TẠP CHÍ KHOA HỌC
TRƯỜNG ĐẠI HỌC QUY NHƠN

Phương pháp mới dự báo theo chuỗi thời gian mờ dựa trên Đại số gia tử

Lê Xuân Việt[1,*], Dương Hoàng Huyên[1], Lê Xuân Vinh[2], Lê Thị Thu Vân[3]

[1]Khoa Công nghệ thông tin, Trường Đại học Quy Nhơn, Việt Nam
[2]Phòng Đào tạo Đại học, Trường Đại học Quy Nhơn, Việt Nam
[3]Tổ Tin học, Trường THPT chuyên Lê Khiết, Việt Nam

Ngày nhận bài: 19/01/2020; Ngày nhận đăng: 01/03/2020

TÓM TẮT

Cho đến nay đã có nhiều tác giả đề xuất các phương pháp dự báo theo chuỗi thời gian mờ. Trong bài báo này, chúng tôi đề xuất một phương pháp mới dự báo chuỗi thời gian mờ dựa trên đại số gia tử. Để kiểm chứng tính hiệu quả của phương pháp, chúng tôi dự báo dựa trên dữ liệu về lượng khách du lịch quốc tế đến Việt Nam và so sánh với phương pháp ARIMA. Kết quả cho thấy phương pháp mới đề xuất cho sai số nhỏ hơn. Hơn nữa, với bộ dữ liệu là số lượng sinh viên nhập học trường Đại học Alabama, phương pháp vừa đề xuất cũng cho kết quả dự báo tốt hơn so với các phương pháp của các tác giả khác.

Từ khóa: *Dự báo, chuỗi thời gian mờ, đại số gia tử, hạng từ ngôn ngữ.*

[*]*Tác giả liên hệ chính.*
Email: lexuanviet@qnu.edu.vn

JOURNAL OF SCIENCE
QUY NHON UNIVERSITY

A Novel Fuzzy Time Series Method for Forecasting Based on Hedge Algebras

Le Xuan Viet[1,*], **Duong Hoang Huyen**[1], **Le Xuan Vinh**[2], **Le Thi Thu Van**[3]

[1]*Department of Information Technology, Quy Nhon University, Vietnam*
[2]*Undergraduate Training Office, Quy Nhon University, Vietnam*
[3] *Department of Information Technology, Le Khiet high school for gifted student, Vietnam*

Received: 19/01/2020; Accepted: 01/03/2020

ABSTRACT

So far, many methods have been proposed to deal with forecasting problems using fuzzy time series. In this paper, we have proposed a new fuzzy time series forecasting method based on hedge algebras. To examine efficiency of our method, we employed the data of the Vietnam's outbound tourists in order to predict and compare with ARIMA. It is clear that our method has smaller error than ARIMA method. Furthermore, experimenting our method with the data of the enrollment of Alabama University, our forecasting results are better than ones presented in other studies.

Keywords: *Forecasting, fuzzy time series, hedge algebras, linguistic term.*

1. INTRODUCTION

The forecasting activities play an important role in our daily life. Accurate forecasting will be used to help people making more suitable decisions. The classical time series methods can not deal with forecasting problems in which the values of time series are linguistic terms represented by fuzzy sets. So, Song and Chissom[1] presented the theory of fuzzy time series to overcome the drawback of the classical time series methods. Over the years, fuzzy time series has been widely used for forecasting data. A lot of studies have been discussed for forecasting used fuzzy time series such as enrollment,[2-9] the stock index,[10] foreign tourists[11] and financial forecasting,[12] etc.

As we know, when solving problems using fuzzy sets, many authors used hedge algebra instead.[13-17] The advantage of using the hedge algebra is that the linguistic values are arranged in semantic order. Moreover, we can easily calculate on these linguistic values because they are quantified into real values by using semantic quantitative function. In this paper, we shall propose a new method based on hedge algebra to solve the forecasting problem. The advantage of this method is that from the fuzziness of linguistic terms of hedge algebras we can determine the corresponding intervals of the discourse of linguistic variable. Moreover, thanks to the parameters of hedge algebras we can adjust the lengths of the intervals, thereby helping the forecasting process to obtain better results. This paper is organized by 5 sections. The first is the introduce. In Section 2, the basic concepts of fuzzy sets and fuzzy time series are

*Corresponding author.
Email: lexuanviet@qnu.edu.vn

described. In Section 3, we have proposed a new method to solve the forecasting problem based on hedge algebras and given some examples to handle. The details are presented in Section 4. The verification and comparison of our method are done with some models.[18-23] Finally, the conclusions and comments are mentioned in Section 5.

2. FUZZY TIME SERIES

2.1. Basic concepts of fuzzy time series

Fuzzy time series model was firstly given by Q. Song and B.S Chissom.[1,7,8] Then, it is improved by S.M Chen[4] to process some arithmetic calculations. From that points, they can get more exactly forecasting results. In this session, we briefly review the concepts of fuzzy time series as in Chen.[19]

Let U be the universe of discourse, where $U = \{u_1, u_2,..., u_n\}$. A fuzzy set defined in the universe of discourse U can be represented as follows:

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \cdots + f_A(u_n)/u_n ,$$

where f_A denotes the membership function of the fuzzy set A, $f_A : U \to [0, 1]$, and $f_A(u_i)$ denotes the degree of membership of u_i belonging to the fuzzy set A, and $f_A(u_i) \in [0, 1]$, and $1 \le i \le n$.

Definition 2.1. Let $Y(t)$ ($t = ...,0,1,2,...$) be the universe of discourse and be a subset of R. Assume $f_i(t)$ ($i = 1,2,...$) are defined on $Y(t)$, and assume that $F(t)$ is a collection of $f_1(t), f_2(t), ...,$ then $F(t)$ is called a fuzzy time series definition $Y(t)$ ($t = ...,0,1,2,...$).

Definition 2.2. Assume that $F(t)$ is caused by $F(t-1)$ only, denoted as $F(t-1) \to F(t)$, then this relationship can be expressed as $F(t) = F(t-1)\circ R(t, t-1)$, where $F(t) = F(t-1)\circ R(t, t-1)$ is called the first-order model of $F(t)$, $R(t, t-1)$ is the fuzzy relationship between $F(t-1)$ and $F(t)$, and "$\circ$" is the Max-Min composition operator.

Definition 2.3. Let $R(t, t-1)$ be a first-order model of $F(t)$. If 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.

Definition 2.4. Assume that the fuzzified input data of the i^{th} year/month is A_j and the fuzzified input data of the i+1th year/month is A_k, where A_j and A_k are two fuzzy sets defined in the universe of discourse U, then the fuzzy logical relationship can be represented by $A_j \to A_k$, where A_j is called the current state of the fuzzy logical relationship.

If we have $A_i \to A_{j1}, A_i \to A_{j2}, ..., A_i \to A_{jk}$ then we can write $A_i \to A_{j1}, A_{j2}, ..., A_{jk}.$

2.2. Rules for fuzzy time series forecasting

Assume that A_j is the value of $F(t-1)$, the forecasted output $F(t)$ be defined with some rules[2]:

i) If there exist a relation 1-1 within group of the relations where A_j on the left of rule, suppose that $A_j \to A_k$, and the maximum membership value of A_k occurs at interval u_k, then the output of $F(t)$ is middle point of u_k.

ii) If $A_k = \varnothing$, that mean $A_j \to \varnothing$ and the maximum membership value of A_j occurs at interval u_j, then the output of $F(t)$ is middle point of u_j.

iii) If we have $A_j \to A_1, A_2, ..., A_n$, and the maximum membership values of $A_1, A_2, ..., A_n$ occur at intervals $u_1, u_2, ..., u_n$ respectively, then the output of $F(t)$ is average of the middle points $m_1, m_2, ..., m_n$ of $u_1, u_2, ..., u_n$, that is $(m_1 + m_2 + \dots + m_n)/n$.

3. THE MODEL OF FORECASTING TIME SERIES BASED ON HEDGE ALGEBRAS

In this section, we shall give a short overview on the algebraic approach to the semantics of vague words in natural languages investigated in papers[13-17] and construct a new method to forecast rely on hedge algebras theory.

3.1. Hedge Algebras: A Short Overview

Given linguistic variable X, every word-domain $\mathcal{X}$ of a variable X can be considered as an abstract algebra AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathcal{H}$, $\le$), where

• $\mathcal{H}$ is the set of linguistic hedges or modifiers considered as *1-ary operations* of the algebra AX;

• $\mathbb{C} = \{\mathbf{0}, W, \mathbf{1}\}$ is a set of *special words* which are, respectively, the least, the medium and the greatest elements of $\mathcal{X}$ and regarded as *constants* of AX since they are fixed points;

• $\mathbb{G} = \{c^-, c^+\}$ is a set of the primary or atomic words of the variable X, the first one is called the *negative* word, say "*young*" of AGE, and the second, the *positive* one, say "*old*".

• $\mathbb{G} \cup \mathbb{C}$ is the set of the generators of the algebra AX that is $\mathcal{H}(\mathbb{G} \cup \mathbb{C}) = \mathcal{X} = \mathbb{C} \cup \mathcal{H}\,\mathbb{G})$, the underlying set of AX where for a subset Z of $\mathcal{X}$, the set $\mathcal{H}(Z)$ denotes the set of all elements freely generated from the words in Z. I.e. $\mathcal{H}(Z) = \{\sigma x : x \in Z \text{ and } \sigma \in \mathcal{H}^*\}$, where $\mathcal{H}^*$ is the set of all strings of hedges in $\mathcal{H}$, including the empty string ε. Note that for $\sigma = \varepsilon$, $\varepsilon x = x$ and, hence, $Z \subseteq \mathcal{H}(Z)$. In the case $Z = \{x\}$ we shall write $\mathcal{H}(x)$ instead of $\mathcal{H}(\{x\})$.

• $\leq$ is a Semantical Order Relation (SOR) upon $\mathcal{X}$.

If a word $x \in \mathcal{X}$ is generated from a given element u by means of hedges h_i, $i = 1 \ldots, n$, then we will write $x = h_n \ldots h_1 u$ and it is called a string representation of x with respect to (w.r.t.) u.

For illustration, let us consider an HA of the variable AGE, AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathcal{H}$, $\leq$), where $\mathbb{G} = \{c^-, c^+\}$, $\mathcal{H}^- = \{R, L\}$ and $\mathcal{H}^+ = \{V, E\}$, where c^- and c^+ stand for "young" and "old" and R, L, V and E for "Rather", "Little", "Very" and "Extremely", respectively. The order and the generality-specificity relation and the set-containing relation of the fuzziness models of the words can be represented by a labeled graph given in Figure 1, in which every *arrow* means that the word at its root generates the word at its peak and the relationships between the words and between the fuzziness models at the respective positions of the arrow are specified by an inequality and a set-containing relation associated with the arrow.

Consider an HA AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathcal{H}$, $\leq$) of an attribute X with numeric reference interval domain U normalized to be [0,1], for convenience in a unified presentation of the quantification of the hedge algebras. Formally, the numeric semantics of the words of X can be determined by a so-called Semantically Quantifying Mapping (SQM), $f : \mathbb{X} \to [0, 1]$, defined as follows.

Definition 3.1. A mapping f: $\mathcal{X} \to [0, 1]$ is said to be an SQM of AX, if we have:

(SQM1) f is an *order isomorphism*, i.e. it is one-to-one and for $\forall x, y \in \mathcal{X}$, $x \leq y \Rightarrow f(x) \leq f(y)$.

(SQM2) The image of $\mathcal{X}$ under f, $f(\mathcal{X})$, is topologically dense in the universe [0, 1].

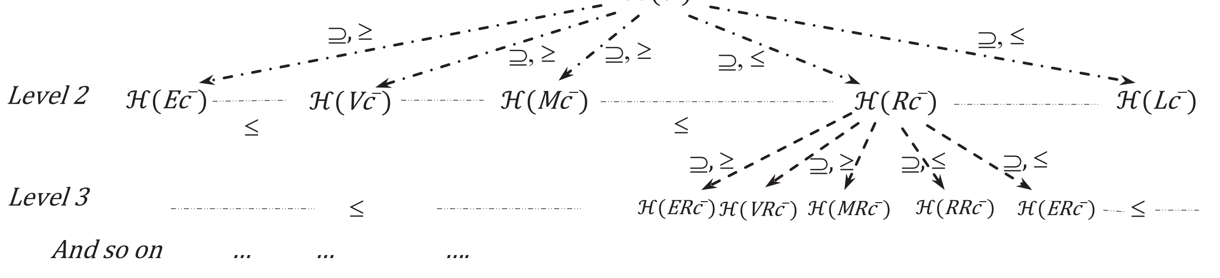

Figure 1. A piece of the graph representation of the semantic structure of the words of X generated from the atomic word c^- and their fuzziness models, which involves the order-based relationships and the set-containing relationships between the fuzziness models of words and the word generality-specificity relationships

Definition 3.2. A function fm: $\mathcal{X} \rightarrow [0, 1]$ is said to be a fuzziness measure of the HA AX associated with the given variable X, if it satisfies the following axioms, for any $x \in \mathcal{X}$ and $h \in \mathcal{H}$:

($fm1$) $fm(c^-) + fm(c^+) = 1$.

($fm2$) $\sum_{-q \le j \le p, j \ne 0} fm(h_j x) = fm(x)$.

($fm3$) $fm(hx) = \mu(h)fm(x)$, where $\mu(h)$ is called for convenience the fuzziness measure of h as well.

($fm4$) For $x = h_n h_{n-1} \ldots h_1 c$, $fm(x) = fm(h_n h_{n-1} \ldots h_1 c) = \mu(h_n)\mu(h_{n-1}) \ldots \mu(h_1)fm(c)$, $c \in \mathbb{G} = \{ c^-, c^+\}$.

($fm5$) Setting $\sum_{-q \le j \le -1} \mu(h_j) = \alpha$ & $\sum_{1 \le j \le p} \mu(h_j) = \beta$, we have $\alpha + \beta = \sum_{-q \le j \le p, j \ne 0} \mu(h_j) = 1$.

In the general case, for given values of the fuzziness parameters of X we can stablish a recursive expression to compute the SQM υ_{fm}, called the SQM induced by the given fm, as follows:

• $\upsilon_{fm}(W) = \kappa = fm(c^-)$, $\upsilon_{fm}(c^-) = \kappa - \alpha fm(c^-) = \beta fm(c^-)$, $\upsilon_{fm}(c^+) = \kappa + \alpha fm(c^+)$;

• $\upsilon_{fm}(h_j x) = \upsilon_{fm}(x) + sign(h_j x) \times$

$$\left(\sum_{i = sign(j)}^{j-1} \mu_i(h_i) + (1 - \omega(h_j x))\,\mu(h_j)\right) fm(x)$$

where

$$\omega(h_j x) = \frac{1}{2}[1 + sign(h_j x)\,sign(h_p h_j x)\,(\beta - \alpha)] \in \{\alpha, \beta\}$$

for all $j \in [-q \ldots p]$, $j \ne 0$, and $sign()$ function is defined as in Ho N.C.[16]

3.2. Semantization and desemantization

To convert the values from the reference domain to semantic domain of a variable X and vice versa, we synthesize some transformations as below. Assume that $[a, b]$ is a reference domain of the variable X, and $[a_s, b_s]$ is semantic domain. The linear conversion from $[a, b]$ to $[a_s, b_s]$ is called *linear semantization* and the conversion from $[a_s, b_s]$ to $[a, b]$ is called *linear desemantization*. The semantic domain by using hedge algebras is usually [0, 1], thus the linear semantization is named *normalization* and linear desemantization is named *denormalization*. We have some functions as below:

LinearSemantization (x) =

$$x_s = a_s + (b_s - a_s)(x - a)/(b - a)$$

LinearDesemantization(x_s) =

$$x = a + (b - a)(x_s - a_s)/(b_s - a_s)$$

Normalization (x) = $x_s = (x - a)/(b - a)$

Denormalization (x_s) = $x = a + (b - a)x_s$

For flexibility in semantization or desemantization, we can expand from linear to nonlinear by adding some parameters $sp, dp \in [-1, 1]$, for example:

NonlinearSemantization (x) = $f(x_s, sp)$, satisfy the conditions $0 \le f(x_s, sp) \le 1$, $f(x_s = 0, sp) = 0$, $f(x_s = 1, sp) = 1$.

NonlinearDesemantization (x_s) = $g(x, dp)$, satisfy conditions $a \le g(x, dp) \le b$, $g(x = a, dp) = a$, $g(x = b, dp) = b$.

In this paper, we shall use the functions:

NonlinearNormalization $(x) = f(x_s, sp) = sp \times x_s(1 - x_s) + x_s$

NonlinearDenormalization$(x_s) = dp \times (Denormalization(f(x_s, sp) - a) \times (b - Denormalization(f(x_s, sp)))/(b - a) + Denormalization(f(x_s, sp))$, where $Denormalization(f(x_s, sp)) = (sp \times x(1 - x) + x) \times (b - a) + a$.

The *NonlinearDenormalization* be denoted *ND* for short.

3.3. A new method on forecasting

Inputs: n values of data $\{y(t_1), y(t_2), \ldots, y(t_n)\}$ with $t_1, t_2, \ldots, t_n$ are point times.

Output: $y(t_{n+1})$ at the time t_{n+1}.

Step 1. Define the discourse U

Put $U = [D_{min}, D_{max}]$ where $D_{min} = \min\{y(t_1), y(t_2), \ldots, y(t_n)\}$ and $D_{max} = \max\{y(t_1), y(t_2), \ldots, y(t_n)\}$.

Step 2. Building the intervals upon U by using fuzziness model of hedge algebras

We assume that U is a discourse of an abstract variable $\mathcal{X}$ and the linguistic domain of $\mathcal{X}$ to be considered as an algebra AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathcal{H}$, ≤). Then, by chosen k, the discourse U is divided into k intevals u_1, u_2, …, u_k w.r.t level k as in Figure 1. The inteval u_i is labeled A_i, i = 1, 2,..., k satifying $A_1 < A_2 < \ldots < A_k$. We caculate the length of intervals of u_i denoted f_{ui}, $f_{ui} = fm(A_i)\times(D_{max} - D_{min})$, i = 1, 2, …, k. So we have $u_1 = [u_{1d}, u_{1c}] = [D_{min}, D_{min} + f_{u1}]$, $u_2 = [u_{2d}, u_{2c}] = [u_{1c}+1, u_{2d}+f_{u2}]$, …, $u_k = [u_{kd}, u_{kc}] = [u_{(k-1)c}+1, u_{kd}+f_{uk}]$.

Step 3. Quantifying semantic of the linguistic values A_1, A_2, ..., A_k.

To quantify the semantic of A_1, A_2, ..., A_k, we use SQM υ_{fm} as $SA_1 = \upsilon_{fm}(A_1)$, $SA_2 = \upsilon_{fm}(A_2)$, ..., $SA_k = \upsilon_{fm}(A_k)$. By properties of hedge algebras, it is clear that $SA_1 < SA_2 < ... < SA_k$.

Step 4. Constructing the relationships

Suppose that, $F(t-1)$ is A_i, $F(t)$ is A_j, and $F(t)$ is caused by $F(t-1)$. Clearly, we have relation between A_i and A_j, denoted $A_i \rightarrow A_j$.

Step 5. Grouping relationship

If $A_i \rightarrow A_{j1}$, A_i A_{j2},..., $A_i \rightarrow A_{jm}$, then we estalish the relation by grouping all of them to unique relation $A_i \rightarrow A_{j1}, A_{j2}, ..., A_{jm}$.

Step 6. Caculating output value

From group of the relations in Step 5, applying the rules similarity in the way of Section 2.2 we get the results of $F(t)$, scilicet:

• If there is a relation $A_i \rightarrow A_j$, then $F(t) = ND(SA_j)$ upon $u_j = [u_{jd}, u_{jc}]$.

• If $A_i \rightarrow \varnothing$ then $F(t) = ND(\varnothing)$ upon $u_i = [u_{id}, u_{ic}]$.

• If $A_i \rightarrow A_{j1}, A_{j2}, ..., A_{jk}$ then $F(t) = ND(W_{i,j1}\times SA_{j1} + W_{i,j2}\times SA_{j2} + ... + W_{i,jk}\times SA_{jk})$ upon $[\min\{u_{j1d}, u_{j2d}, ..., u_{jkd}\}, \max\{u_{j1c}, u_{j2c}, ..., u_{jkc}\}]$ where $W_{i,j}$ is the weights measured in the ratio number of times of real data in the interval u_i to sum of number of times of real data in the intervals u_{j1}, u_{j2}, ..., u_{jk}.

4. EXPERIMENTAL RESULTS

Now, we would like to forecast the number of tourists going to Vietnam on Jan 2012 based on the real data from January 2010 to December 2011 (the data are given in column "Real Data" of Table 1). This data are also on website of Vietnam National Administration of Tourism *(http://vietnamtourism.gov.vn)*.

For illustration, we shall apply the algorithm step by step to solve this problem.

• The discourse U = [D_{min}, D_{max}] = [286618, 611864].

• Chosen a hedge algebra AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathcal{H}$, ≤), where $\mathbb{G}$ = *{Small, Large}*, $\mathbb{C}$ = {**0, W, 1**}, $\mathcal{H}$ = *{Little, Very}*, *μ(Little)* = 0.5, *μ(Very)= 0.5*, *κ = 0.5*. In this case, level k be defined is 3, so we clearly get 8 linguistic values $A_1 < A_2 < A_3 < A_4 < A_5 < A_6 < A_7 < A_8$ with respect to *VeryVerySmall* < *LittleVerySmall* < *LittleLittleSmall* < *VeryLittleSmall* < *VeryLittleLarge* < *LittleLittleLarge* < *LittleVeryLarge* < *VeryVeryLarge*. Obviously, $fm(A_1) = fm(A_2) = ... = fm(A_8) = 0.125$. Consequently, $f_{u1} = f_{u2} = ... = f_{u8} = 40656$. Finally, the discouse U be divided into 8 intervals such as u_1 = [286618, 327274], u_2 = [327275, 367931], u_3 = [367932, 408588], u_4 = [408589, 449245], u_5 = [449246, 489902], u_6 = [489903, 530559], u_7 = [530560, 571216], u_8 = [571217, 611864].

• Quantifying semantic of the linguistic values A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8 by using SQM υ_{fm} of AX, we have SA_1 = 0.0625, SA_2 = 0.1875, SA_3 = 0.3125, SA_4 = 0.4375, SA_5 = 0.5625, SA_6 = 0.6875, SA_7 = 0.8125, SA_8 = 0.9375.

• Constructing the relations from data

Table 1. The relations from data

Order	Time	Real data	Label	The relations
1	1/2010	416249	A_4	
2	2/2010	446323	A_4	$A_4 \rightarrow A_4$
3	3/2010	473509	A_5	$A_4 \rightarrow A_5$
4	4/2010	432608	A_4	$A_5 \rightarrow A_4$

5	5/2010	350982	A_2	$A_4 \rightarrow A_2$
6	6/2010	375707	A_3	$A_2 \rightarrow A_3$
7	7/2010	410000	A_4	$A_3 \rightarrow A_4$
8	8/2010	427935	A_4	$A_4 \rightarrow A_4$
9	9/2010	383463	A_3	$A_4 \rightarrow A_3$
10	10/2010	440071	A_4	$A_3 \rightarrow A_4$
11	11/2010	428295	A_4	$A_4 \rightarrow A_4$
12	12/2010	449570	A_5	$A_4 \rightarrow A_5$
13	1/2011	506424	A_6	$A_5 \rightarrow A_6$
14	2/2011	542671	A_7	$A_6 \rightarrow A_7$
15	3/2011	475733	A_5	$A_7 \rightarrow A_5$
16	4/2011	460000	A_5	$A_5 \rightarrow A_5$
17	5/2011	480886	A_5	$A_5 \rightarrow A_5$
18	6/2011	446966	A_4	$A_5 \rightarrow A_4$
19	7/2011	460000	A_5	$A_4 \rightarrow A_5$
20	8/2011	490000	A_6	$A_5 \rightarrow A_6$
21	9/2011	286618	A_1	$A_6 \rightarrow A_1$
22	10/2011	518477	A_6	$A_1 \rightarrow A_6$
23	11/2011	611864	A_8	$A_6 \rightarrow A_8$
24	12/2011	593408	A_8	$A_8 \rightarrow A_8$

- Group of relations from the Table 1

$A_1 \rightarrow (A_6)$

$A_2\ (A_3)$

$A_3\ (A_4, A_4)$

$A_4\ (A_2, A_3, A_4, A_4, A_4,, A_5, A_5, A_5)$

$A_5\ (A_4, A_4, A_5, A_5, A_6\ , A_6)$

$A_6\ (A_1, A_7, A_8)$

$A_7\ (A_5)$

$A_8\ (A_8)$

Table 2. The weights of group relations

Intv.	Month/ Year	Amount of tourists	Group relations and weights $W_{i,j}$	Num. of values in u_i
u_1	9/2011	286618	$A_1 \rightarrow (A_6)$ $W_{1,6} = 3/3=1$	1
u_2	5/2010	350982	$A_2 \rightarrow (A_3)$ $W_{2,3} = 2/2=1$	1
u_3	6/2010 9/2010	375707 383463	$A_3 \rightarrow (A_4, A_4)$ $W_{3,4} = 8/(2\times8) = 1/2$	2
u_4	1/2010 2/2010 4/2010 7/2010 8/2010 10/2010 11/2010 6/2011	416249 446323 432608 410000 427935 440071 428295 446966	$A_4 \rightarrow (A_2, A_3, A_4, A_4, A_4, A_5, A_5, A_5)$ $W_{4,2} = 1/(1+2+8\times3+6\times3) = 1/45$ $W_{4,3} = 2/(1+2+8\times3+6\times3) = 2/45$ $W_{4,4} = 8/(1+2+8\times3+6\times3) = 8/45$ $W_{4,5} = 6/(1+2+8\times3+6\times3) = 6/45$	8
u_5	3/2010 12/2010 3/2011 4/2011 5/2011 7/2011	473509 449570 475733 460000 480886 460000	$A_5 \rightarrow (A_4, A_4, A_5, A_5, A_6, A_6)$ $W_{5,4} = 8/(8\times2+6\times2+3\times2) = 8/34$ $W_{5,5} = 6/(8\times2+6\times2+3\times2) = 6/34$ $W_{5,6} = 3/(8\times2+6\times2+3\times2) = 3/34$	6
u_6	1/2011 8/2011 10/2011	506424 490000 518477	$A_6 \rightarrow (A_1, A_7, A_8)$ $W_{6,1} = 1/(1+1+2)=1/4$ $W_{6,7} = 1/(1+1+2)=1/4$ $W_{6,8} = 2/(1+1+2) = 2/4 = 1/2$	3
u_7	2/2011	542671	$A_7 \rightarrow (A_5)$ $W_{7,5} = 6/6 = 1$	1
u_8	11/2011 12/2011	611864 593408	$A_8 \rightarrow (A_8)$ $W_{8,8} = 2/2 = 1$	2

For example, to calculate number of tourists at the time February 2010, because the current state is A_4, we consider the relation $A_4 \rightarrow (A_2, A_3, A_4, A_4, A_4,, A_5, A_5, A_5)$. According to Table 2, the data in this group belong to intervals u_2, u_3 u_4, u_5 where u_2 contains 1 value, u_3 contains 2 values, u_4 contains 8 values, but A_4 appears three times, thus the number of values of u_4 is (3×8). Similarly, u_5 contains (3×6) values. So, sum of them equal to (1+2+3×8+3×6 = 45) and quantified semantic x_s is 0.47639.

With the manual chosen sp=0.6 and dp=-0.2, x = $Denormalization(x_s)$ = $f(0.47639, 0.6)$ = (0.6× 0.47639×(1- 0.47639)+ 0.47639) × (489902 – 327275) + 327275 = 429089 and

$ND(x)$ = g(429089,-0.2) = -0.2 × (429089-327275) × (489902-429089)/(489902–327275) + 429089 = 421474.

We find out the forecasted value on February 2010 is 421474. Completely similar as above, all forecated values are represented in Figure 2.

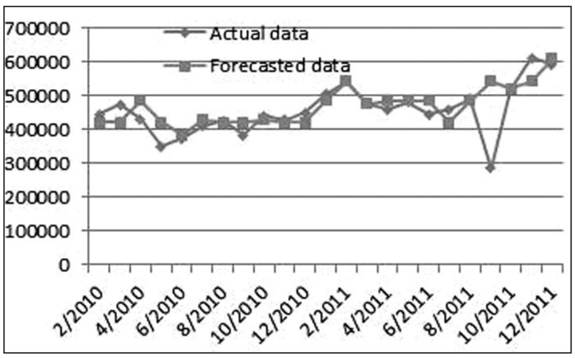

Figure 2. Curves of forecasted data and actual data

In the following, we use the mean square error (MSE) to compare the forecasting results of different forecasting methods, where the mean square error is calculated as follows:

$$MSE = \frac{\sum_{i=1}^{n}\left(Actual\_Data_i - Forecasted\_Data_i\right)^2}{n}$$

where *Actual_Data i* denotes the actual data of month *i*, and *Forecasted_Data i* denotes the forecasted data of month *i*. In Table 3, we compare the forecasting results of the proposed method with the one of the existing method[4] through five month from 1/2011 to 5/2011.

Table 3. A comparison results between our method and ARIMA method[18]

Month/Year	Real data	ARIMA method[18]	The proposed method
1/2011	506424	443853	485611
2/2011	542671	426238	542401
3/2011	475733	611361	476444
4/2011	460000	558587	485611
5/2011	480886	609786	485611
MSE		**249402**	**33347**

To further affirm the effectiveness of our method, we experiment with data of enrollment students of Alabama University from 1971 to 1992.

With chosen hedge algebra AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathcal{H}$, ≤), $\mathbb{G}$ = *{Small, Large}*, $\mathbb{C}$ = {**0, W, 1**}, $\mathcal{H}$ = *{Little, Very}*, where the parameters are set *μ(Little)* = 0.44, *μ(Very)* = 0.56, *κ* = 0.4 and sp = 0.2, dp = 0.0, the calculation is completely similar to the one above. We get the result as shown in Table 4.

Table 4. MSE of some methods with admission data of Alabama University

Year	Num. of students	Chen Method[19]	Lee Method[20]	Manikandan Method[21]	Qiu Method[22]	Huarng Method[23]	The proposed Method
1971	13055						
1972	13563	14000	13833	13739	14195	14000	13233
1973	13867	14000	13833	14639	14424	14000	13233
1974	14696	14000	13833	15539	14593	14000	14608
1975	15460	15500	15500	15557	15589	15500	15192
1976	15311	16000	15722	15539	15645	15500	15611
1977	15603	16000	15722	15557	15634	16000	15611
1978	15861	16000	15722	16739	16100	16000	16159
1979	16807	16000	15722	17639	16188	16000	16159
1980	16919	16833	16750	16439	17077	17500	17155
1981	16388	16833	16750	15539	17105	16000	17155
1982	15433	16833	16750	15557	16369	16000	16159
1983	15497	16000	15722	15557	15643	16000	15611
1984	15145	16000	15722	15539	15648	15500	15611
1985	15163	16000	15722	15539	15622	16000	15611
1986	15984	16000	15722	16739	15623	16000	15611
1987	16859	16000	15722	17639	16231	16000	16159
1988	18150	16833	16750	19139	17090	17500	17155
1989	18970	19000	19000	19439	18325	19000	19235
1990	19328	19000	19000	19289	19000	19000	19235
1991	19337	19000	19000	19289	19000	19500	19235
1992	18876	19000	19000	19439	19000	19000	19235
MSE		**407507**	**397537**	**325158**	**261473**	**226611**	**220401**

5. CONCLUSIONS

In this paper, we have proposed a fuzzy time series method for forcasting based on hedge algebras. The proposed method is both simple and effective. Especially, the intervals of the discourse are made upon the k-level of hedge algebras. We have experimented our method with the data of tourists going to Vietnam and data of enrollment student of University of Alabama. The forecasted results obtained by our method are better than ones of the other methods. It can be proven with MSE as in Table 3&4.

The errors of the forecasted results of our method are directly influenced by the parameters of hedge algebras, but these were intuitively chosen. In the future, we shall put forward a way to set up these parameters automatically, for instance using particle swarm optimization to reach a higher forecasting accuracy rate.

REFERENCES

1. Song Q, Chissom B.S. Fuzzy time series and its models. *Fuzzy Sets and Syst.*, **1993**, *54*, 269 - 277.

2. Chen, S. M. Forecasting Enrollments Based on Fuzzy Time Series. *Fuzzy Sets and Syst.* **1996**, *81*, 311 - 319.

3. Chen S. M. Forecasting Enrollments based on High Order Fuzzy Time Series. *Cybernetics and Systems An International Journa*l, **2002**, *33*, 1 - 16.

4. Chen S. M and Chung N. Y. Forecasting enrollments using high-order fuzzy time series and genetic algorithms, *Int. Journal of Intelligent Systems*, **2006**, *21*, 485 - 501.

5. Lee M. H, Efendi R, Ismad Z. Modified Weighted for Enrollments Forecasting Based on Fuzzy Time Series. *MATEMATIKA*, **2009**, *25* (1), 67 - 78.

6. Lin, C.J.; Chen, H.F.; Lee, T.S. Forecasting tourism demand using time series, artificial neural networks and multivariate adaptive regression splines. *Evidence from Taiwan. Int. J. Bus. Adm*., **2011**, *2*, 14 - 24.

7. Song Q, Chissom B.S. Forecasting enrollments with fuzzy time series – part 1. *Fuzzy Sets and Syst.*, **1993**, *54*, 1 - 9.

8. Song Q, Chissom B.S. Forecasting enrollments with fuzzy time series - part 2. *Fuzzy Sets and Syst.*, **1994**, *62*, 1 - 8.

9. Huarng, K. Heuristic Models of Fuzzy Time Series for Forecasting, *Fuzzy Sets and Syst.* **2001**, *123*, 369 - 386.

10. C.H. Cheng, T.L. Chen & C.H. Chiang. Trend-weighted fuzzy time series model for TAIEX forecasting, *ICONIP, Part III, LNNC*, **2006**, *4234*, 469 - 477.

11. Hu. Predicting Foreign Tourists for the Tourism Industry Using Soft Computing-Based Grey–Markov Models, *Sustainability,* **2017**, *9*(7), 1-12.

12. C.H.L.Lee, A.Liu & W.S.Chen. Pattern Discovery of Fuzzy time series for financial prediction, *IEEE Transactions on Knowledge and data Engineering*, **2006**, *18*, 613 - 625.

13. N.C. Ho, W. Wechler. Hedge algebras: An algebraic approach to structures of sets of linguistic domains of linguistic truth variable, *Fuzzy Sets and Systems*, **1990**, *35* (3), 281 - 293.

14. N. Cat Ho and W. Wechler, Extended hedge algebras and their application to Fuzzy logic, *Fuzzy Sets and Systems*, **1992**, *52*, 259 - 281.

15. N.C. Ho H.V. Nam. Towards an Algebraic Foundation for a Zadeh Fuzzy Logic, *Fuzzy Sets and System*, **2002**, *129*, 229 - 254.

16. N.C. Ho. A Topological Completion of Refined Hedge Algebras and a Model of Fuzziness of Linguistic words, *Fuzzy Sets and Systems*, **2007**, *158* (4), 436 - 451.

17. N.C Ho, V.N. Lan, L.X Viet. Optimal hedge-algebras-based controller: Design and Application, *Fuzzy Sets and Systems*, **2008**, *159*, 968 - 989.

18. D.Q Giam, V.T Han, L.T.L. Phuong, N.T. Thuy. Building ARIMA model to forecast Vietnam outbound tourists, *Journal of Science and Development Hanoi Agricultural University*, **2012**, *10* (2), 364 - 370.

19. Chen, S.M. Forecasting Enrollments Based on Fuzzy Time Series, *Fuzzy Sets and Syst*, **1996**, *81*, 311 - 319.

20. Lee M H, Efendi R, Ismad Z. Modified Weighted for Enrollments Forecasting Based on Fuzzy Time Series. *MATEMATIKA*, **2009**, *25*(1), 67 - 78.

21. Manikandan. M, Dr. Senthamarai Kannan. K, Deneshkumar. V. Computational Method Based on Distribution in Fuzzy Time Series Forecasting, **2013**, *2* (8).

22. Qiu W, Liu X, Li H. Generalized Method for Forecasting Based on Fuzzy Time Series, *Expert Systems with Applications*, **2011**, *38*, 10446 - 10453.

23. Huarng, K. Heuristic Models of Fuzzy Time Series for Forecasting, *Fuzzy Sets and Syst*, **2001**, *123*, 369 - 386.