Forecasting Enrollment Model Based on First-Order Fuzzy Time Series

Melike Şah and Konstantin Y. Degtiarev, *MIEEE*

***Abstract* —** This paper proposes a novel improvement of forecasting approach based on using time-invariant fuzzy time series. In contrast to traditional forecasting methods, fuzzy time series can be also applied to problems, in which historical data are linguistic values. It is shown that proposed time-invariant method improves the performance of forecasting process. Further, the effect of using different number of fuzzy sets is tested as well. As with the most of cited papers, historical enrollment of the University of Alabama is used in this study to illustrate the forecasting process. Subsequently, the performance of the proposed method is compared with existing fuzzy time series time-invariant models based on forecasting accuracy. It reveals a certain performance superiority of the proposed method over methods described in the literature.

***Keywords* —** Forecasting, fuzzy time series, linguistic values, student enrollment, time-invariant model.

I. Introduction

FORECASTING plays a notable role in making both crucial and day-to-day decisions about the future. Weather prediction, staff scheduling, business and production planning and multistage management decision analysis are among distinctive examples of forecasting areas where people want to foresee, within existing limits, as closely as possible. Although, there are many well-known forecasting methods, they cannot solve forecasting problems, in which the historical data are available in linguistic form. Fuzzy time series allows to overcome this drawback [4]. However, fuzzy time series are not just limited to linguistic values, and can be used for the prediction of numerical values too.

For the last decade the problem of forecasting based on fuzzy time series has been studied by several authors [1]-[6]. Based on definitions [1], Song and Chissom introduced time-invariant and time-variant models for forecasting with fuzzy time series [2], [3]. Despite the fact that all necessary concepts and definitions are provided later on in section II, we need to clarify these two important notions. Fuzzy time series $F(t)$ with finite number of elements is called time-variant, if for any moment of time t, $F(t) = F(t-1)$; otherwise it is called a time-variant fuzzy time series [4]. In their studies Song and Chissom used the University of Alabama enrollment data to demonstrate the forecasting process based on model:

$$F(t) = F(t\text{-}1) \circ R \,, \qquad (1)$$

where $F(t\text{-}1)$ is the enrollment of year $t-1$, $F(t)$ is the forecasted enrollment of year t expressed by fuzzy sets, R is a union of first-order fuzzy relations, i.e. relations that represent the relationship between enrollments of two consequent years (for more details see Section II), and, finally, symbol $\circ$ denotes max-min composition operator. As it was shown in [2] and [3], the average error of time-invariant and time-variant model turned out to be 3.18% and 4.37%, correspondingly.

Subsequently, Chen proposed a new model to simplify the computational complexity of forecasting process by means of using simple arithmetic operations instead of max-min composition operator on the same set of historical enrollment data [4]. The average forecasting error of Chen's model was 3.23%. Apart from the fact that the result obtained improves a similar figure of Song-Chissom's time-variant model [3], it appears to be more efficient as compared to both time-invariant and time-variant models of Song and Chissom in respect to more simple computations.

This paper is devoted to the description of a new (modified) time-invariant method to deal with forecasting problems. Unlike Song-Chissom and Chen approaches, the proposed method utilizes variations of the available historical data as fuzzy time series instead of direct usage of *raw* numeric values. Furthermore, the effect of changes in the number of fuzzy sets in the model is investigated. Results obtained are compared with those of Chen and Song-Chissom models for the purpose of forecasting accuracy.

The rest of the paper is organized as follows. Section II recalls those basic concepts and definitions [1]-[2] that are directly relevant to fuzzy time series. Section III discusses the proposed time-invariant method by the example of university enrollment together with its advantages over existing models. Finally, the concluding remarks are drawn in the Section IV.

II. Fuzzy Time Series Definitions

This section briefly summarizes basic fuzzy time series concepts [1], [2] needed for the subsequent text.

Definition 1. Assume $Y(t) \subset \mathbf{R}$ (real line), $t = ..., 0, 1, 2, ...$, to be a universe of discourse defined by the fuzzy set $f_i(t)$. $F(t)$ consisting of $f_i(t)$, $i = 1, 2, ...$, is defined as a fuzzy time series on $Y(t)$. At that, $F(t)$ can be understood as a *linguistic variable*, whereas $f_i(t)$, $i = 1,2,...$, are possible *linguistic values* of $F(t)$.

Definition 2. If there exists a fuzzy relationship $R(t,t-1)$, such that $F(t) = F(t-1) \times R(t,t-1)$, where symbol $\times$ is an operator, then $F(t)$ is said to be caused (or, induced) by $F(t-1)$. The existing relationship between $F(t)$ and $F(t-1)$ can be denoted by the expression $F(t-1) \rightarrow F(t)$.

Definition 3. Denoting $F(t-1)$ by A_i and $F(t)$ by A_j, the relationship between $F(t-1)$ and $F(t)$ can be defined by a logical relationship $A_i \rightarrow A_j$.

Definition 4. Fuzzy logical relationships, which have the same left-hand sides, can be grouped together into fuzzy logical relationship groups. For example, for the identical left-hand side A_i such grouping can be depicted as follows:

$$\left.\begin{array}{l} A_i \rightarrow A_{j1} \\ A_i \rightarrow A_{j2} \\ \dots\dots\dots\dots \end{array}\right\} \Rightarrow A_i \rightarrow A_{j1}, A_{j2}, \dots$$

Definition 5. If $F(t)$ is a time-invariant fuzzy time series, then the logical relationship $F(t-1) \rightarrow F(t)$ is called a first-order logical relationship.

III. Forecasting Enrollments With A New Method Of Time-Invariant Fuzzy Time Series

A. Forecasting

The aim of this study is to propose a method that is aimed to attain better forecasting accuracy by using time-invariant fuzzy time series. It should be emphasized that for forecast it uses only historical data in the numerical form (number of students) without any additional pieces of knowledge.

Based on *actual* historical data of enrollments of the University of Alabama, Song and Chissom set up models, i.e. relationships among values of interests at different moments of time [1]-[3]. Method developed by Chen [4] also provides for construction of fuzzy sets A_i being values of the linguistic variable *(actual) enrollments*. We propose modifications that mainly deal with two key aspects; (a) usage of *variations* of historical data instead of actual enrollment characteristics, and (b) calculation of relationship R utilized for the prediction of future enrollments. In addition, the method is tested on different number of fuzzy sets for the purpose of examination of forecasting accuracy.

Finally, step-by-step forecasting process looks as follows:

Step 1: Define the universe of discourse (universal set U) starting from variations of the historical enrollment data,
Step 2: Partition U into equally length intervals,
Step 3: Define fuzzy sets A_i,
Step 4: Fuzzify variations of the historical enrollment data,
Step 5: Determine fuzzy logical relationships $A_i \rightarrow A_j$,
Step 6: Group fuzzy logical relationships (see Step 5) having the same left-hand side and calculate R_i for each *i*-th fuzzy logical relationship group,
Step 7: Forecast and deffuzify the forecasted outputs,
Step 8: Calculate the forecasted enrollments.

Hence, the approach that uses enrollments of the University of Alabama can be represented more comprehensively in the following way:

Step 1: In accordance with the problem domain, universal set U is defined – on this occasion yearly variations of the enrollments are used. Actual data and corresponding variations are listed in Table I (minimum and maximum variations are $V_{min} = -955$ and $V_{max} = 1291$, respectively). With the object of simplifying division of U into equally length intervals, accept U as $[V_{min} - V_1, V_{max} + V_2]$, where V_1 and V_2 are positive numbers 45 and 109, accordingly. As a result, $U = [-1000, 1400]$.

TABLE I
ACTUAL ENROLLMENTS AND VARIATIONS OF HISTORICAL DATA

Years	Actual enrollments	Variations	Years	Actual enrollments	Variations
1971	13055		1982	15433	– 955
1972	13563	+ 508	1983	15497	+ 64
1973	13867	+ 304	1984	15145	– 352
1974	14696	+ 829	1985	15163	+ 18
1975	15460	+ 764	1986	15984	+ 82
1976	15311	– 149	1987	16859	+ 875
1977	15603	+ 292	1988	18150	+ 1291
1978	15861	+ 258	1989	18970	+ 820
1979	16807	+ 946	1990	19328	+ 358
1980	16919	+ 112	1991	19337	+ 9
1981	16388	– 531	1992	18876	– 461

Step 2: We use 6 (six) fuzzy sets, i.e. U is partitioned into six equal intervals u_i, $i = \overline{1,6}$, namely: $u_1 = [-1000, -600]$, $u_2 = [-600, -200]$, ... , $u_6 = [1000, 1400]$ (the number of fuzzy sets is not necessarily coincides with the number of intervals),

Step 3: We assume that linguistic variable *variations of enrollments* can take as fuzzy values: A_1 (*big decrease*), A_2 (*decrease*), A_3 (*no change*), A_4 (*increase*), A_5 (*big increase*), A_6 (*too big increase*), for example. Regardless of the fact that six fuzzy sets are listed here, with the purpose of comparison we also conducted extra experiments with different number of fuzzy sets.

For 6 intervals given $u_i, i = \overline{1,6}$, the fact that each u_i belongs to a particular A_j, $j = \overline{1,6}$, is expressed by the real value from the range [0,1]:

$A_1 = \{1/u_1,\ 0.5/u_2,\ 0/u_3,\ 0/u_4,\ 0/u_5,\ 0/u_6\}$
$A_2 = \{0.5/u_1,\ 1/u_2,\ 0.5/u_3,\ 0/u_4,\ 0/u_5,\ 0/u_6\}$
$A_3 = \{0/u_1,\ 0.5/u_2,\ 1/u_3, 0.5/u_4,\ 0/u_5,\ 0/u_6\}$
$A_4 = \{0/u_1,\ 0/u_2,\ 0.5/u_3,\ 1/u_4,\ 0.5/u_5,\ 0/u_6\}$
$A_5 = \{0/u_1,\ 0/u_2,\ 0/u_3,\ 0.5/u_4,\ 1/u_5,\ 0.5/u_6\}$
$A_6 = \{0/u_1,\ 0/u_2,\ 0/u_3,\ 0/u_4,\ 0.5/u_5,\ 1/u_6\}$

where $u_i \subset U$ are elements of the universal set, and the number that precedes slash symbol "/" is the membership degree $\mu(u_i)$ to respective A_j, $j = \overline{1,6}$.

Step 4: Find a proper fuzzy set for each year's variation. In other words, if the variation of the year t is $p \in u_i$, and there is a value represented by a fuzzy set A_j with the maximum membership value falling on u_i, then p is fuzzified as A_j. Fuzzification results are summarized in Table II.

Step 5: Determine first-order relations, i.e. obtain a set of logical relationship between two consequent variations as shown in Table III. Following [2], we assume that $D = B \to C$, or using product operator $\times$ defined on two 1-vectors, $D = [d_{ij}] = B_i^T \times C_j$. Fuzzy relationship (element of matrix D) is calculated as $d_{ij} = min(B_i, C_j)$, where B_i and C_j are the i^{th} and j^{th}, $i, j = \overline{1,n}$, elements of vectors B and C, respectively.

Step 6: Combine fuzzy relationships (FR) into FR groups starting from identical left-hand sides (Table IV). After that calculate R_i, $i = \overline{1,6}$, as a union of logical relationships in each group (we operate with six fuzzy sets). Thus,

$$R_1 = A_1^T \times A_3$$
$$R_2 = A_2^T \times A_1 \cup A_2^T \times A_3$$
$$R_3 = A_3^T \times A_2 \cup A_3^T \times A_4 \cup A_3^T \times A_5$$
$$R_4 = A_4^T \times A_3 \cup A_4^T \times A_4 \cup A_4^T \times A_5$$
$$R_5 = A_5^T \times A_3 \cup A_5^T \times A_4 \cup A_5^T \times A_5 \cup A_5^T \times A_6$$
$$R_6 = A_6^T \times A_5$$

where $\cup$ is a union operator.

Step 7: Determine fuzzy logical relationship group based upon known variation A_{i-1} of the previous year as follows

If $A_{i-1} = A_j$, then $R_i = R_j$, $j = \overline{1,6}$.

As a result, R_i obtained is used in the definition of forecasting compositional model

$$A_i = A_{i-1} \circ R_i, \qquad (2)$$

where A_i is a forecasted variation of year i in terms of fuzzy set.

For example, consider forecasting of variation for the year 1973 (University of Alabama data) in the presence of known variation of 1972. Data of the Table II makes it clear that $R_i = R_4$. From (2) it follows that $F(1973) = A_4 \circ R_4$, or $F(1973) = [0\ \ 0.5\ \ 1\ \ 1\ \ 1\ \ 0.5]$. Remaining forecasted fuzzy outputs are calculated in a similar manner (third column of the Table V).

TABLE II
FUZZIFIED HISTORICAL ENROLLMENTS BASED ON VARIATIONS

Years	Variations	Fuzzified variations	Years	Variations	Fuzzified variations
1971			1982	– 955	A_1
1972	+ 508	A_4	1983	+ 64	A_3
1973	+ 304	A_4	1984	– 352	A_2
1974	+ 829	A_5	1985	+ 18	A_3
1975	+ 764	A_5	1986	+ 82	A_5
1976	– 149	A_3	1987	+ 875	A_5
1977	+ 292	A_4	1988	+ 1291	A_6
1978	+ 258	A_4	1989	+ 820	A_5
1979	+ 946	A_5	1990	+ 358	A_4
1980	+ 112	A_3	1991	+ 9	A_3
1981	– 531	A_2	1992	– 461	A_2

TABLE III
VARIATIONS FUZZY LOGICAL RELATIONSHIPS

$A_4 \to A_4$	$A_1 \to A_3$
$A_4 \to A_5$	$A_2 \to A_3$
$A_5 \to A_5$	$A_3 \to A_5$
$A_5 \to A_3$	$A_5 \to A_6$
$A_3 \to A_4$	$A_6 \to A_5$
$A_3 \to A_2$	$A_5 \to A_4$
$A_2 \to A_1$	$A_4 \to A_3$

TABLE IV
VARIATIONS FUZZY LOGICAL RELATIONSHIPS GROUPS

$A_1 \to A_3$
$A_2 \to A_1, A_3$
$A_3 \to A_2, A_4, A_5$
$A_4 \to A_3, A_4, A_5$
$A_5 \to A_3, A_4, A_5, A_6$
$A_6 \to A_5$

Step 8: Results (fuzzy forecasted variations) of the previous step are summarized to obtain crisp integer value (forecasted enrollment) for each year under consideration. This process is known as *defuzzification*. In this paper, the defuzzification approach as it is proposed by Song and Chissom [2]-[3] is used – its essential principles can be brought to the following: (a) If all membership values of the output are 0 (zeros), then the forecasted variation is 0 too,

(b) If the membership of output has exactly one maximum, then midpoint of interval, on which this value is reached, is the forecasted variation,
(c) If the membership of output has two or more consecutive maximums, the midpoint of corresponding conjunct intervals is taken for the forecasted variation,
(d) Otherwise, standardize the fuzzy output and use the midpoint of each interval to apply centroid method for calculation of defuzzified forecasted variation.

When the fuzzy variation is obtained, it is summed up with actual enrollment of the last (previous) year. For instance, if the calculated forecasted variation (year 1979) is 400, and the actual enrollment (year 1978) is 15861, then the forecasted enrollment (year 1979) is $15861+400=16261$. The results for the University of Alabama are shown in Table V.

TABLE V
FORECASTED OUTPUTS AND ENROLLMENTS FROM 1973 TO 1993

Year	Actual enrollments	Fuzzy Outputs	Forecasted enrollments
1973	13867	0 0.5 1 1 1 0.5	*13963*
1974	14696	0 0.5 1 1 1 0.5	*14267*
1975	15460	0 0.5 1 1 1 1	*15296*
1976	15311	0 0.5 1 1 1 1	*16060*
1977	15603	0.5 1 0.5 1 1 0.5	*15530*
…	…	… … …	…
1988	18150	0 0.5 1 1 1 1	*17459*
1989	18970	0 0 0 0.5 1 0.5	*18950*
1990	19328	0 0.5 1 1 1 1	*19570*
1991	19337	0 0.5 1 1 1 0.5	*19728*
1992	18876	0.5 1 0.5 1 1 0.5	*19556*
1993		1 0.5 1 0.5 0 0	*18663*

B. Discussion

The proposed time-invariant method achieved better results in comparison with Song-Chissom and Chen's models. The average forecasting errors and the forecasted enrollments of these methods based on using six fuzzy sets are given in Table VI/Figure 1, respectively.

TABLE VI
AVERAGE FORECASTING ERRORS OF TIME-INVARIANT METHODS

	Song and Chissom time-invariant model	Chen's time-invariant model	Proposed time-invariant method
Average forecasting errors	*3.18%*	*3.23%*	*2.42%*

$$Actual\ forecasting\ error\ = \frac{\left| act.enrollment\ \ - forecast.enrollment \right|}{act.\ enrollment} \times 100$$

Furthermore, it is worth mentioning that the number of fuzzy sets (N_{FS}) used in the model affects the average forecasting error. Table VII shows it clearly that forecasting accuracy increases with the growth of N_{FS}. In particular, the increase of N_{FS} value from 5 to 9 results in decrease of average forecasting error by more than 25% in relative units.

TABLE VII
AVERAGE FORECASTING ERRORS FOR DIFFERENT N_{FS}

Proposed time-invariant method	5 fuzzy sets	6 fuzzy sets	7 fuzzy sets	8 fuzzy sets	9 fuzzy sets
Average forecasting errors	*2.75%*	*2.42%*	*2.50%*	*2.02%*	*2.02%*

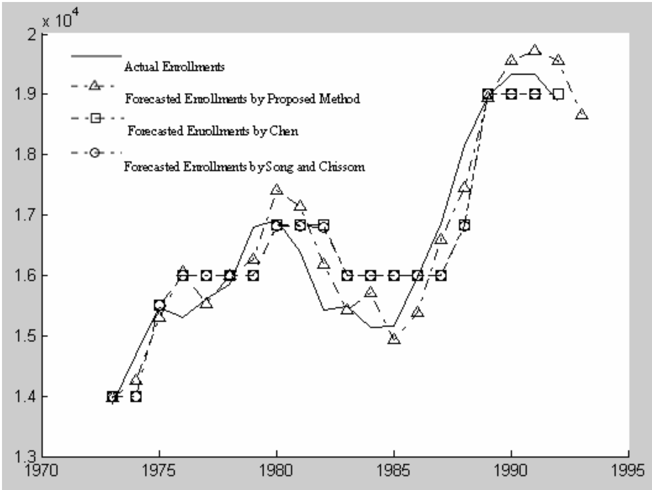

Figure 1. Forecasted enrollments of different models with actual enrollments

IV. Conclusion

In this paper, we presented a novel time-invariant fuzzy time series method for forecasting university enrollments. To illustrate the forecasting process, historical data of the University of Alabama were used as they are summarized in [2]-[4] and [6]. The advantage of the proposed modification lies in utilization of automated forecasting method that operates on sorely available historical data (variations). To study the performance of the method, which significantly improves forecasting accuracy as against [2] and [4], we conducted experiments with different number of linguistic terms (7 ± 2 information span of immediate human memory). As appears from Table VII, the method described turns down average forecasting error below 3% for all cases examined.

References

[1] Q. Song and B.S. Chissom, "Fuzzy time series and its models", Fuzzy Sets and Systems, vol. 54, pp. 269-277, 1993.
[2] Q. Song and B.S. Chissom, "Forecasting enrollments with fuzzy time series – part 1", Fuzzy Sets and Systems, vol. 54, pp. 1-9, 1993.
[3] Q. Song and B.S. Chissom, "Forecasting enrollments with fuzzy time series – part 2", Fuzzy Sets and Systems, vol. 62, pp. 1-8, 1994.
[4] S.-M. Chen, "Forecasting Enrollments Based on Fuzzy Time Series", Fuzzy Sets and Systems, vol. 81, pp. 311-319, 1996.
[5] S.-M. Chen, "Temperature Prediction using Fuzzy Time Series", IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, vol. 30, pp. 263-275, 2000.
[6] K.Huarng, "Heuristic Models of Fuzzy Time Series for Forecasting", Fuzzy Sets and Systems, vol. 123, pp. 369-386, 2001.