzzification, as shown in Eq. (11). The final crisp output of the example is obtained by

$$\frac{0.5 \times m_6 + m_7 + 0.5 \times m_8}{2} = \frac{0.5 \times 29.4 + 29.5 + 0.5 \times 30.2}{2} \approx 29.7.$$

The forecasting result and error of the daily average temperature in Taipei in June, 1996 are listed in Table 5.

Table 5
Historical and fuzzified daily average temperature and cloud density in Taipei, forecasting result and error for June, 1996

Day	Temperature	Fuzzified temperature	Cloud density	Fuzzified cloud density	Forecasting state	Defuzzified result	Forecasting error
1	26.1	A_1	36	B_4			
2	27.6	A_2	23	B_2	A_2	27.8	−0.2
3	29.0	A_5	23	B_2	A_5	29	0
4	30.5	A_8	10	B_1	A_8	30.2	0.3
5	30.0	A_8	13	B_1	A_8	30.2	−0.2
6	29.5	A_7	30	B_4	A_7	29.7	−0.2
7	29.7	A_7	45	B_5	A_7	29.7	0
8	29.4	A_6	35	B_4	A_6	29.3	0.1
9	28.8	A_4	26	B_3	A_4	28.7	0.1
10	29.4	A_6	21	B_2	A_6	29.3	0.1
11	29.3	A_6	43	B_5	A_6	29.3	0
12	28.5	A_3	40	B_5	A_3	28.3	0.2
13	28.7	A_4	30	B_4	A_4	28.7	0
14	27.5	A_1	29	B_3	A_1	27.5	0
15	29.5	A_7	30	B_4	A_7	29.7	−0.2
16	28.8	A_4	46	B_5	A_4	28.7	0.1
17	29.0	A_5	55	B_6	A_5	29.0	0
18	30.3	A_8	19	B_2	A_8	30.2	0.1
19	30.2	A_8	15	B_1	A_8	30.2	0
20	30.9	A_9	56	B_6	A_9	30.6	0.3
21	30.8	A_9	60	B_7	A_9	30.6	0.2
22	28.7	A_4	96	B_7	A_4	28.7	0
23	27.8	A_2	63	B_7	A_2	27.8	0
24	27.4	A_1	28	B_3	A_1	27.5	−0.1
25	27.7	A_2	14	B_1	A_2	27.8	−0.1
26	27.1	A_1	25	B_2	A_1	27.5	−0.4
27	28.4	A_3	29	B_3	A_3	28.3	0.1
28	27.8	A_2	55	B_6	A_2	27.8	0
29	29.0	A_5	29	B_3	A_5	29.0	0
30	30.2	A_8	19	B_2	A_8	30.2	0

Table 6
Certain transition rules of fuzzy time series of daily average temperature and cloud density in Taipei in June, 1996

$(A_1, B_2) \rightarrow (A_3, B_3)$	$(A_5, B_3) \rightarrow (A_8, B_2)$
$(A_4, B_4)\ (A_1, B_3) \rightarrow (A_7, B_4)$	$(A_5, B_6) \rightarrow (A_8, B_2)$
$(A_2, B_7)\ (A_1, B_3) \rightarrow (A_2, B_1)$	$(A_6, B_2) \rightarrow (A_6, B_5)$
$(A_1, B_4) \rightarrow (A_2, B_2)$	$(A_6, B_4) \rightarrow (A_4, B_3)$
$(A_2, B_1) \rightarrow (A_1, B_2)$	$(A_6, B_5) \rightarrow (A_3, B_5)$
$(A_2, B_2) \rightarrow (A_5, B_2)$	$(A_8, B_1)\ (A_7, B_4) \rightarrow (A_7, B_5)$
$(A_2, B_6) \rightarrow (A_5, B_3)$	$(A_1, B_3)\ (A_7, B_4) \rightarrow (A_4, B_5)$
$(A_2, B_7) \rightarrow (A_1, B_3)$	$(A_7, B_5) \rightarrow (A_6, B_4)$
$(A_3, B_3) \rightarrow (A_2, B_6)$	$(A_8, B_1)\ (A_8, B_1) \rightarrow (A_7, B_4)$
$(A_3, B_5) \rightarrow (A_4, B_4)$	$(A_5, B_2)\ (A_8, B_1) \rightarrow (A_8, B_1)$
$(A_4, B_3) \rightarrow (A_6, B_2)$	$(A_8, B_2)\ (A_8, B_1) \rightarrow (A_9, B_6)$
$(A_4, B_4) \rightarrow (A_1, B_3)$	$(A_8, B_2) \rightarrow (A_8, B_1)$
$(A_1, B_5) \rightarrow (A_5, B_6)$	$(A_9, B_6) \rightarrow (A_9, B_7)$
$(A_4, B_7) \rightarrow (A_2, B_7)$	$(A_9, B_7) \rightarrow (A_4, B_7)$
$(A_5, B_2) \rightarrow (A_8, B_1)$	

4.2. Model reliability

Since the resulting clusters of FCM depend on the initial random assignments of membership degrees, the forecasting result for each experiment might be different. In order to analyze model reliability and distribution of forecasting values, we use box plot supported by Monte Carlo simulation. The Monte Carlo method performs stochastic sampling experiments to provide approximate solutions of problems based on random numbers and probability statistics. It allows increasing the sample size and if the number of simulations is large enough then the sample average will be much closer to the true mean distribution. With the availability of Monte Carlo simulation results, the reliability of the forecasting model can be justified by box plot.

Box plot is widely used in exploratory data analysis for conveying location and variation information in data sets [33,34]. It provides a convenient way of graphically depicting a five-number summary: the lowest, lower quartile, median, upper quartile, and largest values. The summary conveys the level, spread and symmetry of a distribution of data values. Whiskers extend from the box out to the most extreme data values in $1.5 \times$ IQR (IQR = the upper quartile − the lower quartile). Any observation which lies outside the whiskers is considered an outlier.

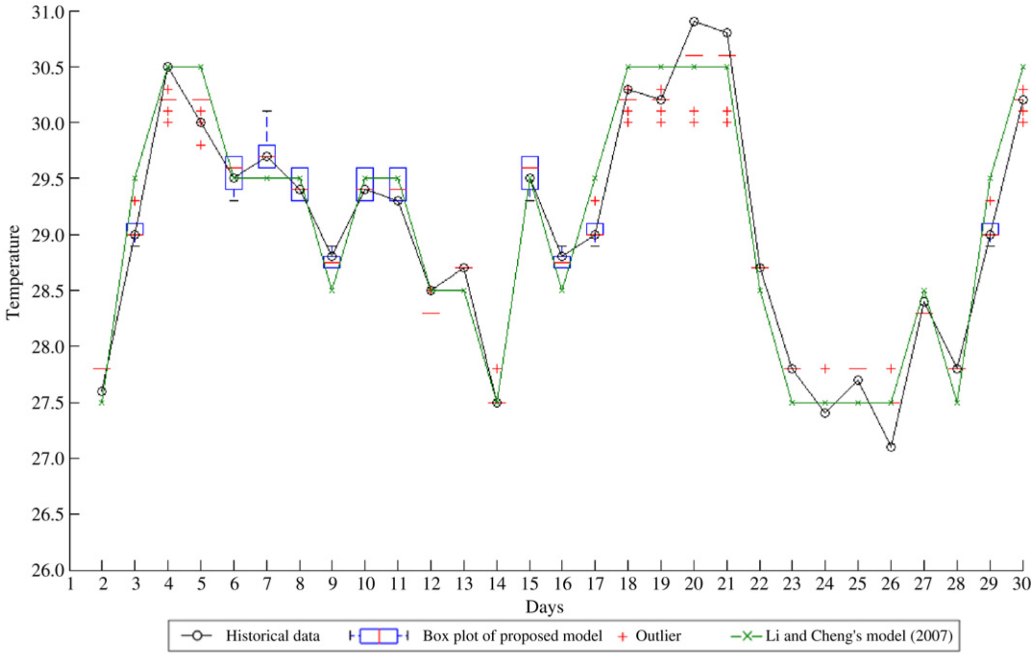


Fig. 3. Box plot by the proposed model for June, 1996.

Figs. 3–6 illustrate the box plots of the forecasting results for June, July, August, and September of 1996 obtained by the proposed model after conducting 30 Monte Carlo simulations. They provide some interesting observations: first, these boxes represent the middle 50% of the distribution of the forecast values in a day and all figures show that most of the medians are nearer to the real values, compared to the forecasted values of [1]. Next, most real values fell inside the boxes, thus generating an acceptable forecasting accuracy. All spreads were very narrow except for days 29 and 30 of August which were slightly larger, indicating that the proposed model is reasonably reliable. One notes that some box plots exhibit a single line, like the one in June 20 representing that the 28 forecasting results were all the same and two outliers were identified (denoted as '+') in the 30 simulations. Finally, there were a number of outliers on the low sides and/or high sides, however the spread at the spot in which outliers occurred was quite narrow (for example, June 19th, July 6th, Aug. 19th, Sept. 15th, etc.). These outliers had no significant impact on model reliability.

Taken as a whole, the forecasting result was not sensitive to the initial membership degrees assigned in FCM and thus the proposed model is reliable.

4.3. Performance comparison

The effectiveness of the proposed model is also validated by conducting modeling performance comparisons with the models of [1,16,22]. The proposed model proves to be superior in performance to any other existing models. The performance metrics used in the comparisons include mean square error (*MSE*) and average forecasting error rate (*AFER*), defined as follows:

$$\mathrm{AFER} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\mathrm{Forecasting\_Value}_i - \mathrm{Actual\_Value}_i|}{\mathrm{Actual\_Value}_i} \times 100\% \tag{12}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Forecasting\_Value}_i - \mathrm{Actual\_Value}_i)^2\,, \tag{13}$$

where n is the number of data to be forecasted.

As indicated by Sullivan and Woodall [6], it is important to distinguish between modeling accuracy and forecasting accuracy, which are respectively concerned with justifying the performance with the data used to build the forecasting model and with the data not used in estimating the parameters of the model. Modeling accuracy is typically better than forecasting accuracy; however it is not necessarily a good indicator of the latter. Thus, in this study the performance will be compared in terms of modeling and forecasting accuracies, using a range between one and eight orders for evaluating Lee's model.

Table 7 summarizes the comparison of modeling accuracy in terms of the average *AFER* of four months. It shows that the proposed model outperforms the models reported in [1,16,22], of which the former two use even-length interval partitioning and the latter GA.

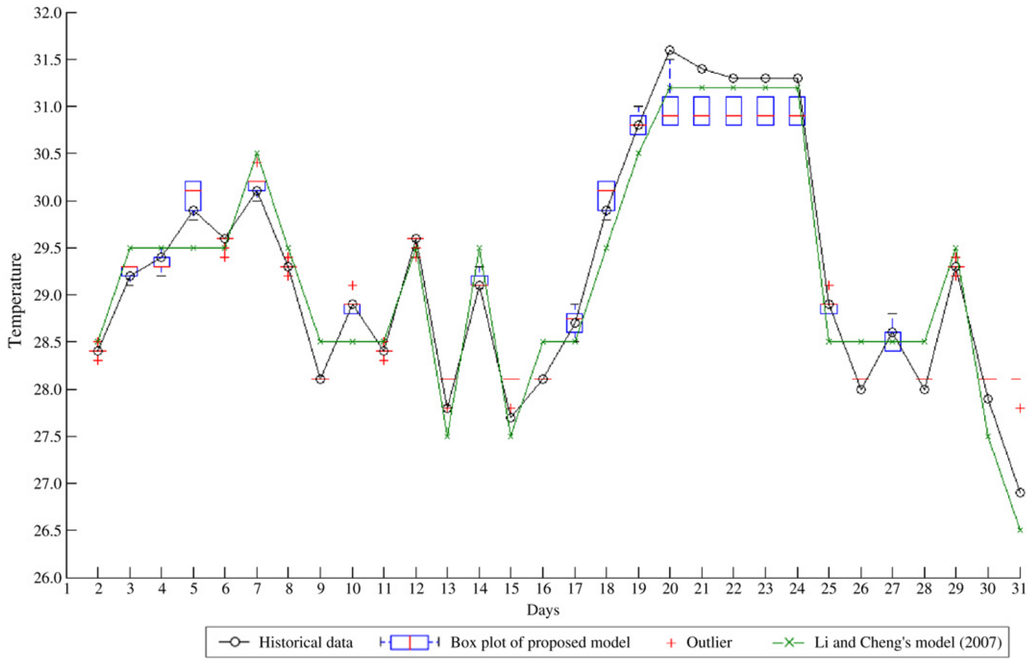


Fig. 4. Box plot by the proposed model for July, 1996.

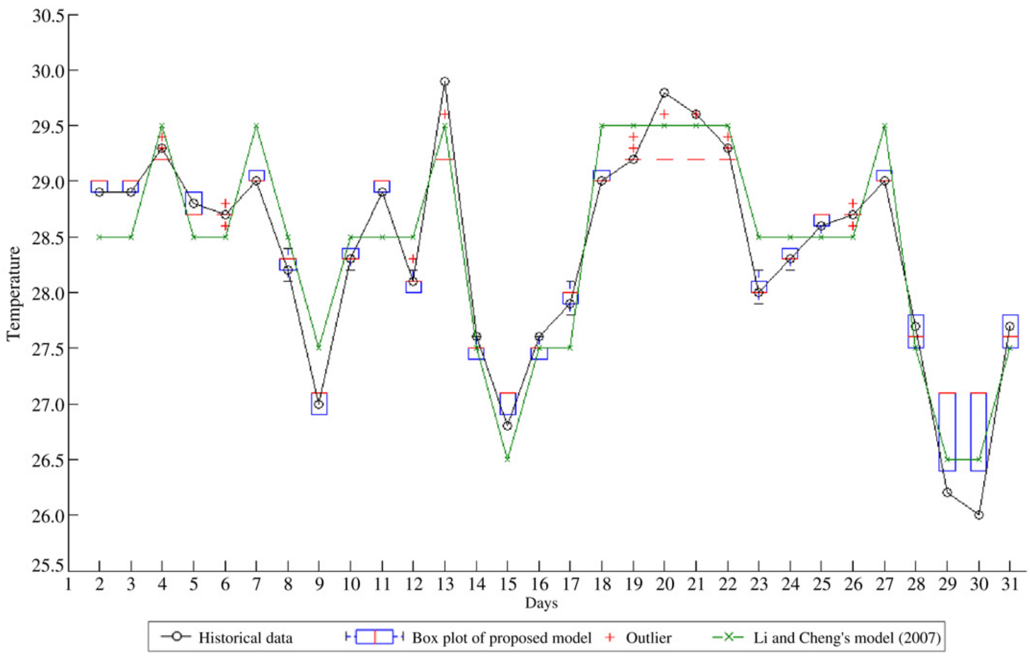


Fig. 5. Box plot by the proposed model for August, 1996.

Next, we compare the forecasting accuracy with other models, where the data set from June to September of 1991–1995 is used to build the underlying forecasting model, and the set from June to September of 1996 is used to assess forecasting accuracy. The comparisons of forecasting accuracy in terms of average *AFER* and *MSE* of four months are shown in Table 8. It demonstrates the superiority of the proposed model in forecasting accuracy, which also achieved better performance than our previous model with *k*-means clustering as the interval partitioning method. One notes that Lee's model obtains the best results for order two, however the forecasting accuracy worsens as the order increases.

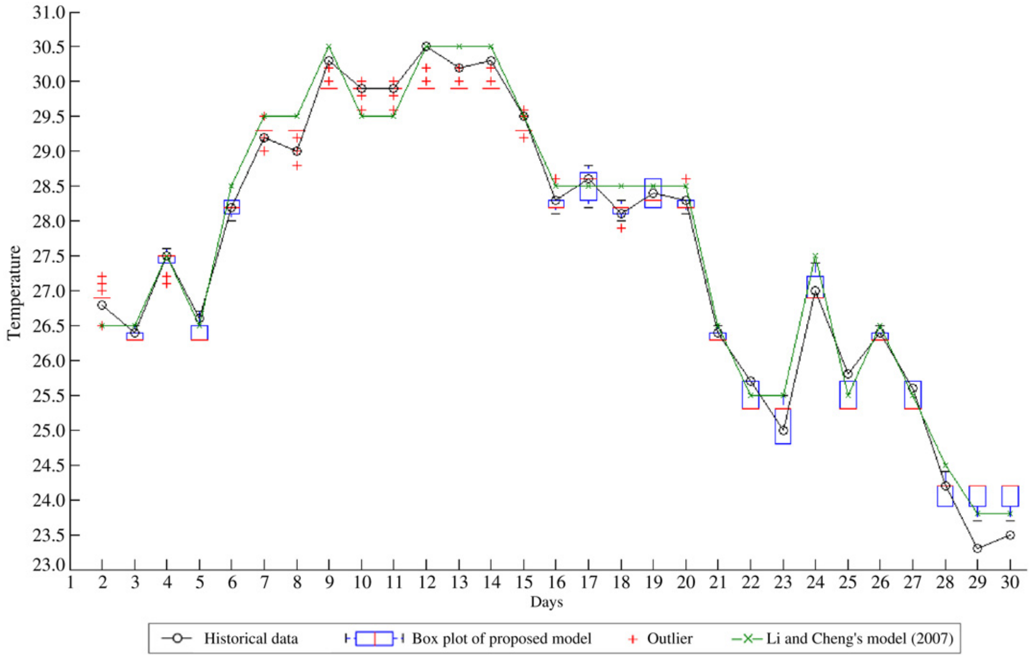


Fig. 6. Box plot by the proposed model for September, 1996.

Table 7
Modeling performance comparisons by Avg. *AFER*

Lee et al. [16]								Lee, Wang, and Chen [22]								Li and Cheng [1]	Proposed model
Order1	Order2	Order3	Order4	Order5	Order6	Order7	Order8	Order1	Order2	Order3	Order4	Order5	Order6	Order7	Order8		
1.55%	0.96%	0.92%	0.94%	0.94%	0.96%	0.95%	0.95%	1.21%	0.84%	0.78%	0.91%	0.74%	0.78%	0.73%	0.72%	0.92%	0.55%

Table 8
Forecasting performance comparisons

	Lee et al. [16]								Li and Cheng [1]	*k*-means deterministic	Proposed model
	Order1	Order2	Order3	Order4	Order5	Order6	Order7	Order8			
Avg. *AFER*	3.13%	2.87%	2.92%	3.10%	3.31%	3.53%	3.70%	3.91%	3.30%	3.15%	2.67%
Avg. *MSE*	1.2199	1.0451	1.0584	1.1594	1.3047	1.4436	1.5618	1.7095	1.3084	1.2345	0.9021

5. Conclusions and future works

In this paper, we proposed a FCM-based deterministic model to handle the forecasting problem of fuzzy time series, an extension of our previous work with two enhancements. The first improvement results from applying FCM clustering to tackle the issue of interval partitioning of the discourse universe so that the distribution of historical data can be taken into account and unequal-sized intervals can be derived. Next, the previous forecasting model is generalized to allow the manipulation of two-factor problems. The benchmark experiment on forecasting the daily average temperature in Taipei, Taiwan is conducted for evaluating the performance of our model and related forecasting models. Furthermore, we used Monte Carlo simulation and box plots to verify the reliability of the proposed model. The experimental results and analysis confirm the superiority of the proposed model in forecasting accuracy and reliability. Future works will aim at generalizing the proposed model to handle multi-factor forecasting problems; however the ratio of rule matching for forecasting becomes lower when more factors are taken into consideration, bringing out the issue of rule redundancy addressed in [35]. Other interesting future works involve applying the proposed model to deal with more complicated real-world problems.

Acknowledgments

This study was supported in part by NSC96-2416-H-006-011-MY2, Taiwan, ROC. The authors appreciate Shih-Min Chien for his dedicated help in the algorithm revision.

References

[1] S.-T. Li, Y.-C. Cheng, Deterministic Fuzzy time series model for forecasting enrollments, Computers and Mathematics with Applications 53 (2007) 1904–1920.
[2] D. Rafiei, On similarity-based queries for time series data, in: Proceedings of the 15th International Conference on Data Engineering, Sydney, Australia, 1999, pp. 410–417.
[3] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time series-part I, Fuzzy Sets Systems 54 (1993) 1–9.
[4] Q. Song, B.S. Chissom, Fuzzy time series and its models, Fuzzy Sets and Systems 54 (1993) 269–277.
[5] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time series-part II, Fuzzy Sets and Systems 62 (1994) 1–8.
[6] J. Sullivan, W.H. Woodall, A comparison of fuzzy forecasting and Markov modeling, Fuzzy Sets and Systems 64 (1994) 279–293.
[7] S.-M. Chen, Forecasting enrollments based on fuzzy time series, Fuzzy Sets and Systems 81 (1996) 311–319.
[8] K. Huarng, Heuristic models of fuzzy time series for forecasting, Fuzzy Sets and Systems 123 (2001) 369–386.
[9] S.-M. Chen, Forecasting enrollments based on high-order fuzzy time series, Cybernetics and Systems: An International Journal 33 (2002) 1–16.
[10] Y.-Y. Hsu, S.-M. Tse, B. Wu, A new approach of bivariate fuzzy time series analysis to the forecasting of a stock index, International Journal of Uncertainty Fuzziness and Knowledge-Based Systems 11 (6) (2003) 671–690.
[11] C.-M. Own, P.-T. Yu, Forecasting fuzzy time series on a heuristic high-order model, Cybernetics and Systems: An International Journal 36 (2005) 705–717.
[12] K. Huarng, H.-K. Yu, A Type 2 fuzzy time series model for stock index forecasting, Physica A 353 (2005) 445–462.
[13] K. Huarng, T.H.-K. Yu, The application of neural networks to forecast fuzzy time series, Physica A 363 (2006) 481–491.
[14] H.-K. Yu, A refined fuzzy time-series model for forecasting, Physica A 346 (2005) 657–681.
[15] H.-K. Yu, Weighted fuzzy time series models for TAIEX forecasting, Physica A 349 (2005) 609–624.
[16] L.-W. Lee, L.-H. Wang, S.-M. Chen, Y.-H. Leu, Handling forecasting problems based on two-factors high-order fuzzy time series, IEEE Transactions on Fuzzy Systems 14 (3) (2006) 468–477.
[17] K. Huarng, Effective lengths of intervals to improve forecasting in fuzzy time series, Fuzzy Sets Systems 123 (2001) 387–394.
[18] S.-T. Li, Y.-P. Chen, Natural partition-based forecasting model for fuzzy time series, in: IEEE International Conference on Fuzzy Systems, Budapest, Hungary, 2004, pp. 25–29.
[19] S.-M. Chen, C.-C. Hsu, A new method to forecast enrollments using fuzzy time series, International Journal of Applied Science and Engineering 2 (3) (2004) 234–244.
[20] K. Huarng, T.H.-K. Yu, Ratio-based lengths of intervals to improve fuzzy time series forecasting, IEEE Transactions on Systems, Man, Cybernetics-part B: Cybernetics 36 (2) (2006) 328–340.
[21] S.-M. Chen, N.-Y. Chung, Forecasting enrollments using high-order fuzzy time series and genetic algorithms, International Journal of Intelligent Systems 21 (2006) 485–501.
[22] L.-W. Lee, L.-H. Wang, S.-M. Chen, Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms, Expert Systems with Applications: An International Journal 33 (2007) 539–550.
[23] S.-M. Chen, J.-R. Hwang, Temperature prediction using fuzzy time series, IEEE Transactions on Systems, Man, Cybernetics-part B: Cybernetics 30 (2000) 263–275.
[24] H. Liu, F. Hussain, C.L. Tan, M. Dash, Discretization: An enabling technique, Data Mining and Knowledge Discovery 6 (4) (2002) 393–423.
[25] J.-C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, 1981.
[26] T.A. Jilani, S.M.A. Burney, C. Ardil, Multivariate high order fuzzy time series forecasting for car road accidents, International Journal of Computational Intelligence 19 (1) (2008) 15–50.
[27] T.H.-K. Yu, K.-H. Huarng, A bivariate fuzzy time series model to forecast the TAIEX, Expert Systems with Applications 34 (2008) 2945–2952.
[28] N. Metropolis, S. Ulam, The Monte Carlo method, Journal of the American Statistical Association 44 (247) (1949) 335–341.
[29] K. Mosegaard, A. Tarantola, Monte Carlo sampling of solutions to inverse problems, Journal of Geophysical Research 100 (B7) (1995) 12431–12447.
[30] J. Dougherty, R. Kohavi, M. Sahami, Supervised and unsupervised discretization of continuous features, in: The 12 International Conference on Machine learning, 1995, pp. 194–202.
[31] C.-H. Leon Lee, Alan Liu, W.-S. Chen, Pattern discovery of fuzzy time series for financial prediction, IEEE transactions on Knowledge and Data Engineering 18 (5) (2006) 613–625.
[32] M. Versaci, F.C. Morabito, Fuzzy time series approach for disruption prediction in Tokamak reactors, IEEE transactions on Magnetics 39 (3) (2003) 1503–1506.
[33] J.M. Chambers, W.S. Cleveland, B. Kleiner, P.A. Tukey, Graphical methods for data analysis, Duxtury (1983) 21–24, 57–60.
[34] D.F. Williamson, R.A. Parker, J.S. Kendrick, The box plot: A simple visual method to interpret data, Annals of Internal Medicine 110 (1989) 916–921.
[35] S.-T. Li, Y.-C. Cheng, An enhanced deterministic fuzzy time series forecasting model (submitted for publication).