ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 18 (2013) 1155 – 1162

Procedia Computer Science

2013 International Conference on Computational Science

A novel stock forecasting model based on fuzzy time series and genetic algorithm

QiSen Cai[a], Defu Zhang[a,]*, Bo Wu[a], Stehpen C.H. Leung[b]

[a]Department of Computer Science, Xiamen University, Xiamen, 361005, China
[b]Department of Management Sciences, City University of Hong Kong, Hong Kong

Abstract

Stock market has been developed for over twenty years, and has gone deeply into all aspects of daily economic life and attracted more and more investors' attentions. Therefore, researches on finding internal rules and establishing an efficient stock forecast model to help investors minimize risks and maximize returns are very practical and amazing. In this paper, a hybrid model FTSGA based on fuzzy time series and genetic algorithm is proposed. FTSGA improves the performance by applying the operations of genetic algorithm such as selection, crossover and mutation to iteratively search a good discourse partition. TAIEX is selected as the experimental data set. And experimental results show that comparing with other models based on fuzzy time series FTSGA can greatly reduce the root mean square error and improve accuracy.

Keywords:Stock forecasting; Fuzzy time series; Genetic algorithm

1. Introduction

For time series, the observed uncertain value can be modeled as a fuzzy variable, which results in the so-called fuzzy time series (FTS). Fuzzy time series is an application of fuzzy mathematics in the field of time series. In the case that data is incomplete and contains noise, using fuzzy theory to help forecast can generally obtain better results.

Fuzzy time series was first proposed by Song and Chissom [1] in 1993. Subsequently, the model is widely researched and improved by many scholars, and is applied to solving the prediction problem of various fields, such as enrollment [2-3], temperature [4-5], the demand for tourism [6], stocks [4, 7, 13] and so on. In recent

* Corresponding author. Tel.: +8613860491711.

E-mail address: dfzhang@xmu.edu.cn.

Selection and peer review under responsibility of the organizers of the 2013 International Conference on Computational Science
doi:10.1016/j.procs.2013.05.281

years, fuzzy time series is gradually applied to predicting the stock market, which has achieved good results. For example, Yu [7] predicted the Taiwan weighted index; Huarng [11] combined with back-propagation neural network and fuzzy time series to forecast the Taiwan weighted index; Cheng [12] adjusted economics Adaptive Expectation to correct prediction results and so on.

The partition of the universe of discourse plays an important role in improving the forecasting results of fuzzy time series model. Huarng [9, 10] proposed two heuristic methods to partition isometric universe, one is based on distribution and the other is based on the average. Kuo [15] used particle swarm optimization method to search for a suitable partition of universe. In order to find better partition of universe to improve the results of the fuzzy time series models, a hybrid method combined with genetic algorithm is proposed in this paper. Genetic algorithm is a well-known search heuristic that mimics the process of natural evolution. This heuristic is widely used to generate useful solutions to optimization and search problems including the partition problem in fuzzy time series. Our model FTSGA adjusts genetic algorithm to obtain suitable partition of the universe so that the proposed model can achieve certain better results compared with existing conventional fuzzy time series models.

2. Basic Concepts

In this section, we briefly describe the underlying concepts of fuzzy time series and Genetic Algorithm.

2.1. Fuzzy Time Series

In the past few decades, researches on time series have made a progress in dealing with precise figures. But in real life, people tend to encounter a lot of random fuzzy sequences containing noise. The prediction based on traditional time series appears to be powerless in addressing these situations. But fuzzy mathematics has a large advantage in solving such problems [8]. Therefore, Song and Chissom [1] introduced the concept of fuzzy mathematics into time series and proposed the concept of fuzzy time series. Below are the related concepts of fuzzy time series.

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

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

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

Definition 2. Assume that $F(t)$ is caused by $F(t\text{-}1)$ only, denoted as $F(t\text{-}1) \rightarrow F(t)$, then this relationship can be expressed as $F(t) = F(t\text{-}1) \text{ o } R(t, t\text{-}1)$, where $F(t) = F(t\text{-}1) \text{ o } R(t, t\text{-}1)$ is called the first-order model of $F(t)$, $R(t, t\text{-}1)$ is the fuzzy relationship between $F(t\text{-}1)$ and $F(t)$, and "o" is the Max-Min composition operator.

Definition 3. Let $R(t, t\text{-}1)$ be a first-order model of $F(t)$. If for any t, $R(t, t\text{-}1) = R(t\text{-}1, t\text{-}2)$, then $F(t)$ is called a time-invariant fuzzy time-series. Otherwise, it is called a time-variant fuzzy time-series.

Definition 4. Let $F(t\text{-}1) = A_i$ and $F(t) = A_j$, it can be denoted by $A_i \rightarrow A_j$, where A_i is called the left-hand side (LHS), A_j is the right-hand side (RHS) and be called the current state of the fuzzy logical relationship (FLR).

Song and Chissom [1-2] established a four-step framework to manipulate the forecasting problem: 1) determine and partition the universe of discourse into intervals; 2) define fuzzy sets on the universe of discourse and fuzzify the time series; 3) derive the fuzzy relationships existing in the fuzzified time series; and

4) forecast and defuzzify the forecasting outputs. In the literature, the fuzzy relation $R_{ij}(t, t-1)$ is usually represented by a fuzzy logical relationship rule, i.e., IF–THEN rule in the sense that forecasting is facilitated by grouping fuzzy logical relationships into rules and then applying a “table-look-up” method when forecasting.

2.2. Genetic Algorithms

Genetic algorithm begins with a set of candidate solutions (chromosomes) called population. The new offspring solution is generated using crossover and mutation operators based on the old solutions. Then in order to get better population, the new population is formed by selecting solutions according to their fitness. The more suitable the solutions are the bigger chances they survive. This process is repeated until some termination condition is satisfied. More detailed operators combined with the specific problem characters are explained in the next section.

Table1. Taiwan's weighted index in January 2000

$Date_t$	$Index_t$	$close_t$	Percentage change	Fuzzy value
2000-1-4	1	8756.55	——	——
2000-1-5	2	8849.87	1.07%	A_4
2000-1-6	3	8922.03	0.82%	A_3
2000-1-7	4	8845.47	-0.86%	A_2
2000-1-10	5	9102.6	2.91%	A_5
2000-1-11	6	8927.03	-1.93%	A_1
2000-1-12	7	9144.65	2.44%	A_5
2000-1-13	8	9107.19	-0.41%	A_2
2000-1-14	9	9023.24	-0.92%	A_2
2000-1-15	10	9191.37	1.86%	A_4
2000-1-17	11	9315.43	1.35%	A_4
2000-1-18	12	9250.19	-0.70%	A_2
2000-1-19	13	9151.44	-1.07%	A_1
2000-1-20	14	9136.95	-0.16%	A_2
2000-1-21	15	9255.94	1.30%	A_4
2000-1-24	16	9387.07	1.42%	A_4
2000-1-25	17	9372.37	-0.16%	A_2
2000-1-26	18	9581.96	2.24%	A_5
2000-1-27	19	9628.98	0.49%	A_3
2000-1-28	20	9696.91	0.71%	A_3
2000-1-29	21	9636.38	-0.62%	A_2
2000-1-31	22	9744.89	1.13%	A_4

3. The stock forecasting method based on fuzzy time series and genetic algorithm

In this section, we present our work about the stock forecasting method based on fuzzy time series and genetic algorithm. In order to make the model easier to understand, we use the Taiwan's weighted index in January 2000 (Table 1) as example. The proposed method is presented as follows:

Step 1: Firstly we calculate daily percentage change [14] of stock market closing price of training data and determine the scope of the universe. Let D_{max} and D_{min} be the maximum and minimum values of percentage change respectively, then universe of discourse $U = [D_{min}, D_{max}]$, which is shown in Table 1.

Step 2：Encoding. The population is encoded as $Group_k = \{C_j|j=1,2,...,M;\ k=1,2,...,T\}$, where C_j is the *j-th* chromosome in the population，M is the size of population, and k is the number of iterations. The chromosome is encoded as $C_j=\{g_{j,i}|i=1,2,..., n_j\}$, where $g_{j,i}$ is the gene in the *j-th* chromosome and n_j is the total number of genes in *j-th* chromosome. The gene is encoded as $g_{j,i} = (v_{j,i}, v_{j,i+1}]$, where the $v_{j,i}$ is the boundary of interval.

Assume that there are $n+1$ genes in a chromosome, the universe of discourse U is divided into $n+1$ intervals, then $U = (-\infty, v_1] \cup (v_1, v_2] \cup ... \cup (v_{n-1}, v_n] \cup (v_n, +\infty)$. So that, every chromosome can be represented for one kind of partition of universe of discourse and the population can be represented for the set of partitions. The encoding of genetic algorithm is described in Figure 1.

Table 2. FLRGs Table

left	FLRG			
A1	A1 →A2 1	A1 →A5 1		
A2	A2 →A1 1	A2 →A2 1	A2 →A4 3	A2 →A5 2
A3	A3 →A2 2	A3 →A3 1		
A4	A4 →A2 2	A4 →A3 1	A4 →A4 2	
A5	A5 →A1 1	A5 →A2 1	A5 →A3 1	

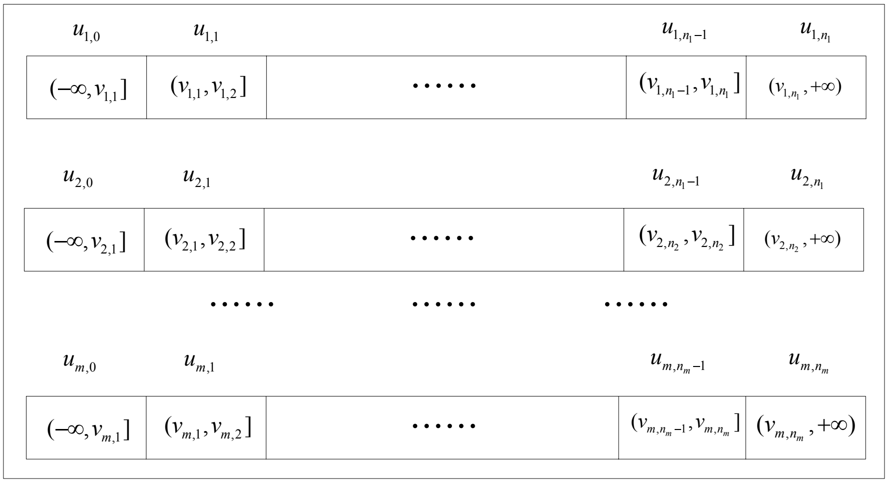

Figure 1. The encoding of genetic algorithm

Step 3: Fitness calculation. The model uses the root mean square error between the predicted value and the actual value of the training data as the fitness of individual. It is calculated as follows:

Assume there are n split points (i.e., $n+1$ gene fragments) in the chromosome, then the fuzzy set can be defined as $A = \{A_0, A_1, A_2,\ldots, A_n\}$. According to the fuzzy set, we transform the daily percentage change into fuzzy values, which are shown in the fifth column of Table 1. After that, we extract the fuzzy logic relationship FLRs which are classified and form the FLRGs. The normalized weight matrix W can be calculated from the FLRGs. The FLRGs and matrix W are shown in Table 2 and Figure 2.

$$
\begin{array}{c}
\begin{array}{cccccc} A_0 & A_1 & A_2 & A_3 & A_4 & A_5 \end{array} \\
\begin{array}{c} A_0 \\ A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{array}
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/2 & 0 & 0 & 1/2 \\
0 & 1 & 1/4 & 1/2 & 1/2 & 1/4 \\
0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\
0 & 0 & 3/5 & 0 & 2/5 & 0 \\
0 & 1/3 & 2/3 & 0 & 0 & 0
\end{bmatrix}
\end{array}
$$

Figure 2. Normalized weight matrix $\boldsymbol{W}$

Predicted values of the training data can be calculated according to the FLRGs and the weight matrix W. Then the root mean square error between the predicted value and the actual value of stock price is employed as the fitness of GA.

Step 4: Selection process. The proposed model uses tournament method to select superior individuals. The specific approach is to select randomly m individuals from the population, and select the *top 2* individuals as the parents to process the crossing. Through the comparison of fitness among the population, the selection process using tournament method can make full use of the advantages of excellent individual.

Step 5: Crossover process. The proposed model adopts single-point crossover operator to recombine the solution. The specific method is described as below: assume there are two parent individuals, Parent 1 and Parent 2. We select a split point of them randomly as the crossover point. Then the first half of Parent 1 and the latter half of Parent 2 are merged into a new individual Child 1 and the latter half of Parent 1 and the first half of Parent 2 are merged into another new individual Child 2. Thereafter, sort all the split point on the Child 1 and Child 2 in ascending order. Then two new partitions of universe are formed. The process is shown in Figure 3.

Step 6: Mutation process. The mutation process is to implement the variation to the offspring after crossover with the probability P_M. This process is able to avoid premature convergence of the genetic algorithm. There are three mutation ways of this model as follows:

- Insertion. Randomly generate one split point within $[D_{min}, D_{max}]$ and insert it into the chromosome.
- Deletion. Randomly delete one split point from the chromosome if there is one more split point in the chromosome.
- Variation. Randomly select a split point in the chromosome and change its value. Then adjust the position of the split point to ensure all the split points are sorted in ascending order.

Step 7: Termination condition. Our model determines whether the genetic algorithm should be terminated by adopting two rules: 1. the iteration times exceeds the default maximum iteration. 2. The fitness of the best individual no longer changes after several iterations. Once meeting one of above-mentioned rules, the model ends the iterative process of the genetic algorithm and goes to Step 8; otherwise continues the iterative process.

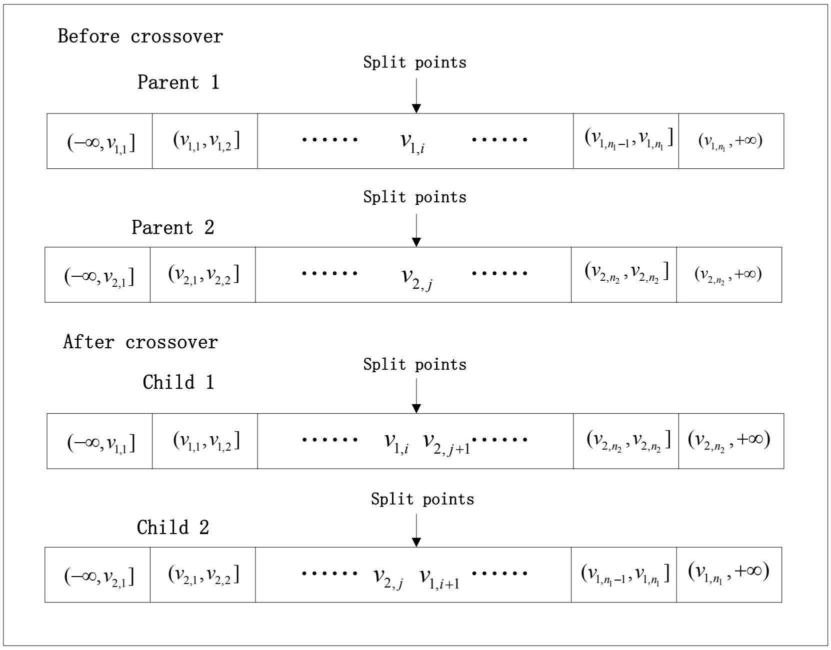

Figure 3. Crossover process

Step 8: Let C_{best} be the best individual selected by the genetic algorithm, then partition the universe of discourse according to the setting stored in C_{best}. Then fuzzify the training time series and derive the fuzzy relationships existing in the fuzzified time series. Finally we process forecasting and calculate the root mean square error between the predicted and actual values.

4. Experiment results.

4.1. Data description and setup

To demonstrate the effectiveness of the FTSGA models, large amounts of data are needed. For this reason we use the daily TAIEX closing prices covering the period from 1990 to 1999 as the verification dataset, and compare its performance with that of existing conventional fuzzy time series models. The data from January to October for each year are used to perform the estimation while those for November and December are used for forecasting. To inspect forecasting performance of the proposed model, the indicator root of mean squared error (RMSE) is employed as evaluation criterion for the forecasting performance of the proposed model and the comparison models, which is defined as Eq. (2):

$$RMSE = \sqrt{\sum_{t=1}^{n} \left(actual(t) - predicted(t)\right)^2 / n} \tag{2}$$

In the experiment, the parameter settings in the genetic algorithm are showed as below: the size of the population is set as 200; the maximum number of iterations is set as 100; the crossover probability is set as

80%; the mutation probability is set as 1%. Because of the randomness, we run our program 100 times and use the average value as the result to be compared.

Table 3. Comparisons of the forecast results with different models(RMSE).

	1991	1992	1993	1994	1995	1996	1997	1998	1999	average
Chen	80	60	110	112	79	54	148	167	149	106.6
Yu	61	67	105	135	70	54	133	151	142	102
Huarng	54	54	107	79	74	73	141	121	109	90.2
Cheng	**42**	43	105	**75**	**53**	51	134	113	109	80.6
FTSGA	47	**40**	**104**	76	56	**47**	**133**	**111**	**103**	**79.7**

Table 4. The Directional accuracy of the forecast results.

	1991	1992	1993	1994	1995	1996	1997	1998	1999	average
FTSGA	0.6012	0.7002	0.5709	0.4493	0.5452	0.5072	0.5406	0.5514	0.6308	0.5830

4.2. Overall performance

In this paper, we choose four models based on fuzzy time series to be compared with FTSGA model. These four models are proposed by Chen [3], Yu [7], Hurang [11] and Cheng [12] respectively. And the forecasted errors RMSE of five models are listed in Table 3. On comparison of the three proposed models, the experimental results show that our proposed model bears all the smallest RMSE in seven testing period. In regard to the average RMSE, the proposed model obtains the smallest value 79.7 and is far exceed any of the other compared models. In terms of the yearly comparison, the FTSGA model performs better than other models in 6 out of 9 years. From these results, it is obvious that FTSGA model significantly outperforms the models proposed by Chen [3], Yu [7] and Hurang [11] and is slightly better than the one proposed by Cheng [12]. In terms of the directional accuracy of the forecast results, the FTSGA model also achieves good results, as shown in Table 4. In fact, the directional accuracy of forecasting model is more important than RMSE when generating the trading strategy to invest [16].

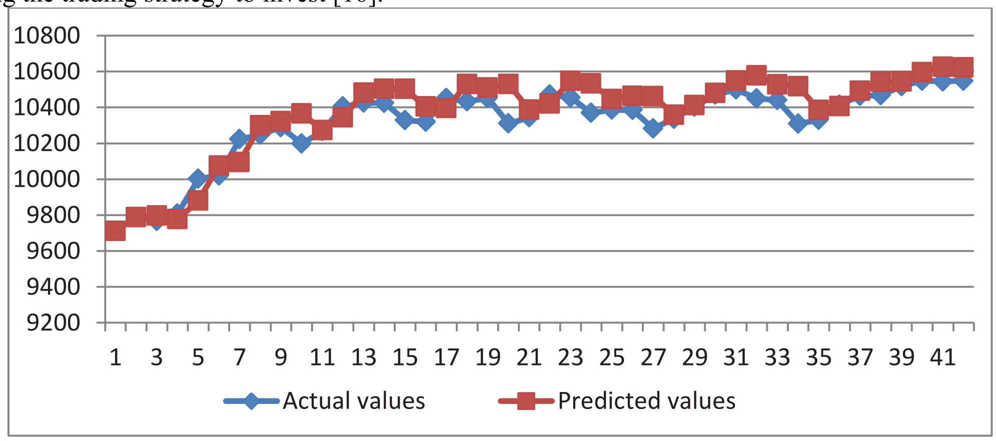

Figure 4. Comparison of the predicted value and the actual value of the Dow

4.3. Forecasting detail

After knowing the overall performance, we further probed into the details of the forecasting. We take the Dow Jones Indexes in year 2009 as an example. From Figure 4 we can see that the predicted values and the actual values are very close and the trends of stock price are well predicted.

5. Conclusion

In this paper, we have proposed a hybrid model based on fuzzy time series and genetic algorithm. By adopting genetic algorithm, our model can obtain more suitable partition of the universe, which can improve the forecasting results significantly. Furthermore, the proposed model is compared with three different conventional fuzzy time-series models proposed earlier by Chen [3], Yu [7], Hurang [11] and Cheng [12], and the comparison shows that the proposed model surpasses in testing dataset of the TAIEX stock markets.

Reference

1. Song Q and Chissom BS. Fuzzy time series and its models. Fuzzy Sets and Systems, 1993, (54), 269-277.
2. Song Q and Chissom BS. Forecasting enrollments with fuzzy time series. Fuzzy Sets and Systems, 1993, (54), 1-9.
3. Chen SM. Forecasting enrollments based on fuzzy time series. Fuzzy Sets and Systems, 1996,(81), 311–319.
4. Lee LW, Wang LH and Chen SM. Temperature prediction and TAIEX forecasting based on high-order fuzzy logical relationships and genetic simulated annealing techniques. Expert Systems with Applications, 2008, (34), 328–336.
5. Chen SM and Hwang JR. Temperature prediction using fuzzy time series. IEEE Transactions on Systems, Man, Cybernetics—Part B:Cybernetics, 2000, vol. 30, no. 2, 263–275.
6. Tsaur RC and Kuo TC. The adaptive fuzzy time series model with an application to Taiwan's tourism demand. Expert Systems with Applications, 2011, Volume 38, Issue 8, 9164-9171.
7. Yu HK. Weighted fuzzy time-series models for TAIEX forecasting. Physica A, 2004, (349), 609-624.
8. Hwang JR, Chen SM and Lee CH. Handling forecasting problems using fuzzy time series. Fuzzy Sets and Systems, 1998, (100), 217–228.
9. Huarng K. Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets and Systems, 2001,(123), 387–394.
10. Huarng K and Yu TH. Ratio-Based lengths of intervals to improve fuzzy time series forecasting. IEEE Transactions on Systems, Man, Cybernetics—Part B:Cybernetics, 2006, 36(2), 328-340.
11. Huarng, KH, Yu, THK. The application of neural networks to forecast fuzzy time series. Physica A, 2006, (336),481-491.
12. Cheng CH, Chen TL, Teoh HJ and Chiang CH. Fuzzy time-series based on adaptive expectation model for TAIEX forecasting. Expert Systems with Applications, 2009, (34), 1126-1132.
13. Cheng CH, Chen TL and Chiang CH. Trend-Weighted Fuzzy Time-Series Model for TAIEX Forecasting. Neural Information Processing, Lecture Notes in Computer Science, 2006, Volume 4234, 469-477.
14. Stevenson M and John EP. Fuzzy time series forecasting using percentage change as the universe of discourse world academy of science. Engineering and Technology, 2009, (55), 154-157.
15. Kuo IH, Horng SJ , Kao TW, Lin TL, Lee CL and Pan L. An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. Expert Systems with Applications, 2009, (36), 6108–6117.
16. Zhang D, Jiang Q, Li X. Application of neural networks in financial data mining. International Journal of Computational Intelligence, 2004, 1(2), 116-119.