2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

An Improvement in Forecasting Interval based Fuzzy Time Series

Shanoli Samui Pal, Tandra Pal, Samarjit Kar

***Abstract*— In this paper, we have proposed a fuzzy interval time series model using a new strategy to replace the conventional defuzzification step, where genetic algorithm has been used to optimize the interval parameters and neural network has been used to learn the trend of the time series. First order fuzzy time series with equal time interval has been used on two data sets, enrollments of the University of Alabama and gold exchange traded fund. We compare the proposed model with two other existing models. The results of the comparisons show that the proposed model performs better.**

Keywords: Fuzzy time series, genetic algorithm, Halton sequence, neural network.

I. Introduction

Several time series forecasting models are used for forecasting the future based on historical observations. Time series data may contain uncertainty due to the error that may occur during data collection process. Fuzzy logic is used to manage this uncertainty. Song and Chissom [1], [2] proposed the concept of fuzzy time series. They defined fuzzy time series as well as the time-variant and time-invariant models. After that, various fuzzy time series models have been proposed by many researchers. These models have been extensively used in various domains of applications, such as university enrollment, exchange rates, stock market, temperature prediction, etc.

The model, proposed by Song and Chissom, considered four basic steps for forecasting: (1) partitioning the universe of discourse, (2) defining fuzzy sets and fuzzifying the crisp time series, (3) establishing fuzzy logical relationship, and (4) defuzzifying the output before forecasting. Many researchers have tried to improve one or more of these steps, which has either improved the accuracy or reduced the computational overhead. Most of them tried to improve Step (1), Step (3), and Step (4). Many people worked on deciding the length of interval, equal or unequal, weighted or unweighted mostly for better accuracy. Various fuzzy membership functions have been used for fuzzification of the time series data. Chen [5] proposed an efficient procedure that groups fuzzy logical relationships into rules to reduce the computational overhead. Yu [6] proposed the weighted fuzzy logical relationships.

Lee et al. [7] modified the weighted fuzzy time series forecasting. For most of these approaches, defuzzifying and forecasting the output are almost same as Song and Chissom's model [1], [2].

Chen and Chung [3] have used real coded genetic algorithm (GA) for adjusting the length of intervals which results unequal interval length. They have used higher-order fuzzy logical relationships.

In this paper, equal length intervals and first-order fuzzy logical relationships have been considered. The proposed method does not need to use any particular fuzzy membership function. It looks for the interval in which the time series values lie and establish the first-order fuzzy logical relationship. We do not use the mid-points of intervals at the time of defuzzification. Rather, we adjust mid-points according to the actual values in the respective intervals using real coded GA. We have used Halton Sequence [4] to initialize the population of GA.

For forecasting, one need to learn the trend of the time series data. In this paper, multilayer perceptron (MLP) with feed forward error back propagation (EBP) learning has been used for this purpose. First order differences of the mid-points of the intervals in which each data point lies are used for learning the trend. Once learnt, it gives the future mid-points, which are able to forecast.

The remaining of this paper is organized as follows. We discuss the preliminaries in Section II and the proposed model has been described in Section III. Experimentation, results, and discussion are presented in Section IV. We conclude in Section V.

II. Preliminaries

In this section, we discuss basics of fuzzy time series, fuzzy relationship, genetic algorithm (GA), and neural network (NN).

1) Fuzzy Set and Its Membership Function: A fuzzy set A is defined in the universe of discourse $U = \{u_1, u_2, u_3, ..., u_n\}$ as follows:

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + ... + f_A(u_n)/u_n, \quad (1)$$

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

2) Fuzzy relationship: Assume that there is a fuzzy relationship $R(t, t-1)$ between $F(t)$ and $F(t-1)$, such that $F(t) = F(t-1)oR(t, t-1)$, where "o" is a composition operator. Then we can say, $F(t)$ is caused by $F(t-1)$ and it is denoted by a fuzzy logical relationship $F(t-1) \to F(t)$, where both $F(t)$ and $F(t-1)$ are fuzzy sets. This relation R is called first-order model of $F(t)$.
Similarly, if $F(t)$ is caused by previous n-terms, i.e., $F(t-1)$,

Department of Mathematics, National Institute of Technology, Durgapur, (email: shanoli.pal@gmail.com), Contact no. +919830877042
Department of Computer Science, National Institute of Technology, Durgapur, (email: tandra.pal@gmail.com)
Department of Mathematics, National Institute of Technology, Durgapur, (email: kar_s_k@yahoo.com).

$F(t-2)$,..., and $F(t-n)$, then the fuzzy logical relationship can be denoted as follows:

$$F(t-n), ..., F(t-2), F(t-1) \rightarrow F(t) \tag{2}$$

It is called the n^{th} order model, where $n \geq 2$.

Let $F(t-1) = A_i$ and $F(t) = A_j$. The relationship between two consecutive observations, $F(t-1)$ and $F(t)$, is referred as a fuzzy logical relationship and is denoted by $A_i \rightarrow A_j$.

A. Genetic Algorithm (GA)

GA is an optimization search procedure. It follows the natural genetics and Darwin's principle of survival of the fittest. In GA, a point in the search space which represents a solution is called *chromosome* and a set of chromosomes is called *population*. Initial population generation for the evolutionary algorithms is another important area of research. Researcher are trying to improve this area by using some low-discrepancy sequence generator, some optimization algorithm, etc.

In GA, an *objective function* (or *fitness function*) is used to compute the *fitness* of a chromosome. This function assigns a value or a set of values to each chromosome. Fitness value is used in selection. In time series forecasting, the error between the original and forecasted values is considered as the objective function. GA tries to minimize the fitness value to find the best chromosome. Error metric, like - mean square error, absolute error, root mean square error, average error, etc., can be used as convergence criteria. Sometimes the maximum number of generations are used as termination criteria.

There are three basic operations in GA. Here, we briefly describe them.

1) *Selection:* GA selects the chromosomes as parents from current population according to their fitness values for mating (crossover) to produce chromosomes (as children) for the population of next generation. More fitted chromosome is selected more than a less fitted chromosome.
2) *Crossover:* There are several variations of the crossover operator. The most common trend is to select two chromosomes as two parents. These parents then exchange some of their genes two generate two off springs. The amount of exchange of gene is determined by a parameter, called *probability of crossover*.
3) *Mutation:* Several mutation operators are present in the literature. Yet, a common trend is to modify each gene of a chromosome with a low probability, called *probability of mutation*.

B. Neural Network (NN)

This soft-computing tool mimics human brain's ability of adaption, i.e., the act of fitting itself in evolving environment. Neural Networks are widely used in many applications and also in the field of time series forecasting. It is a good framework for modeling nonlinear problems. It works better with large sample points. Here, we have used MLP with EBP learning, which is supervised. Training data set is the mid-points of fuzzy intervals and simple feed forward network is used to get the further mid-points. One input layer with five nodes, one hidden layer with four nodes, and one output layer with one node are used in our approach.

Error Back Propagation (EBP) Learning Algorithm: EBP is a supervised learning process. It learns from error corrections which are propagated backward from output layer to input layer and the weights between output layer and hidden layer as well as between hidden layer and input layer are updated accordingly. EBP requires differentiable activation functions. We have used sigmoid activation function.

Steps for EBP learning are as follows.

Step 1. Initialize the weights to small random numbers.
Step 2. Present an input and desired output.
Step 3. Propagate the input forward through the network to produce actual output.
Step 4. Calculate the error and back propagate it to the weights between output and hidden layers and then between hidden and input layers as follows.

$$\Delta w_{kj} = \alpha \delta_k o_j, \tag{3}$$

$$\Delta w_{ji} = \alpha \delta_j o_i, \tag{4}$$

where α is the learning rate,

$$\delta_k = (t_k - o_k) * g'_k(h_k) \tag{5}$$

for the output layer,

$$\delta_j = g'_j(h_j) * \sum_k \delta_k w_{kj} \tag{6}$$

for the hidden layer.
Here, g' is the derivative of the activation function g, h_k and h_j are respectively the net input to the node k and j of the output and hidden layer.
Step 5. Repeat by going to Step 2 until a maximum number of iteration or error in output layer is below a pre-specified threshold.

III. Proposed Approach with GA and NN on Fuzzy Time Series

We have used GA and NN on first order fuzzy time series to improve the forecasting results. Steps of our approach are given below:

Step 1: Decide the universe of discourse. For example, minimum and maximum values of the enrollment data of the University of Alabama are respectively 5,127 and 27,052. So, we take the domain of it, i.e., $U =$ [5,100 27,100].

Step 2: Decide the length of the interval. If it is l, then dividing whole length of U by l, we get the number of intervals. For enrollment data, length of U, i.e., 22,000, is divided into 220 intervals, each of interval length being 100. We have used the first order fuzzy logical relationship between fuzzy sets. We have applied our method on $(n-t)$ data points where $t = 1$ for first order fuzzy relationship.

Step 3: Calculate the mid-points of the intervals in which actual data point lies. There may be intervals having more than one data ponits as well as intervals having no data points.

Step 4: Apply genetic algorithm with $(n-t)$ data points and their respective mid-points. Basically, we have tried to fit each mid-point linearly, i.e., $y = mx + c$, where m and c are optimized by GA. Here, x is the mid-point and y is the forecasted value for each of the $(n-t)$ data points.

a. Initialization of the population: Halton sequence has been used to initialize the population. It is a quasi-random or low discrepancy sequence. It is less random than pseudo-random number sequence and covers the whole region more uniformly. It generates the random number in the interval [0 1], which fills the search region better. A base of prime numbers 2, 3, 5, 7,... is associated with each dimension. If we have the index I of the required Halton number, then the method for producing the sequence $H(I)$ with base 2 is, (i) write each I in base 2, (ii) reverse the digits including the decimal sign, and (iii) convert it again in base 10. This is same for bases 3, 5, 7,...,etc. We have considered the base 3 for m and 2 for c.

b. Fitness function: After initialization, we can calculate each y_i, where $n-t \leq i \leq n$. Here, the fitness function is the absolute value of minimum distance of y_i from the original data point. The chromosome corresponding to minimum distance is the most fitted chromosome for that population. For each y_i, we have generated $(n-t)$ set of values of m_is and c_is. Finally, we select those m_is and c_is for which y_i has maximum fitness.

c. Crossover: It randomly selects a number and if it is less than crossover rate, then two chromosomes say, $x1$, $x2$ are selected randomly for crossover and are modified accordingly as $\alpha x1 + (1-\alpha)x2$ and $\alpha x2 + (1-\alpha)x1$ respectively, where α is selected randomly.

d. Mutation: One chromosome is selected randomly and it is modified adding two random numbers with the values of m and c respectively.
Repeat from Step b to Step d until the optimal solution is produced.

Step 5: GA gives the forecasted values corresponding to $(n-t)$ actual values. Now, we have the original and forecasted values. The Pearson's correlation coefficient (R), coefficient of variation (CV), normalized mean square error (NMSE), and mean square error (MSE) are evaluated. They are defined in the appendix.

Step 6: (For prediction further) In this step, EBP in MLP has been used. We already have the mid-points of each interval. We calculate the differences between the successive mid-points. If there are p mid-points m_1, m_2, m_3,..., and m_p, then there are $(p-1)$ first differences, which are $m_2 - m_1$, $m_3 - m_2$,..., and $m_p - m_{p-1}$. If some of the differences are negative, then find the minimum (must be negative) and it's absolute value, say k, add it with all the first differences including itself and then normalize them within [0 1] dividing by the maximum value of the first differences, say d. We have such p =61 points in the example of enrollment data set. In our approach $(p-1)$ is 60 after first differencing. Then we generate the input data, for training, from this 60 first differences, say d_1, d_2,..., d_{60}. We have considered 5 successive differences, say, d_i, d_{i+1},..., d_{i+4}, as an input to the NN and d_{i+5} is the corresponding expected value of output. So, the NN is trained by $(p-1-t) = 55$ data points where $t = 5$ (number of inputs to the NN). The training is terminated when output error is less than 0.0005. Then, calculate the y-value as $(y \times d + (p-1-t)$th mid-value - $k)$. After completion of training, we have a set of weights in between *Input Layer- Hidden Layer* and *Hidden Layer- Output Layer*. We can feed forward the network with those weights and get further values from where we can say in which interval those values lie and get their respective mid-values. Using those mid-values and most fitted values of m_i and c_i from **Step 4**, we can get the further forecasted values.

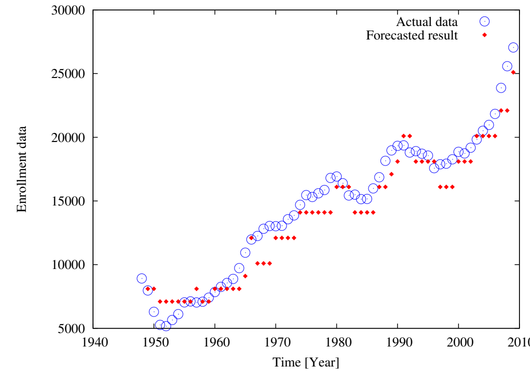

Fig. 1. Result of Song and Chissom's model [1], [2] for enrollment data

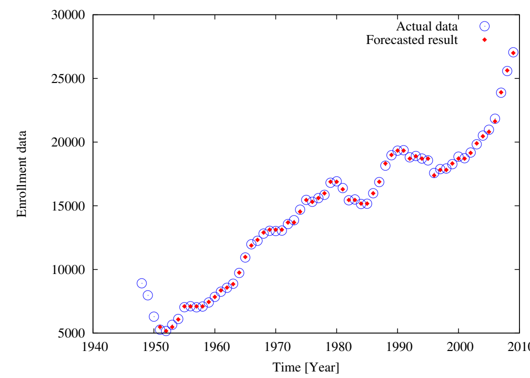

Fig. 2. Result of Chen and Chung's model [3] for enrollment data

IV. Result and Discussion

Two data sets are used in this paper. First one is the yearly enrollment data of the University of Alabama for the years 1948-2009 [10]. Second one is daily data on Gold ETF of India.

Gold ETF is Gold exchange-traded product that aim to track the price of gold. Gold exchange-traded products

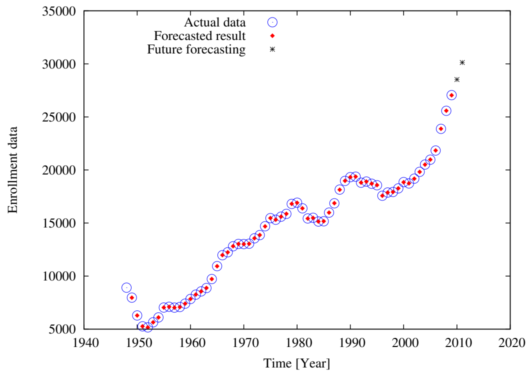

Fig. 3. Result of proposed approach for enrollment data

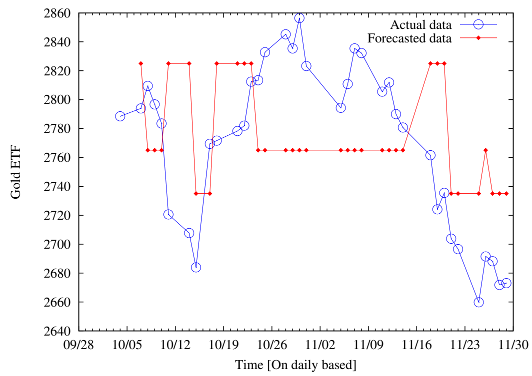

Fig. 4. Result of Song and Chissom's model [1], [2] for Gold ETF data

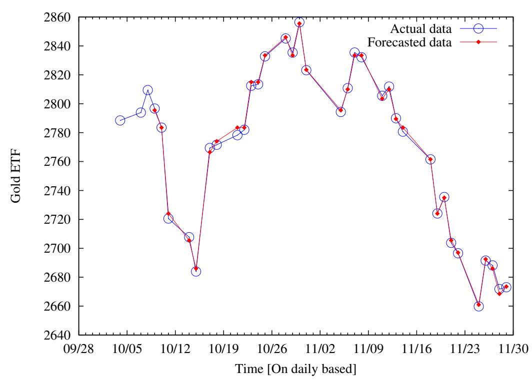

Fig. 5. Result of Chen and Chung's model [3] for Gold ETF data

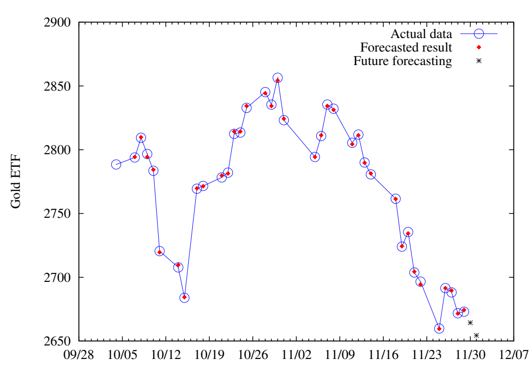

Fig. 6. Result of proposed approach for Gold ETF data

	Song and Chissom	Chen and Chung	Proposed approach
R-value	0.976933	0.999859	0.999988
CV	0.097914	0.006047	0.001890
NMSE	0.072130	0.000295	0.000027
MSE	2040502.75	7789.34	760.61

TABLE I: Values of R-value, CV, NMSE, and MSE for three models of enrollment data set

are traded on the major stock exchanges including Zurich, Mumbai, London, Paris, and New York. The idea of Gold ETF was first introduced in India by Benchmark Asset Management Company Private Ltd., when they proposed it with the Securities and Exchange Board of India (SEBI) in May, 2002. But it was fully approved in March, 2007. In India, people can gain additional benefit by investing through mutual funds tax savings scheme. Gold is a symbol of wealth and good fortune in India. In recent years, many gold based financial instruments have been introduced including Gold ETF. Gold ETF can be easily sold in secondary market, need no storage and security, has a transparent pricing mechanism, assures purity of Gold as per SEBI regulations, the purity of gold in Gold ETF should be 0.995 fineness [9]. Still there are some drawbacks in this kind of mutual funds. So, we need to study the nature of this kind of time series. In this study, we try to forecast the time series data on SBI mutual fund Gold ETF data collected daily from Oct, 2013 to Nov, 2013.

Fuzzy relation for Song and Chissom's model [1], [2] is the max-min composition operation which will take more computation time with the increase of the number of fuzzy sets and the size of the data set. In this paper, $U = [5{,}100\ 27{,}100]$, number of intervals is 11 for enrollment data set and $U = [2{,}600\ 2{,}900]$, number of intervals is 10 for Gold ETF data set.

In our proposed model, the same universe of discourse has been considered for both enrollment data and Gold ETF data as mentioned in case of Song and Chissom's model [1], [2].

Chen and Chung's model [3] gives better result when number of intervals and order of the fuzzy relationship are increased. The R-value, CV, NMSE, MSE are given in Table I and Table II respectively considering 100 intervals for enrollment data and 50 for Gold ETF data for their model. Third order fuzzy relationship is considered for both the cases.

In our proposed model, number of intervals is 220 for enrollment data and 30 for Gold ETF data. Number of chromosomes in a population is 1800, number of generations is 10, crossover rate is 0.05, and mutation rate is 0.02 for our proposed model.

Figure 1 and Figure 4 show the results of Song and Chissom's model for enrollment data and Gold ETF data respectively. Figure 2 and Figure 5 show the results of Chen and Chung's model for enrollment data and Gold ETF

	Song and Chissom	Chen and Chung	Proposed approach
R-value	0.177918	0.999735	0.999885
CV	0.021932	0.000695	0.000467
NMSE	1.14	0.001108	0.000518
MSE	3688.08	3.70	1.67

TABLE II: Values R-value, CV, NMSE, and MSE for three models on Gold ETF data set

data respectively. Figure 3 and Figure 6 show the results of proposed approach for enrollment data and Gold ETF data respectively.

A comparison on R, CV, NMSE, and MSE values of the three models on two data sets are given in Table I and Table II respectively. It is observed that our approach is better compared to other two models.

V. Conclusion

We have proposed a new model for fuzzy time series using GA and NN. Our approach gives better result when the number of intervals is not too large or not too small. In Chen and Chung's work, GA is used to decide the intervals directly, whereas, we have applied it for forecasting to replace the defuzzification step. In future, these two models can be combined and GA can also be used for learning using different operators.

Appendix

If A_i is the i-th actual value, F_i is the i-th forecasted value, x and σ^2 are the mean and variance of the actual data set respectively and m denotes the number of historical data, then Pearson's correlation coefficient (R), coefficient of variation (CV), normalized mean square error (NMSE), and mean square error (MSE) are described below. R, CV, and NMSE has been used in [8].

A. Pearson's Correlation Coefficient (R)

It measures the sufficiency of the proposed model with the data in the range from -1 to +1. A positive value means a positive linear correlation and a negative value means negative linear correlation. Value of R near zero means no correlation between actual data and adjusted model. It can be calculated as follows.

$$R = \frac{m\sum_{i=1}^{m}(\mathrm{A}_i\mathrm{F}_i) - \left(\sum_{i=1}^{m}\mathrm{A}_i\right).\left(\sum_{i=1}^{m}\mathrm{F}_i\right)}{\sqrt{\left[m\sum_{i=1}^{m}\mathrm{A}_i^2 - \left(\sum_{i=1}^{m}\mathrm{A}_i\right)^2\right]\left[m\sum_{i=1}^{m}\mathrm{F}_i^2 - \left(\sum_{i=1}^{m}\mathrm{F}_i\right)^2\right]}} \quad (7)$$

B. Coefficient of Variation (CV)

It is a metric which shows the relative scatter in data with respect to the mean. If the value of CV is smaller, then the corresponding model is better. It is evaluated by (8), given below. It can be calculated as follows.

$$CV = \frac{1}{x}\sqrt{\left[\frac{1}{m}\sum_{i=1}^{m}(\mathrm{A}_i - \mathrm{F}_i)^2\right]} \quad (8)$$

C. Normalized Mean Square Error (NMSE)

It is used to compare the mean of the actual time series data values with the forecasted values. If the NMSE is greater than unity, then the forecasting is worse than the actual time series mean. If NMSE is less than unity, then the forecasting is better than the actual time series mean. It can be calculated as follows.

$$NMSE = \frac{1}{\sigma^2}\left[\frac{1}{m}\sum_{i=1}^{m}(\mathrm{A}_i - \mathrm{F}_i)^2\right] \quad (9)$$

D. Mean Square Error (MSE)

It measures the average of the squares of the errors where error is the difference between forecasted and actual value.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\mathrm{F}_i - \mathrm{A}_i)^2 \quad (10)$$

References

[1] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series, Part I," Fuzzy Sets and Systems, vol. 54, pp. 1-9, 1993.

[2] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series, Part II," Fuzzy Sets and Systems, vol. 62, pp. 1-8, 1994.

[3] S. M. Chen and N. Y. Chung, "Forecasting enrollments using high-order fuzzy time series and Genetic Algorithms," International Journal of Intelligent Systems, vol. 21, pp. 485-501, 2006.

[4] J. Halton, "Algorithm 247: Radical-inverse quasi-random point sequence," Association for Computing Machinery (ACM), vol. 7, pp. 701-702, 1964.

[5] S. M. Chen, "Forecasting enrollments based on fuzzy time series," Fuzzy Sets and Systems, vol. 81, pp. 311-319, 1996.

[6] H. K. Yu, "Weighted fuzzy time series models for Taiex forecasting," Physica A, vol. 349, pp. 609-624, 2005.

[7] M. H. Lee, R. Efendi, and Z. Ismail, "Modified Weighted for Enrollment Forecasting Based on Fuzzy Time Series," Matematika, vol. 25, pp. 67-78, 2009.

[8] S. L. Heitor and R. W. Wagner, "A gene expression programming system for time series modeling," In: Proceedings of XXV Iberian Latin American Congress on Computational Methods in Engineering (CILAMCE), Recife (Brazil), November 10-12, 2004.

[9] P. Athma and Suchitra K., "Gold ETFs: An emerging investment option," Asia Pacific Journal of Research in Business Management (APJRBM), vol.2, pp. 66-78, 2011.

[10] H. T. Jasim, A. G. J. Salim, and K. I. Ibraheem, "A novel algorithm to forecast enrollment based on fuzzy time series," Applications and Applied Mathematics (AAM), vol. 7, pp. 385-397, 2012.