Thiết kế mô hình dự báo chuỗi thời gian mờ tối ưu dựa trên đại số gia tử

TÓM TẮT

Dự báo chuỗi thời gian mờ đã thu hút được sự chú ý đáng kể nhờ khả năng xử lý sự không chắc chắn và thiếu chính xác trong dữ liệu chuỗi thời gian. Các mô hình chuỗi thời gian mờ truyền thống thường gặp hạn chế trong việc nắm bắt các mối quan hệ phức tạp giữa các biến. Để giải quyết thách thức này, chúng tôi đề xuất một mô hình tiếp cận mới gọi là mô hình dự báo chuỗi thời gian mờ dựa trên đại số gia tử (OHAM). Đầu tiên, chúng tôi giới thiệu khái niệm về đại số gia tử và ứng dụng của chúng trong phân tích chuỗi thời gian mờ. Sau đó, chúng tôi trình bày các bước xây dựng mô hình, bao gồm việc xác định các nhãn ngôn ngữ trong đại số gia tử, xây dựng các quan hệ mờ từ dữ liệu, chia đoạn cho không gian tham chiếu. Tiếp đó, chúng tôi đề xuất một thuật toán tối ưu hóa để tinh chỉnh các tham số của OHAM, nhằm đạt được hiệu quả dự báo tối ưu. Cuối cùng là thử nghiệm trên một số bộ dữ liệu cụ thể để đánh giá tính hiệu quả của mô hình. Kết quả thử nghiệm cho thấy mô hình mới đề xuất ít sai số hơn so với nhiều mô hình khác.

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

Design of Optimal Hedge-Algebras-based Model for Fuzzy Time Series Forecasting

ABSTRACT

Fuzzy time series forecasting has garnered significant attention due to its ability to handle uncertainty and imprecision in time series data. Traditional fuzzy time series models often face limitations in capturing complex relationships between variables. To address this challenge, we propose a novel approach called the Optimal Hedge-Algebras-based Model (OHAM). First, we introduce the concepts of hedge algebra and its application in fuzzy time series analysis. Subsequently, we present the model construction steps, including defining linguistic labels in hedge algebra, constructing fuzzy relations from data, and partitioning the universe of discourse. Following this, we propose an optimization algorithm to fine-tune the parameters of OHAM, aiming to achieve optimal forecasting performance. Finally, experiments are conducted on several specific datasets to evaluate the effectiveness of the model. The experimental results demonstrate that the newly proposed model exhibits better accuracy than many others.

Keywords: *Forecasting, Fuzzy Time Series, Hedge Algebras, Vague Words, Linguistic Terms.*

1. INTRODUCTION

The proposed hedge algebra by N. C. Ho[1,2,3] has been tested in various applications, yielding positive results in problems such as fuzzy control, classification, fuzzy clustering, and fuzzy time series forecasting,[4,5] among others.

Forecasting plays a crucial role in numerous fields such as finance, weather prediction, and stock market analysis.[6,7] In recent years, fuzzy time series forecasting models have gained attention due to their ability to handle the uncertainty and vagueness present in real-world data. One such model is the hedge-algebras-based forecasting model.[8]

The hedge-algebras-based forecasting model utilizes an algebraic structure to capture the relationships between historical data and future predictions. However, the performance of this model heavily relies on parameter calibration. Determining optimal parameters is a challenging task that requires an efficient optimization algorithm.

In this paper, we propose the application of the Artificial Bee Colony (ABC) algorithm to optimize the parameters of the hedge-algebras-based forecasting model for fuzzy time series. The ABC algorithm is a metaheuristic optimization technique inspired by the foraging behavior of honey bees. It has been successfully applied to various optimization problems and showcases robustness and convergence efficiency.

By employing the ABC algorithm, this research aims to enhance the accuracy and reliability of the hedge-algebras-based forecasting model. The ABC algorithm will efficiently search the parameter space, finding the optimal combination of parameters for the model. This process will help in achieving improved forecast accuracy, reduced error rates, and enhanced decision-making capabilities in diverse applications.

To evaluate the proposed approach, extensive experiments will be conducted using real-world datasets from different domains. Comparative analyses will be carried out, comparing the performance of the optimized hedge-algebras-based forecasting model with other well-established optimization techniques. The results obtained will provide insights into the effectiveness and efficiency of the ABC algorithm in parameter optimization for fuzzy time series forecasting models.

2. PROBLEM OF FUZZY TIME SERIES FORECASTING

The problem in time series forecasting is to accurately predict future values or trends based on historical data. This involves addressing challenges such as identifying and modeling trends, handling seasonality and noise, accounting for non-linear relationships and non-stationarity, and choosing the optimal model and parameters. The goal is to develop a robust forecasting method that can generalize well beyond the training data and provide reliable predictions for effective decision-making. Overcoming these challenges requires a combination of statistical techniques, machine learning algorithms, and domain expertise to achieve accurate and meaningful forecasts.

Fuzzy time series, a concept derived from fuzzy set theory, is a powerful tool for modeling and forecasting time-dependent data with inherent uncertainty and imprecision. Unlike traditional time series analysis, which assumes crisp values, fuzzy time series allows for the representation of vague and uncertain information through linguistic terms and membership functions. By incorporating fuzzy logic into the modeling process, fuzzy time series enables the handling of complex data, making it particularly suitable for real-world scenarios where uncertainty is prevalent. This approach has found applications in various domains, including finance, economics, weather prediction, and decision-making systems, providing valuable insights and accurate predictions in situations where conventional methods may fall short.

The problem is stated as follows: Given n values $y(t_1)$, $y(t_2)$,…, $y(t_n)$ where t_1, t_2, …, t_n are point times. How to predict the next value?

2.1. Some basic definitions

The fuzzy time series model was first proposed by Q. Song and B. S. Chissom.[9] Then, it is improved by S.M Chen[10,11] to process some arithmetic calculations. From that point, they can get more exact forecasting results. In this session, we briefly review the concepts of fuzzy time series as in Q.Song.[9]

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, $f_A : U \rightarrow [0, 1]$, 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 1. [9] 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. [9] Assume that $F(t)$ is caused by $F(t-1)$ only, denoted as $F(t-1) \rightarrow F(t)$, 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 "∘" is the Max-Min composition operator.

Definition 3. [9] Assume that the fuzzified input data of the i^{th} year/month is A_j and the fuzzified input data of the i+1[th] 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 $A_j \rightarrow A_k$, where A_j is called the current state of the fuzzy logical relationship.

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

2.2. Rules for calculating output value

Assume that A_j is the value of $F(t-1)$, the forecasted output $F(t)$ be defined as in research:[10]

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

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

b) If we have $A_j \rightarrow A_1, A_2 ,…, A_n$, 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 $(m_1 + m_2 + … + m_n)/n$.

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

In this section, we provide a brief overview of the algebraic approach to the semantics of vague words in natural languages, as explored in previous studies,[1-4] and introduce a new forecasting method based on hedge algebra theory.

3.1. Algebraic structure of vague term domain

Hedge algebras, denoted as AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathscr{H}$, ≤), are a mathematical structure to handle uncertainty and vagueness. In hedge algebras, $\mathcal{X}$ represents a set of words $\mathscr{H}$ is the set of linguistic hedges or modifiers considered as 1-*ary* operations of the algebra AX; $\mathbb{C}$ = {0, W, 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 linguistic variable X, the first one is called the negative word, and the second, the positive one. $\mathbb{G} \cup \mathbb{C}$ is the set of the generators of the algebra AX that is $\mathscr{H}(\mathbb{G} \cup \mathbb{C}) = \mathcal{X} = \mathbb{C} \cup \mathscr{H}(\mathbb{G})$, the underlying set of AX where for a subset Z of $\mathcal{X}$, the set $\mathscr{H}$(Z) denotes the set of all elements freely generated from the words in Z. I.e. $\mathscr{H}$(Z) = {σx : x ∈ Z and σ ∈ $\mathscr{H}$*}, where $\mathscr{H}$* is the set of all strings of hedges in $\mathscr{H}$, including the empty string ε. Note that for σ = ε, εx = x and, hence, Z ⊆ $\mathscr{H}$(Z). In the case Z = {x} we shall write $\mathscr{H}$(x) instead of $\mathscr{H}$({x}). ≤ is a semantical order relation upon $\mathcal{X}$.

Consider a hedge algebra AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathscr{H}$, ≤) 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 $\mathcal{X}$ can be determined by a so-called Semantically Quantifying Mapping (SQM), *f* : $\mathcal{X}$ → [0, 1], defined as follows.

Definition 4. [3] A mapping *f*: $\mathcal{X}$ → [0, 1] is said to be an SQM of AX, if we have:

- *f* is an *order isomorphism*, i.e. it is one-to-one and for $\forall x, y \in \mathcal{X}, x \le y \Rightarrow f(x) \le f(y)$.
- The image of $\mathcal{X}$ under *f*, *f*($\mathcal{X}$), is topologically dense in the universe [0, 1].

Definition 5. [3] A function *fm*: $\mathcal{X}$ → [0, 1] is said to be a fuzziness measure of the hedge algebra AX associated with the given variable X, if it satisfies the following axioms, for any $x \in \mathcal{X}$ and $h \in \mathscr{H}$:

- $fm(c-) + fm(c+) = 1$.
- $\sum_{-q \le j \le p, j \ne 0} fm(h_j x) = fm(x)$.
- *fm(hx) = μ(h)fm(x), where μ(h) is called for convenience the fuzziness measure of h as well.*

- *For* $x = h_n h_{n-1} \dots h_1 c$, $fm(x) = fm(h_n h_{n-1} \dots h_1 c) = \mu(h_n)\mu(h_{n-1}) \dots \mu(h_1) fm(c)$, $c \in \mathbb{G} = \{c-, c+\}$.
- *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 establish a recursive expression to compute the SQM υ_{fm}, called the SQM induced by the given *fm*, as follows:[3]

- $\upsilon_{fm}(W) = \kappa = fm(c^-)$, $\upsilon_{fm}(c^-) = \kappa - \alpha fm(c^-) = \beta fm(c^-)$, $\upsilon_{fm}(c^+) = \kappa + \alpha fm(c^+)$;
- $\upsilon_{fm}(h_j x) = \upsilon_{fm}(x) + sign(h_j x) \times \left(\sum_{i=sign(j)}^{j-1} \mu_i(h_i) + \left(1 - \omega(h_j x)\right)\mu(h_j)\right) fm(x)$

where
$\omega(h_j x) = \frac{1}{2}\left[1 + sign(h_j x) sign(h_p h_j x)(\beta - \alpha)\right] \in \{\alpha, \beta\}$, for all $j \in [-q \dots p]$, $j \ne 0$, and *sign*() function is defined as in research[3,4].

3.2. Converting values between semantic and reference domains

To convert the values from the reference domain to the semantic domain of a variable X and vice versa, we synthesize some transformations as: Assume that [*a*, *b*] is a reference domain of the variable X, and $[a_s, b_s] \subseteq [0, 1]$ is the semantic domain. The conversion value *x* from [*a*, *b*] to $[a_s, b_s]$ is called *semantization*, denoted *S*(*x*) and the conversion value *y* from $[a_s, b_s]$ to [*a*,*b*] is called *desemantization*, denoted *D*(*y*).

For flexibility in semantization or desemantization, we add some parameters *sp, dp* ∈[-1, 1] then: $S(x) = f(x, sp)$, satisfy $0 \le f(x, sp) \le 1$, $f(x{=}a, sp) = 0$, $f(x{=}b, sp) = 1$. And, $D(y) = g(y, dp)$, satisfy $a \le g(y, dp) \le b$, $g(y = 0, dp) = a$, $g(y = 1, dp) = b$.

In this paper, we use the functions: $S(x) = f(x,sp) = (sp \times x(1-x) + x)/(b-a)$ and $D(y) = g(y, dp) = dp \times (f(y, sp) - a) \times (b - f(y, sp))/(b - a) + f(y, sp)$.

Figure 1 illustrates the hedge algebra AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathscr{H}$, ≤) with the hedge set $\mathscr{H}$ ={Very-V, More-M, Rather-R, Less-L} and the transfer of values from the semantic domain to the reference domain and vice versa.

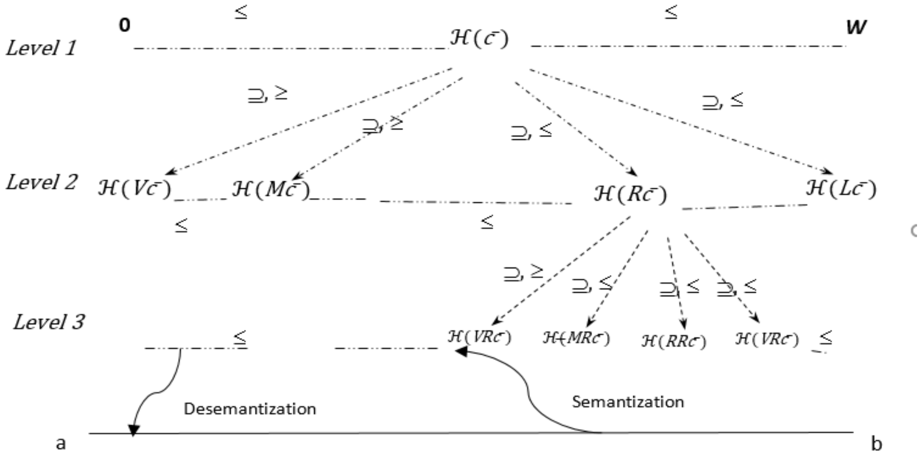

Figure 1. A graph representation of $\mathscr{H}(Z) \subseteq \mathscr{H}(c^-)$ and transform a value from [0, 1] to [a, b] and vice versa.

3.3. Hedge-Algebras-based Model (HAM) for time series forecasting

We consider each reference domain in the forecasting problem to correspond to a hedge algebra. Let PAR be a set of parameters, including the fuzzy measures of the hedges and the values sp and dp. Given that PAR has been determined, in this section, we present the fuzzy time series forecasting algorithm as follows.

Algorithm 1. HAM(*PAR*)

INPUT:

- *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.
- System of the parameters of hedge algebras and sp, dp for semantization and desemantization, denoted *PAR.*

OUTPUT: the forecasted value $F(t_i)$.

Step 1. Define the discourse $\boldsymbol{U}$

Put $\boldsymbol{U} = [D_{min}, D_{max}]$ *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 $\boldsymbol{U}$ by using the fuzziness model of hedge algebra.

Based on an algebra $AX = (\mathcal{X}, \mathbb{G}, \mathbb{C}, \mathscr{H}, \leq)$ *we divide* $\boldsymbol{U}$ *into k intervals* $u_1, u_2, \ldots, u_k$ *w.r.t level l* (*see Figure* 1). *The interval* u_i *is labeled* A_i, $i = 1, 2,\ldots, k$ *satisfying* $A_1 < A_2 < \ldots < A_k$. *We calculate* $f_{ui} = fm(A_i)\times(D_{max} - D_{min})$, $i = 1, 2, \ldots, k$. *So we have* $u_1 = [u_{1d}, u_{1c}] = [D_{min}, D_{min} + f_{u1}]$, $u_2 = [u_{2d}, u_{2c}] = [u_{1c}+1, u_{2d}+f_{u2}]$, $\ldots$, $u_k = [u_{kd}, u_{kc}] = [u_{(k-1)c}+1, u_{kd}+f_{uk}]$.

Step 3. Quantifying semantics of the linguistic values $A_1, A_2, \ldots, A_k$.

To quantify the semantic of $A_1, A_2, \ldots, A_k$, *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* $SA_1 < SA_2 < \ldots < SA_k$.

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 a relation between* A_i *and* A_j, *denoted* $A_i \rightarrow A_j$.

Step 5. Grouping relationships

If $A_i \rightarrow A_{j1}$, $A_i \rightarrow A_{j2}$,..., $A_i \rightarrow A_{jm}$, *then we establish the relation by grouping all of them into a unique relation* $A_i \rightarrow A_{j1}, A_{j2}, \ldots, A_{jm}$.

Step 6. Calculating output value

From a group of the relations in Step 5, applying the rules as in Section 2.2 we get the results of $F(t)$, *scilicet: If there is a relation* $A_i \rightarrow A_j$, *then* $F(j) = D(SA_j)$ *upon* $u_j = [u_{jd}, u_{jc}]$. *If* $A_i \rightarrow \varnothing$ *then* $F(j) = D(\varnothing)$ *upon* $u_i = [u_{id}, u_{ic}]$. *If* $A_i \rightarrow A_{j1}, A_{j2}, \ldots, A_{jk}$ *then* $F(j) = D(W_{i,j1}\times SA_{j1} + W_{i,j2}\times SA_{j2} + \ldots + W_{i,jk}\times SA_{jk})$ *upon interval* $[min\{u_{j1d}, u_{j2d}, \ldots, u_{jkd}\}, max\{u_{j1c}, u_{j2c}, \ldots, u_{jkc}\}]$ *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* $u_{j1}, u_{j2}, \ldots, u_{jk}$.

Step 7. Return the values $F(t_i)$, $i = 1,\ldots, n$.

4. The Optimal Hedge-Algebras-based Model

In Step 2 of the HAM model above**,** we assume that each point at a time will belong to a word in the hedge algebra AX = ($\mathcal{X}$, $\mathbb{G}$, $\mathbb{C}$, $\mathscr{H}$, ≤), $\mathbb{C}$ = {c^-, c^+}, $\mathscr{H}$ = {h_{-q}, …, h_{-1}, h_1,…h_p} with given parameters $\mu(h_i)$, $h_i \in \mathscr{H}$. Obviously, all parameters to be used in HAM contain $n = p+q+2$ parameters, which are $\mu(h_{-q})$, $\mu(h_{-q+1})$,… $\mu(h_{-1})$, $\mu(h_1)$,…, $\mu(h_p)$, *and sp, dp* for semantization and desemantization. So we can present that by the vector *PAR* = (x_1, x_2, …, x_n) where $x_1 = \mu(h_{-q})$, $x_2 = \mu(h_{-q+1})$,… $x_{n-2} = \mu(h_p)$, $x_{n-1} = sp$, $x_n = dp$. Vector *PAR* is also an artificial bee in the OHAM below.

To optimize the parameters, we choose the fitness function to be the Root Mean Square Error (RMSE), where a smaller value indicates better fitness. The root mean squared error can be expressed as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}$$

where y_t is the actual data point at time *t*, and $\hat{y}_t$ is the predicted value at time *t*.

The fitness function can be written:

Algorithm 2. Fitness(*PAR*)

INPUT: A system of parameters *PAR* represented for a bee; a real dataset $\{y_t\}_{t=1}^{n}$.

OUTPUT: Value of fitness of *PAR.*

Step 1. Generate language lattice of HA and quantify those values based on parameters in PAR.

Step 2. Calculate forecast values $\hat{y}_t$ ($t = 1,\dots, n$) by HAM(PAR).

Step 3. Set Err = 0.

Step 4. For each real value y_t and forecasted value $\hat{y}_t$ at *t* time, we put:

$$\text{Err} = \text{Err} + (y_t - \hat{y}_t)^2.$$

Step 5. RMSE = $\sqrt{\frac{Err}{n}}$.

Step 6. Return value RMSE.

The model is built as:

Algorithm 3. OHAM()

INPUT: *n* values of data {$y(t_1)$, $y(t_2)$,…, $y(t_n)$} with t_1, t_2, …, t_n are point times.

OUTPUT: the best system of parameters for solving optimization forecast problems.

Step 1. Initialization

Start by randomly initializing a population of artificial bees, where each bee represents a potential solution to the optimization problem. The population size is typically defined beforehand.

Step 2. Employed Bees' Phase

Each employed bee explores a new solution by adjusting its current position based on information shared with a randomly selected neighbor bee. The new solution is generated by modifying the position using specific search operators or strategies. After generating the new solution, the fitness of both the current and new solutions is evaluated.

Step 3. Onlooker Bees' Phase

Onlooker bees probabilistically choose a solution to explore based on the fitness values of employed bees. The better the fitness value, the higher the probability of being chosen. This phase allows good solutions to be shared among the population and improves the overall search process.

Step 4. Scout Bees' Phase

If an employed bee exhausts its exploration resources without finding a better solution, it becomes a scout bee. Scout bees generate a new random solution to diversify the search space and prevent the algorithm from getting stuck in local optima.

Step 5. Memorize the best solution (*BestPAR*) achieved so far.

Step 6. Termination

The algorithm will be stopped if a termination condition is satisfied. If not, go back to *Step* 2.

Step 7**.** Return *BestPAR.*

---

5. EXPERIMENTAL RESULTS

In this section, the proposed approach is applied to forecast the price of State Bank of India (SBI) shares at BSE India from April 2008 to March 2010, the enrollments at the University of Alabama from years 1971 to 1992, and the TAIEX Index of November and December 2004. The result will then be compared with different published methods. To measure the accuracy of the forecasting methods, the following metrics are used for comparison with RMSE.

For each test dataset, we used hedge algebra consisting of four hedge operators: Very, More, Possible, and Little, along with two parameters, *sp* and *dp*, for semantization and desemantization. The OHAM model was implemented using the ABC optimization algorithm with a maximum of 3000 iterations, and the number of employed and onlooker bees was set to 50. The optimal parameters obtained correspond to the experimental datasets: the SBI price, student enrollment at the University of Alabama, and the TAIEX stock index, as presented in Table 1.

Using the RMSE metric to evaluate forecasting performance, it is evident that the OHAM model produces less error than other models (see the last column of Tables 2-4). Visually, from Figures 2-4, the forecasted data curves generated by the proposed method follow the actual trend more closely compared to other models. Notably, at points with large amplitude variations, the OHAM model's predictions remain closer to the actual values, further demonstrating the high adaptability of the proposed model.

Table 1. The optimal parameters obtained by OHAM.

Parameters / Forecasting problems	*μ*(*Little*)	*μ*(*Possible*)	*μ*(*More*)	*μ*(*Very*)	*sp*	*dp*
SBI	0.316	0.286	0.204	0.194	0.467	-0.457
Enrollment	0.205	0.213	0.395	0.187	0.066	-0.167
TAIEX	0.194	0.239	0.149	0.418	0.113	-0.449

5.1. SBI prices Forecasting

Table 2. Results of the forecasting models for SBI data.

Months	Actual SBI Prices	Chen[10] (1996)	Huarng[12] (2001)	Pathak and Singh[13] (2011)	Joshi and Kumar[14] (2012)	Kumar and Gangwar[15] (2015)	Bisht and Kumar[6] (2016)	OHAM
April-08	1819.95	-	-	-	-	-	-	-
May-08	1840.00	1900	1855	1770.00	1777.80	1725.98	1877.657	1867.00
June-08	1496.70	1900	1855	1832.50	1865.71	1725.98	1877.657	1583.00
July-08	1567.50	1500	1575	1470.00	1531.50	1512.39	1466.360	1583.00
August-08	1638.90	1500	1505	1570.00	1531.50	1512.39	1466.360	1583.00
September-08	1618.00	1600	1610	1670.00	1777.80	1574.35	1533.504	1583.00
October-08	1569.90	1600	1610	1603.33	1531.50	1574.35	1533.504	1583.00
November-08	1375.00	1500	1505	1670.00	1531.50	1512.39	1466.360	1366.00
December-08	1325.00	1433	1482	1382.50	1504.23	1305.52	1520.652	1366.00
January-09	1376.40	1433	1365	1332.50	1504.23	1665.90	1520.652	1366.00
February-09	1205.90	1433	1482	1332.50	1504.23	1305.52	1520.652	1192.00
March-09	1132.25	1433	1155	1195.00	1258.23	1294.27	1144.718	1192.00
April-09	1355.00	1300	1365	1145.00	1258.23	1294.27	1322.446	1366.00
May-09	1891.00	1433	1482	1357.50	1504.23	1665.90	1520.652	1867.00
June-09	1935.00	1900	1890	1882.50	1865.71	2006.51	1877.657	1867.00
July-09	1840.00	1900	1890	1970.00	1883.93	2006.51	1895.491	1867.00
August-09	1886.90	1900	1855	1470.00	1865.71	1725.98	1877.657	1867.00
September-09	2235.00	1900	1855	1970.00	1865.71	2006.51	1877.657	2259.00
October-09	2500.00	2300	2485	2245.00	2142.04	2520.00	2311.382	2498.00
November-09	2394.00	2300	2415	2470.00	2245.65	2420.00	2374.204	2384.00
December-09	2374.75	2300	2345	2395.00	2191.75	2365.99	2352.723	2384.00
January-10	2315.25	2300	2205	2395.00	2191.75	2365.99	2352.723	2384.00

February-10	2059.95	2300	2205	2295.00	2142.04	2020.00	2311.382	2083.00
March-10	2120.05	2100	2135	2070.00	1883.93	2120.00	2166.247	2083.00
RMSE		**187.26**	**164.04**	**205.96**	**200.17**	**131.28**	**179.03**	**36.50**

Figure 2. Line chart of forecast method results for SBI data.

5.2. Enrollment student Forecasting

Table 3. Results of the forecasting models for enrollment student.

Actual data	Song and Chissom[9] (1993)	Chen[10] (1996)	Huarng[12] (2001)	Lee and Chou[16] (2004)	SC_time variant[17] (1994)	Cheng et al.[18] (2006)	Cheng et al.[19] (2008)	Yolcu et al.[20] (2009)	Qiu et al.[21] (2011)	Joshi and Kumar[14] (2012)	Kumar and Gangwar[15] (2015)	Bisht and Kumar[6] (2016)	OHAM
13055	-	-	-	-	-	-	-	-	-	-	-	-	-
13563	14000	14000	-	14025	-	14230	14242	14031.35	14195	14250	-	13595.67	13752
13867	14000	14000	-	14568	-	14230	14242	14795.36	14424	14246	13693	13814.75	13752
14696	14000	14000	14000	14568	-	14230	14242	14795.36	14593	14246	13693	14929.79	14753
15460	15500	15500	15500	15654	14700	15541	15474.3	14795.36	15589	15491	14867	15541.27	15341
15311	16000	16000	15500	15654	14800	15541	15474.3	16406.57	15645	15491	15287	15540.62	15341
15603	16000	16000	16000	15654	15400	15541	15474.3	16406.57	15634	15491	15376	15540.62	15341
15861	16000	16000	16000	15654	15500	16196	15474.3	16406.57	16100	16345	15376	15540.62	16040
16807	16000	16000	16000	16197	15500	16196	16146.5	16406.57	16188	16345	15376	16254.5	16879
16919	16813	16833	17500	17283	16800	16196	16988.3	17315.29	17077	15850	16523	17040.41	16879
16388	16813	16833	16000	17283	16200	17507	16988.3	17315.29	17105	15850	16066	17040.41	16040
15433	16789	16833	16000	16197	16400	16196	16146.5	17315.29	16369	15850	17519	16254.5	15341
15497	16000	16000	16000	15654	16800	15541	15474.3	16406.57	15643	15450	16606	15540.62	15341
15145	16000	16000	15500	15654	16400	15541	15474.3	16406.57	15648	15450	15376	15540.62	15341
15163	16000	16000	16000	15654	15500	15541	15474.3	16406.57	15622	15491	15376	15541.27	15341
15984	16000	16000	16000	15654	15500	15541	15474.3	16406.57	15623	15491	15376	15541.27	16040
16859	16000	16000	16000	16197	15500	16196	16146.5	16406.57	16231	16345	15287	16254.5	16879

18150	16813	16833	17500	17283	16800	17507	16988.3	17315.29	17090	17950	16523	17040.41	18283
18970	19000	19000	19000	18369	19300	18872	19144	19132.79	18325	18961	17519	18902.3	19291
19328	19000	19000	19000	19454	17800	18872	19144	19132.79	19000	18961	19500	19357.3	19291
19337	19000	19000	19000	19454	19300	18872	19144	19132.79	19000	18961	19000	19168.56	19291
18876	-	19000	19000	-	19600	18872	19144	19132.79	19000	18961	19500	19168.56	19291
RMSE	**650.40**	**880.73**	**638.36**	**476.97**	**501.28**	**511.04**	**478.45**	**805.17**	**511.33**	**433.76**	**493.56**	**428.63**	**178.21**

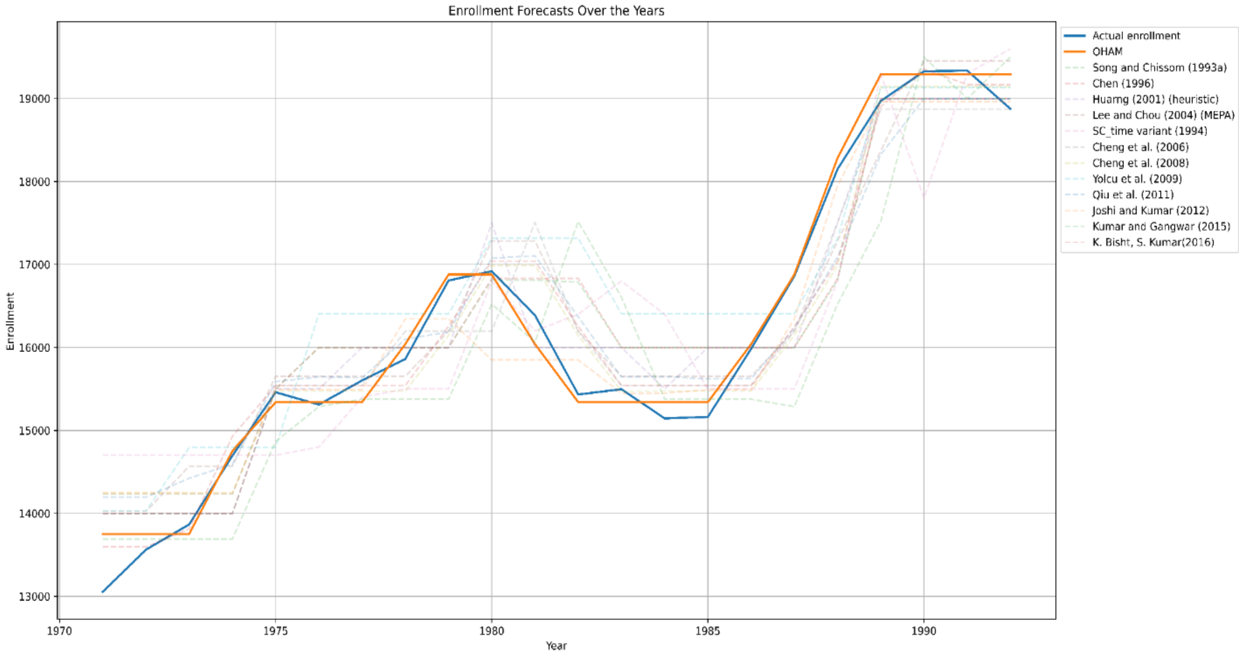

Figure 3. Line chart of forecast method results for enrollment student data.

5.3. TAIEX index Forecasting

Table 4. Results of the forecasting models for TAIEX index.

Date	**Actual Index**	**Chen' Forecasted Index[10]**	**Loc' Forecasted Index[7] (a)**	**Loc' Forecasted Index[8] (b)**	**OHAM**
2/11/2004	5759.61	5674.81	5743.00		5768.00
3/11/2004	5862.85	5768.14	5852.00	5886.00	5863.00
4/11/2004	5860.73	5854.81	5876.04	5886.00	5863.00
5/11/2004	5931.31	5908.26	5876.04	5934.00	5942.00
8/11/2004	5937.46	5934.81	5912.05	5934.00	5942.00
9/11/2004	5945.2	5943.81	5912.05	5934.00	5942.00
10/11/2004	5948.49	5934.81	5912.05	5978.00	5942.00
11/11/2004	5874.52	5937.12	5912.05	5886.00	5863.00
12/11/2004	5917.16	5908.26	5919.27	5934.00	5903.00
15/11/2004	5906.69	5934.81	5919.27	5934.00	5903.00
16/12/2004	5910.85	5934.81	5919.27	5934.00	5903.00
17/11/2004	6028.68	5937.12	5919.27	5978.00	6038.00

18/11/2004	6049.49	6068.14	5979.18	5978.00	6038.00
19/11/2004	6026.55	6068.14	5979.18	5978.00	6038.00
22/11/2004	5838.42	5976.47	5979.18	5886.00	5833.00
23/11/2004	5851.10	5854.81	5876.04	5886.00	5833.00
24/11/2004	5911.31	5934.85	5876.04	5934.00	5903.00
25/11/2004	5855.24	5934.81	5919.27	5886.00	5863.00
26/11/2004	5778.65	5854.81	5876.04	5768.00	5768.00
29/11/2004	5785.26	5762.12	5797.89	5768.00	5768.00
30/11/2004	5844.76	5762.12	5852.00	5886.00	5833.00
1/12/2004	5798.62	5834.85	5876.04	5768.00	5768.00
2/12/2004	5867.95	5803.26	5797.89	5886.00	5863.00
3/12/2004	5893.27	5854.81	5876.04	5886.00	5903.00
6/12/2004	5919.17	5854.81	5919.27	5934.00	5903.00
7/12/2004	5925.28	5937.12	5912.05	5934.00	5942.00
8/12/2004	5892.51	5876.47	5912.05	5886.00	5903.00
9/12/2004	5913.97	5854.81	5919.27	5934.00	5903.00
10/12/2004	5911.63	5934.81	5919.27	5934.00	5903.00
13/12/2004	5878.89	5937.12	5919.27	5863.00	5863.00
14/12/2004	5909.65	5854.81	5919.27	5903.00	5903.00
15/12/2004	6002.58	5934.81	5919.27	5994.00	5994.00
16/12/2004	6019.23	6068.14	5979.18	6038.00	6038.00
17/12/2004	6009.32	6062.12	5979.18	5994.00	5994.00
20.12.2004	5985.94	6062.12	5979.18	5994.00	5994.00
21/12/2004	5987.85	5937.12	5979.18	5994.00	5994.00
22/12/2004	6001.52	5934.81	5979.18	5994.00	5994.00
23/12/2004	5997.67	6068.14	5979.18	5994.00	5994.00
24/12/2004	6019.42	5934.81	5979.18	6038.00	6038.00
27/12/2004	5985.94	6068.14	5979.18	5994.00	5994.00
28/12/2004	6000.57	5937.12	5979.18	5994.00	5994.00
29/12/2004	6088.49	6068.14	5979.18	6125.00	6125.00
30/12/2004	6100.86	6062.12	6119.36	6125.00	6125.00
31/12/2004	6139.69	6137.12	6143.57	6125.00	6125.00
RMSE		**56.86**	**48.02**	**26.88**	**12.73**

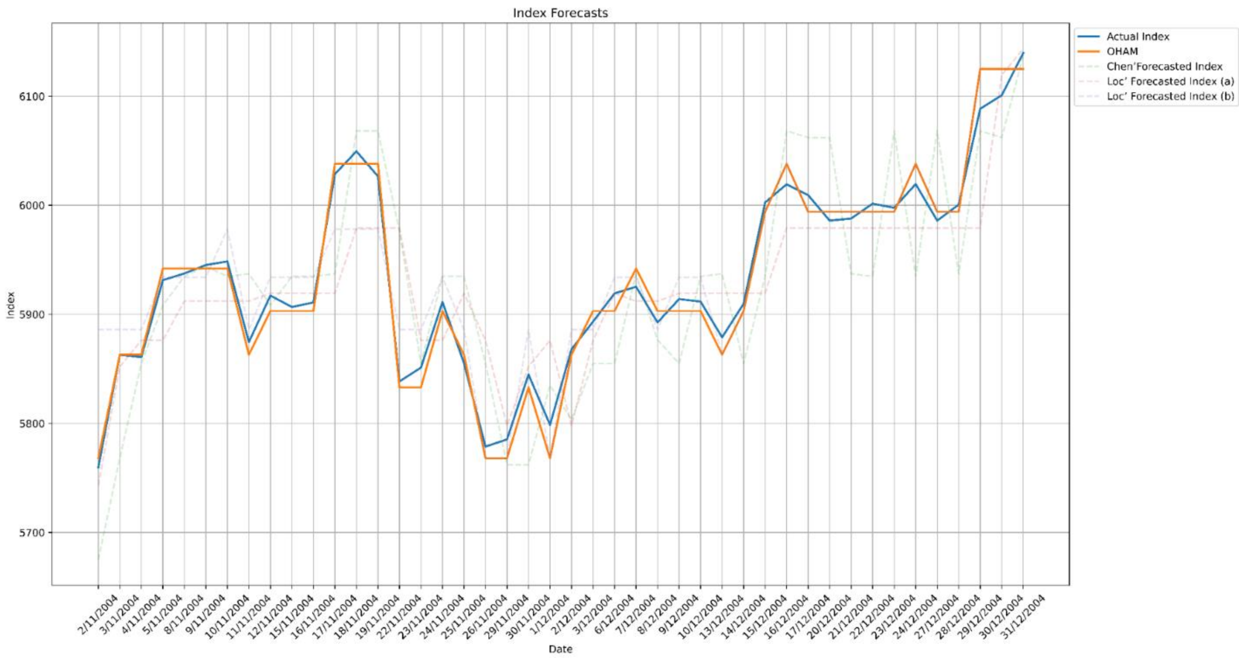

Figure 4. Line chart of forecast method results for TAIEX index data.

6. CONCLUSIONS

In this study, we propose a new fuzzy time series forecasting method using hedge algebra. We also introduce a segmentation approach for the reference space based on *k-level* and the fuzziness measure of linguistic terms of hedge algebra.

The effectiveness of this fuzzy time series forecasting method is demonstrated by applying it to the benchmark problem of forecasting the enrollment numbers at the University of Alabama. The relatively small RMSE value indicates that the proposed model outperforms other methods. Moreover, financial time series exhibit intrinsic characteristics such as relatively high volatility and frequent fluctuations in individual time series data over time, making forecasting more challenging compared to other types of time series data. Even well-established time series forecasting methods tend to produce high forecasting errors. However, the proposed OHAM model proves to be highly suitable and effective for forecasting financial time series, where nonlinearity, intrinsic characteristics, and fuzziness complicate the forecasting process.

REFERENCES

1. 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.
2. N. C. Ho, W. Wechler. Extended hedge algebras and their application to Fuzzy logic, *Fuzzy Sets and Systems*, **1992**, *52*, 259-281.
3. N. C. Ho, N. V. Long, Fuzziness measure on complete hedge algebras and quantifying semantics of terms in linear hedge algebras, *Fuzzy Sets and Systems*, **2007,** 158 (4), 452–471.
4. 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.
5. N. C. Ho, W. Pedrycz, D. T. Long, T. T. Son. A genetic design of linguistic terms for fuzzy rule based classifiers, *International Journal of Approximate Reasoning*, **2012**, *54*, 1-21.
6. K. Bisht, S. Kumar. Fuzzy time series forecasting method based on hesitant fuzzy sets, *Expert Systems with Applications*, **2016**, *64*, 557-568.
7. V. M. Loc, V. H. Dung, P. T. Yen. The Context-Aware Calculating Method in Language Environment Based on Hedge Algebras Approach to Improve Result of Forecasting Time Series, *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, (LNICST 217 Springer), **2017**, *217*, 110-123.
8. V. M. Loc, V. H. Dung. Hedge Algebra Approach for Fuzzy Time series To Improve Result Of Time Series Forecasting, *EAI Endorsed Transactions on Context-aware Systems and Applications*, **2018**, *4*, 1-11.
9. Q. Song, B. S. Chissom. Fuzzy time series and its models, *Fuzzy Sets and Systems*, **1993**, *54*, 269–277.
10. S. M. Chen. Forecasting Enrollments Based on Fuzzy Time Series, *Fuzzy Sets and Systems*, **1996**, *81*, 311–319.

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

12. K. Huarng. Effective length of intervals to improve forecasting in fuzzy time series, *Fuzzy Sets and Systems*, **2001**, *123*, 387–394.

13. H. K. Pathak, P. Singh. A new bandwidth interval based forecasting method for enrollments using fuzzy time series, *Applied Mathematics*, **2011**, *2*, 504 507.

14. B. P. Joshi, S. Kumar. Intuitionistic fuzzy sets based method for fuzzy time series forecasting, *Cybernetics and Systems*, **2012**, *43*, 34–47.

15. S. Kumar, S. Gangwar. Intuitionistic fuzzy time series: An approach for handling non-determinism in time series forecasting, *IEEE Transactions on Fuzzy Systems*, **2015**, doi: 10.1109/TFUZZ.2015.2507582.

16. H. S. Lee, M. T. Chou. Fuzzy forecasting based on fuzzy time series, *International Journal of Computer Mathematics*, **2004**, *81*, 781–789.

17. Q. Song, B. S. Chissom. Forecasting enrollments with fuzzy time series, Part II, *Fuzzy Sets and Systems*, **1994**, *64*, 1–8.

18. C. Cheng, J. Chang, C. Yeh. Entropy-based and trapezoid fuzzification based fuzzy time series approaches for forecasting IT project cost, *Technological Forecasting and Social Change*, **2006**, *73*, 524–542.

19. C. H. Cheng, G. W. Cheng, J. W. Wang. Multi-attribute fuzzy time series method based on fuzzy clustering, *Expert Systems with Applications*, **2008**, *34*, 1235–1242.

20. U. Yolcu, E. Egrioglu, V. R. Uslu, M. A. Basaran, C. H. Aladag. A new approach for determining the length of intervals for fuzzy time series, *Applied Soft Computing*, **2009**, *9*, 647–651.

21. W. Qiu, X. Liu, H. Li. A generalized method for forecasting based on fuzzy time series, *Expert Systems with Applications*, **2011**, *38*, 10446–10453.