Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (16): 3335-3339, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.3335.3339

A Fuzzy Time Series Model Based on Genetic Discretization Approach

Jing-Rong Chang and Chung-Chi Liu
Department of Information Management, Chaoyang University of Technology,
Taichung, Taiwan

Abstract: There are many uncertainty problems in the Human society, such as the forecasting of economic growth rate, financial crisis, etc. Since Song and Chissom (1993) proposed the concept of fuzzy time series, many scholars have proposed different models to deal with these problems. However, previous studies usually do not consider the transfer original data to the fuzzy linguistic value by the subjective opinions in fuzzy process, which cannot objectively show the characteristics of the data. Based on above concepts, the purpose of this study is to explore ways of determining the objective lengths of intervals in fuzzy time series. This study proposed a high-order weighted fuzzy time series model based on Genetic Discretization Approach (GDA). In order to verify the proposed method, the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) from the (http://www.twse.com.tw) are used in the experiment and the experiment results are compared with other methods in with this study. The forecasting performance shows that the proposed method having better forecasting ability.

Key words: Fuzzy linguistic, fuzzy time series, high-order, genetic discretization approach, Taiwan stock exchange capitalization weighted stock index

INTRODUCTION

Forecasting is not only frequent activities but also an important role in our life. The traditional time series methods can predict sequential problems but fail to deal with the problem with linguistic historical data. Song and Chissom (1993) proposed the theory of fuzzy time series to solve the restrictions of the traditional time series methods. Recently, many researchers presented their fuzzy time series methods to deal with the forecasting problems. They usually partitioned the universe of discourse into several intervals with equal length and utilized the triangular fuzzy number to fuzzify historical data. However, there is an issue of these methods is that they do not consider assigning weight to fuzzy relations and characteristics of observations in the interval. Partitioning the universe of discourse and determining effective intervals are critical for forecasting in fuzzy time series. Equal length intervals used in most existing literatures are convenient but subjective to partition the universe of discourse. In order to solve above problems, this study proposes a new method to determine the length of intervals according to the Genetic Discretization Approach (GDA). It's calculates the cut-points by Chen and Chung (2006). Besides, the concept of determining feasible weights for fuzzy relations by Chang *et al*. (2007) is also adopted in this study. The step-by-step forecasting procedure is developed to present proposed method. In this study, the daily data on the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) are adopted to illustrate the steps of proposed method and utilize to evaluate the performance of proposed method. The forecasting accuracies of the proposed method are better than other fuzzy time series methods.

FUZZY TIME SERIES

Song and Chissom (1993) first proposed a forecasting model called Fuzzy Time Series (Chen and Chung, 2006), which provided a theoretic framework to model a special dynamic process whose observations are linguistic values. In the following, some basic concepts of fuzzy time series are briefly reviewed.

Definition 1 fuzzy time series: Assume that Y (t) (t = …, 0, 1, 2, …) is a subset of R. Let Y (t) be the universe of discourse defined by fuzzy set f_i (t). If F (t) consists of f_i (t) (t = 1, 2, …), F (t) is defined as a fuzzy time series on Y (t) (t = …, 0, 1, 2, …).

Corresponding Author: Jing-Rong, Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan

Definition 2 fuzzy time series relationship: If there is a fuzzy relationship R (t-1, t) such that F (t) = F (t-1)×R (t-1, t). Where × represents an operator, then F (t) is said to be caused by F (t-1). (Note that the operator can be another arithmetic operator).

When F (t-1) = A_i and F (t) = A_j, the relationship between F (t-1) and F (t) (called a Fuzzy Logical Relationship (FLR) is denoted by $A_i \rightarrow A_j$ where A_i is called the Left-hand Side (LHS) and A_j the Right-hand Side (RHS) of the FLR.

PROPOSED FUZZY TIME SERIES MODEL

In this section, the procedure of proposed model is introduced and the daily data on TAIEX are adopted to illustrate the steps of this procedure.

A theoretical model is proposed, for which the various steps are explained as follows:

Step 1: To calculate the momentum. This study forecasts the momentum in the observations instead, by obtaining the momentum V between every two consecutive observations at Actual (t) and Actual (t-1):

$$V(t) = \text{Actual}(t) - \text{Actual}(t-1),\ t = 2, 3, \ldots, h \quad (1)$$

where, Actual (t) and Actual (t-1) are two consecutive observations at t and t-1, h denotes total number of training data and V (t) is their momentum respectively.

Step 2: To define the universe of discourse U (Chang *et al*., 2007). The universe of discourse can be partitioned by Genetic algorithms. Let U = (D_{min}-D_1, D_{max}+D_2), where D_{min} and D_{max} are the minimum value and the maximum value, respectively. D_1 and D_2 are two proper positive numbers, respectively

Step 3: To randomly generates chromosomes as the initial population (Chen and Chung, 2006)

The first gene is upper bound and the final gene is lower bound on each chromosome in the population.

The first gene on each chromosome in the population is generated by:

$$\text{Initial gene} = \text{random (upper bound, lower bound)} \quad (2)$$

Then, we need to sorted gene value on each chromosome in an ascending sequence.

Step 4: To calculate the fitness value on each chromosome

Step 4.1: To fuzzify the historical data on each chromosome

The gene on each chromosome can be defined as a cut point (Chen and Chung, 2006) and the chromosome in the population can be defined as a solution. Based on the corresponding membership functions of the intervals, we can fuzzify the historical data of the TAIEX.

Step 4.2: To find all one period fuzzy relations from the training data (Chang *et al*., 2007)

The one period fuzzy relation can be extracted from the historical data (i.e., training data). The one period fuzzy relation can be defined as (Chang *et al*., 2007):

$$\begin{aligned}
&(L_i^{O_1}, \mu(L_i^{O_1})) \rightarrow (L_j^{O_1+1}, \mu(L_j^{O_1+1})) \\
&(L_i^{O_2}, \mu(L_i^{O_2})) \rightarrow (L_j^{O_2+1}, \mu(L_j^{O_2+1})) \\
&\quad \vdots \\
&(L_m^{O_n}, \mu(L_m^{O_n})) \rightarrow (L_j^{O_n+1}, \mu(L_j^{O_n+1}))
\end{aligned} \quad (3)$$

where, $1 \le i \le m$, $1 \le j \le m$, $1 \le v \le n$, n denotes the total occurrence times of the same relation. O_v denotes v times for occurrence of the same relation. $L_i^{O_1}$ denotes the antecedent of the rule $L_i \rightarrow L_j$ for first time occurrence and $L_j^{O_1+1}$ denotes the next linguistic value of $L_i^{O_1}$. In this study, L_i and L_j represent the linguistic values of antecedent and consequent part for relational rules, respectively.

Step 4.3: To calculate the cardinality of each fuzzy relation (Chang *et al*., 2007)

First, the general equation is derived to calculate the cardinality of k-th periods fuzzy relation by Eq. 4 and when k = 1, the cardinality of each one period fuzzy relation can be computed. The cardinality of k periods fuzzy relation could be defined as:

$$\begin{aligned}
&W(L_i^{O_s-k+1}, \ldots, L_i^{O_s-1}, L_i^{t_s}, L_i^{O_s+1}) = \\
&\sum_{p=1}^{n} \min(\mu(L_i^{O_p-k+1}), \ldots, \mu(L_i^{O_p-1}), \mu(L_i^{O_p}), \mu(L_i^{O_p+1}))
\end{aligned} \quad (4)$$

where, k denotes k-th periods fuzzy relation and s denotes sth relation, respectively.

Step 4.4: To forecast training data based on selecting fuzzy rules from step 4.3 (Chang *et al*., 2007)

From one period fuzzy rules, if L (t) = $L_i^{t_s}$, then the fuzzy rules, which antecedent equals $L_i^{t_s}$, are judged as fitting fuzzy rules, where L (t) denotes the previous linguistic value of forecasted value. The one period forecasting value can be calculated by Eq. 5:

$$F\ (t+1) = \sum_{p=1}^{n_1} W^{p}_{(L_i^{0s}, L_i^{0s+1})} * D(L_j) \Big/ \sum_{p=1}^{n_1} W^{p}_{(L_i^{0s}, L_i^{0s+1})} \quad (5)$$

where, F (t+1) denotes forecasted momentum value, n_1 denotes the total number of one period fitting fuzzy rules and W is the weight of fitting fuzzy rule. D (Lj) is defuzzified value and determined by consequent of the fitted fuzzy rule by Eq. 6:

$$D\ (L_j) = (a_{L_j} + b_{L_j} + c_{L_j})/3 \quad (6)$$

where, a_{Lj}, b_{Lj} and c_{Lj} denote the lower bound, upper bound and midpoint of interval of L_j, respectively.

Once we obtain the forecasted momentum between t-1 and t, we can calculate the forecast for t+1 as follows:

$$F'\ (t+1) = F\ (t+1) + Actual\ (t) \quad (7)$$

where, F' (t+1) denotes forecasted value, F (t+1) is calculated by Eq. 6 and Actual (t) is observations at t.

Step 4.5: To calculate the fitness value of each chromosome in the initial population

The fitness value is calculated by:

$$Fintness\_Valuei = 1/\sqrt{\frac{\sum_{t=1}^{h-1}(Acture\ (t) - F'(t))^2}{h-1}} \quad (8)$$

where, Acture (t) is observations at t and F' (t) is calculated by Eq. 7, respectively.

Step 4.6: To determine whether stop criteria is true or false

If the stop criteria become valid then select the top chromosomes from the population whose fitness value is smaller than the other chromosomes; go to Step 7. Otherwise, go to Step 5.

Step 5: To form a new population Select the chromosomes from the population by roulette wheel selection method (Gen and Cheng, 1997), to form a new population

Step 6: To perform the crossover operation and mutation operation. This step is calculates the crossover and mutation operation by Chen and Chung (2006)

Step 7: To fuzzify the training data on each chromosome. Based on the corresponding membership functions of the intervals, we can fuzzify the historical data of the TAIEX by the top chromosomes from the population whose fitness value is smaller than the other chromosomes.

Step 8: To find all one period fuzzy relations from the training data (Chang *et al*., 2007). The one period fuzzy relation can be extracted from the historical data. The one period fuzzy relation can be defined as equation (3).

Step 9: To calculate the cardinality of each fuzzy relation (Chang *et al*., 2007)

First, the general equation is derived to calculate the cardinality of k-th periods fuzzy relation by Eq. 3 and when k = 1, the cardinality of each one period fuzzy relation can be computed. The cardinality of k periods fuzzy relation could be defined as Eq. 4.

Step 10: To forecast testing data based on selecting fuzzy rules from step 10 (Chang *et al*., 2007)

The one period forecasting value can be calculated by Eq. 5. Once we obtain the forecasted momentum between t-1 and t, we can calculate the forecast for t+1 by Eq. 7.

EXPERIMENTS AND COMPARISON

The five-year period of TAIEX data, from 2000 to 2004, is selected as the experimental dataset. The testing data is from January to October of each year and the last two months is for testing. Previous studies on fuzzy time series often use the daily data on TAIEX (2000~2004) to evaluate their models. We define ten linguistic values in the U.

The forecast accuracy is compared by root mean square error (RMSE). The RMSE is computed by:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{h-1}(Acture\ (t) - F'(t))^2}{h-1}} \quad (9)$$

where, Acture (t) is observations at t and F' (t) is calculated by Eq. 7, respectively. The top chromosome's membership functions as show in Fig. 1. Forecasting results are compared with other fuzzy time series methods,

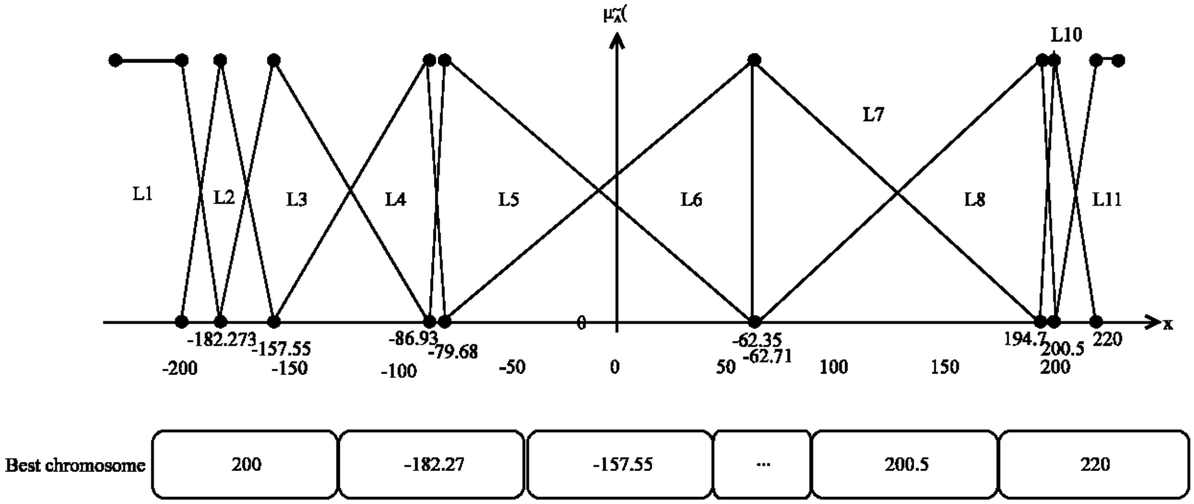

Fig. 1: Top chromosome's membership functions

Table 1: Comparison with previous works

Methods	Years 2000	2001	2002	2003	2004	Avg.
Chen and Tanuwijaya (2011) NASDAQ and M1b	128.34	113.56	66.51	53.41	55.11	83.39
Chen and Chen (2011) TAIEX and NASDAQ and Dow	124.06	125.12	72.25	57.14	56.95	87.1
Wei (2012)	146	114	66	61	N/A	96.75
Cheng *et al*. (2013)	130	120	66	55	N/A	92.75
Su *et al*. (2013)	126	114	65	N/A	N/A	101.67
Proposed model						
1-order	129.25	122.82	66.01	56.36	57.87	86.462
2-order	121.92	121.92	66.67	56.46	51.74	83.742
3-order	123.61	125.49	62.2	52.21	52.02	83.122

Table 2: Parameter settings

Parameter	Parameter settings
Population Size	100
Crossover Rate	0.6-0.9
Mutation Rate	0.01-0.1
Selection	Roulette wheel selection
Stop floating parameters	If the best fitness value of 5 generations is less than 0.0001
Stop criteria	(1) If the generation of the evolution is 1000
	(2) If the best fitness value of 30 generations is less than 0.0001

as Table 1. It shows the proposed method gets higher forecasting accuracy than other methods in the daily data on TAIEX (2000~2004). The parameter settings as show in Table 2.

CONCLUSION

This study proposes a genetic discretization approach. It is a more objective and theoretical approach for data fuzzification. The cut-point of each interval is given by genetic algorithms. Experiment results show that the proposed method can improve the forecasting performance and the average RMSE outperforms than conventional methods. The improvement reveals that proposed model having better ability for dealing with the forecasting problems of timely data.

In future work, the proposed model can be applied to other stock markets, such as in Dow Jones, S and P and ETF. Other important technology indicators could be incorporated into the proposed model to enhance accuracy and different rule-based artificial intelligence techniques can be used to generate useful decision rules for investors.

ACKNOWLEDGMENT

The research STUDY was supported by National Science Council of Taiwan (R.O.C) under the Grant NSC 101-2410-H-324-002.

REFERENCES

Chang, J.R., Y.T. Lee, S.Y. Liao and C.H. Cheng, 2007. Cardinality-based fuzzy time series for forecasting enrollments. Proceedings of the 20th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, June 26-29, 2007, Kyoto, Japan, pp: 735-744.

Chen, S.M. and N.Y. Chung, 2006. Forecasting enrollments using high-order fuzzy time series and genetic algorithms. Int. J. Intell. Syst., 21: 485-501.

Chen, S.M. and K. Tanuwijaya, 2011. Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques. Expert Syst. Appl., 38: 10594-10605.

Chen, S.M. and C.D. Chen, 2011. TAIEX forecasting based on fuzzy time series and fuzzy variation groups. IEEE Trans. Fuzzy Syst., 19: 1-12.

Cheng, C.H., L.Y. Wei, J.W. Liu and T.L. Chen, 2013. OWA-based ANFIS model for TAIEX forecasting. Econ. Modell., 30: 442-448.

Gen, M. and R. Cheng, 1997. Genetic Algorithms and Engineering Design. 1 Edn., John Wiley and Sons, New York, USA.

Song, Q. and B.S. Chissom, 1993. Forecasting enrollments with fuzzy time series: Part I. Fuzzy Sets Syst., 54: 1-9.

Su, C.H., C.H. Cheng and W.L. Tsai, 2013. Fuzzy time series model based on fitting function for forecasting TAIEX index. Int. J. Hybrid Inform. Technol., 6: 111-122.

Wei, L.Y., 2012. An adaptive expectation genetic algorithm based on anfis and multinational stock market volatility causality for TAIEX forecasting. Cybern. Syst. Int. J., 43: 410-425.