Knowledge-Based Systems 74 (2015) 61–68

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression

Qisen Cai [a], Defu Zhang [a,*], Wei Zheng [a,*], Stephen 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

ARTICLE INFO

Article history:
Received 3 May 2013
Received in revised form 20 October 2014
Accepted 6 November 2014
Available online 15 November 2014

Keywords:
Fuzzy time series
Ant colony
Auto-regression
Stock forecasting
Levenberg–Marquardt algorithm

ABSTRACT

This paper presents a new fuzzy time series model combined with ant colony optimization (ACO) and auto-regression. The ACO is adopted to obtain a suitable partition of the universe of discourse to promote the forecasting performance. Furthermore, the auto-regression method is adopted instead of the traditional high-order method to make better use of historical information, which is proved to be more practical. To calculate coefficients of different orders, autocorrelation is used to calculate the initial values and then the Levenberg–Marquardt (LM) algorithm is employed to optimize these coefficients. Actual trading data of Taiwan capitalization weighted stock index is used as benchmark data. Computational results show that the proposed model outperforms other existing models.

1. Introduction

Relationships between sequential set of past data measured over time to forecast future values, are investigated by time series forecasting. Many statistical tools such as regression analysis, moving average, exponential moving average and autoregressive moving average have been used in traditional forecasting. However, these analysis models highly rely on historical data, and the data are required to follow normal distribution to get a relatively good forecasting performance. Moreover, traditional crisp time-series forecasting methods are normally not applicable when the historical data are represented by linguistic values. In order to deal with these kinds of problems, the fuzzy time-series approach has been developed as an alternative forecasting method. It has been proved that the fuzzy time-series can be appropriately applied to datasets of linguistic values to generate forecasting rules with high accuracy.

Past decades have witnessed the development of fuzzy time series approach, since it was first introduced by Song and Chissom [1] in 1993. Different fuzzy time-series models have been applied to solve problems arising in various domains. For example, Song and Chissom [1,2] and Chen [3] developed fuzzy time series models to forecast enrollments of the University of Alabama. To enhance the accuracy of forecast values, Chen [4] proposed high-order fuzzy time series models for forecasting enrollments. Song [5] introduced an autocorrelation function to measure the dependency between the fuzzy data for selecting suitable order for fuzzy time series. Yu [6] proposed a weighted fuzzy time series model to forecast Taiwan capitalization weighted stock index (TAIEX). Own and Yu [7] presented a heuristic higher order model by introducing a heuristic function to incorporate the heuristic knowledge so as to improve TAIEX forecasting. Chen and Chang [8] presented a method for multivariable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques. Huarng [9,10] pointed out that the length of intervals affects forecast accuracy in fuzzy time series and proposed a method with distribution-based length and average-based length to reconcile this issue. Keles et al. [11] proposed a model for forecasting the domestic debt by adaptive neuro-fuzzy inference system. Chen et al. [12] presented a new method to forecast TAIEX using fuzzy time series and automatically generated weights of multiple factors. Gangwar and Kumar [13] proposed a computational method of forecasting based on multiple partitioning and higher order fuzzy time series. Chen and Tanuwijaya [14] adopted an automatic clustering technique to overcome the drawback of partition of the universe of discourse. Chen and Kao [15] proposed a new method for forecasting the TAIEX, based on fuzzy time series, particle swarm optimization techniques and support vector machines. Pritpal and Bhogeswar [16] presented a new model based on hybridization of fuzzy time series theory with artificial neural network (ANN). Cheng and Li [17] proposed an enhanced HMM-based forecasting model by developing a novel fuzzy smoothing method to overcome the

* Corresponding authors.
E-mail addresses: dfzhang@xmu.edu.cn (D. Zhang), zhengw@xmu.edu.cn (W. Zheng).

http://dx.doi.org/10.1016/j.knosys.2014.11.003
0950-7051/

problem of rule redundancy and achieve better results. Recently, Wei et al. [18] developed A hybrid ANFIS based on n-period moving average model to forecast TAIEX stock. Chen and Chen [19] presented a hybrid fuzzy time series model based on granular computing for stock price forecasting.

While adopting the fuzzy time series model, a reasonable partition of the universe of discourse can significantly enhance the accuracy of the forecasts. Cai et al. [20] presented a genetic algorithm to partition the universe of discourse. According to our review of literature, models [6–15] proposed in recent years also attached great importance to this issue. From another perspective, high-order fuzzy logical relationship methods [4,7] and multiple-factor forecasting methods [8,12,13] are widely employed in hybrid models. However, these two methods have their own drawbacks. On one hand, the high-order method, according to the shape proposed by Chen [4], cannot achieve the expected coverage of the high-order fuzzy logical relationships when the training data do not reach the desired size. On the other hand, the multiple-factor forecasting method is highly limited by the selection of the secondary factor. In addition, such a multiple-factor method also introduces extra noise when more factors are considered. To the best of our knowledge, compared with several univariate models [6,21], the multiple-factor forecasting method can hardly achieve significant improvement in accuracy of forecasting results.

Taking into account the discussion above, this paper focuses on two issues: (1) searching a suitable partition of universe of discourse and (2) taking better advantage of the high-order data. We propose a hybrid model to address these two issues in order to improve forecasting accuracy. For partitioning the universe of discourse, although many statistical methods have been used to achieve efficient results, we believe there still remains a plenty of space for improvement by adopting the Meta-heuristic optimization algorithms, such as the ant colony optimization (ACO) algorithm. ACO is a well-known probabilistic technique for solving computational problems [22,23]. Since finding partition can be defined as a graph search problem, and ACO is commonly thought to be good at solving this type of problem, we adopt ACO in this paper, where the ants are employed to search boundary of each interval and finally obtain a better partition in the research. In order to enhance the application of the historical high-order data, we propose a fuzzy time series model combined with the auto-regression method. The proposed model uses the percentage change as the universe of discourse, the same as that proposed by Stevenson and Porter in [24]. To verify the performance of the proposed model, the TAIEX is used as the experimental dataset and several fuzzy time series models [3,6,8,12,25–27] are used as competitors. The experimental results show that the proposed model gets higher average forecasting accuracy rates than other existing models.

The rest of this paper is organized as follows. Section 2 briefly reviews the definitions of fuzzy time series, ACO algorithm and Levenberg–Marquardt (LM) algorithm. Section 3 presents a novel high-order fuzzy time series model which adopts the classic concept of the auto-regression model. Section 4 presents how ACO is employed to search the suitable partition of fuzzy time series model. Section 5 compares and discusses the forecast results of the proposed model with those of existing models and shows the details of forecast results. The conclusion is provided in Section 6.

2. Brief review of basic concepts

As shown in Fig. 1, the proposed model employs ACO algorithm to search the best partition for high-order fuzzy time series model. Then the LM algorithm is adopted to optimize the coefficients of the fuzzy time series model. In this section, the underlying concepts of fuzzy time series, ACO, and the LM algorithm are introduced.

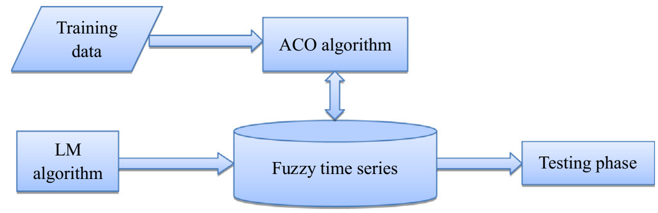

Fig. 1. Structure of the proposed model.

2.1. Fuzzy time series

In the past few decades, research on time series has made progress in dealing with precise figures. However, in real world, people tend to encounter a lot of random fuzzy sequences containing noise. Prediction based on traditional time series appears to be powerless in such situations. Fortunately, it was realized that fuzzy mathematics has a great advantage in solving such problems. As a result, Song and Chissom [1,2] introduced the concept of fuzzy mathematics into time series and proposed the concept of fuzzy time series. We briefly introduce the concepts and notations related to fuzzy time series as follows.

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:

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \cdots + f_A(u_n)/u_n \tag{1}$$

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

Definition 1. Let $Y(t)$ ($t = 0, 1, 2, \ldots$) be the universe of discourse, which is a subset of real numbers. Assume $f_i(t)$ ($i = 0, 1, 2, \ldots$) are defined on $Y(t)$, and $F(t)$ 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 $F(t-1)$ only, denoted as $F(t-1) \to F(t)$, then this relationship can be expressed as $F(t) = F(t-1) \text{ o } R(t, t-1)$, where $F(t) = F(t-1) \text{ o } R(t, t-1)$ 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 "o" is the Max–Min composition operator.

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

Definition 4. If $F(t)$ is caused by $F(t-1), F(t-2), \ldots, F(t-k)$, then the fuzzy logical relationship between them can be represented by a high-order fuzzy logical relationship. For example, a k-order fuzzy logical relationship can be expressed as follows:

$$F(t-k), \ldots, F(t-2), F(t-1) \to F(t) \tag{2}$$

Definition 5. If $F(t)$ is caused by $F(t-1), F(t-2), \ldots, F(t-k)$ and $G(t-1), G(t-2), \ldots, G(t-j)$, where $G(t)$ is another fuzzy time series. Then the fuzzy logical relationship between them is defined as a multivariable high-order fuzzy logical relationship.

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, fuzzy relationship $R_{ij}(t, t-1)$ 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. Ant colony optimization algorithm

Meta-heuristic optimization algorithm is a relatively new approach for problem solving. Such algorithms usually take inspiration from social behaviors of insects or other animals. Use of ACO for solving combinatorial optimization problems was first proposed by Dorigo and Stützle [22], where it was used to solve the traveling salesman problem (TSP): given that a set of n cities must be visited by one single traveling salesman, how to minimize the total distance length, which is a well-known NP-hard problem.

The ACO algorithm is inspired by and based on the way ants search for food and find their way back to the nest. During their trips ants leave a chemical trail called pheromone on the ground. The role of pheromone is to guide other ants toward the target point. An ant chooses the path according to the quantity of pheromone.

The ACO algorithm manages scheduling of three activities [23,28], as shown in Fig. 2. The first step consists mainly of initialization of the pheromone trail. In the iteration (second) step, each ant constructs a complete solution to the problem, according to a probabilistic state transition rule. The state transition rule depends mainly on the state of the pheromone. The third step updates the quantity of pheromone. A global pheromone updating rule is applied in two phases: in the first phase, a fraction of the pheromone evaporates; and in the second phase each ant deposits an amount of pheromone which is proportional to the fitness of its solution. This process is iterated until a stopping criterion is met.

More details of each step are described in the next section, combined with the specific problem characteristics.

2.3. Levenberg–Marquardt algorithm

Levenberg–Marquardt (LM) algorithm [29] is a popular alternative to the Gauss–Newton method of finding the minimum of a function $F(x, \beta)$ that is a sum of squares of nonlinear functions $f_i(x, \beta)$, such as

$$F(x, \beta) = \sum_{i=1}^{m} [f_i(x, \beta)]^2 \tag{3}$$

To start the minimization process, the user has to provide an initial guess for the parameter vector $\boldsymbol{\beta}$. In cases with only one minimum, an uninformed standard guess like $\boldsymbol{\beta}^{\mathrm{T}} = (1, 1, \ldots, 1)$ will work fine; in cases with multiple minima, the algorithm converges only if the initial guess is already somewhat close to the final solution. Let the Jacobian of $f_i(x, \beta)$ be denoted by J_i, where $J_i = \frac{\partial f_i(x,\beta)}{\partial \beta}$. Then the LM method searches in the direction given by solution p to the equation

$$\left(J_k^T J_k + \lambda_k l\right)\rho_k = -J_k^T f_k \tag{4}$$

where λ_k are nonnegative scalars, l is the identity matrix, giving as the increment ρ_k, to the estimated parameter vector $\boldsymbol{\beta}$. It is worth mentioning that this method has a nice property: for some scalar -ε related to λ_k, the vector ρ_k is the solution of the constrained subproblem of minimizing

$$\|J_k \rho + f_k\|^2 / 2 \tag{5}$$

subject to

$$\|\rho\|_2^2 \leqslant \varepsilon \tag{6}$$

The reduction of sum of squares from the latest parameter vector $\boldsymbol{\beta} + \boldsymbol{\delta}$, falls below predefined limits, iteration stops and the last parameter vector, $\boldsymbol{\beta}$ is considered to be the best parameter. LM algorithm is recognized as more robust than the Gauss–Newton method and needs fewer iterations in general, for finding an optimum [29].

Step 1: Initialization.

Initialize the pheromone trail.

Step 2: Iteration.

For each ant repeat:

- Solution construction using the pheromone trail.
- Update the pheromone trail.

Until stopping criteria is met.

Fig. 2. A generic ant algorithm.

3. The high-order AR fuzzy time series model

In this section, details of the novel high-order fuzzy time series model are described and compared with the traditional high-order model. The workflow of experiment is depicted in Fig. 3. The details of each step are demonstrated as follows.

Step 1:

Firstly, the daily change (%) in stock market closing price of training data is calculated and used as the universe of discourse. The daily percentage is calculated as follows:

$$\text{Percentage Change}_t = \frac{\text{close}_t - \text{close}_{t-1}}{\text{close}_{t-1}} \tag{7}$$

Then the scope of the universe is determined by calculating the maximum and minimum values of the universe. Let D_{max} and D_{min} be the maximum and minimum values of percentage change respectively, then the scope of universe $U = [D_{min}, D_{max}]$.

Step 2:

ACO is employed to generate a suitable partition of universe. Details of how ACO works are described in the next subsection. Here, it is assumed that a partition of universe has been obtained, n is the number of intervals and each interval is denoted as u_1, u_2, u_3, ..., u_n.

It is worthy to notice that we define two intervals u_0 and u_{n+1} while applying the model for the test data set if some testing data exceed the boundaries D_{min} and D_{max}, where u_0 is defined as $(-\infty, D_{min})$ and u_{n+1} is defined as $(D_{max}, +\infty)$. And the values of those fuzzy terms belonged to these two intervals, m_0 and m_{n+1}, are D_{min} and D_{max} respectively.

Step 3:

Then linguistic terms $A_1, A_2, A_3, \ldots, A_n$ of the time series represented by fuzzy sets, are defined as follows:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_n$$
$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_n$$
$$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \cdots + 0/u_{n-1} + 0/u_n$$
$$\cdots$$
$$A_n = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \cdots + 0.5/u_{n-1} + 1/u_n$$

Each historical datum is fuzzified into a fuzzy set. If a datum belongs to u_i, then this datum is fuzzified into A_i, where $1 \leqslant i \leqslant n$. For instance, when a partition of discourse is given as follows:

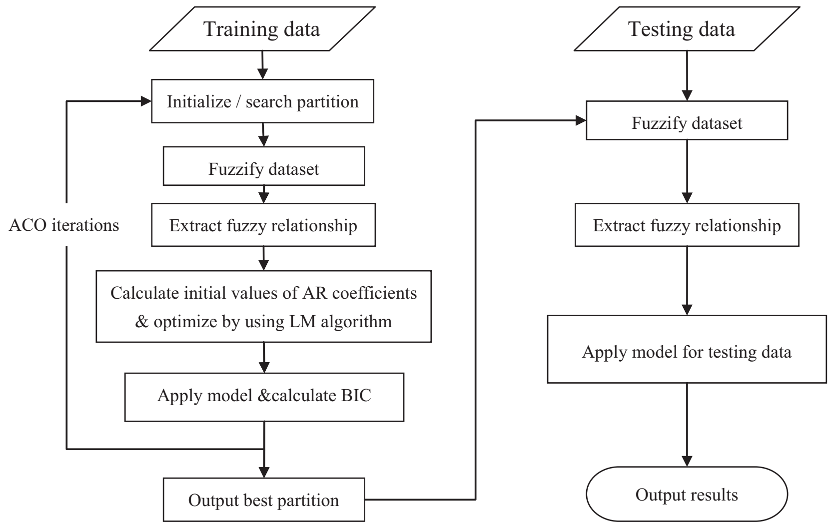

Fig. 3. The flowchart of the presented algorithm.

$$U = [-1.93\%, -1\%] \cup (-1\%, 0\%] \cup (0\%, 1\%] \cup (1\%, 2\%] \cup (2\%, 2.91\%],$$

then the Taiwan's weighted index in January 2000 is fuzzified as in Table 1.

Step 4:

The previous steps are similar to traditional fuzzifying methods used on fuzzy time series. Next the novel high-order fuzzy time series method is presented and explained. According to the brief review of the fuzzy time series method, traditional high-order methods are then used to build the high-order fuzzy logical relationship and forecast as follows.

If k-th order fuzzy time series is used, fuzzified values at time $t-k, \ldots, t-2$ and $t-1$ are $A_{i,k}, \ldots, A_{i,2}$ and $A_{i,1}$, respectively, and there is only one fuzzy logical relationship in the fuzzy logical relationship groups, shown as follows:

$$A_{i,k}, \ldots, A_{i,2}, A_{i,1} \rightarrow A_{j,1}(x_1), A_{j,2}(x_2), A_{j,3}(x_3), \ldots, A_{j,p}(x_p)$$

where x_l denotes the weight of the fuzzy logical relationship $A_{i,k}, \ldots, A_{i,2}, A_{i,1} \rightarrow A_{j,l}$ generated by number of occurrences. Then the forecasted value at time t is calculated as:

Table 1
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

$$\frac{x_1 \times m_{j,1} + x_2 \times m_{j,2} + \cdots + x_p \times m_{j,p}}{x_1 + x_2 + \cdots + x_p} \tag{8}$$

where $m_{j,1}, m_{j,2}, \ldots, m_{j,p}$ denotes middle values of intervals $u_{j1}, u_{j2}, \ldots, u_{jp}$ respectively.

Assume that there are n fuzzy sets, and then there will be n^2 possible fuzzy logical relationships and n possible left side rules (the formula in the left of $\rightarrow$ in the fuzzy relationship) in the first-order fuzzy time series model. To ensure that every possible left side rule occurs at least not less than once, the times series is required to be of at least $n+1$ length. The second-order fuzzy time series model, it requires at least n^2+2 length and in k-th order model it requires n^k+k length. Even if a dataset can cover every left side rule, it is essential to generate more accuracy weight of the logical relationship to get better forecasting results, which requires a bigger dataset. As a result, it is hard to employ a traditional third or higher order fuzzy time series model to achieve reasonable forecasting results.

To address the drawback of the traditional high-order fuzzy time series model discussed above and take advantage of using historical information, the concept of auto-regression model is introduced: this model is named as high-order AR fuzzy time series model.

The high-order AR fuzzy time series model is described as follows.

According to the traditional first-order fuzzy time series model, we can generate the fuzzy logical relationships as $F(t-1) \rightarrow F(t)$ at each time t and calculate their respective weights. Then we can calculate the forecasted value $\hat{y}_{t,1}$ according to the formula (7). If we consider the relationship between $F(t-2)$ and $F(t)$, it is similar to generating the fuzzy logical relationships as $F(t-2) \rightarrow F(t)$ and obtaining another forecast value $\hat{y}_{t,2}$. Considering the k-th order relationship $F(t-k) \rightarrow F(t)$, we can get forecast value $\hat{y}_{t,k}$ likewise.

While using k-th order AR fuzzy time series $\hat{y}_{t,1}, \hat{y}_{t,2}, \ldots, \hat{y}_{t,k}$, are calculated; and then the final forecast value of the proposed model is calculated as follows:

$$\text{predicted}_t = \phi_1 \hat{y}_{t,1} + \phi_2 \hat{y}_{t,2} + \cdots + \phi_k \hat{y}_{t,k} \tag{9}$$

where ϕ_i denotes the coefficient of $\hat{y}_{t,k}$ and $\phi_1 + \phi_2 + \cdots + \phi_k = 1$. The l-th lag coefficient represents l-th lag historical data's contributing to the forecast. In order to obtain the opportune coefficient, the autocorrelations are calculated as follows:

$$\hat{\rho}_l = \frac{\sum_{t=1}^{n-1} (y_t - \bar{y})(y_{t+1} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2} \tag{10}$$

where $\hat{\rho}_l$ denotes the l-th lag autocorrelation, y_t denotes value of time series and $\bar{y}$ denotes the mean of the time series. After $\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_k$ are calculated, the initial value of the coefficient ϕ_i is calculated as follows:

$$\phi_i = \frac{\hat{\rho}_i}{\hat{\rho}_1 + \hat{\rho}_2 + \cdots + \hat{\rho}_k} \tag{11}$$

After obtaining the initial value of the coefficient, the proposed model employs LM algorithm on a few iterations to optimize these coefficients as in Fig. 3. For details about how the LM algorithm works, please refer to the literature [29].

4. ACO for searching suitable partition

This section explains how the ACO algorithm works when searching a suitable partition of the universe of discourse.

As mentioned in the previous section, the universe of discourse can be represented as $U = [D_{min}, D_{max}]$, where D_{max} and D_{min} are the maximum and minimum values of percent change, respectively. Assume the universe of discourse can be partitioned into n intervals: u_1, u_2, u_3, ..., u_n, which can be denoted as:

$$U = [D_{min}, v_1] \cup (v_1, v_2] \cup \cdots \cup (v_{n-2}, v_{n-1}] \cup (v_{n-1}, D_{max}] \tag{12}$$

To encode the partition of the universe of discourse, we have used a representation, which is based on an ascending permutation of n float values:

$$s = (v_1, v_2, \ldots, v_{n-1}) \tag{13}$$

where v_i denotes the boundary of the interval.

In order to apply the ACO algorithm, a searching graph is built by partitioning the universe of discourse into m intervals of equal length and then $m - 1$ split points are obtained:

$$M = (p_1, p_2, \ldots, p_{m-1}) \tag{14}$$

where p_i denotes i-th split point and $m \gg n$. Hence, every solution of the partition s can be a subset of M. In the ant system, as shown in Fig. 4, the solution can be constructed as: every ant starts from D_{min} point, through $n - 1$ spilt points, reach the D_{max} point, where n is not a fixed value.

Details of the procedure of the ACO algorithm will be described as follows:

4.1. Initialization

The discussion above, shows that there are $m + 1$ points (include D_{min} and D_{max}) in the searching graph so that a $(m + 1) \times (m + 1)$ matrix of pheromones can be built. Before the iteration the matrix of pheromones F is initialized as follows:

$$\tau_{i,j}^0 = \tau_0 \quad i,j \in \{1, \ldots, m+1\} \tag{15}$$

where τ_0 denotes a constant term.

4.2. Solution construction

Assume that an ant (solution) has passed j points, D_{min}, v_1, v_2, ..., v_{j-1}, which are in an ascending order, and then the ant selects the next point in two ways: (1) select the next point randomly or (2) select the next point with a probability proportional to the associated pheromone:

$$\rho_{v_{j-1},l} = \frac{\tau^k_{v_{j-1},l}}{\sum_{r \neq v_{j-1}} \tau^k_{v_{j-1},r}} \quad l \in \{1, \ldots, m+1\} \tag{16}$$

An ant will select the first way with probability 0.1 and the second way with probability 0.9. It is worth noting that, in the selection process, an ant is allowed to select a point with a value lower than the points it has passed. Nevertheless, when an ant has selected a point with less value, the construction process will end, and then the D_{max} will be added to the solution automatically. Through these considerations, the solution can be constructed evenly and meaningfully.

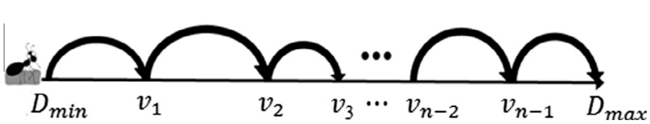

Fig. 4. How an ant constructs a partition.

4.3. Update of the pheromone matrix

Firstly, we update the pheromone matrix to simulate the evaporation process, which consists of reducing the matrix values F with the following formula:

$$\tau_{i,j}^{k+1} = (1 - \alpha)\tau_{i,j}^k \quad i,j \in \{1, \ldots, m+1\} \tag{17}$$

where $0 < \alpha < 1$ ($\alpha = 0.1$ in our experiments, it is noted that the experimental results show that it is not an important factor.) and k denotes the number of iteration. If α is close to 0, influence of the pheromone will last for a long time. Otherwise its action will be short-lived.

Secondly, the pheromone is reinforced function of the solution found. The fitness of the solution found needs to be calculated. The solution (a kind of partition of universe of discourse) is applied to the high-order AR fuzzy time series model to forecast the training data. After forecasting, the MSE (Mean Squared Error) is calculated as follows:

$$\text{MSE} = \frac{\sum_{t=1}^{N} (\text{actual}_t - \text{predicted}_t)^2}{\text{N}} \tag{18}$$

where actual_t denotes the actual value, predicted_t denotes the forecast value and N denotes the length of training data. Instead of taking into account the goodness of fit, the generalization capability is considered. It is obvious that a fuzzy time series model with more intervals is easier to get over fitting while a model with fewer intervals has stronger generalizability. So the BIC (Bayesian Information Criterion) is adopted and calculated as follows:

$$\text{BIC} = \ln(\text{MSE}) + n\frac{\ln(N)}{N} \tag{19}$$

where n denotes the number of intervals of the solution (partition). As is well known, the smaller the BIC is, the better is the solution. So far, the update formula of pheromones matrix F is:

$$\tau_{i,j}^{k+1} = (1 - \alpha)\tau_{i,j}^k + \frac{\tau_0}{\text{BIC}} \quad i,j \in \{1, \ldots, m+1\} \tag{20}$$

4.4. Stopping criterion

The iteration process ends when the best solution is not updated or the process reaches the maximum number of iterations. Then the solution (partition) with the smallest BIC ever found is adopted to the high-order AR fuzzy time series model.

5. Experiment results

5.1. Data description and setup

To demonstrate the effectiveness of the proposed models, large amounts of data are needed. For this reason, the daily TAIEX closing prices covering the period from 1990 to 2004 are used as the verification dataset, which can be downloaded from http://finance.yahoo.com/. Then the proposed model's performance is compared with that of existing conventional fuzzy time series models. The data from January to October for each year are used for 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 the evaluation metric that is used by the comparison models, which is defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{N}(\mathrm{actual}_t - \mathrm{predicted}_t)^2}{N}} \tag{21}$$

where N denotes the number of dates needed to be forecast, "$\mathrm{predicted}_t$" denotes the forecast value of day t, "actual_t" denotes the actual value of day t and $1 \leqslant t \leqslant \mathrm{N}$.

To determine the max order of the high-order AR time series model, we have made a comparison of the performance of different order. The universe of discourse was divided into 7 intervals with equal length, and then the average RMSE of TAIEX forecast result was calculated. From Fig. 5, one can see that, with the increasing of the order from 1 to 4, the average RMSE decreases. However, for the models with the order higher than 4, we cannot achieve the promising results because they cause overfitting. Moreover, the higher the order is, the easier the overfitting. Consequently, the fourth-order AR fuzzy time series model is adopted to forecast the TAIEX.

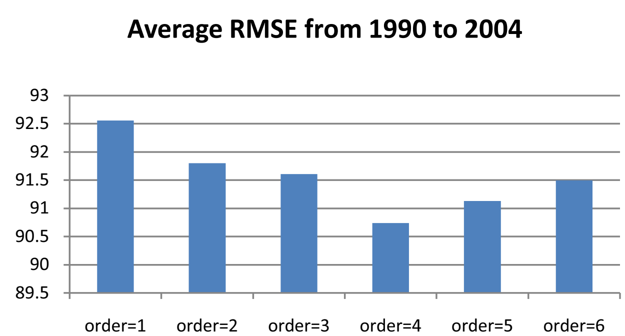

Fig. 5. The comparison of average RMSE of different orders.

5.2. Overall performance

The proposed model is applied for forecasting the TAIEX from 1990 to 1999 and the forecasted results of the proposed model are compared with conventional models [6], weighted models [6], Chen and Chen's model [27], Chen et al.'s model [12] and Chen and Kao's model [15], as shown in Table 2. Where, the smallest RMSE obtained by different models is in bold. All the RMSEs in Table 2 are taken from [12,15]. Comparison of these models shows that the proposed model bears the smallest RMSE in all four testing periods. With regard to the average RMSE, the proposed model obtains the smallest value of 89.44 and far exceeds any of the other compared models.

In Table 3, a comparison of the RMSE and the average RMSE of the proposed model with Yu's model [26], Huarng's model [25], Yu's model [30], Chen and Chang's model [8] and Chen and Chen's models [27] is made. It is worth noting that the model with the specific secondary factors which can achieve the smallest average RMSE are selected while comparing with the multiple factors fuzzy time models [8,25,27]. All the RMSEs in Table 3 are taken from [12,15]. Similarly, the smallest RMSE is in bold. From Table 3, one can see that the proposed model outperforms the existing models for forecasting the TAIEX from 1999 to 2004.

To further reveal the advantage of proposed model, we carried out a t-test experiment to see if our improvement is significant. Let d_j denote the improvement rate:

$$d_j = \frac{\widehat{\mathrm{RMSE}}_j - \mathrm{RMSE}_j}{\widehat{\mathrm{RMSE}}_j}$$

where RMSE_j denotes the RMSE of year j of the proposed model and $\widehat{\mathrm{RMSE}}_j$ denotes the RMSE of year j of the model to be compared. Then the overall variance in the observed improvement rates is estimated using the following formula:

$$\sigma_d^2 = \frac{\sum_{j=1}^{k}(d_j - \bar{d}_t)}{k(k-1)}$$

where $\bar{d}_t$ denotes the average improvement rate. For this approach, we need to use a t-distribution to compute the confidence interval of d_t:

Table 2
Comparison of the RMSEs and the average RMSE for different methods (from 1990 to 1999).

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	Average
Conventional models [6]	220	80	60	110	112	79	54	148	167	149	117.9
Weighted models [6]	227	61	67	105	135	70	54	133	151	142	114.5
Chen and Chen [27]	172.89	72.87	43.44	103.21	78.63	66.66	59.75	139.68	124.44	115.47	97.70
Chen et al. [12]	174.62	43.22	42.66	104.17	94.6	**54.24**	50.5	138.51	**117.87**	101.33	92.17
Chen and Kao [15]	**156.47**	56.50	**36.45**	126.45	**62.57**	105.52	51.50	125.33	104.12	**87.63**	91.25
The proposed model	187.10	**39.58**	39.37	**101.80**	76.32	56.05	**49.45**	**123.98**	118.41	102.34	**89.44**

Table 3
Comparison of the RMSEs and the average RMSE for different methods (from 1999 to 2004).

	1999	2000	2001	2002	2003	2004	Average
Yu and Huarng [26] (use U_FTS)	120	176	148	101	74	84	117.4
Huarng et al. [25] (use NASDAQ and DOW Jones and MIB)	N/A	154.42	124.02	95.73	70.76	72.35	103.46
Yu et al. [30] (use B_NN_FTS_S)	112	131	130	80	58	67	96.4
Chen and Chang [8] (use NASDAQ and DOW Jones and MIB)	111.70	129.42	113.67	66.82	56.10	64.76	90.41
Chen and Chen [27] (use NASDAQ and DOW Jones)	116.64	123.62	123.85	71.98	58.06	57.73	91.98
Chen et al. [12] (use NASDAQ and DOW Jones)	101.33	**121.27**	114.48	67.18	52.72	52.27	84.88
Chen and Kao [15] (use PSO and SVM)	**87.63**	125.34	114.57	76.86	54.29	58.17	86.14
The proposed model	102.22	131.53	**112.59**	**60.33**	**51.54**	**50.33**	**84.75**

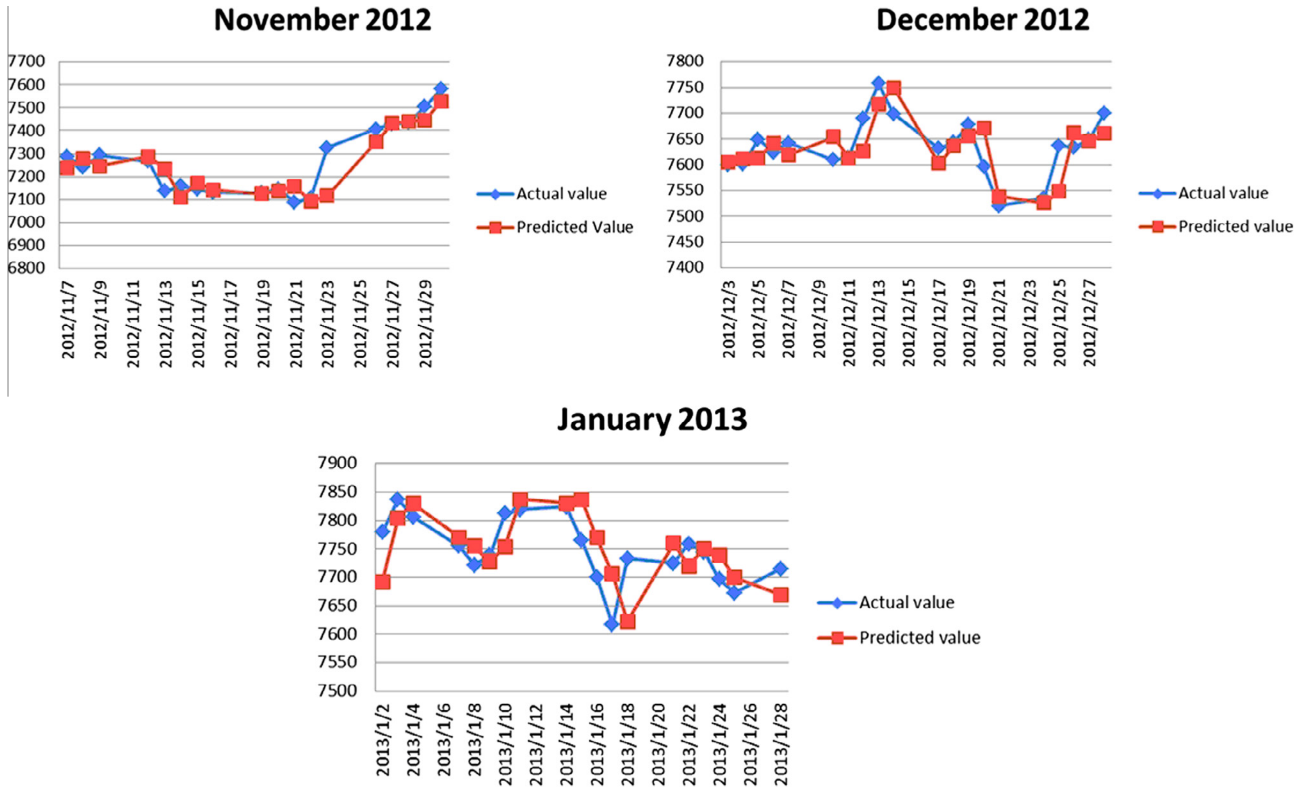

Fig. 6. Comparison of actual value and predicted value obtained by the proposed model.

$$d_t = \bar{d}_t \pm t_{1-\alpha,k-1} * \sigma_d$$

The coefficient $t_{1-\alpha,k-1}$ is obtained from the t-distribution probability table with $(1-\alpha)$ confidence level and $k-1$ degrees of freedom.

According to above formulas, at a 95% confidence level, the confidence interval of d_t can be calculated as:

$$d_{1990-1999} = 0.038 \pm 2.262 * 0.005$$
$$d_{1999-2004} = 0.017 \pm 2.571 * 0.006$$

Notes that, for the years from 1990 to 1999, the compared model is set as Chen and Kao [15] which is the best among the compared models, and for the years from 1999 to 2004, the compared model is Chen et al. [12] which is the best among the compared models. Since two confidence intervals are both above zero, the improvement of the proposed model is statistically significant.

5.3. Forecasting details

After knowing the overall performance, the details of the forecasting were explored further. The TAIEX close prices from January to October on year 2012 are used as the training data and then the forecasting on the three month price from November 2012 to January 2013 is performed. The RMSEs of these 3 months obtained are 64.11, 47.48 and 48.94 respectively. And the average of three months' RMSE is 53.51. In Fig. 6, one can see that the predicted values and the actual values (the detailed results can be obtained from the authors) are very close and the trends of stock prices are well predicted.

6. Conclusions

In this paper, a new model for forecasting the TAIEX based on fuzzy time series has been presented. To address the drawback of the conventional high-order fuzzy time series model, an AR high-order fuzzy time series model is proposed in this paper. While adopting the concept of the classical AR model, the novel high-order model makes a more effective use of the historical data and has been proved to be more suitable for practical use. The paper also proposes a new heuristic method based on ACO algorithm to partition the universe of discourse, which has been the main issue of fuzzy time series models. Compared with multiple factors model, the proposed model, as an univariate model, achieved better results by combining with ACO and AR. The experimental results show that the proposed model gets higher average forecasting accuracy rates than other existing models.

Acknowledgments

The work described in this paper was partially supported by the National Nature Science Foundation of China (jointly by Grant No. 61272003 and Grant No. 61202361), the Major Program of the National Social Science Foundation on China (Grant No. 13&ZD148) and a grant from City University of Hong Kong (Project No. 7002907).

References

[1] Q. Song, B.S. Chissom, Fuzzy time series and its models, Fuzzy Sets Syst. 54 (3) (1993) 269–277.
[2] Q. Song, B.S. Chissom, Forecasting enrollments with fuzzy time series—part I, Fuzzy Sets Syst. 54 (1) (1993) 1–9.
[3] S.M. Chen, Forecasting enrollments based on fuzzy time series, Fuzzy Sets Syst. 81 (3) (1996) 311–319.
[4] S.M. Chen, Forecasting enrollments based on high-order fuzzy time series, Cybernet. Syst.: Int. J. 33 (1) (2002) 1–16.
[5] Q. Song, A note on fuzzy time series model selection with sample autocorrelation functions, Cybernet. Syst.: Int. J. 34 (2) (2003) 93–107.
[6] H.K. Yu, Weighted fuzzy time series models for TAIEX forecasting, Physica A 349 (2005) 609–624.

[7] C.M. Own, P.T. Yu, Forecasting fuzzy time series on a heuristic high-order model, Cybernet. Syst.: Int. J. 36 (7) (2005) 705–717.
[8] S.M. Chen, Y.C. Chang, Multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques, Inform. Sci. 180 (24) (2010) 4772–4783.
[9] K. Huarng, Effective lengths of intervals to improve forecasting in fuzzy time series, Fuzzy Sets Syst. 123 (3) (2001) 387–394.
[10] K. Huarng, Heuristic models of fuzzy time series for forecasting, Fuzzy Sets Syst. 123 (3) (2001) 369–386.
[11] A. Keles, M. Kolcak, A. Keles, The adaptive neuro-fuzzy model for forecasting the domestic debt, Knowl.-Based Syst. 21 (2008) 951–957.
[12] S.M. Chen, H.P. Chu, T.W. Sheu, TAIEX forecasting using fuzzy time series and automatically generated weights of multiple factors, IEEE Trans. Syst. Man Cybernet. Part A Syst. Hum. 42 (6) (2012) 1485–1495.
[13] S.S. Gangwar, S. Kumar, Partitions based computational method for high-order fuzzy time series forecasting, Expert Syst. Appl. 39 (15) (2012) 12158–12164.
[14] S.M. Chen, K. Tanuwijaya, Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques, Expert Syst. Appl. 38 (8) (2011) 10594–10605.
[15] S.M. Chen, P.Y. Kao, TAIEX forecasting based on fuzzy time series, particle swarm optimization techniques and support vector machines, Inform. Sci. 247 (2013) 62–71.
[16] S. Pritpal, B. Bhogeswar, High-order fuzzy-neuro expert system for time series forecasting, Knowl.-Based Syst. 46 (2013) 12–21.
[17] Y.C. Cheng, S.T. Li, Fuzzy time series forecasting with a probabilistic smoothing hidden Markov model, IEEE Trans. Fuzzy Syst. 20 (2) (2012) 291–304.
[18] L.Y. Wei, C.H. Cheng, H.H. Wu, A hybrid ANFIS based on *n*-period moving average model to forecast TAIEX stock, Appl. Soft Comput. 19 (2014) 86–92.
[19] M.Y. Chen, B.T. Chen, A hybrid fuzzy time series model based on granular computing for stock price forecasting, Inform. Sci. 294 (2) (2015) 227–241.
[20] Q. Cai, D. Zhang, B. Wu, S.C.H. Leung (Eds.), A novel stock forecasting model based on fuzzy time series and genetic algorithm, The 13th International Conference on Computational Science (2013 ICCS)Procedia Computer Science 18 (2013) 1155–1162.
[21] K. Huarng, T.H.K. Yu, The application of neural networks to forecast fuzzy time series, Physica A 363 (2) (2006) 481–491.
[22] M. Dorigo, T. Stützle, Ant Colony Optimization, MIT Press, 2004.
[23] M. Dorigo, M. Birattari, T. Stützle, Ant colony optimization – artificial ants as a computational intelligence technique, IEEE Comput. Intell. Mag. 1 (4) (2006) 28–39.
[24] M. Stevenson, J.E. Porter, Fuzzy time series forecasting using percentage change as the universe of discourse, Eng. Technol. 55 (2009) 154.
[25] K.H. Huarng, T.H.K. Yu, Y.W. Hsu, A multivariate heuristic model for fuzzy time-series forecasting, IEEE Trans. Syst. Man Cybernet. B Cybernet. 37 (4) (2007) 836–846.
[26] T.H.K. Yu, K.H. Huarng, A bivariate fuzzy time series model to forecast the TAIEX, Expert Syst. Appl. 34 (4) (2008) 2945–2952.
[27] S.M. Chen, C.D. Chen, TAIEX forecasting based on fuzzy time series and fuzzy variation groups, IEEE Trans. Fuzzy Syst. 19 (1) (2011) 1–12.
[28] E.G. Talbi, O. Roux, C. Fonlupt, D. Robillard, Parallel ant colonies for the quadratic assignment problem, Future Gener. Comput. Syst. 17 (4) (2001) 441–449.
[29] K. Levenberg, A method for the solution of certain non-linear problems in least squares, Quart. Appl. Math. (1944) 164–168.
[30] T.H.K. Yu, K.H. Huarng, Corrigendum to "A bivariate fuzzy time series model to forecast the TAIEX" [Expert Systems with Applications 34 (4) (2010) 2945-2952], Expert Syst. Appl. 37 (2010) 5529.