Forecasting enrollments with fuzzy time series — Part I

Article *in* Fuzzy Sets and Systems · February 1993

DOI: 10.1016/0165-0114(93)90355-L

Qiang Song

Atlanta Forecasting Systems, LLC

Fuzzy Sets and Systems 54 (1993) 1–9
North-Holland

Forecasting enrollments with fuzzy time series – Part I

Qiang Song and Brad S. Chissom
College of Education, The University of Alabama, Tuscaloosa, AL, USA

Received October 1991
Revised January 1992

Abstract: There have been a good many methods to forecast university enrollments in the literature. However, none of them could be applied when the historical data are linguistic values. Fuzzy time series is an effective tool to deal with such problems. In this paper, as an application of fuzzy time series in educational research, the forecast of the enrollments of the University of Alabama is carried out. In so doing, a fuzzy time series model is developed using historical data. A complete procedure is proposed which includes: fuzzifying the historical data, developing a fuzzy time series model, and calculating and interpreting the outputs. To evaluate the forecasting model, the robustness of the fuzzy time series model is tested. Advantages and problems of the forecasting method are also discussed.

Keywords: Fuzzy time series; enrollments; memberships; forecasting.

1. Introduction

It is very important to make reasonably accurate estimates of the future enrollments for a university because many decisions of the university will be made from them. For this reason, many researchers have proposed various methods to forecast enrollments.

The enrollments of a university are influenced by many factors. Therefore, the main point of developing a forecasting model is to try to make clear these influencing factors and their relationships with enrollments. Different factors and relationships will make up different models or methods. Shaw [8] made very good comments on some methods and Gardner [3] provided a classification of various methods. Many successful forecasting methods can be found in the literature.

Weiler [11] utilized the growth curve models to forecast the enrollments for the short-period between the beginning of the application and the start of the fall semester. These models were used to forecast the enrollments of the University of Minnesota with forecasting errors[1] ranging from 2.4% to 16% and an average error[2] being 9.7%.

Chatman [2] applied the regression method to the accumulative number of accepted students to forecast enrollments. His study yielded the forecasting errors from 2.8% to 23%.

Pope and Evans [7] developed a Decision Support System for forecasting enrollments. In their study, the enrolling freshmen were classified into four divisions. For each division, a different model was derived, but the forecasting errors were not reported.

Warrack and Russell [10] developed a forecasting model using the motivational index and the demand index. The forecasted enrollments of four universities are found in their paper and the forecasting errors range from 3.7% to 14.4%. The main drawback of this model is the costly survey which was the foundation of the method.

Hoenack and Weiler [4] developed a forecasting model which had 11 equations and considered 19 factors. The model was used for a case application with good accuracy. This model was, in essence, econometrics models mostly used in economic analysis.

[1] Forecasting error = (|forecasted value − actual value|/actual value) * 100%.
[2] Average forecasting error = (sum of forecasting errors)/(total # of errors).

The double exponential smoothing method can be used to forecast enrollments. Gardner [3] discussed how to select the weight factor of the method. Obviously, the main weakness of this method is the question of how to determine an appropriate value for the weight factor.

Paulsen [6] proposed a step-by-step method. This method was conceptually similar to those by Weiler [11], Pope and Evans [7], and Chatman [2]. But this method could achieve an average forecasting error of 2.4% which is the smallest of all the short-term forecasting methods reviewed in this paper.

From the literature, we can determine that researchers have been using many different methodologies or models for forecasting enrollments, but all these methods fail when the historical enrollment data are composed of linguistic values. Therefore, it is our aim to deal with such forecasting problems in this paper by developing a mid-term forecasting model with fuzzy time series to forecast the enrollments of the University of Alabama.

Fuzzy time series is a new concept proposed by the authors to deal with forecasting problems where the historical data are linguistic values. In a previous study, based upon fuzzy set theory, fuzzy logic and approximate reasoning, Song and Chissom [9] put forth the definition of fuzzy time series and the outline of its modeling by means of fuzzy relational equations and approximate reasoning. Although in the literature, one can easily find research papers introducing successful applications of fuzzy logic, approximate reasoning, fuzzy control, fuzzy linear regression and many other aspects of fuzzy set theory and applications, a theoretical framework based on Zadeh's work [13, 14, 15] to cope with such forecasting problems in which the historical data are not real numbers but linguistic values has not been delineated. Bintley [1] has successfully applied fuzzy logic and approximate reasoning to a practical case of forecasting, but he did not apply the concept of fuzzy time series or other relative concepts.

As an application of fuzzy time series, the forecast of the enrollments of the University of Alabama is carried out in this paper. To do so, a first-order, time-invariant model will be developed and applied. In Section 2, some basic concepts of fuzzy time series will be introduced; in Section 3, a step-by-step forecasting procedure will be explained and carried out; in Section 4, the robustness of the forecasting model will be tested; and in Section 5, both advantages and disadvantages of the forecasting method will be discussed.

2. Some concepts of fuzzy time series and its models

For a better appreciation of the following contents, it is beneficial to introduce briefly some concepts and conclusions in fuzzy time series. For more details, refer to Song and Chissom [9].

Definition 1. Let $Y(t)$ $(t = \ldots, 0, 1, 2, \ldots)$, a subset of R^1, be the universe of discourse on which fuzzy sets $f_i(t)$ $(i = 1, 2, \ldots)$ are defined and $F(t)$ is a collection of $f_1(t), f_2(t), \ldots$. Then $F(t)$ is called a fuzzy time series defined on $Y(t)$ $(t = \ldots, 0, 1, 2, \ldots)$.

In the above definition, $F(t)$ can be understood as a linguistic variable and $f_i(t)$ $(i = 1, 2, \ldots)$ as the possible linguistic values of $F(t)$. Since at different times, the values of $F(t)$ can be different, $F(t)$, as a set of fuzzy sets, is a function of time t. Also, the universes of discourse can be different at different times. So, we use $Y(t)$ for the universe at time t.

The main difference between the conventional time series and fuzzy time series is that the values of the former are real numbers while the values of the latter are fuzzy sets.

For simplicity, only the first-order model will be used to forecast the enrollments. Therefore, just the definition of the first-order model is needed and given below.

Definition 2. Suppose $F(t)$ is caused by $F(t-1)$ only, i.e., $F(t-1) \rightarrow F(t)$. Then this relation can be expressed as $F(t) = F(t-1) \circ R(t, t-1)$ where $R(t, t-1)$ is the fuzzy relationship between $F(t-1)$ and $F(t)$, and $F(t) = F(t-1) \circ R(t, t-1)$ is called the first-order model of $F(t)$.

There are two main categories of fuzzy time series. The classification is determined by their models in terms of their relationships with time t. In the case of the first-order models, we have:

Definition 3. Suppose $R(t, t-1)$ is a first-order model of $F(t)$. If for any t, $R(t, t-1)$ is independent of t, i.e., for any t, $R(t, t-1) = R(t-1, t-2)$, then $F(t)$ is called a time-invariant fuzzy time series otherwise it is called a time-variant fuzzy time series.

To appreciate the method used below, the following theorems might be useful.

Theorem 1. *Let $F(t)$ be a fuzzy time series. If for any t, $F(t) = F(t-1)$ and $F(t)$ has only finite elements, then $F(t)$ is a time-invariant fuzzy time series.*

We can see that just as the determination of $f_i(t)$ $(i = 1, 2, \ldots)$ is subjective, so is that of $F(t)$. This implies that we can define either a time-invariant or a time-variant fuzzy time series on the same universe. But since the determination of the fuzzy memberships of $f_i(t)$ is not arbitrary, the concepts of time-invariant and time-variant fuzzy times series are meaningful, which can be seen from the following theorem.

Theorem 2. *If $F(t)$ is a fuzzy time series, $F(t) = F(t-1)$ for any t and $F(t)$ has only finite elements $f_i(t)$ $(i = 1, \ldots, n)$, then*

$$R(t, t-1) = \cdots f_{i_1}(t-1) \times f_{j_0}(t) \cup f_{i_2}(t-2) \times f_{j_1}(t-1) \cup \cdots \cup f_{i_m}(t-m) \times f_{j_{m-1}}(t-m+1) \cdots \tag{1}$$

where $m > 0$ and all pairs of fuzzy sets are different.

This theorem implies that in the case of time-invariant fuzzy time series, it is very easy and convenient to calculate the first-order model. As a matter of fact, since we can hardly define infinite fuzzy sets on any universe of discourse, once $F(t) = F(t-1)$, i.e., at any two successive time points t_1 and t_2, we have the same fuzzy sets, then we can define a time-invariant fuzzy time series. Therefore, Theorem 2 is very useful.

In the following section, we will describe the procedure in detail to forecast enrollments using fuzzy time series models.

3. Forecasting enrollments with fuzzy time series

Fuzzy time series was proposed in order to deal with such situations in which the traditional time series model is no longer applicable. For example, in a forecasting problem, the historical data are not real numbers but are linguistic values. In this case, no models of conventional time series can be applied but a fuzzy time series model can be applied more properly. This can be seen from the definitions and theorems of fuzzy time series in Section 2.

There are two possibilities in any application cases. One is that the historical data are in the forms of linguistic values. The other is that the historical data are real numbers. In the second case, the data should be fuzzified first. In our study, the historical data are real numbers. Therefore, in this application problem, there is one more step than when the data are linguistic value – the fuzzification of data. The other steps are the same.

Like time series, we use historical data to set up the relationship among values of interests at different times, but the relationship in fuzzy time series is different from that in traditional time series. These relationships are models. In fuzzy time series, we apply the past experience knowledge into the model and the knowledge has the form of "IF . . . THEN . . .". So, the main point of modelling fuzzy time series is to identify the historical law with "IF . . . THEN . . ." forms. If the historical data are real numbers, we should first fuzzify them. The procedures of forecasting the enrollments using fuzzy time series are as follows:

Step 1. Define the universe of discourse U within which the historical data are and upon which the fuzzy sets will be defined. Usually, when defining the universe, first find the minimum enrollment $D_{\min}$ and the maximum enrollment $D_{\max}$ of known historical data. Based on $D_{\min}$ and $D_{\max}$, define the universe U as $[D_{\min} - D_1, D_{\max} + D_2]$ where D_1 and D_2 are two proper positive numbers. In our study, we have on hand the enrollments of the university from 1971 to 1990 with $D_{\min} = 13055$ and $D_{\max} = 19328$. For simplicity, we choose $D_1 = 55$ and $D_2 = 672$. Thus, the universe is the interval of $U = [13000, 20000]$.

Step 2. Partition the universe U into several even lengthy intervals. In our study, we divide U into 7 intervals with equal lengths. We use $u_1, u_2, u_3, u_4, u_5, u_6$ and u_7 for each interval, i.e., $u_1 = [13000, 14000]$, $u_2 = [14000, 15000]$, $u_3 = [15000, 16000]$, $u_4 = [16000, 17000]$, $u_5 = [17000, 18000]$, $u_6 = [18000, 19000]$ and $u_7 = [19000, 20000]$.

Step 3. Define fuzzy sets on the universe U. First, determine some linguistic values. For the linguistic variable 'enrollments', let $A_1 =$ (not many), $A_2 =$ (not too many), $A_3 =$ (many), $A_4 =$ (many many), $A_5 =$ (very many), $A_6 =$ (too many), and $A_7 =$ (too many many) be the possible values. There is no restriction on the number of the fuzzy sets defined. Second, define fuzzy sets on U. All the fuzzy sets will be labeled by the possible linguistic values. In our study, $u_1, u_2, \ldots,$ and u_7 are chosen as the elements of each fuzzy set. To determine the memberships of $u_1, u_2, \ldots,$ and u_7 to each A_i $(i = 1, \ldots, 7)$, make a judgement of how well each u_k $(k = 1, \ldots, 7)$ belonging to A_i. If a u_k completely belongs to A_i, the membership will be 1; if u_k does not belong to A_i at all, the membership will be 0; otherwise, choose a number from $(0, 1)$ as the degree to which u_k belongs to A_i. With our own experience, the authors have determined the memberships for each element in the respective fuzzy sets. Thus, all the fuzzy sets A_i $(i = 1, \ldots, 7)$ are expressed as follows:

$$
\begin{aligned}
A_1 &= \{u_1/1,\ u_2/0.5,\ u_3/0,\ u_4/0,\ u_5/0,\ u_6/0,\ u_7/0\},\\
A_2 &= \{u_1/0.5,\ u_2/1,\ u_3/0.5,\ u_4/0,\ u_5/0,\ u_6/0,\ u_7/0\},\\
A_3 &= \{u_1/0,\ u_2/0.5,\ u_3/1,\ u_4/0.5,\ u_5/0,\ u_6/0,\ u_7/0\},\\
A_4 &= \{u_1/0,\ u_2/0,\ u_3/0.5,\ u_4/1,\ u_5/0.5,\ u_6/0,\ u_7/0\},\\
A_5 &= \{u_1/0,\ u_2/0,\ u_3/0,\ u_4/0.5,\ u_5/1,\ u_6/0.5,\ u_7/0\},\\
A_6 &= \{u_1/0,\ u_2/0,\ u_3/0,\ u_4/0,\ u_5/0.5,\ u_6/1,\ u_7/0.5\},\\
A_7 &= \{u_1/0,\ u_2/0,\ u_3/0,\ u_4/0,\ u_5/0,\ u_6/0.5,\ u_7/1\},
\end{aligned}
\tag{2}
$$

where u_i $(i = 1, \ldots, 7)$ is the element and the number below '/' is the membership of u_i to A_j $(j = 1, \ldots, 7)$.

For simplicity, we will also use $A_1, A_2, \ldots, A_7$ as row vectors whose elements are the corresponding memberships in (2).

Step 4. Fuzzify the historical data, i.e., find out an equivalent fuzzy set to each year's enrollment. The commonly used method is to define a cut set for each A_i $(i = 1, \ldots, 7)$. If at year t, the enrollment is within the cut set of A_k, then the enrollment of year is A_k [12]. The problem with this method is that there is the possibility that the enrollment at year t may fall within more than one cut set. To avoid this, a different method is employed in our study. Instead of defining cut sets, the degree of each year's enrollment belonging to each A_i $(i = 1, \ldots, 7)$ is determined. The process is the same as that of determining the memberships of u_i to A_j in Step 3. The equivalent fuzzy sets to each year's enrollment are shown in Table 1 and each fuzzy set has seven elements.

Step 5. Obtain the historical knowledge from Table 1 about the evolution of the enrollments of this university to set up the forecasting model. To do so, assume that if the maximum membership of one year's enrollment is under A_k, then we treat this year's enrollment as A_k. For example, for 1982, the maximum membership is under A_3, then we say that the enrollment of 1982 is A_3, or many. We can do the same for the rest. Thus, we can transform the historical data into linguistic values. Since we are

Table 1

Year	A1	A2	A3	A4	A5	A6	A7
1990	0	0	0	0.3	0.5	0.8	1
1989	0	0	0	0.25	0.55	1	0.8
1988	0	0	0.1	0.5	0.8	1	0.7
1987	0	0.1	0.5	1	0.8	0.1	0
1986	0	0.2	1	0.7	0.2	0	0
1985	0.2	0.8	1	0.2	0	0	0
1984	0.2	0.8	1	0.2	0	0	0
1983	0.2	0.8	1	0.2	0	0	0
1982	0.2	0.8	1	0.2	0	0	0
1981	0	0.2	0.8	1	0.5	0	0
1980	0	0.1	0.5	1	0.9	0.2	0
1979	0	0.1	0.5	1	0.9	0.2	0
1978	0	0.5	1	0.7	0.2	0	0
1977	0	0.6	1	0.6	0.1	0	0
1976	0.2	0.8	1	0.2	0	0	0
1975	0.2	0.8	1	0.2	0	0	0
1974	0.8	1	0.8	0.1	0	0	0
1973	1	0.9	0.2	0	0	0	0
1972	1	0.8	0.1	0	0	0	0
1971	1	0.5	0	0	0	0	0

looking for the laws governing any two successive years' enrollments in terms of fuzzy sets and fuzzy conditional statements, we will develop such logical relationships as 'If the enrollment of year i is A_k then that of year $i+1$ is A_j'', and so on. Using the symbols of Song and Chissom [9], we can obtain all the fuzzy logical relationships from Table 1 as follows (Note: the repeated relationships are counted only once):

$$A_1 \to A_1,\ A_1 \to A_2,\ A_2 \to A_3,\ A_3 \to A_3,\ A_3 \to A_4,$$
$$A_4 \to A_4,\ A_4 \to A_3,\ A_4 \to A_6,\ A_6 \to A_6 \quad \text{and} \quad A_6 \to A_7.$$

By definition, we know that we have defined a time-invariant fuzzy times series. Let us define an operator '$\times$' of two vectors. Suppose C and B are row vectors of dimension m and $D=(d_{ij})=C^{\mathrm{T}} \times B$. Then the element d_{ij} of matrix D at row i and column j is defined as $d_{ij}=\min(C_i, B_j)$ $(i, j=1, \ldots, m)$ where C_i and B_j are the i-th and the j-th elements of C and B respectively.

Let $R_1=A_1^{\mathrm{T}} \times A_1$, $R_2=A_1^{\mathrm{T}} \times A_2$, $R_3=A_2^{\mathrm{T}} \times A_3$, $R_4=A_3^{\mathrm{T}} \times A_3$, $R_5=A_3^{\mathrm{T}} \times A_4$, $R_6=A_4^{\mathrm{T}} \times A_4$, $R_7=A_4^{\mathrm{T}} \times A_3$, $R_8=A_4^{\mathrm{T}} \times A_6$, $R_9=A_6^{\mathrm{T}} \times A_6$ and $R_{10}=A_6^{\mathrm{T}} \times A_7$. Then, according to Theorem 2, we get

$$R(t, t-1)=R=\bigcup_{i=1}^{10} R_i \tag{3}$$

where R is a 7×7 matrix and $\bigcup$ is the union operator.

Using formula (3), some calculation yields:

$$R=\begin{bmatrix} 1 & 1 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 1 & 1 & 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & 1 & 1 & 0.5 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0.5 \end{bmatrix}$$

Table 2

Year	Output membership	Standardized membership	Predicted value
1972	1,1, 0.5, 0.5, 0.5, 0, 0	0.286, 0.286, 0.143, 0.143, 0.143, 0, 0	14000
1973	1, 1, 0.8, 0.5, 0.5, 0.1, 0.1	0.25, 0.25, 0.2, 0.125, 0.125, 0.025, 0.025	14000
1974	1, 1, 0.9, 0.5, 0.5, 0.2, 0.2	0.2325, 0.2325, 0.209, 0.116, 0.116, 0.047, 0.047	14000
1975	0.8, 0.8, 1, 0.8, 0.5, 0.5, 0.5	0.163, 0.163, 0.204, 0.163, 0.102, 0.102, 0.102	15500
1976	0.5, 0.5, 1, 1, 0.5, 0.5, 0.5	0.111, 0.111, 0.222, 0.222, 0.111, 0.111, 0.111	16000
1977	0.5, 0.5, 1, 1, 0.5, 0.5, 0.5	0.111, 0.111, 0.222, 0.222, 0.111, 0.111, 0.111	16000
1978	0.5, 0.5, 1, 1, 0.5, 0.6, 0.5	0.109, 0.109, 0.217, 0.217, 0.109, 0.13, 0.109	16000
1979	0.5, 0.5, 1, 1, 0.5, 0.7, 0.5	0.106, 0.106, 0.213, 0.213, 0.106, 0.149, 0.106	16000
1980	0.1, 0.5, 1, 1, 0.5, 1, 0.5	0.0217, 0.108, 0.217, 0.217, 0.108, 0.217, 0.108	16813
1981	0.1, 0.5, 1, 1, 0.5, 1, 0.5	0.0217, 0.108, 0.217, 0.217, 0.108, 0.217, 0.108	16813
1982	0.2, 0.5, 1, 1, 0.5, 1, 0.5	0.0425, 0.106, 0.213, 0.213, 0.106, 0.213, 0.106	16789
1983	0.5, 0.5, 1, 1, 0.5, 0.5, 0.5	0.111, 0.111, 0.222, 0.222, 0.111, 0.111, 0.111	16000
1984	0.5, 0.5, 1, 1, 0.5, 0.5, 0.5	0.111, 0.111, 0.222, 0.222, 0.111, 0.111, 0.111	16000
1985	0.5, 0.5, 1, 1, 0.5, 0.5, 0.5	0.111, 0.111, 0.222, 0.222, 0.111, 0.111, 0.111	16000
1986	0.5, 0.5, 1, 1, 0.5, 0.5, 0.5	0.111, 0.111, 0.222, 0.222, 0.111, 0.111, 0.111	16000
1987	0.2, 0.5, 1, 1, 0.5, 0.7, 0.5	0.045, 0.114, 0.227, 0.227, 0.114, 0.159, 0.114	16000
1988	0.1, 0.5, 1, 1, 0.5, 1, 0.5	0.027, 0.108, 0.217, 0.217, 0.108, 0.217, 0.108	16813
1989	0, 0.5, 0.5, 0.5, 0.5, 1, 1	0, 0.125, 0.125, 0.125, 0.125, 0.25, 0.25	19000
1990	0, 0.5, 0.5, 0.5, 0.5, 1, 1	0, 0.125, 0.125, 0.125, 0.125, 0.25, 0.25	19000
1991	0, 0.5, 0.5, 0.5, 0.5, 0.8, 0.8	0, 0.138, 0.138, 0.138, 0.138, 0.222, 0.222	19000

Using R, define the forecasting model as

$$A_i = A_{i-1} \circ R \tag{4}$$

where A_{i-1} is the enrollment of year $i-1$ and A_i the forecasted enrollment of year i in terms of fuzzy sets and '$\circ$' is the 'max-min' operator.

Step 6. Calculate the forecasted outputs. Suppose the enrollment of year t is known and can be found from Table 1, to forecast the enrollment of year $t+1$, let A_{i-1} in (4) be the enrollment at year t and apply formula (4). Then A_i will be the forecasted enrollment of year $t+1$. For 1972 to 1991, the forecasted outputs are shown in Table 2.

Step 7. Interpret the forecasted outputs. The calculation results of formula (4) are actually all fuzzy sets. If the results in the form of fuzzy sets can satisfy the requirement for the forecasting job, just stop here. But in many cases, an equivalent scalar is desired. Therefore, translating the fuzzy output into a regular number is indeed a necessary step. Sometimes, this step is called defuzzification [5]. For defuzzification, the most popular method is to use the centroid of a fuzzy set as the equivalent scalar. But, in this study, it has been found that the centroid method can not yield satisfactory forecasting results. Therefore, we will utilize a combination of several methods. Since it is difficult to have a unique method for defuzzification, at this stage, we can only propose some principles to interpret the forecasted results. The principles are:

(1) If the membership of an output has only one maximum, then select the midpoint of the interval corresponding to the maximum as the forecasted value.

(2) If the membership of an output has two or more consecutive maximums, then select the midpoint of the corresponding conjunct intervals as the forecasted value.

(3) Otherwise, standardize the fuzzy output and use the midpoint of each interval to calculate the centroid of the fuzzy set as the forecasted value.

Following the above principles, we have obtained the predicted values for the enrollments from 1972 to 1991. The results are listed in Table 2 and shown in Figure 1 where the solid line is the actual enrollment and the dashed line is the forecasted enrollment. Note that we did not use the enrollment

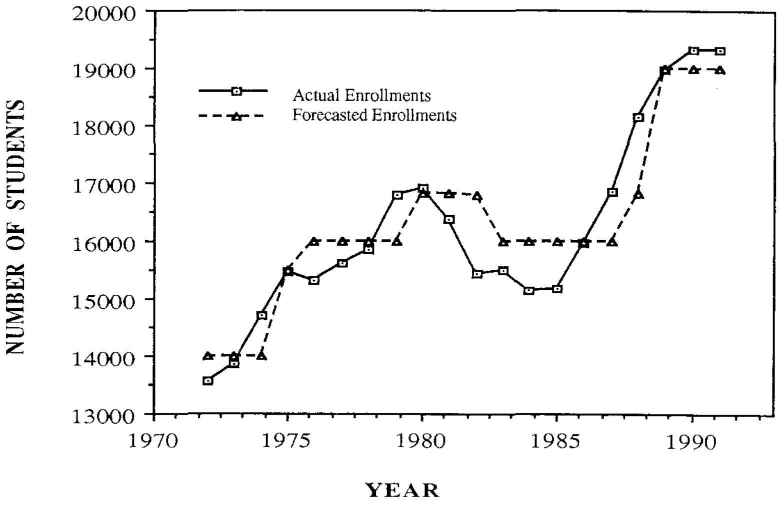

Fig. 1. Forecasted enrollments and actual enrollments.

data of 1991 to develop the forecasting model. The forecasting errors range from 0.1% to 8.7% with the average error being 3.18%. For 1991, the forecasting error is 1.7%. For a mid-term forecasting model, an average error of 3.18% is quite satisfactory.

4. Robustness of fuzzy time series model

To test if the fuzzy time series model can still yield good forecasting results when the historical data are not accurate, the robustness of the model has been checked. To do so, increase a few years' enrollment data by 5% with the rest of the data unchanged, and then repeat steps 1 to 7 in Section 3.

In our study, we randomly selected the enrollments of 1974, 1978, 1985 and 1990 and increased them by 5%. Thus, the fuzzy relationships obtained from the changed historical enrollment data are different from those in Section 3. The relationships are (note: the repeated relationships are counted only once):

$$A_1 \to A_1,\ A_1 \to A_3,\ A_3 \to A_3,\ A_3 \to A_4,\ A_4 \to A_4,$$
$$A_4 \to A_3,\ A_4 \to A_6,\ A_6 \to A_6 \quad \text{and} \quad A_6 \to A_7.$$

The forecasted results calculated and interpreted from the outputs of the model made up of the above relationships are shown in Figure 2 where the solid line is the actual enrollment and the dashed line is the forecasted enrollment. The forecasting errors range from 0.1% to 11% and the average error is 3.9%. From Figure 2, it can be seen that as time t increases, the forecasting error decreases. This indicates that the fuzzy time series model has good robustness. This test shows that even if the historical data are not accurate, because human beings' experience knowledge has been included in the model, we can still make good forecasts. This is consistent with our ultimate goal in proposing fuzzy time series.

5. Discussions and conclusion remarks

In this paper, we have applied fuzzy time series model to forecast the enrollments of The University of Alabama and have provided a step-by-step procedure. Compared with any other forecasting methods mentioned in Section 1, the method applied here has at least the following 5 advantages:

1. The average forecasting error of this method is smaller than that of most of other methods. Although Paulsen's method [6] can achieve a smaller average error (2.4%), his model is a short-term

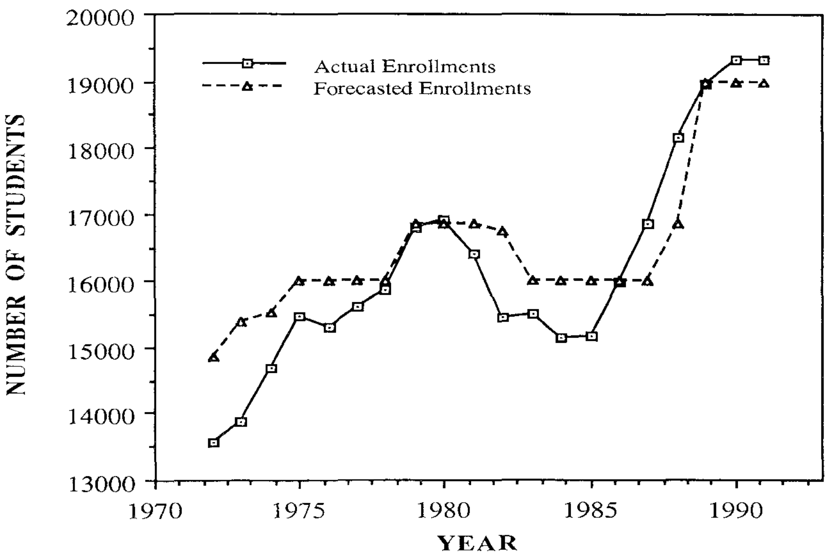

Fig. 2. Curves of forecasted enrollments and actual enrollments.

forecasting model while ours is a mid-term one. In addition, the difference between the two average errors (0.78%) can be ignored.

2. Human experience knowledge can be utilized from the very beginning until the end of the forecasting process. If the historical data cannot be obtained but the knowledge of the evolution of the university's enrollments of the past can, i.e., the evolution laws in the form of "IF... THEN..." can be obtained, none of the other methods can be applicable. Nevertheless, in this case, with fuzzy time series, we are still able to establish the forecasting model and make good forecasts. Our method is better than the others in this aspect.

3. Historical data have less important roles in our method. Note that in Section 3, the data were only used in Step 4 of the procedure where the data were fuzzified. Once fuzzified, they were never used. Although in Step 1, the minimum and the maximum of the historical data were used to define the universe, it is not necessary to do so. In effect, we could define a universe with only our experience. The principle is that the universe should cover the possible range of the enrollments.

4. Even if the historical data are real numbers, since the model has good robustness, the requirement for the historical data is not very strict.

5. When the historical data are linguistic values, the procedure is almost the same for forecasting enrollments except that Step 4 is omitted and Table 1 contains the corresponding fuzzy sets of (2) as row vectors.

Of course, there is a major problem with this method: the forecasted values depend largely on our interpretations of the outputs of the forecasting model in Step 7. Different interpretations may lead to different forecasted results. This makes the process quite subjective. To overcome this shortcoming, an objective method (not the centroid method) should be applied.

In spite of the problem with this method, the advantages make fuzzy time series model very competitive in forecasting university enrollments. This also indicates that fuzzy sets theory is a successful tool for solving complex practical problems.

In future studies, we will work with other models to verify that fuzzy time series is a useful tool to make good forecasts in fuzzy environments.

References

[1] H. Bintley, Time series analysis with REVEAL, *Fuzzy Sets and Systems* **23** (1987) 97–118.

[2] S.P. Chatman, Short-term forecasts of the number and scholastic ability of enrolling freshman by academic divisions, *Res. Higher Educ.* **25** (1986) 68–81.

[3] D.E. Gardner, Weight factor selection in double exponential smoothing enrollment forecasts, *Res. Higher Educ.* **14** (1981) 49–56.
[4] S.A. Hoenack and W.C. Weiler, The demand for higher education and institutional enrollment forecasting, *Economic Inquiry* **17** (1979) 89–113.
[5] B. Kosko, *Neural Networks and Fuzzy Systems* (Prentice Hall, Englewood Cliffs, NJ, 1992).
[6] M.B. Paulsen, A practical model for forecasting new freshmen enrollment during the application period, *College and University* **64** (1989) 379–391.
[7] J.A. Pope and J.P. Evans, A forecasting system for college admissions, *College and University* **60** (1985) 113–131.
[8] R.C. Shaw, Enrollment forecasting: what methods work best? *NASSP Bull.* **68** (1984) 52–58.
[9] Q. Song and B.S. Chissom, Fuzzy time series and its models, *Fuzzy Sets and Systems* **54** (1993) 269–277.
[10] B.J. Warrack and C.N. Russell, Forecasting demand for postsecondary education in Manitoba: the motivational index and the demand index as an enrollment forecasting tool, *Res. Higher Educ.* **19** (1983) 335–349.
[11] W.C. Weiler, A model for short-term institutional enrollment forecasting, *J. Higher Education* **51** (1980) 314–327.
[12] W. Wu et al., *Applied Fuzzy Sets Methodology* (Beijing Normal University Press, 1985).
[13] L.A. Zadeh, Fuzzy sets, *Inform. and Control* **8** (1965) 338–353.
[14] L.A. Zadeh, Fuzzy logic and approximate reasoning, *Synthese* **30** (1975) 407–428.
[15] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Part 1, 2, 3, *Inform. Sci.* **8** (1975) 199–249; 301–357; **9** (1975) 43–80.