Available online at www.sciencedirect.com
ScienceDirect
Computers and Mathematics with Applications 53 (2007) 1904–1920
An International Journal
computers & mathematics with applications
www.elsevier.com/locate/camwa

Deterministic fuzzy time series model for forecasting enrollments

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

[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 Trade, Tainan University of Technology, Taiwan, ROC

Received 7 November 2005; received in revised form 22 March 2006; accepted 27 March 2006

Abstract

The fuzzy time series has recently received increasing attention because of its capability of dealing with vague and incomplete data. There have been a variety of models developed to either improve forecasting accuracy or reduce computation overhead. However, the issues of controlling uncertainty in forecasting, effectively partitioning intervals, and consistently achieving forecasting accuracy with different interval lengths have been rarely investigated. This paper proposes a novel deterministic forecasting model to manage these crucial issues. In addition, an important parameter, the maximum length of subsequence in a fuzzy time series resulting in a certain state, is deterministically quantified. Experimental results using the University of Alabama's enrollment data demonstrate that the proposed forecasting model outperforms the existing models in terms of accuracy, robustness, and reliability. Moreover, the forecasting model adheres to the consistency principle that a shorter interval length leads to more accurate results.

Keywords: Fuzzy time series; Forecasting; Fuzzy logical relationship; State transition; Interval partitioning

1. Introduction

The forecasting problem of time series data, consisting of time-dependent sequences of continuous values, is important and interesting in a great variety of applications such as monitoring air pollution in environmental protection, predicting stock prices in the stock market, estimating blood pressure in a hospital, and so on. This problem has been widely studied in areas of statistics, signal processing, and neural networks in past decades. In 1993, Song and Chissom introduced fuzzy logic to the classic problem and proposed a new paradigm of time series forecasting, namely the fuzzy time series, which is capable of dealing with vague and incomplete data represented as linguistic values under uncertain circumstances [1–3]. They studied the problem of forecasting fuzzy time series using the enrollment data of the University of Alabama and proposed a forecasting model which is mainly composed of four steps: (1) partitioning the universe of discourse into even lengthy intervals, (2) defining fuzzy sets on the universe of

* 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).

doi:10.1016/j.camwa.2006.03.036

discourse and fuzzifying the time series, and deriving fuzzy logical relationships existing in the fuzzified time series, (3) forecasting, and (4) defuzzifying the forecasting outputs. Song and Chissom solved the forecasting problem using fuzzy relational equations and approximate reasoning, which takes a large amount of computation time in deriving the fuzzy relationship [1]. Since the work of Song and Chissom, a number of researches have been conducted to improve the forecasting accuracy or reduce the computation overhead. Firstly, to alleviate the overhead of computation time in deriving the fuzzy relationship in Song and Chissom's model, Sullivan and Woodall proposed the 'Markov-based model' [4] by using conventional matrix multiplication. Subsequently, Chen presented an efficient forecast procedure for enrollments at the University of Alabama using simplified arithmetic operations and improved the forecasting accuracy [5]. Chen later extended the previous work and proposed a high-order fuzzy time series model, in order to reduce the forecasting error [6]. Unfortunately, the issue of how to determine the order in the high-order forecasting model was not discussed. In [7], Chen and Hwang developed two algorithms for temperature prediction to deal with forecasting problems, and obtained good forecasting results. The work of Hwang, Chen, and Lee [8] showed that the variation of enrollments for the next year is related to the enrollment trend of past years. Huarng proposed heuristic models by integrating problem-specific heuristic knowledge with Chen's model to improve forecasting [9]. Song and Chissom applied first-order time-variant models in forecasting the enrollment and discussed the difference between time-invariant and time-variant models [3]. Recently, Tsaur, Yang, and Wang applied the concept of entropy to measure the degrees of fuzziness when a time-invariant relation matrix is derived [10]. Other similar work on fuzzy time series can be found in [11,12]. All the work reviewed primarily focused on improving steps (3) and (4) in Song and Chissom's framework.

In the previous work, the universe of discourse was defined with arbitrary selected parameters and was decomposed into even length intervals. Nevertheless, the forecasting performance could be affected significantly by the partition of the universe of discourse [13]. Huarng investigated the impact of interval length on the forecasting results and proposed two heuristic approaches, namely distribution and average-based, to determine the length of the interval [13]. However, the reason behind how the so-called 'base-mapping table' was not specified. Li and Chen proposed a natural partitioning-based forecasting model in order to substitute the 'base-mapping table' and obtained a similar performance [14]. On the other hand, a university enrollments domain expert in this study believes that the interval length should be decided by the experts. It is more important that the interval length reflects the sensitivity of the investigated data. Using an enrollment of 5000 students at a university as an example, if the amount of 500 enrollments is expected to be reduced, which could result in a crisis in running the school, forecasting enrollments with 1000 of the interval length is then meaningless. In price statistics, an economist believes that the price index will rise 0.02, which is sufficient to significantly influence the economy's decision-making; hence the length of interval should remain 0.02, which means the interval length's sensitivity should stay at a constant of 0.02.

Another issue is the consistency of the forecasting accuracy with the interval length. In general cases, better accuracy can be achieved with a shorter interval length [13]. However, the work presented in [13] conflicts with this general rule. It is expected that an effective forecasting model should adhere to the consistency principle.

In this study, we focus on the enhancement of steps (1), (3) and (4) in Song and Chissom's framework. We devote ourselves to tackling the issues of improving forecasting accuracy by controlling uncertainty and determining the length of intervals effectively. By extending our preliminary work presented in [15], which outlines the issues, this paper pays special attention to establishing theoretical foundations for dealing with such issues. We propose a new forecasting model based on the state-transition analysis, which overcomes the hurdle of determining the 'k-order' from Chen's model. More importantly, we quantify a deterministic maximum of length of subsequence in the fuzzy time series which leads to a certain state. Such quantification can help with the derivation of a new forecasting step in the framework. We conduct experiments in forecasting the enrollments at the University of Alabama. The result is quite encouraging because its accuracy is better than those in the literature and it is consistent with the length of interval as well. In addition, it is robust when the historical data are contaminated and is more reliable through an analysis of residual scatter.

There are seven sections in this paper. In Section 2, we briefly introduce the basic concept of fuzzy time series and give an outline of related work. Section 3 presents the issues of designing an effective forecasting model. In Section 4, the deterministic forecasting model is proposed by illustrating the example of forecasting the university's enrollments. In Section 5, the performance evaluation and comparison in accuracy, robustness, reliability and consistency are given and discussed. The last section is a conclusion and future work.

2. Fuzzy time series and related work

Let $Y(t)$ $(t = \ldots, 0, 1, 2, \ldots)$, a subset of R, 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 $Y(t)$ $(t = \ldots, 0, 1, 2, \ldots)$. Let $F(t)$ and $F(t-1)$ be fuzzy time series on $Y(t)$ and $Y(t-1)$ $(t = \ldots, 0, 1, 2, \ldots)$. For any $f_j(t) \in F(t)$, there exists an $f_i(t-1) \in F(t-1)$ such that there is a first-order fuzzy relation $R(t, t-1)$ and $f_j(t) = f_i(t-1) \circ R_{ij}(t, t-1)$, then $F(t)$ is said to be caused by $F(t-1)$ only. Denote this as $f_i(t-1) \rightarrow f_j(t)$ or equivalently $F(t-1) \rightarrow F(t)$. Song and Chissom derived the first-order model based on the first-order relation and extended to mth-order model [2].

Definition 1. Suppose $F(t)$ is caused by $F(t-1)$ or $F(t-2)$ or ... or $F(t-m) (m > 0)$ only. This relation can be expressed as the following fuzzy relational equation:

$$F(t) = F(t-1) \circ R(t, t-1) \quad \text{or} \quad F(t) = F(t-2) \circ R(t, t-2) \quad \text{or} \quad \ldots \quad \text{or}$$
$$F(t) = F(t-m) \circ R(t, t-m)$$

or

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

where '$\cup$' is the union operator, and '$\circ$' is the composition. $R(t, t-m)$ is a relation matrix to describe the fuzzy relationship between $F(t-m)$ and $F(t)$. This equation is called the first-order model of $F(t)$.

Definition 2. Suppose that $F(t)$ is caused by $F(t-1), F(t-2), \ldots,$ and $F(t-m) (m > 0)$ simultaneously. This relation can be expressed as the following fuzzy relational equation:

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

The equation is called the mth-order model of $F(t)$, and $R_a(t, t-m)$ is a relation matrix to describe the fuzzy relationship between $F(t-1), F(t-2), \ldots, F(t-m)$ and $F(t)$.

It was reported [5,6] that the steps of Chen's first-order and high-order forecasting models are similar to the ones of Song and Chissom's framework except for Steps 3 and 4. The following is Chen's approach.

Step 1. Partitioning the universe of discourse U into several even length intervals. Let $D_{\min}$ and $D_{\max}$ be the minimum enrollment and the maximum enrollment of historical data. Let $U = [D_{\min} - D_1, D_{\max} + D_2]$ be the universe of discourse, where D_1 and D_2 are two proper positive numbers, then U is partitioned into n equal intervals with length l defined as $l = \frac{1}{n}[(D_{\max} + D_2) - (D_{\min} - D_1)]$.

For the enrollment data of the University of Alabama, $U = [13\,000, 20\,000]$ is partitioned into seven intervals $u_1, u_2, u_3, u_4, u_5, u_6,$ and u_7, where $u_1 = [13\,000, 14\,000]$, $u_2 = [14\,000, 15\,000]$, $u_3 = [15\,000, 16\,000]$, $u_4 = [16\,000, 17\,000]$, $u_5 = [17\,000, 18\,000]$, $u_6 = [18\,000, 19\,000]$, and $u_7 = [19\,000, 20\,000]$.

Step 2. Defining fuzzy sets on the universe of discourse U and fuzzifying the time series.
A fuzzy set A_i of U is defined by

$$A_i = f_{A_i}(u_1)/u_1 + f_{A_i}(u_2)/u_2 + \cdots + f_{A_i}(u_n)/u_n \tag{3}$$

where f_{A_i} is the membership function of fuzzy set A_i, $f_{A_i} : U \rightarrow [0, 1]$, and $f_{A_i}(u_j)$ indicates the grade of membership of u_j in A_i. By finding out the degree of each value belonging to each A_i $(i = 1, 2, \ldots, n)$, the fuzzified time series for that time t was treated as A_i, which the maximum membership degree of some time t occurred at:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7,$$
$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7,$$
$$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + 0/u_6 + 0/u_7,$$
$$A_4 = 0/u_1 + 0/u_2 + 0.5/u_3 + 1/u_4 + 0.5/u_5 + 0/u_6 + 0/u_7,$$

Table 1
The fuzzy logical relationships for Chen's first-order model

$A_1 \to \{A_1, A_2\}$	$A_2 \to \{A_3\}$	$A_3 \to \{A_3, A_4\}$
$A_4 \to \{A_3, A_4, A_6\}$	$A_6 \to \{A_6, A_7\}$	$A_7 \to \{A_6, A_7\}$

Table 2
The fuzzy logical relationships for Chen's high-order model

2nd-order	3rd-order	4th-order	5th-order
$A_1, A_1 \to \{A_1, A_2\}$	$\#, A_1, A_1 \to \{A_1\}$	$\#, A_1, A_1, A_1 \to \{A_2\}$	$\#, A_1, A_1, A_1, A_2 \to \{A_3\}$
$A_1, A_2 \to \{A_3\}$	$A_1, A_1, A_1 \to \{A_2\}$	$A_1, A_1, A_1, A_2 \to \{A_3\}$	$A_1, A_1, A_1, A_2, A_3 \to \{A_3\}$
$A_2, A_3 \to \{A_3\}$	$A_1, A_1, A_2 \to \{A_3\}$	$A_1, A_1, A_2, A_3 \to \{A_3\}$	$A_1, A_1, A_2, A_3, A_3 \to \{A_3\}$
$A_3, A_3 \to \{A_3, A_4\}$	$A_1, A_2, A_3 \to \{A_3\}$	$A_1, A_2, A_3, A_3 \to \{A_3\}$	$A_1, A_2, A_3, A_3, A_3 \to \{A_3\}$
$A_3, A_4 \to \{A_4, A_6\}$	$A_2, A_3, A_3 \to \{A_3\}$	$A_2, A_3, A_3, A_3 \to \{A_3\}$	$A_2, A_3, A_3, A_3, A_3 \to \{A_4\}$
$A_4, A_4 \to \{A_3, A_4\}$	$A_3, A_3, A_3 \to \{A_3, A_4\}$	$A_3, A_3, A_3, A_3 \to \{A_3, A_4\}$	$A_3, A_3, A_3, A_3, A_4 \to \{A_4, A_6\}$
$A_4, A_3 \to \{A_3\}$	$A_3, A_3, A_4 \to \{A_4, A_6\}$	$A_3, A_3, A_3, A_4 \to \{A_4, A_6\}$	$A_3, A_3, A_3, A_4, A_4 \to \{A_4\}$
$A_4, A_6 \to \{A_6\}$	$A_3, A_4, A_4 \to \{A_4\}$	$A_3, A_4, A_4, A_4 \to \{A_3\}$	$A_3, A_3, A_4, A_4, A_4 \to \{A_3\}$
$A_6, A_6 \to \{A_7\}$	$A_4, A_4, A_4 \to \{A_3\}$	$A_4, A_4, A_4, A_3 \to \{A_3\}$	$A_3, A_4, A_4, A_4, A_3 \to \{A_3\}$
$A_6, A_7 \to \{A_7\}$	$A_4, A_4, A_3 \to \{A_3\}$	$A_4, A_4, A_3, A_3 \to \{A_3\}$	$A_4, A_4, A_4, A_3, A_3 \to \{A_3\}$
$A_7, A_7 \to \{A_6\}$	$A_4, A_3, A_3 \to \{A_3\}$	$A_4, A_3, A_3, A_3 \to \{A_3\}$	$A_4, A_4, A_3, A_3, A_3 \to \{A_3\}$
$A_7, A_6 \to \#$	$A_3, A_4, A_6 \to \{A_6\}$	$A_3, A_3, A_4, A_6 \to \{A_6\}$	$A_4, A_3, A_3, A_3, A_3 \to \{A_3\}$
	$A_4, A_6, A_6 \to \{A_7\}$	$A_3, A_4, A_6, A_6 \to \{A_7\}$	$A_3, A_3, A_3, A_3, A_3 \to \{A_4\}$
	$A_6, A_6, A_7 \to \{A_7\}$	$A_4, A_6, A_6, A_7 \to \{A_7\}$	$A_3, A_3, A_3, A_4, A_6 \to \{A_6\}$
	$A_6, A_7, A_7 \to \{A_6\}$	$A_6, A_6, A_7, A_7 \to \{A_6\}$	$A_3, A_3, A_4, A_6, A_6 \to \{A_7\}$
	$A_7, A_7, A_6 \to \#$	$A_6, A_7, A_7, A_6 \to \#$	$A_3, A_4, A_6, A_6, A_7 \to \{A_7\}$
			$A_4, A_6, A_6, A_7, A_7 \to \{A_6\}$
			$A_6, A_6, A_7, A_7, A_6 \to \#$

$$A_5 = 0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5 + 0.5/u_6 + 0/u_7,$$
$$A_6 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0.5/u_5 + 1/u_6 + 0.5/u_7,$$
$$A_7 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0.5/u_6 + 1/u_7.$$

The fuzzy time series for enrollments is thus as follows: $A_1, A_1, A_1, A_2, A_3, A_3, A_3, A_3, A_4, A_4, A_4, A_3, A_3, A_3, A_3, A_3, A_4, A_6, A_6, A_7, A_7, A_6$.

Step 3. Deriving fuzzy logical relationships. By using Definition 1, the fuzzy logical relationships are further grouped based on the same $F(t-1)$ value, $A_i \to \text{Group}(A_i)$, where $\text{Group}(A_i)$ is a subset of $\{A_1, A_2, \ldots, A_n\}$. The fuzzy logical relationships for Chen's first-order model are illustrated in Table 1.

From Definition 2, mth-order relationships are grouped based on the same $F_{j,k}(t-1) = A_{j_1} A_{j_2} \ldots A_{j_k} \to \text{Group}(A_{j_1} A_{j_2} \ldots A_{j_k})$, where $\text{Group}(A_{j_1} A_{j_2} \ldots A_{j_k})$ is a subset of $\{A_1, A_2, \ldots, A_n\}$. For the example of enrollments, the fuzzy logical relationships are shown in Table 2.

Step 4. Forecasting and defuzzifying the forecasting outputs. The forecasting result of the first-order forecasting model is based on the following heuristic rules:

Rule 1: If $F(t-1) = A_i$ and the number of $\text{Group}(A_i)$, $|\text{Group}(A_i)| = 0$, then the predicted result at time t, m_i, is the midpoint of interval u_i in which the maximum membership degree of A_i locates.

Rule 2: If $F(t-1) = A_i$ and $\text{Group}(A_i) = \{A_{j_1}, A_{j_2}, \ldots, A_{j_p}\}$, $p \geq 1$, then the predicted result at time t is

$$\frac{1}{p} \sum_{i=1}^{p} m_{j_i} \tag{4}$$

where $m_{j_1}, m_{j_2}, \ldots, m_{j_p}$, is the midpoint of the interval $u_{j_1}, u_{j_2}, \ldots, u_{j_p}$ in which the maximum membership degree of $A_{j_1}, A_{j_2}, \ldots, A_{j_p}$ locates, respectively.

The forecasting result of the high-order forecasting model is calculated by the following principles:

Choose k, $k \geq 2$; there exists a fuzzy logical relationship $A_{i_1} A_{i_2} \dots A_{i_k} \rightarrow \text{Group}(A_{i_1} A_{i_2} \dots A_{i_k})$, where $\text{Group}(A_{i_1} A_{i_2} \dots A_{i_k})$ is a subset of $\{A_1, A_2, \dots, A_n\}$,

Rule 1: If $F(t-k) = A_{i_1} A_{i_2} \dots A_{i_k}$, and $|\text{Group}(A_{i_1} A_{i_2} \dots A_{i_k})| = 0$, then the predicted result at time t is

$$\frac{\sum\limits_{j=1}^{k} (k+1-j) \times m_{i_{(k+1-j)}}}{\sum\limits_{j=1}^{k} j} \tag{5}$$

where $m_{i_1}, m_{i_2}, \dots, m_{i_k}$, is the midpoint of the interval $u_{i_1}, u_{i_2}, \dots, u_{i_k}$, respectively, in which the maximum membership degree of $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ locates.

Rule 2: If $F(t-k) = A_{i_1} A_{i_2} \dots A_{i_k}$ and $\text{Group}(A_{i_1} A_{i_2} \dots A_{i_k}) = \{A_j\}$, then the predicted result at time t, m_j, is the midpoint of interval u_j in which the maximum membership degree of A_j locates.

Rule 3: If $F(t-k) = A_{i_1} A_{i_2} \dots A_{i_k}$ and $\text{Group}(A_{i_1} A_{i_2} \dots A_{i_k}) = \{A_{j_1}, A_{j_2}, \dots, A_{j_p}\}$, then $k = k+1$; find the fuzzy logical relationship $A_{i_0} A_{i_1} A_{i_2} \dots A_{i_k} \rightarrow \text{Group}(A_{i_0} A_{i_1} A_{i_2} \dots A_{i_k})$, until $|\text{Group}(A_{i_0} A_{i_1} A_{i_2} \dots A_{i_k})| = 1$.

3. Issues of designing a deterministic forecasting model

In this section, we discuss two important issues in developing an effective forecasting model for fuzzy time series, which are ignored by most literature.

3.1. Controlling uncertainty

There are several interesting observations worth noting when investigating Chen's model in forecasting enrollments at the University of Alabama from 1971 to 1992. Firstly, Table 3 shows the forecasting outputs and errors for the first-order model, where the forecasting error = (first-order forecasting enrollment) − (actual enrollment). The most forecasting errors occur at 1400 in 1982 and −1317 in 1988, which result from $A_4 \rightarrow A_3$ and $A_4 \rightarrow A_6$, respectively. Indeed, there exists a fuzzy relationship $A_4 \rightarrow \text{Group}(A_4) = \{A_3, A_4, A_6\}$, which indicates the degree of uncertainty of A_4.

Secondly, when one considers the standard deviation of forecasting errors as shown in Table 4, it is noted that the larger the number of items included in a group of fuzzy relationships, the more likely there is a larger standard deviation. This confirms the previous observation on the uncertainty issue.

Finally, we verify Chen's high-order forecasting model as illustrated in Table 5. Chen gave the opinion that order = 3 achieved the optimal accuracy. However, from Table 5, one notes that the 2-, 3-, 4-, and 5-order forecasting values are all the same, except that the predictions of previous $(k-1)$ years are unavailable for the k-order case. As a result, it should not be concluded that order = 3 will be the best-forecasting outcome. In this sense, using a higher-order forecasting model is not necessary.

Therefore, the number of orders should not be the only factor that decides the model's accuracy. Instead, the issue of controlling uncertainty should be also taken into consideration. These stimulate the emergence of a deterministic forecasting approach based on the following heuristic rule by extending Chen's high-order forecasting model. If an initial fuzzy set has more than one fuzzy logical relationship, its fuzzy logical relationship will be constructed in an incremental order, backtracking to its previous $k+1$ time. The construction process continues until the fuzzy set has only one or no corresponding fuzzy logical relationship in the group, i.e., a certain state is reached.

3.2. Consistent accuracy with interval lengths

Huarng investigated the impact of interval length on the forecasting results and indicated that there will be no fluctuations when the length of intervals is too large, whereas the meaning of fuzzy time series will be diminished when the length is too small [13]. He proposed two heuristic approaches in determining the length of intervals, namely distribution and average-based [13]. Although the improvement of forecasting accuracy over Chen's model had been demonstrated, there are two flaws that exist in Huarng's method. First, there was no explanation of why and how the group determined the 'base-mapping table', which they rely on. Second, the numbers of intervals identified by the two

Table 3
The forecasting error on Chen's first-order forecasting model

Year	Actual enrollment	Fuzzy set	First-order forecasting enrollment	Forecasting error
1971	13 055	A_1		
1972	13 563	A_1	14 000	437
1973	13 867	A_1	14 000	133
1974	14 696	A_2	14 000	−696
1975	15 460	A_3	15 500	40
1976	15 311	A_3	16 000	689
1977	15 603	A_3	16 000	397
1978	15 861	A_3	16 000	139
1979	16 807	A_4	16 000	−807
1980	16 919	A_4	16 833	−86
1981	16 388	A_4	16 833	445
1982	15 433	A_3	16 833	1400
1983	15 497	A_3	16 000	503
1984	15 145	A_3	16 000	855
1985	15 163	A_3	16 000	837
1986	15 984	A_3	16 000	16
1987	16 859	A_4	16 000	−859
1988	18 150	A_6	16 833	−1317
1989	18 970	A_6	19 000	30
1990	19 328	A_7	19 000	−328
1991	19 337	A_7	19 000	−337
1992	18 876	A_6	19 000	124

Table 4
The standard deviation of forecasting error on Chen's model

Item	Group(A_i)	Standard deviation
A_1	$\{A_1, A_2\}$	478.81
A_2	$\{A_3\}$	–
A_3	$\{A_3, A_4\}$	612.60
A_4	$\{A_3, A_4, A_6\}$	981.29
A_6	$\{A_6, A_7\}$	179.00
A_7	$\{A_6, A_7\}$	230.50

approaches (18 and 24 intervals in the historical enrollments at the University of Alabama, respectively) are far more than the traditional Song and Chissom or Chen's models (seven intervals). However, too many intervals could result in fewer fluctuations in the fuzzy time series, as Huarng indicated. It also complicates the task of defuzzification. Moreover, as the author declares, the experimental result shown in [13] is inconsistent with the general principle that the more intervals identified, the better the accuracy that can be achieved.

To effectively control the uncertainty and to eliminate the inconsistency of interval partitioning, we now propose the deterministic forecasting model for fuzzy time series.

4. Deterministic forecasting model

By solving the aforementioned two important issues, the details of the deterministic forecasting model for fuzzy time series is described in the following subsections.

4.1. Interval partition and fuzzification

The first step in the model is determining the universe of discourse and partition intervals. For forecasting enrollment, Song and Chissom choose 1000 as the length of intervals without any reason [1]. Huarng [13] and Li and Chen [14] get 500 as the length of intervals by distribution-based and natural partitioning-based, respectively. The resulting forecasting accuracy is almost equivalent except the interval partitioning. This implies that the length of the

Table 5
Chen's high-order forecasting model on fuzzy time series

Year	Actual enrollment	2-order forecasting enrollment	3-order forecasting enrollment	4-order forecasting enrollment	5-order forecasting enrollment
1971	13 055				
1972	13 563	13 750			
1973	13 867	13 750	13 750		
1974	14 696	14 750	14 750	14 750	
1975	15 460	15 250	15 250	15 250	15 250
1976	15 311	15 250	15 250	15 250	15 250
1977	15 603	15 750	15 750	15 750	15 750
1978	15 861	15 750	15 750	15 750	15 750
1979	16 807	16 750	16 750	16 750	16 750
1980	16 919	16 750	16 750	16 750	16 750
1981	16 388	16 250	16 250	16 250	16 250
1982	15 433	15 250	15 250	15 250	15 250
1983	15 497	15 250	15 250	15 250	15 250
1984	15 145	15 250	15 250	15 250	15 250
1985	15 163	15 250	15 250	15 250	15 250
1986	15 984	15 750	15 750	15 750	15 750
1987	16 859	16 250	16 250	16 250	16 250
1988	18 150	18 250	18 250	18 250	18 250
1989	18 970	18 750	18 750	18 750	18 750
1990	19 328	19 250	19 250	19 250	19 250
1991	19 337	19 250	19 250	19 250	19 250
1992	18 876	18 750	18 750	18 750	18 750

universe of discourse U will also influence the forecasting accuracy and subsequently will affect the strategic decision-making. Song and Chissom [1] and Tsaur, Yang, and Wang [10] defined the universe U as $[D_{\min} - D_1, D_{\max} + D_2]$, where D_1 and D_2 are two proper positive numbers, but did not explain the reason of how to determine the 'proper positive numbers', which they rely on. As stated in Section 1, one may appeal to the advice from an expert of the domain under study. Therefore, initially determining the appropriate length of interval l, then the lower and upper bound of the universe of discourse U, D_{low} and D_{up}, and the number of fuzzy sets n, can be derived by

$$D_{\text{low}} = \left\lfloor \frac{D_{\min}}{l} \right\rfloor \times l, \qquad D_{\text{up}} = \left\{ \left\lfloor \frac{D_{\max}}{l} \right\rfloor + 1 \right\} \times l, \qquad n = \frac{D_{\text{up}} - D_{\text{low}}}{l} \tag{6}$$

where $D_{\min}$ and $D_{\max}$ are the minimal value and maximal values of the known historical data, respectively, and $\lfloor \bullet \rfloor$ is the floor function.

In the enrollments data of the University of Alabama, $D_{\min} = 13\,055$ and $D_{\max} = 19\,337$. Generally, the runner of university predicts enrollments in thousands per year, and in order to compare accuracy with the prior method, we choose $l = 1000$; then $D_{\text{low}} = \lfloor 13\,055/1000 \rfloor \times 1000 = 13\,000$, $D_{\text{up}} = \{\lfloor 19\,337/1000 \rfloor + 1\} \times 1000 = 20\,000$, and the universe of discourse is thus $U = [13\,000, 20\,000]$, and $n = (20\,000 - 13\,000)/1000 = 7$. By partitioning the universe of discourse U into seven intervals with equal length, we obtain $u_1 = [13\,000, 14\,000)$, $u_2 = [14\,000, 15\,000)$, $u_3 = [15\,000, 16\,000)$, $u_4 = [16\,000, 17\,000)$, $u_5 = [17\,000, 18\,000)$, $u_6 = [18\,000, 19\,000)$, $u_7 = [19\,000, 20\,000]$.

The second step of the proposed model is defining fuzzy sets on the universe of discourse and fuzzifying the time series. To fuzzy the enrollment time series, fuzzy sets A_i $(i = 1, 2, \ldots, 7)$ have to be defined on the linguistic variable, and the membership degree of each interval u_j $(j = 1, 2, \ldots, 7)$ in A_i. For this, seven linguistic values can be defined as follows: A_1 = (not many), A_2 = (not too many), A_3 = (many), A_4 = (many many), A_5 = (very many), A_6 = (too many), A_7 = (too many many). In this way, all the fuzzy sets, A_i $(i = 1, 2, \ldots, 7)$, are expressed as:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7,$$
$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7,$$
$$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + 0/u_6 + 0/u_7,$$

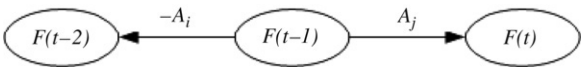

Fig. 1. State transition and backtracking.

$$A_4 = 0/u_1 + 0/u_2 + 0.5/u_3 + 1/u_4 + 0.5/u_5 + 0/u_6 + 0/u_7,$$
$$A_5 = 0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5 + 0.5/u_6 + 0/u_7,$$
$$A_6 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0.5/u_5 + 1/u_6 + 0.5/u_7,$$
$$A_7 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0.5/u_6 + 1/u_7.$$

After finding out the degree of each year's enrollment belonging to an appropriate A_i ($i = 1, 2, \ldots, 7$) in Table 3, the fuzzified enrollment for that year was treated as A_i, which is where the maximum membership degree of some year's enrollment occurred. Therefore, the fuzzified time series of enrollments is represented as $A_1, A_1, A_1, \ldots, A_7, A_7, A_6$.

4.2. *Identifying all certain transitions*

The third step in the forecasting model is identifying all certain transitions. Conceptually, we use the state transition diagram as shown in Fig. 1 to model the casual relationship between two fuzzy time series, $F(t)$ and $F(t-1)$, in which $F(t)$ is caused by $F(t-1)$.

It indicates that state $F(t)$ is reached when state $F(t-1)$ moves forward one time step with edge A_j. A state s for some fuzzy time series can have more than one state transition leaving that state. In this situation, state s is called an uncertain state; otherwise, s is a certain state. The transition which leads to a certain state is named as 'certain transition'. To eliminate uncertainties which could result in larger prediction errors as analyzed in Section 3, a backtracking scheme can be conducted. Backtracking means finding the previous state of s, i.e., a fuzzy time series begins at A_i followed by $F(t-1)$. We use the negation sign on the edge leaving state s to indicate that a backtracking action is required to be performed on it (see Fig. 1).

The backtracking process will generate new states which can be uncertain, and thus needs to be processed in the same way. Given a fuzzy time series $F(t-1) = f_1(t-1)f_2(t-1)\cdots f_q(t-1)$, where $f_i(t-1) \in A, i = 1, \ldots, q$, one needs to identify all certain transitions in order to facilitate the forecasting in the next step. Table 6 illustrates the algorithm of identifying all certain transitions.

Set P is the resulting fuzzy logical relationship set, of which the fuzzy logical relationship is in the form of $c \rightarrow S$, where c and S are the cause and effect of the state transition, respectively. Therefore, set P is also named as a 'cause–effect' set.

For the example of enrollments at the University of Alabama, we have the fuzzy set $A = \{A_1, A_2, \ldots, A_7\}$, and fuzzy time series $F(t-1) : A_1\ A_1\ A_1\ A_2\ A_3\ A_3\ A_3\ A_3\ A_4\ A_4\ A_4\ A_3\ A_3\ A_3\ A_3\ A_3\ A_4\ A_6\ A_6\ A_7\ A_7\ A_6$, where $f_0(t-1) = \%$, $f_1(t-1) = A_1$, $f_2(t-1) = A_1, \ldots, f_{22}(t-1) = A_6$. Initially, $F_{1,1}(t-1) = A_1$, $F_{2,1}(t-1) = A_1, \ldots, F_{21,1}(t-1) = A_7$, and $F_{22,1}(t-1) = A_6$; the candidate set C is thus $C = \{A_1, A_2, A_3, A_4, A_6, A_7\}$.

After constructing the initial set C, all possible state transitions will be derived and analyzed. For example, when $c = A_1$, the possible current states are $F_{1,1}(t-1) = A_1$, $F_{2,1}(t-1) = A_1$, and $F_{3,1}(t-1) = A_1$, whose next state set can be $S = \{f_2(t-1), f_3(t-1), f_4(t-1)\} = \{A_1, A_2\}$, which in turn leads to backtracking to the previous state set $R = \{\%A_1, A_1A_1\}$ due to $F_{0,2}(t-1) = \%A_1$, $F_{1,2}(t-1) = A_1A_1$ and $F_{2,2}(t-1) = A_1A_1$. Fig. 2 illustrates the state transition diagram for the example. Therefore, the updated candidate set is $C = \{A_2, A_3, A_4, A_6, A_7, \%A_1, A_1A_1\}$.

When taking into consideration $c = \%A_1$, since $F_{0,2}(t-1) = \%A_1$ and $f_2(t-1) = A_1$, $S = \{A_1\}$ and $|S| = 1$, a certain state transition $\%A_1 \rightarrow \{A_1\}$ is obtained and added into the cause–effect set P. The candidate set now becomes $C = \{A_2, A_3, A_4, A_6, A_7, A_1A_1\}$.

For $c = A_1A_1$, thanks to the facts of $F_{1,2}(t-1) = A_1A_1$, $f_3(t-1) = A_1$, $F_{0,3}(t-1) = \%A_1A_1$, $F_{2,2}(t-1) = A_1A_1$, $f_4(t-1) = A_2$, and $F_{1,3}(t-1) = A_1A_1A_1$, the next state set is $S = \{A_1, A_2\}$. Therefore, $|S| = 2$ and $R = \{\%A_1A_1, A_1A_1A_1\}$, which make the candidate set to be modified as $C = \{A_2, A_3, A_4, A_6, A_7, \%A_1A_1, A_1A_1A_1\}$.

In the case of $c = \%A_1A_1$, due to $F_{0,3}(t-1) = \%A_1A_1$ and $f_3(t-1) = A_1$, $S = \{A_1\}$ and a certain transition $\%A_1A_1 \rightarrow \{A_1\}$ is added to the cause–effect set P. The resulting candidate set becomes $C = \{A_2, A_3, A_4, A_6, A_7, A_1A_1A_1\}$ and the cause–effect set is $P = \{\%A_1A_1 \rightarrow \{A_1\}, \%A_1 \rightarrow \{A_1\}\}$.

Table 6
The algorithm of generating the set of certain transitions

Input: A fuzzy set $A = \{A_i \mid i = 1, 2, \ldots, n\}$, and a fuzzy time series $F(t-1)$. $f_i(t-1) \in A$, $i = 1, \ldots, q$. $F(t-1) = f_0(t-1)F(t-1)$, where $f_0(t-1) = \%$, representing the beginning of the fuzzy time series.
Output: A set of certain transitions P in the fuzzy time series $F(t-1)$.
Algorithm:
Let fuzzy time series $F_{j,k}(t-1)$ be a subsequence of $F(t-1)$ with length k which starts from $f_j(t-1)$. The candidate set C is a set of subsequences of $F(t-1)$, representing the states whose certainty property needs to be examined.

Let S be the subset of fuzzy set A that were caused from $F_{j,k}(t-1)$ and R be the set of subsequences of fuzzy time series that backtracks $F_{j,k}(t-1)$ one time. $|S|$ is the number of elements in set S.

```
P = Ø
C = Ø

for  j = 1  to  q
  if  F_{j,1}(t − 1) ∉ C  then  C = C ∪ {F_{j,1}(t − 1)}
next j

for each element c in C
  begin
    C = C − {c}
    k = length(c)
    S = Ø
    R = Ø
    for  j = 0  to  q − k + 1
      begin
        if  F_{j,k}(t − 1) = c  then
            if  f_{j+k}(t − 1) ∉ S  then
            S = S ∪ {f_{j+k}(t − 1)}
        if  j > 0  and  F_{j−1,k+1}(t − 1) ∉ C  then
            R = R ∪ {F_{j−1,k+1}(t − 1)}
      end
    if  |S| = 0  then  P = P ∪ {c → Ø}
    if  |S| = 1  then  P = P ∪ {c → S}
    if  |S| > 1  then  C = C ∪ R
  end
return P
```

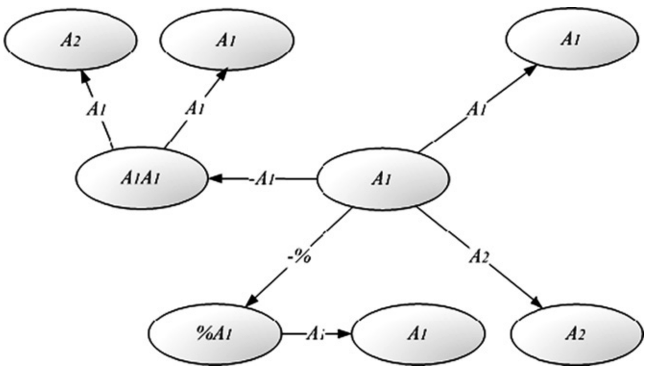

Fig. 2. The state transition diagram for $c = A_1$.

For $c = A_1A_1A_1$, $S = \{A_2\}$,owing to $F_{1,3}(t-1) = A_1A_1A_1$ and $f_4(t-1) = A_2$. Therefore a new certain transition $A_1A_1A_1 \to \{A_2\}$ is added to P and $C = \{A_2, A_3, A_4, A_6, A_7\}$.

For the example of $c = A_2$, the facts of $F_{4,1}(t-1) = A_2$ and $f_5(t-1) = A_3$ result in $S = \{A_3\}$ and the certain transition $A_2 \to \{A_3\}$. The candidate set is accordingly updated as $C = \{A_3, A_4, A_6, A_7\}$.

Following the algorithm in Table 6, the final cause–effect set is as follows:

$$\begin{aligned} P = \{ & \%A_1 \to \{A_1\}, .\%A_1A_1 \to \{A_1\}, A_1A_1A_1 \to \{A_2\}, A_2 \to \{A_3\}, \\ & A_2A_3 \to \{A_3\}, A_4A_3 \to \{A_3\}, A_2A_3A_3 \to \{A_3\}, A_4A_3A_3 \to \{A_3\}, A_2A_3A_3A_3 \to \{A_3\}, \\ & A_4A_3A_3A_3 \to \{A_3\}, A_2A_3A_3A_3A_3 \to \{A_4\}, A_3A_3A_3A_3A_3 \to \{A_4\}, \\ & A_4A_3A_3A_3A_3 \to \{A_3\}, A_3A_4A_4 \to \{A_4\}, A_4A_4A_4 \to \{A_3\}, \\ & A_2A_3A_3A_3A_3A_4 \to \{A_4\}, A_3A_3A_3A_3A_3A_4 \to \{A_6\}, A_4A_6 \to \{A_6\}, \\ & A_6A_6 \to \{A_7\}, A_7A_6 \to \{\}, A_6A_7 \to \{A_7\}, A_7A_7 \to \{A_6\}\}. \end{aligned}$$

In addition to identifying certain transitions given a fuzzy time series $F(t-1)$, it is of great interest to deterministically quantify the maximum length of subsequence in the fuzzy time series which leads to a certain state. We denote the length as w. The quantification can help the analysis of the best and worst cases in the above algorithm.

Theorem 1. *For the best case, the maximum length of subsequence in the fuzzy time series which leads to a certain state is one, i.e.,* $w = 1$.

Proof. Apparently, the best case occurs when no backtracking action is needed in identifying all certain transitions in a fuzzy time series. In this situation, if $q > 1$, it means for each A_i, $i = 1, \ldots, n$, there, at least, exists a certain transition $A_i \to S$, where $|S| = 0$ or 1. Thus, $w = 1$. If $q = 1$, then $S = \emptyset$, thus $w = 1$. ■

Theorem 2. *For the worst case, the maximum length of subsequence in the fuzzy time series which results in a certain state is* $w = q - 1$.

Proof. The worst case is encountered when there is a need for backtracking and the maximal backtracking length is theoretically q. However, it is not infeasible; instead, it should be $q - 1$, as follows.

Given a fuzzy time series $F(t-1) = f_1(t-1)f_2(t-1)\cdots f_q(t-1)$, the condition of $w = q$ implies that $F(t-1) \to \{\}$ must be in the set of certain transitions P. It indicates that there exist, at least, two equivalent subsequences with length $q - 1$,

$$\begin{aligned} & f_2(t-1)f_3(t-1)\cdots f_q(t-1) = f_{k_1-q+2}(t-1)f_{k_1-q+3}(t-1)\cdots f_{k_1}(t-1) \\ & \quad = \cdots = f_{k_i-q+2}(t-1)f_{k_i-q+3}(t-1)\cdots f_{k_i}(t-1) = \cdots = f_{k_l-q+2}(t-1)f_{k_l-q+3}(t-1)\cdots f_{k_l}(t-1), \end{aligned}$$

where $k_i < q$, $i = 1, \ldots, l$ and the following constraint is satisfied:

$$f_{q+1}(t-1), f_{k_1+1}(t-1), \ldots, \text{ and } f_{k_l+1}(t-1) \text{ are not all equal.}$$

From the definition of fuzzy time series, $f_i(t-1)$, $\forall i > 0$ and we let $f_0(t-1) = \%$, hence, $k_i - q + 2 \geq 0$; then $q - 2 \leq k_i < q$. Thus, k_i is equal to either $q - 1$ or $q - 2$. That is, there are only the following three subsequences that are equivalent:

$$f_2(t-1)f_3(t-1)\cdots f_q(t-1) = f_1(t-1)f_2(t-1)\cdots f_{q-1}(t-1) = f_0(t-1)f_1(t-1)\cdots f_{q-2}(t-1).$$

However, the beginning of the algorithm is defined as $f_0(t-1) = \%$; therefore, the subsequence $f_0(t-1)f_1(t-1)\cdots f_{q-2}(t-1)$ is removed from the equation, i.e.

$$f_2(t-1)f_3(t-1)\cdots f_q(t-1) = f_1(t-1)f_2(t-1)\cdots f_{q-1}(t-1).$$

Since the next states of $f_1(t-1)f_2(t-1)\cdots f_{q-1}(t-1)$ and $f_2(t-1)f_3(t-1)\cdots f_q(t-1)$ are $f_q(t-1)$ and $\{\}$, respectively, a certain state transition $f_1(t-1)f_2(t-1)\cdots f_{q-1}(t-1) \to \{f_q(t-1)\}$ is obtained according to the algorithm of generating the set of certain transitions (Table 6). In either case, w is equal to $q - 1$. Hence the maximum length of subsequence in the fuzzy time series which leads to a certain state will be $w = q - 1$.

For example, assume a fuzzy set $A = \{A_1, A_2\}$, and a fuzzy time series $F(t-1) : A_1\ A_1\ A_1\ A_1\ A_1\ A_1\ A_1\ A_1\ A_1\ A_2$, $q = 10$. The final certain transition set P is as follows:

$$\begin{aligned} P = \{ & A_2 \to \{\}, \%A_1 \to A_1, \%A_1A_1 \to A_1, \%A_1A_1A_1 \to A_1, \%A_1A_1A_1A_1 \to A_1, \\ & \%A_1A_1A_1A_1A_1 \to A_1, \%A_1A_1A_1A_1A_1A_1 \to A_1, \%A_1A_1A_1A_1A_1A_1A_1 \to A_1, \\ & \%A_1A_1A_1A_1A_1A_1A_1A_1 \to A_1, A_1A_1A_1A_1A_1A_1A_1A_1A_1 \to A_2\}. \end{aligned}$$

The maximum length of subsequence in the fuzzy time series which leads to a certain state is thus $w = 9$. ■

With Theorems 1 and 2, we may further obtain the following important theorem.

Theorem 3. *If* $F_{i,j}(t-1) \to S \in P$ *then* $F_{i+1,j-1}(t-1) \to S \notin P$.

Proof. This theorem holds because if $F_{i+1,j-1}(t-1) \to S \in P$, then $F_{i+1,j-1}(t-1) \to S$ is a certain state transition; there is no need for backtracking. Therefore, $F_{i,j}(t-1) \to S \notin P$. This theorem provides a very useful heuristic in forecasting the output at next time t, as will be described in the following section. ■

After the above analysis, the complexity of backtracking can be analyzed. Assume a fuzzy time series $F(t-1)$ with a fuzzy set $A = \{A_i \mid i = 1, 2, \ldots, n\}$, $f_i(t-1) \in A, i = 1, 2, \ldots, q$, where $q \geq n$. The maximum times of backtracking in the algorithm of generating the set of certain transitions is

$$\sum_{k=1}^{q-2} \min(n^{k+1} + 1, q - k + 1),$$

and its complexity is $O(q^2)$.

The maximum times of backtracking in the algorithm of generating the set of certain transitions occurs when each element c in the candidate set C needs backtracking. Initially, consider the subsequences c in C with length one, i.e. $k = \text{length}(c) = 1$. Set R contains all possible backtracking one step subsequences which are the repeated permutation results, and one unique $\%A_i$, i.e. $R = \{A_j A_i \mid i, j = 1, \ldots, n\} \cup \{\%A_i \mid i = 1, \text{ or } 2, \ldots, \text{ or } n\}$. Since $A_j A_i$ is a subsequence of $F(t-1)$, the actual possible number of $A_j A_i$ in C cannot be larger than the number of subsequences with length two extracted from the fuzzy time series, which is $q - 1$. Therefore, the worst backtracking times for subsequences with length one ($k = 1$) is $\min(n^{k+1} + 1, q - k + 1) = \min(n^2 + 1, q)$. In the following, consider the subsequences c in C with length k and increment the length of subsequence by backtracking one step. By Theorem 2, the maximum length of subsequence which results in a certain state is $q - 1$, whereas such a subsequence is obtained by backtracking from a subsequence with length $k = q-2$. Therefore, the maximum times of backtracking in the algorithm of generating the set of certain transitions is

$$\sum_{k=1}^{q-2} \min(n^{k+1} + 1, q - k + 1),$$

and its complexity is $O(q^2)$.

For the sake of illustration, we take into consideration the experiment with different number of fuzzy sets, $n = 5, 7, 9, 11$, and the length of fuzzy time series, q, ranging from 5 to 100. Fig. 3 shows how the parameter set (n, q) affects the times of backtracking in the algorithm of generating the set of certain transitions. It is clear that the backtracking times grow approximately with q^2. On the other hand, the difference of backtracking times is not significant for different n, which implies that the performance is not as sensitive to n.

4.3. Forecasting and defuzzifying

The last step in the forecasting model is forecasting and defuzzifying the forecasting outputs. Let the historic fuzzy time series be $F(t-1)$, and the length of $F(t-1)$ be q. In addition, let the given query for fuzzy time series be

$$F'(t-1) = f_1'(t-1) \cdots f_i'(t-1) \cdots f_r'(t-1), \tag{7}$$

where $f_i'(t-1) \in A$, $i = 1, 2, \ldots, r$. The key point of forecasting is that if r is larger than or equal to w, one has only to look into the subsequence with length w, which begins at

$$F'_{r-w+1,w}(t-1) = f'_{r-w+1}(t-1) f'_{r-w+2}(t-1) \cdots f_r'(t-1). \tag{8}$$

On the other hand, if r is less than w, the subsequence

$$F'_{0,r+1}(t-1) = f_0'(t-1) f_1'(t-1) f_2'(t-1) \cdots f_r'(t-1), \tag{9}$$

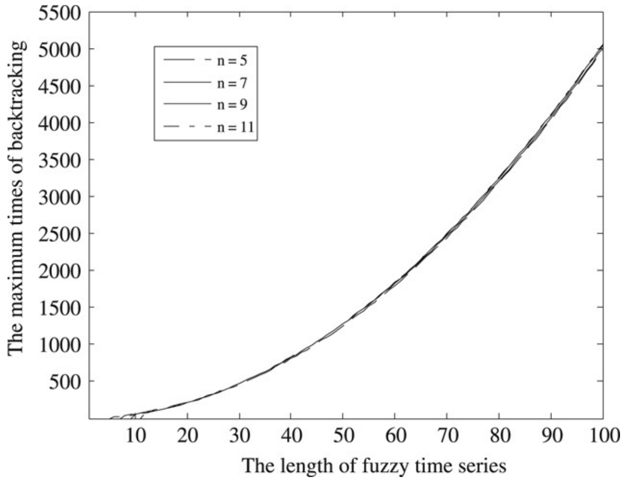

Fig. 3. The maximum times of backtracking.

Table 7
The algorithm of forecasting and defuzzification

Input: $F(t-1)$, w, the cause–effect set P; the query fuzzy time series $F'(t-1)$
Output: the crisp forecasting output of time t
Algorithm:
begin
  Let $q = \text{length}(F(t-1))$, $r = \text{length}(F'(t-1))$
  if $r \geq w$ **then** $(i, k, S) = \text{forecasting}(i, k, F'_{r-w+1,w}(t-1))$
  if $r < w$ **then** $(i, k, S) = \text{forecasting}(i, k, F'_{0,r+1}(t-1))$,
  Let $F_{i,k}(t-1) \to S \equiv f_i(t-1) f_{i+1}(t-1) \cdots f_{i+k-1}(t-1) \to S \equiv A_{i_1} A_{i_2} \cdots A_{i_k} \to S$
  if $S = \{A_e\}$
    and the maximum membership value of A_e occurs at interval u_e,
    and the midpoint of u_e is m_e,
  then return m_e.
  if $S = \emptyset$
    and the maximum membership values of $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ occur at intervals $u_{i_1}, u_{i_2}, \ldots, u_{i_k}$, respectively,
    and the midpoints of $u_{i_1}, u_{i_2}, \ldots, u_{i_k}$ are $m_{i_1}, m_{i_2}, \ldots, m_{i_k}$, respectively,
  then return $\frac{1}{k} \sum_{j=1}^{k} m_{i_j}$
end

function forecasting$(i, k, F_{i,k}(t-1))$

begin
  if a value is bound to the key $F_{i,k}(t-1)$ in P **then return** $(i, k, P[F_{i,k}(t-1)])$
  if $k = 1$
    then return $(i, k, \emptyset)$
    else forecasting$(i+1, k-1, F_{i+1,k-1}(t-1))$
end

where $f'_0(t-1) = \%$, needs to be examined. Then, one recursively searches for the subsequence in the cause–effect pattern set, P, and retrieves the forecasting fuzzy output of time t, if any. Since w is the maximum of length of historic fuzzy time series which leads to a certain state, the search process proceeds by looking for all patterns with length from w down to one in P. Finally, a similar defuzzification procedure as Chen's approach is conducted to obtain the crisp output of time t. The algorithm of forecasting and defuzzification is illustrated in Table 7. One notes that in order to efficiently retrieve the associated effect given a cause, the cause–effect set, P, is implemented in a data structure of an associative array, which provides a handy way to store data in a group. An associate array is created with a set of key/value pairs so that the associated value with a key can be retrieved by simply looking up the array.

Table 8
The forecasting result of the proposed forecasting model

Year	Actual enrollment	Fuzzy set	Forecasting enrollment	Forecasting error
1971	13 055	A_1		
1972	13 563	A_1	13 500	−63
1973	13 867	A_1	13 500	−367
1974	14 696	A_2	14 500	−196
1975	15 460	A_3	15 500	40
1976	15 311	A_3	15 500	189
1977	15 603	A_3	15 500	−103
1978	15 861	A_3	15 500	−361
1979	16 807	A_4	16 500	−307
1980	16 919	A_4	16 500	−419
1981	16 388	A_4	16 500	112
1982	15 433	A_3	15 500	67
1983	15 497	A_3	15 500	3
1984	15 145	A_3	15 500	355
1985	15 163	A_3	15 500	337
1986	15 984	A_3	15 500	−484
1987	16 859	A_4	16 500	−359
1988	18 150	A_6	18 500	350
1989	18 970	A_6	18 500	−470
1990	19 328	A_7	19 500	172
1991	19 337	A_7	19 500	163
1992	18 876	A_6	18 500	−376

In the example of enrollments at the University of Alabama, $w = 6$ is decided deterministically as in the previous section. When forecasting the enrollments in 1972, since $F'(t-1) = A_1$, $r = 1 < w$, $(i, k, S) =$ forecasting$(F'_{0,2}(t-1))$, we have $i = 0, k = 2$, and $S = \{A_1\}$, which indicates that a certain transition pattern $\%A_1 \rightarrow \{A_1\}$ is retrieved. The forecasting output of year 1972 is thus $m_1 = 13\,500$, which is the midpoint of u_1, defined in Step 1, Section 5.

When time $t = 1973$, because of $F'(t-1) = A_1A_1$, $r = 2 < w$, and $(i, k, S) =$ forecasting$(F'_{0,3}(t-1))$, the retrieved pattern is $F_{0,3}(t-1) = \%A_1A_1$, and $S = \{A_1\}$. As a result, the forecasting output of year 1973 is $m_1 = 13\,500$.

If time $t = 1974$, due to $F'(t-1) = A_1A_1A_1$, $r = 3 < w$, and $(i, k, S) = (F'_{0,4}(t-1))$, $F_{0,4}(t-1) = \%A_1A_1A_1$; however, there is not a certain state found in the cause–effect set P. By decrementing k, we search for $F_{1,3}(t-1) = A_1A_1A_1$ and obtain $S = \{A_2\}$; the forecasting output of year 1974 is hence $m_2 = 14\,500$, the midpoint of u_2.

When forecasting 1982, thanks to $F'(t-1) = A_1\ A_1\ A_1\ A_2\ A_3\ A_3\ A_3\ A_3\ A_4\ A_4\ A_4$, $r = 11 > w$, one only has to look into the subsequence $F'_{r-w+1,w}(t-1) = F'_{6,6}(t-1) = A_3\ A_3\ A_3\ A_4\ A_4\ A_4$, in P. The search for a certain state in P continues for $F_{7,5}(t-1)$ and $F_{8,4}(t-1)$ until $F_{9,3}(t-1)$, of which a certain state transition $A_4A_4A_4, \rightarrow \{A_3\}$ is achieved. Consequently, the defuzzied forecast of year 1982 is $m_3 = 15\,500$, which is the midpoint of u_3. The overall yearly forecasting result is shown in Table 8.

5. Performance evaluation and comparison

In this section, we evaluate the forecasting performance of the proposed fuzzy time series model with enrollments at the University of Alabama from 1971 to 1992 and compare it with the previous models, which all used the enrollment data set as the benchmark [1–6,8–10,13,14]. To be fair, we conduct our experimentation using the same benchmark with the same interval length 1000. In addition, the impact of various interval lengths is investigated to validate the consistency issue.

Fig. 4 shows the comparison with Chen's first-order model, which demonstrates that the proposed model with the same length, 1000, is capable of forecasting more accurately. Fig. 5 shows that the forecasting reliability of the proposed model using the analysis of residual scatter is better than Chen's forcasting model. The residues of our model scatter between −500 and 500, included in $(-l/2, l/2)$, where l is the interval length, but the maximal and minimal residues of the latter model are 1400 and −1317, respectively. This achievement is constributed by the fact that the

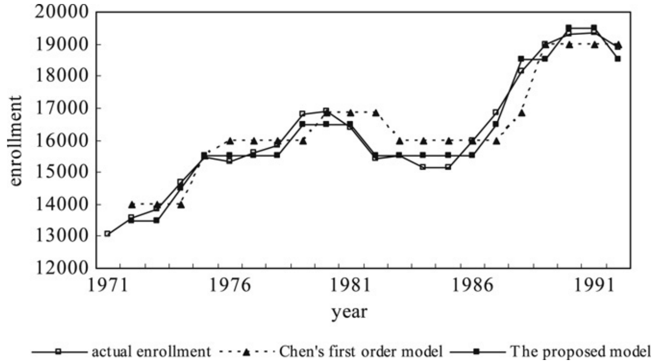

Fig. 4. Curves of forecasting enrollments and actual enrollments with length 1000.

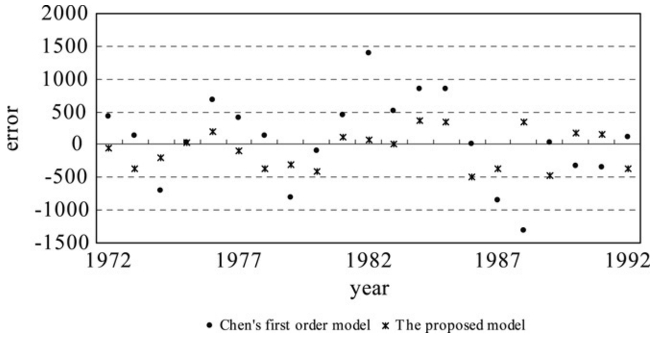

Fig. 5. Comparison of residual scatter.

standard deviation of the proposed model decreases, and therefore the uncertainty grade also decreases, making the precision increase.

The forecasting accuracy is measured in terms of the mean square error (MSE), E_{MSE},

$$E_{\mathrm{MSE}} = \frac{1}{m} \sum_{i=1}^{m} (o_i - d_i)^2, \tag{10}$$

where o_i and d_i are the forecasting and actual enrollments, respectively, and m is the number of records in the enrollment database. The comparison of the proposed model and others in terms of forecasting accuracy is illustrated in Table 9. The E_{MSE} of the proposed method is 85 040, less than all others. It is even better than the Chen's 3-order model. In all the literature reviewed, the forecasting result is the best achieved.

The performance comparison is further validated by forecasting the errors percentage, E_{FEP} and average forecasting errors percentage, E_{AFEP}, which are defined as follows:

$$E_{\mathrm{FEP}} = \frac{|(o - d)|}{d} \times 100\% \tag{11}$$

$$E_{\mathrm{AFEP}} = \frac{1}{m} \sum_{i=1}^{m} \frac{|(o_i - d_i)|}{d_i} \times 100\%. \tag{12}$$

Table 10 illustrates the comparison result, which shows that E_{FEP} of the proposed model ranges from 0.02% to 3.03%, and E_{AFEP} is 1.53%. It is better than Song's study, whose E_{FEP} is from 0.1% to 8.7%, and E_{AFEP} is 3.18%. It outperforms Chen's model as well. In addition, it is superior to Tsaur's model, which achieved E_{FEP} ranging from 0.02% to 4.8%, and 1.86% in terms of E_{AFEP} [10].

To investigate the consistency issue of interval length, we conducted an experiment on the comparison of the proposed method with Huarng's model [13] according to various lengths of intervals. The result is summarized in Table 11. One notes that, for our model, the smaller the length of interval is, the better the achieved forecasting

Table 9
The E_{MSE} comparison of various models with the length 1000

Model	Song and Chissom [9]	Song and Chissom ($w = 4$) [11]	Sullivan and Woodall [12]	Chen [1]	Chen (2-order) [3]	Chen (3-order) [3]	Chen (4-order) [3]	Chen (5-order) [3]	Chen (6-order) [3]	Chen (7-order) [3]	Chen (8-order) [3]	Chen (9-order) [3]	Tsaur [14]	The proposed method
MSE	412 499	775 687	386 055	407 507	89 093	86 694	89 376	94 539	98 215	104 056	102 179	102 789	134 923	85 040

Table 10
The E_{FEP} and E_{AFEP} comparison of various models with the length 1000

Model	Song and Chissom [9]	Chen [1]	Tsaur [14]	The proposed method
E_{FEP} (%)	0.1–8.7	0.1–9.07	0.02–4.8	0.02–3.03
E_{AFEP} (%)	3.18	3.11	1.86	1.53

Table 11
The E_{MSE} comparison with various length of intervals

Length of interval	200	300	400	500	600	700	800	900	1000
Huarng [5]	104 640	78 792	124 707	173 453	254 592	222 557	365 045	246 892	407 521
The proposed method	3 040	7 154	14 916	21 516	26 354	54 735	62 783	72 440	85 040

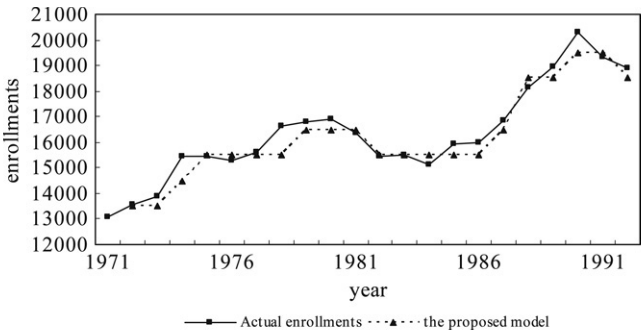

Fig. 6. Curves of forecasting enrollments and actual enrollments with contaminated data.

accuracy. In Huarng's model, the mean square error in the smallest length, 200, does not forecast better than 300; the smaller length, 600, does not forecast better than 700; and the length of 800 does not forecast better than 900, either. However, the forecasting accuracy of the proposed model is consistent with the issue of interval length.

Finally, we take the robustness issue into consideration, which is concerned with whether a forecasting model can still yield good forecasting results when the historical data are not accurate or are contaminated. For this, we adopt the example shown in [1,2,5] to intentionally increase a few years' enrollment data by 5% with the rest of the data unchanged, i.e., the university enrollments of 1974, 1978, 1985, and 1990 are increased by 5%. The curves of the actual enrollments and the forecasting enrollments are depicted in Fig. 6. The proposed model accomplishes E_{FEP} ranging from 0.02% to 3.03%, and 1.45% for E_{AFEP}. It is much better than Song's study [1], of which is from 0.1% to 11%, and 3.9%, respectively. In addition, the result of Chen's range of 0.1% to 9.07%, and 3.23%, respectively [5], demonstrates our model's superiority. It is interesting to observe that, from Fig. 6, as time moves forward, the forecasting error decreases. This indicates that even if the historical data are not accurate, the proposed method can still make good forecasts.

6. Conclusion and future work

In this paper, we have proposed a deterministic forecasting model to deal with the forecasting problem of fuzzy time series. The proposed model is provoked by the need for controlling the uncertainty which exists in the fuzzy relationships groups and removing the inconsistency of partitioning intervals. It not only achieves the best accuracy with the least mean square errors of all related work in the area of forecasting the University of Alabama's enrollment, but it can also make robust forecasts when historical data are contaminated. Moreover, it coincides with the principle that the shorter the interval length is, the less the mean square error is, i.e., the forecasting accuracy of the proposed model is consistent with the length of intervals. In contrast to the major works in the literature, the maximum length of subsequence in a fuzzy time series resulting in a certain state is deterministically quantified, which facilitates the development of the new forecasting model in Song and Chissom's framework. The analysis of residual scatter further confirms the superiority of the proposed model in forecasting reliability. Future work involves applying the proposed model to deal with more complicated applications and extending it to handle the problem of multi-dimensional fuzzy time series.

Acknowledgement

The first author gratefully appreciates the financial support from National Science Council, Taiwan, ROC under contract NSC94-2416-H-006-019.

References

[1] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time series—part I, Fuzzy Sets and Systems 54 (1993) 1–9.
[2] Q. Song, B.S. Chissom, Fuzzy time series and its models, Fuzzy Sets and Systems 54 (1993) 269–277.
[3] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time series—part II, Fuzzy Sets and Systems 62 (1994) 1–8.
[4] J. Sullivan, W.H. Woodall, A comparison of fuzzy forecasting and Markov modeling, Fuzzy Sets and Systems 64 (1994) 279–293.
[5] S.-M. Chen, Forecasting enrollments based on fuzzy time series, Fuzzy Sets and Systems 81 (1996) 311–319.
[6] S.-M. Chen, Forecasting enrollments based on high-order fuzzy time series, Cybernetics and Systems: An International Journal 33 (2002) 1–16.
[7] S.-M. Chen, J.-R. Hwang, Temperature prediction using fuzzy time series, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics 30 (2000) 263–275.
[8] J.R. Hwang, S.M. Chen, C.H. Lee, Handling forecasting problems using fuzzy time series, Fuzzy Sets and System 100 (1998) 217–228.
[9] K. Huarng, Heuristic models of fuzzy time series for forecasting, Fuzzy Sets and Systems 123 (2001) 369–386.
[10] R.-C. Tsaur, J.-C. Yang, H.-F. Wang, Fuzzy relation analysis in fuzzy time series model, Computers & Mathematics with Applications 49 (2005) 539–548.
[11] T.-S. Lee, C.-C. Chiu, F.-C. Lin, Prediction of the unemployment rate using fuzzy time series with Box–Jenkins methodology, International Journal of Fuzzy Systems 3 (2001) 577–585.
[12] C.-C. Tsai, S.-J. Wu, A study for second-order modeling of fuzzy time series, Proceedings of the IEEE International Conference on Fuzzy Systems (1999) 719–725.
[13] K. Huarng, Effective lengths of intervals to improve forecasting in fuzzy time series, Fuzzy Sets and Systems 123 (2001) 387–394.
[14] 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.
[15] S.-T. Li, Y.-C. Cheng, A deterministic forecasting model for fuzzy time series, in: The Fourth IASTED International Conference on Computational Intelligence, Calgary, Alberta, Canada, 2005, pp. 25–30.

Sheng-Tun Li received his B.S. and M.S. degrees in computer engineering from Tamkang University, Taiwan, and his Ph.D. degree in computer science from the University of Houston, University Park, USA.

He is currently a professor of Institute of Information Management at National Cheng Kung University, Taiwan. Prior to joining NCKU, he was on the faculty of the Department of Information Management at National Kaohsiung First University of Science and Technology, Taiwan, during 1996–2003. His research interests include fuzzy systems, soft computing, intelligent decision support systems, and Java interoperability computing.

Yi-Chung Cheng received her B.S. degree in Business Mathematics from Soochow University, Taiwan and her M.S. degree in Statistics from National Central University, Taiwan, in 1989. She is a Ph.D. candidate in the Department of Industrial and Information Management at National Cheng Kung University, Taiwan and is also currently an assistant professor of Department of International Business Management at Tainan University of Technology. Technology. Her research interests include time series forecasting and data mining.