

A modified genetic algorithm for forecasting fuzzy time series

Eren Bas · Vedide Rezan Uslu · Ufuk Yolcu · Erol Egrioglu

Published online: 11 April 2014
© Springer Science+Business Media New York 2014

Abstract Fuzzy time series approaches are used when observations of time series contain uncertainty. Moreover, these approaches do not require the assumptions needed for traditional time series approaches. Generally, fuzzy time series methods consist of three stages, namely, fuzzification, determination of fuzzy relations, and defuzzification. Artificial intelligence algorithms are frequently used in these stages with genetic algorithms being the most popular of these algorithms owing to their rich operators and good performance. However, the mutation operator of a GA may cause some negative results in the solution set. Thus, we propose a modified genetic algorithm to find optimal interval lengths and control the effects of the mutation operator. The results of applying our new approach to real datasets show superior forecasting performance when compared with those obtained by other techniques.

Keywords Genetic algorithm · Forecasting · Fuzzy time series · Mutation operator

1 Introduction

Fuzzy time series procedures do not require the assumptions, such as a large sample, linear model, and stationary and normal distribution, typically needed by classic time series approaches such as autoregressive moving average models [1]. Recent studies have mainly focused on fuzzy time series since they do not require the strict assumptions and generally provide remarkable forecasting performance. Besides, artificial intelligence optimization algorithms such as genetic algorithm (GA) have been used in almost all areas [2–18].

The fuzzy set was first introduced in [19] and since then has been used in many application areas. Fuzzy time series were first introduced by Song and Chissom [20–22]. The proposed fuzzy time series techniques generally consist of three stages: fuzzification, determining fuzzy relations, and defuzzification.

According to the literature, decomposition of the universe of discourse is mostly used in the fuzzification stage with intervals thereof determined arbitrarily in the studies by Song and Chissom [20–22] and Chen [23, 24]. In addition, Huarng [25] suggested the importance of the interval length on forecasting performance and proposed two new techniques for finding intervals based on the mean and distribution.

Egrioglu et al. [26, 27] suggested considering the problem of finding intervals as an optimization problem. The authors in [28–30] used different interval lengths, instead of a fixed interval length, obtained by using a genetic algorithm, while those in [31–37] used particle swarm optimization. Furthermore, the authors in [38–40] used fuzzy c-means clustering, while those in [41, 42] used Gustafson-Kessel fuzzy clustering in this stage.

E. Bas (✉)
Faculty of Arts and Science, Department of Statistics, Giresun University, Giresun 28100, Turkey
e-mail: eren.bas@giresun.edu.tr

V. R. Uslu · E. Egrioglu
Faculty of Arts and Science, Department of Statistics, Ondokuz Mayıs University, Samsun 55139, Turkey

U. Yolcu
Science Faculty, Department of Statistics, Ankara University, Ankara 06100, Turkey

With respect to determining fuzzy relations, several studies have made contributions. Whereas Song and Chissom [20–22] used matrix operations, Chen [23] and some others used a fuzzy logic relations group (FLRG) table, while yet others [43–47] used artificial neural networks to determine fuzzy relations. Various other methods focusing on this stage have also been proposed [44–46, 48–50].

For the defuzzification process, most reported studies use either the centroid method [23, 25, 51] or the adaptive expectation method [38, 50].

The proposed method contributes to the fuzzification stage by using a modified genetic algorithm (MGA). In particular, we aim to check the negative effects of the mutation operation so as to obtain better forecasts. Moreover, by using MGA we avoid subjective judgments in determining the intervals, and to obtain more realistic results, we find dynamic instead of fixed length intervals.

An outline of the rest of the paper is given below. The fundamental definitions of fuzzy time series are given in Section 2. The third section of the article reviews GAs, while a concise explanation of the differential evaluation algorithm (DEA) is given in Section 4. In Section 5, the proposed method is introduced. Section 6 presents the results of applying the proposed method to three real life datasets and finally, Section 7 presents our conclusions.

2 Fuzzy time series

A definition of fuzzy time series was first introduced in [20, 21]. A number of studies in the literature have used fuzzy time series together with GAs. These include a method to optimize fuzzy time series using GAs [52], an efficient non-linear time series prediction system using a GA and fuzzy time series [53], a forecasting model using a GA with fuzzy time series [54], and a time invariant fuzzy time series forecasting method based on a GA [55]. The authors in [30, 54] also used a GA in the fuzzification stage, which is the first stage of fuzzy time series.

In contrast to conventional time series methods, various theoretical assumptions do not need to be checked in fuzzy time series approaches. The most important advantage of the fuzzy time series approaches is their ability to work with a very small set of data.

Let U be the universe of discourse, where $U = \{u_1, u_2, \dots, u_n\}$. A fuzzy set A_i of U can be defined as

$$A_i = \frac{\mu_{A_i}(u_1)}{u_1} + \frac{\mu_{A_i}(u_2)}{u_2} + \dots + \frac{\mu_{A_i}(u_n)}{u_n}, \quad (1)$$

where μ_{A_i} is the membership function of fuzzy set A_i and $\mu_{A_i}; U \rightarrow [0, 1]$. In addition, $\mu_{A_i}(u_j)$, $j = 1, 2, \dots, n$ denotes a generic element of fuzzy set A_i , and $\mu_{A_i}(u_j)$ is the degree of belongingness of u_j to A_i , with $\mu_{A_i}(u_j) \in [0, 1]$.

Definition 1 Fuzzy time series

Let $Y(t) (t = \dots, 0, 1, 2, \dots)$, a subset of real numbers, be the universe of discourse by which fuzzy sets $f_i(t)$ are defined. If $F(t)$ is a collection of $f_1(t), f_2(t), \dots$, then $F(t)$ is called a fuzzy time series defined on $Y(t)$.

Definition 2 Fuzzy time series relationships

Assume that $F(t)$ is caused only by $F(t-1)$; then the relationship can be expressed as $F(t) = F(t-1) * R(t, t-1)$, which is the fuzzy relationship between $F(t)$ and $F(t-1)$, where $*$ denotes an operator. To summarize, let $F(t-1) = A_i$ and $F(t) = A_j$. The fuzzy logical relationship between $F(t)$ and $F(t-1)$ can be denoted as $A_i \rightarrow A_j$, where A_i (current state) refers to the left-hand side and A_j (next state) refers to the right-hand side of the fuzzy logical relationship. Furthermore, these fuzzy logical relationships can be grouped to establish different fuzzy relationships.

3 Genetic algorithms

GAs were proposed by Holland [56] and developed by Goldberg [57]. A GA is a stochastic global search technique that solves problems by imitating processes observed during natural evolution. The GA procedure is a simulation that depends on biological evolution behavior. The first decision to make when solving a problem using a GA is the encoding. There are many types of encoding in GAs including binary, value, and permutation encoding. First, based on the population size an initial population is randomly generated after determining the encoding and GA operators, namely, crossover, mutation and natural selection, to be applied to the population. For the crossover operation, the researcher determines the crossover rate (cor). Then, a random number is generated with the help of uniform distribution $U(0, 1)$. If cor is greater than or equal to the random number, the crossover operation is performed by randomly choosing both a pair of chromosomes and a crossover point and then swapping the genes of the selected chromosomes.

After the crossover operation, the mutation operator is applied if needed. First, a chromosome is randomly selected and then a mutation point is determined. A gene of the selected chromosome is changed by considering the encoding. Thereafter, the researcher determines the mutation rate (mr) and a random number is generated with the help of uniform distribution $U(0, 1)$. If this number is smaller than or equal to mr , the mutation operation is performed. In the natural selection operation, each chromosome of the population is evaluated using an evaluation function. All chromosomes are ordered according to their corresponding evaluation function values and the best chromosomes are



Fig. 1 The structure of a chromosome in MGA

transferred to the next generation. Some of the worst chromosomes are discarded from the population. The natural selection operation may be used at either the beginning or the end of the algorithm.

4 Differential evaluation algorithm

The DEA, proposed by Storn and Price [58], is a heuristic algorithm based on the population like the GA. It has some operators including mutation and crossover, which are used to create new generations. At the end of the process, candidate solutions are found by using some mathematical operations and these solutions are compared with the current solutions in the population. The best solution based on the evaluation function is transferred to the new generation, while the best chromosome is taken as the optimal solution. For more details of this algorithm, consult the study in [58].

5 Proposed method

As is well known, all the stages of fuzzy time series approaches have a marked influence on the forecasting performance of the applied model. Recently, artificial intelligence algorithms have frequently been used in these stages. GAs are the most popular of these algorithms because of their operators and rapid evaluation process. The crossover and mutation operators prevent achieving sub-optimal solutions. The most important GA operator is mutation, which ensures that solutions are found in as yet unsearched areas.

In classic GAs, the only approach to counteract the loss of relevant alleles, thereby avoiding premature convergence, is mutation. However, as mutation is used as an undirected operator, the probability to recover relevant alleles by lucky mutations decreases rapidly when trying to solve larger problems. In addition to mutation, various other ideas to reduce the negative effects of premature convergence have also been discussed in the literature. Among these,

the most common are preselection [59], crowding [60], and fitness-sharing [57]. The main idea of these approaches is to maintain genetic diversity by replacing solutions occupying similar regions of the search space more frequently (preselection, crowding) or to reduce the fitness value of solutions that are located in densely populated regions of the search space (fitness-sharing). All three approaches require the definition of a distance measure to calculate the similarity of solutions in the search space, while fitness-sharing is additionally quite restricted to fitness proportional selection. As such, these approaches are not applicable in all cases. Furthermore, they do not really address the problem of losing relevant alleles, but instead try to reduce the loss of genetic diversity in general [61].

Chromosomes represent the solution sets in GAs. When applying the mutation operator in a GA, a chromosome is first randomly selected and then a random gene of this selected chromosome is changed. However, the changed gene may be a gene that provides a positive contribution to the solution set. In other words, this gene may be a useful gene for the chromosome. Therefore, changing this gene may have a negative contribution. To avoid changing useful genes of the chromosomes, and thus eliminating a good solution, we propose a new mutation algorithm for fuzzy time series.

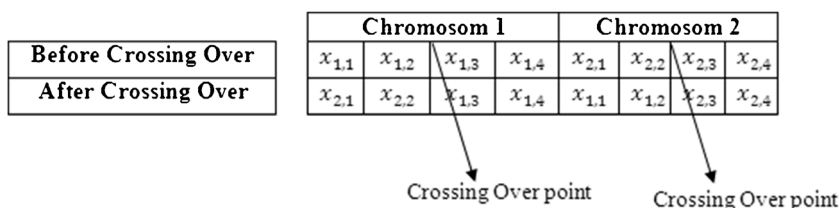
As explained above, applying the mutation operator can be both beneficial and harmful in the GA process. It may alter a harmful gene of a chromosome, but it could also alter a reliable gene thereby moving further from the optimal solution. Thus, to control the effect of the mutation operator we propose a new GA called MGA. In this method, the mutation operator of the GA is inspired by the DEA.

In this study, by combining two population-based artificial intelligence techniques in the fuzzification stage in the form of MGA, we aim to improve the forecasting performance of the model and to determine the interval lengths without resorting to subjective decisions. From this perspective we feel that a more flexible solution process is provided by obtaining dynamic instead of fixed length intervals.

The advantages of our proposed method are as follows:

- ✓ Negative consequences caused by the mutation operator are prevented.
- ✓ More realistic results are obtained by finding dynamic length intervals instead of fixed length ones.

Fig. 2 The crossover operation between two chromosomes



| | Chromosome 1 | | | | Chromosome 2 | | | |
|----------------------|--------------|-------|-------|-------|--------------|-------|-------|-------|
| Before Crossing Over | 13250 | 14000 | 14200 | 16800 | 13700 | 14500 | 17100 | 18000 |
| After Crossing Over | 13700 | 14500 | 14200 | 16800 | 13250 | 14000 | 17100 | 18000 |

Crossover point

Fig. 3 Example of the crossover operation

- ✓ The interval lengths are determined by avoiding subjective decisions through the use of the MGA.

The modified genetic algorithm

Step 1. Define the universe of discourse and number of intervals.

The margins of the universe of discourse and the actual universe are first defined as (D_{min}) and (D_{max}) and

$$U = [D_{min}, D_{max}]. \tag{2}$$

After the number of intervals (m) has been determined, the number of genes in a chromosome in MGA is defined as ($m - 1$). These are represented by $x_i (i = 1, 2, \dots, m - 1)$; the structure of a chromosome in MGA is illustrated in Fig. 1.

Then the margins of all intervals are given by

$$u_1 = [D_{min}, x_1], u_m = [x_{m-1}, D_{max}]. \tag{3}$$

Step 2. Determine the parameters of MGA and generate the initial population.

This step determines the values of the parameters such as the number of chromosomes (cn), cor ($0 < cor \leq 1$), the number of chromosomes to be eliminated in the natural selection stage (dcn), and the number of iterations (itr).

Then, the initial population with ($m - 1$) genes and cn chromosomes is randomly generated from the uniform distribution with parameters (D_{min}, D_{max}). The generated data for each chromosome is ordered from smallest to largest.

Step 3. For each chromosome in the population, the root of the mean squared error ($RMSE$) selected as the evaluation function is calculated by applying steps 3.1 to 3.5.

| | | | | |
|--------------|-------|-------|-------|-------|
| Chromosome 1 | 13700 | 14200 | 14500 | 16800 |
| Chromosome 2 | 13250 | 14000 | 17100 | 18000 |

Fig. 4 The chromosomes after applying the repair operator

| Total vector | Chromosome 2 | Candidate chromosome |
|--------------|--------------|----------------------|
| 13560 | 14600 | 13560 |
| 14840 | 15000 | 14840 |
| 16200 | 15200 | 15200 |
| 17820 | 15600 | 15600 |
| 17960 | 16000 | 16000 |
| 19140 | 17100 | 17100 |

Fig. 5 Example of the mutation operation

Step 3.1. Form fuzzy sets using m intervals based on the values of the genes in the chromosomes as follows:

$$A_i = \frac{a_{i1}}{u_1} + \frac{a_{i2}}{u_2} + \dots + \frac{a_{im}}{u_m} \quad i = 1, 2, \dots, m, \tag{4}$$

where a_{ik} is the membership degree as defined in Eq. (5).

$$a_{ik} = \begin{cases} 1 & k = i \\ 0.5 & , k = i - 1, i + 1, i = 1, 2, \dots, m \\ 0 & , d.d \end{cases} \tag{5}$$

The observations of a crisp time series are converted to fuzzy sets where the interval in which the corresponding observation is included, has the highest membership values.

Step 3.2. Obtain the FLRs and FLRG tables.

When we observe a relation such as $F(t - 1) = A_i$ and $F(t) = A_j$ for any time t , this fuzzy logic relation (FLR) is represented by $A_i \rightarrow A_j$. For the whole series, if we obtain the relation $F(t - 1) = A_i$ and $F(t) = A_k$ for any time t , we express the FLR as $A_i \rightarrow A_j, A_k$. We can also save the number of times an FLR such as $A_i \rightarrow A_j$ occurs, as the weight w_j .

Table 1 Example of a random population

| | | | | | | |
|------|-------|-------|-------|-------|-------|-------|
| Chr1 | 13400 | 14600 | 15800 | 17100 | 17800 | 18500 |
| Chr2 | 14600 | 15000 | 15200 | 15600 | 16000 | 17100 |
| Chr3 | 13500 | 14000 | 14300 | 14700 | 15000 | 16800 |
| Chr4 | 14100 | 15200 | 16300 | 17800 | 18000 | 18900 |
| Chr5 | 13800 | 15100 | 16300 | 1720 | 18300 | 19100 |
| Chr6 | 14000 | 14700 | 15400 | 16100 | 17500 | 18700 |
| Chr7 | 14200 | 15000 | 15900 | 17000 | 17700 | 19500 |

Table 2 Comparative presentation of enrollment forecasts for training set

| Year | Enrollment data | Proposed method | Year | Enrollment data | Proposed method |
|------|-----------------|-----------------|------|-----------------|-----------------|
| 1971 | 13055 | | 1982 | 15433 | 15453,75 |
| 1972 | 13563 | 13697,93 | 1983 | 15497 | 15467,77 |
| 1973 | 13867 | 13697,93 | 1984 | 15145 | 15467,77 |
| 1974 | 14696 | 14621,26 | 1985 | 15163 | 15199,03 |
| 1975 | 15460 | 15453,75 | 1986 | 15984 | 15960,07 |
| 1976 | 15311 | 15467,77 | 1987 | 16859 | 16765,38 |
| 1977 | 15603 | 15467,77 | 1988 | 18150 | 17910,88 |
| 1978 | 15861 | 15467,77 | 1989 | 18970 | 19028,93 |
| 1979 | 16807 | 16765,38 | 1990 | 19328 | 19260,91 |
| 1980 | 16919 | 16935,45 | 1991 | 19337 | 19076,86 |
| 1981 | 16388 | 16367,5 | 1992 | 18876 | 19076,86 |

Table 3 Comparison of enrollment forecasts for training set in terms of RMSE and MAPE

| Methods | RMSE | MAPE |
|-----------------|------|--------|
| [62] | 366 | 1.72 % |
| [63] | 295 | 1.56 % |
| [64] | 668 | 2.75 % |
| [64] | 511 | 2.66 % |
| [65] | 367 | 1.87 % |
| [66] | 501 | 2.67 % |
| [25] | 476 | 2.45 % |
| [23] | 638 | 3.11 % |
| [21] | 650 | 3.22 % |
| [30] | 178 | 0.90 % |
| Proposed method | 160 | 0.73 % |

Step 3.3. Obtain fuzzy forecasts.

Fuzzy forecasts are obtained with respect to the FLR table formed in Step 3.2. For example; if $F(t - 1) = A_i$ and there is a relation such as $A_i \rightarrow A_j$ in the FLR table, the fuzzy forecast will be A_j . If $F(t - 1) = A_i$ and there is a relation such as $A_i \rightarrow A_j, A_k$ in the FLR table, the fuzzy forecast will be A_j, A_k . If $F(t - 1) = A_i$ and $A_i \rightarrow Empty$ in the FLR table, the fuzzy forecast will be A_j .

Step 3.4. Defuzzify the fuzzy forecasts.

The weights w_j obtained from the FLR table are used in the defuzzification stage.

For example, if $F(t - 1) = A_i$ and there exists relation $A_i \rightarrow A_j$ in the table, the defuzzified forecast will be m_j , which is the midpoint of u_j , which is the subinterval of fuzzy set A_j with the largest membership degree. That is, we

do not consider how many times the relation is repeated in the table.

If $F(t - 1) = A_i$ and there exists relation $A_i \rightarrow A_j, A_k$ and w_j denotes the number of times relation $A_i \rightarrow A_j$ is repeated in the whole time series and w_k denotes the number of times relation $A_i \rightarrow A_k$ is repeated, the defuzzified forecast is calculated as

$$\hat{x}_t = \frac{w_j m_j + w_k m_k}{w_j + w_k} \tag{6}$$

If $F(t - 1) = A_i$ and there exists relation $A_i \rightarrow Empty$ in the FLR table, the defuzzified forecast will be m_i , which is the midpoint of subinterval u_i , which is the fuzzy set A_i with the largest membership degree.

Step 3.5. Let x_t be the original time series and \hat{x}_t be its defuzzified forecasts with n observations. $RMSE$ is calculated as

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (x_t - \hat{x}_t)^2}{n}} \tag{7}$$

Step 4. Save the chromosome with the minimum $RMSE$ value in the population.

Step 5. Apply the MGA parameters.

Natural selection, crossover, and mutation operations are applied.

Table 4 Comparison of enrollment forecasts for training set in terms of RMSE

| [48] | [38] | [44] | [27] | [26] | [41] | [67] | Proposed method |
|------|------|------|------|------|------|------|-----------------|
| 621 | 478 | 279 | 258 | 246 | 245 | 215 | 160 |

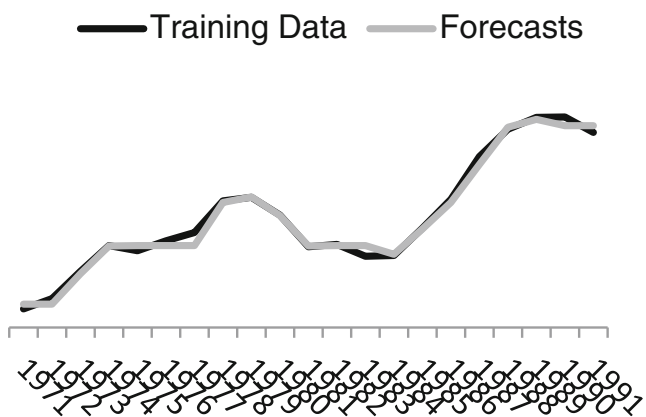


Fig. 6 The forecasts and actual values of enrollment data for training set

Step 5.1. Initially, the first *dcn* chromosomes in the population are removed from the list of the ordered RMSE values from largest to smallest. Then, *dcn* new chromosomes, generated randomly as in Step 2, are added to the population.

Step 5.2. To determine whether the crossover operation should be applied, a number is randomly generated from the uniform distribution with parameters (0,1). If the number is smaller than *cor*, the crossover operation is applied. When applying the crossover operation two chromosomes are randomly chosen from the population. The crossover point at which the genes are swapped is also randomly determined. The crossover operation is illustrated in Fig. 2.

To better understand the crossover operation, consider the example in Fig. 3 with the universe of discourse $U = [13000, 20000]$ and assuming that two random chromosomes in the population have been chosen.

If there is a problem after crossover, such as the first chromosome in Fig. 4 not being sorted, a repair operator must be used to sort the chromosomes in ascending order

Step 5.3. After applying the crossover operator, mutation is performed. First, the chromosome to be mutated is selected randomly. Then three chromosomes, differing from the chromosome to be mutated, are selected randomly. The first two chromosomes are subtracted from each other to form the difference vector. Then the difference vector is multiplied by parameter *F* (this parameter takes values between 0 and 2, but is generally set to 0.8). This new chromosome, called the weighted difference vector is summed with the third chromosome to obtain the total vector. Next, the mutation rate (*mr*) is determined and

a random number is generated between 0 and 1 with the help of the uniform distribution.

If this random number is smaller than *mr*, the gene is taken from the total vector. If not, the gene is taken from the interested chromosome and a candidate chromosome is generated and its fitness value calculated.

The candidate chromosome and interested chromosome are compared in terms of fitness values. The chromosome with the smaller *RMSE* value, used to evaluate the function, is transferred to the population. The mutation operation is illustrated in Fig. 5.

Assume a population with seven chromosomes and universe of discourse $U = [13000, 20000]$ as given in Table 1.

To better understand the mutation operation, consider the example given in Fig. 5 assuming that Chromosome 2 is the chromosome to be mutated and Chromosomes 6, 7, and 1 are the chromosomes randomly selected in addition to Chromosome 2.

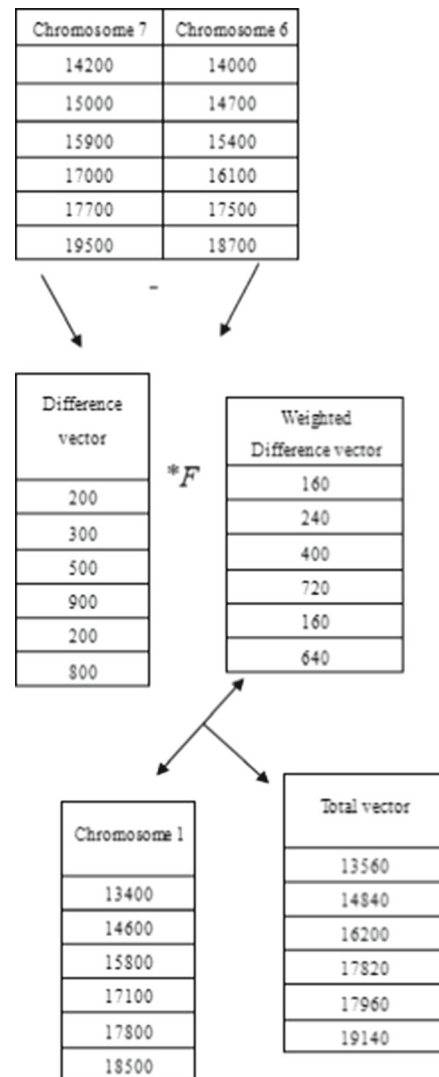


Table 5 Comparison of the results for test set

| Methods | RMSE |
|-----------------|----------|
| [20] | 642.2609 |
| [22] | 880.7309 |
| [48] | 621.3332 |
| [23] | 638.3627 |
| [25] | 280.6991 |
| [68] | 353.1388 |
| [27] | 258.1879 |
| [69] | 267.847 |
| [41] | 245.2346 |
| Proposed method | 215.4553 |

Table 6 Results obtained from conventional time series forecasting methods

| Methods | RMSE |
|-----------------------------|----------|
| Brown exponential smoothing | 2469.727 |
| Logarithmic regression | 2332.806 |
| Inverse regression | 3001.306 |
| Quadratic regression | 2865.191 |
| Cubic regression | 2412.646 |
| Compound regression | 1605.017 |
| Power regression | 3009.670 |
| S-curve regression | 1605.010 |
| Growth regression | 1605.017 |
| Exponential regression | 1605.017 |
| ARIMA(0,1,0) method | 477.468 |

Let the mutation ratio be 0.10, and then we generate random numbers for each gene, (0.01, 0.08, 0.15, 0.20, 0.16), respectively, to obtain the candidate chromosome.

Finally, we calculate the *RMSE* value of the candidate chromosome. If its *RMSE* value is smaller than that of Chromosome 2, the new chromosome is taken as the candidate chromosome and the mutation operation is completed. If not, there is no need to apply the mutation operation to the population, because this mutated chromosome may be the optimal solution at the end of the process.

Table 7 Comparison of the results for test set

| TAIFEX Test data | [54] | [29] | [44] | [37] | [70] | Proposed method |
|------------------|---------|--------|--------|---------|--------|-----------------|
| 6709.75 | 6621.43 | 6917.4 | 6850 | 6745.45 | 6750 | 6760.529 |
| 6726.5 | 6677.48 | 6852.2 | 6850 | 6757.89 | 6750 | 6739.995 |
| 6774.55 | 6709.63 | 6805.7 | 6850 | 6731.76 | 6850 | 6739.995 |
| 6762 | 6732.02 | 6762.4 | 6850 | 6722.54 | 6850 | 6760.529 |
| 6952.75 | 6753.38 | 6793.1 | 6850 | 6753.72 | 6850 | 6760.529 |
| 6906 | 6756.02 | 6784.4 | 6850 | 6761.54 | 6850 | 6869.525 |
| 6842 | 6804.26 | 6970.7 | 6850 | 6857.27 | 6850 | 6869.525 |
| 7039 | 6842.04 | 6977.2 | 6850 | 6898.97 | 6850 | 6869.525 |
| 6861 | 6839.01 | 6874.5 | 6850 | 6853.07 | 6950 | 6869.525 |
| 6926 | 6897.33 | 7126.1 | 6850 | 6951.95 | 6850 | 6869.525 |
| 6852 | 6896.83 | 6862.5 | 6850 | 6896.84 | 6850 | 6869.525 |
| 6890 | 6919.27 | 6944.4 | 6850 | 6919.94 | 6850 | 6869.525 |
| 6871 | 6903.36 | 6831.9 | 6850 | 6884.99 | 6850 | 6869.525 |
| 6840 | 6895.95 | 6843.2 | 6850 | 6894.10 | 6850 | 6869.525 |
| 6806 | 6879.31 | 6858.5 | 6850 | 6866.17 | 6850 | 6760.529 |
| 6787 | 6878.34 | 6825.6 | 6850 | 6865.06 | 6750 | 6760.529 |
| RMSE | 93.49 | 102.96 | 83.58 | 80.02 | 72.55 | 70.42 |
| MAPE | 1.09 % | 1.14 % | 0.96 % | 0.87 % | 0.82 % | 0.66 |

Then we can obtain the intervals given below:

$$\begin{aligned}
 u_1 &= [13000, 13560], u_2 = [13560, 14840], \\
 u_3 &= [14840, 15200], u_4 = [15200, 15600], \\
 u_5 &= [15600, 16000], u_6 = [16000, 17100], \\
 u_7 &= [17100, 20000]
 \end{aligned}$$

Step 6. Steps 3 to 5 are repeated as many times as defined by the iteration number, which was previously determined by the researcher.

6 Application

To verify the performance of the proposed method, we applied it to three different time series datasets. The results are compared with those from other fuzzy time series methods in terms of RMSE and mean absolute percent error (MAPE).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|X_t - \hat{X}_t|}{X_t} * 100. \tag{8}$$

For each of the time series datasets, MGA parameters were set as follows:

- *cn* was varied between 20 and 100 in increments of 10.
- *cor* was varied between 0.1 and 1 in increments of 0.1.
- For each chromosome, *dcn* was set to 7, 10, 13, 17, 20, 23, 26, 30, and 33.

Table 8 Comparison of “killed” forecasts by various methods and the actual observations

| Year | Actual killed | [29] | [71] | [30] | [72] | Proposed method |
|------|---------------|--------|--------|--------|--------|-----------------|
| 1974 | 1574 | | 1497 | | | |
| 1975 | 1460 | | 1497 | 1506 | | 1458 |
| 1976 | 1536 | | 1497 | 1453 | | 1467 |
| 1977 | 1597 | 1500 | 1497 | 1598 | 1497 | 1606 |
| 1978 | 1644 | 1500 | 1497 | 1584 | 1497 | 1592 |
| 1979 | 1572 | 1500 | 1497 | 1584 | 1497 | 1592 |
| 1980 | 1616 | 1500 | 1497 | 1506 | 1598 | 1637 |
| 1981 | 1564 | 1500 | 1497 | 1584 | 1598 | 1592 |
| 1982 | 1464 | 1500 | 1497 | 1506 | 1498 | 1472 |
| 1983 | 1479 | 1500 | 1497 | 1453 | 1498 | 1409 |
| 1984 | 1369 | 1500 | 1497 | 1375 | 1398 | 1409 |
| 1985 | 1308 | 1400 | 1396 | 1383 | 1298 | 1366 |
| 1986 | 1456 | 1300 | 1296 | 1454 | 1498 | 1428 |
| 1987 | 1390 | 1500 | 1497 | 1453 | 1398 | 1467 |
| 1988 | 1432 | 1400 | 1396 | 1383 | 1398 | 1366 |
| 1989 | 1488 | 1400 | 1396 | 1509 | 1498 | 1467 |
| 1990 | 1574 | 1500 | 1497 | 1598 | 1598 | 1606 |
| 1991 | 1471 | 1500 | 1497 | 1506 | 1498 | 1458 |
| 1992 | 1380 | 1500 | 1497 | 1375 | 1398 | 1409 |
| 1993 | 1346 | 1400 | 1396 | 1383 | 1298 | 1366 |
| 1994 | 1415 | 1300 | 1296 | 1383 | 1398 | 1428 |
| 1995 | 1228 | 1400 | 1396 | 1231 | 1198 | 1223 |
| 1996 | 1122 | 1100 | 1095 | 1135 | 1098 | 1144 |
| 1997 | 1150 | 1200 | 1196 | 1180 | 1198 | 1097 |
| 1998 | 1224 | 1200 | 1196 | 1245 | 1198 | 1238 |
| 1999 | 1173 | 1200 | 1196 | 1135 | 1198 | 1144 |
| 2000 | 1253 | 1300 | 1296 | 1245 | 1298 | 1238 |
| 2001 | 1288 | 1300 | 1296 | 1284 | 1298 | 1278 |
| 2002 | 1145 | 1100 | 1095 | 1143 | 1098 | 1114 |
| 2003 | 1035 | 1000 | 995 | 970 | 997 | 1097 |
| 2004 | 953 | 1000 | 995 | 970 | 997 | 929 |
| | RMSE | 85.35 | 83.12 | 41.61 | 46.78 | 37.66 |
| | MAPE | 5.25 % | 5.06 % | 2.29 % | 2.70 % | 2.33 % |

- For all possible cases derived from the above, MGA was executed 100 times in MATLAB.
- m was varied from 5 to 20.

At the end of the process, 1440 different solutions were obtained. The parameters (m , cn , cor , dcn) with the smallest RMSE value were taken as the optimal solution of all these solutions.

6.1 Enrollment data

The performance of the proposed method was evaluated separately for test and training datasets of enrollment data. First all data were used as the training set like almost all

studies in the literature and then the last three observations of the enrollment data were used for the test set. Results obtained by the proposed method were compared with those of other studies, clearly showing that our proposed method has superior forecasting performance.

As the first experiment, the proposed method was applied to the training dataset. The best result was obtained with $m = 17$, $cn = 50$, $cor = 0.01$, and $dcn = 17$. Table 2 presents a comparison of the forecasts together with RMSE and MAPE values obtained from the proposed method and various other methods from the literature. The results from both belong to the best case. The graph of the forecasts of proposed method and real data is given in Fig. 6.

Next, a comparison of enrollment forecasts in terms of RMSE and MAPE values for various methods is given in Table 3.

Finally, a comparison of enrollment forecasts in terms of RMSE values for various methods is given in Table 4.

As the second experiment, the proposed method was applied to the test dataset. The best result was obtained with $m = 12$, $cor = 60$, $mr = 0.05$, and $dcn = 20$. Table 5 presents a comparison of all the RMSE values obtained from the proposed method and various other methods.

In addition, the results obtained from conventional time series forecasting methods are given in Table 6.

6.2 Taifex data

In the second case, the proposed method was applied to TAIFEX data, with observations between 03.08.1998 and 30.09.1998. The last 16 observations were used as the test dataset. The best result was obtained with $m = 11$, $cn = 60$, $cor = 0.05$, and $dcn = 20$. Table 7 gives a comparison of the forecasts as well as RMSE and MAPE values obtained from the proposed method and various other methods from the literature. The results from both belong to the best case.

6.3 Data from vehicle road accidents in Belgium

In the third case, the proposed method was applied to time series data for 'killed in car road accidents in Belgium'. The best result was obtained with $m = 18$, $cn = 70$, $cor = 0.05$, and $dcn = 23$. Table 8 shows a comparison of all the results, including forecasts and RMSE and MAPE values, obtained from the proposed method and various other methods. The results from both belong to the best case.

7 Conclusion

Researchers have recently found that there are many factors affecting forecasting performance of various fuzzy time series approaches. Although the GA is a popular optimization algorithm owing to its process, the mutation operator can cause unexpected solutions in the population. Thus, we proposed a modified genetic algorithm to avoid any unexpected solutions.

In this study, we used the MGA in the fuzzification stage, combining two-population based algorithms, DEA and GA, to control the harmful effects of the mutation operator. By avoiding the harmful effects of the mutation operation, superior forecasting performance is obtained with the proposed method.

We expect that in future studies researchers will concentrate on different optimization techniques in the defuzzification stage.

Acknowledgments The authors would like to thank the anonymous reviewers and editors of journal for their valuable comments and suggestions to improve the quality of the paper.

References

1. Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control. Holden-Day, San Francisco
2. Khan SA, Engelbrecht AP (2012) A fuzzy particle swarm optimization algorithm for computer communication network topology design. Appl Intell 36(1):161–177
3. Kang JG, Kim S, An SY, Oh SY (2012) A new approach to simultaneous localization and map building with implicit model learning using neuro evolutionary optimization. Appl Intell 36(1):242–269
4. Pan WT (2012) The use of genetic programming for the construction of a financial management model in an enterprise. Appl Intell 36(2):271–279
5. Kang MH, Choi HR, Kim HS, Park BJ (2012) Development of a maritime transportation planning support system for car carriers based on genetic algorithm. Appl Intell 36(3):585–604
6. Ali YMB (2012) Psychological model of particle swarm optimization based multiple emotions. Appl Intell 36(3):649–663
7. Shin KS, Jeong YS, Jeoung MK (2012) A two-leveled symbiotic evolutionary algorithm for clustering problems. Appl Intell 36(4):788–799
8. Qu R, Xing HA (2012) A compact genetic algorithm for the network coding based resource minimization problem. Appl Intell 36(4):809–823
9. Szlapczynski R, Szlapczynska R (2012) On evolutionary computing in multi-ship trajectory planning. Appl Intell 37(2):155–174
10. Malek H, Ebadzadeh M, Rahmati M (2012) Three new fuzzy neural networks learning algorithms based on clustering, training error and genetic algorithm. Appl Intell 37(2):280–289
11. Norouzzadeh MS, Ahmadzadeh MR, Palhang M (2012) LADPSO: using fuzzy logic to conduct PSO algorithm. Appl Intell 37(2):290–304
12. Cuevas E, Senci3n F, Zaldivar D, Cisneros MP, Sossa H (2012) A multi-threshold segmentation approach based on Artificial Bee Colony optimization. Appl Intell 37(3):321–336
13. Zheng YJ, Wang K (2012) A new particle swarm optimization algorithm for fuzzy optimization of armored vehicle scheme design. Appl Intell 37(4):520–526
14. Karaboga N, Kockanat S, Dogan H (2013) The parameter extraction of the thermally annealed Schottky barrier diode using the modified artificial bee colony. Appl Intell 38(3):279–288
15. Masoud H, Jalili S, Hasheminejad SMH (2013) Dynamic clustering using combinatorial particle swarm optimization. Appl Intell 38(3):289–314
16. Vera CM, Cano A, Romero C, Ventura S (2013) Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. Appl Intell 38(3):315–330
17. Han MF, Liao SH, Chang JY, Lin CT (2013) Dynamic group-based differential evolution using a self-adaptive strategy for global optimization problems. Appl Intell 39(1):41–56
18. Abbasian R, Mouhoub M (2013) A hierarchical parallel genetic approach for the graph coloring problem. Appl Intell 39(3):510–528
19. Zadeh LA (1965) Fuzzy sets. Inform Control 8:338–353
20. Song Q, Chissom BS (1993a) Fuzzy time series and its models. Fuzzy Sets Syst 54:269–277

21. Song Q, Chissom BS (1993b) Forecasting enrollments with fuzzy time series- Part I. *Fuzzy Sets Syst* 54:1–10
22. Song Q, Chissom BS (1994) Forecasting enrollments with fuzzy time series- Part II. *Fuzzy Sets Syst* 62:1–8
23. Chen SM (1996) Forecasting enrollments based on fuzzy time-series. *Fuzzy Sets Syst* 81:311–319
24. Chen SM (2002) Forecasting enrollments based on high order fuzzy time series. *Cybern Syst* 33:1–16
25. Huarng K (2001a) Effective length of intervals to improve forecasting in fuzzy time-series. *Fuzzy Sets Syst* 123:387–394
26. Egrioglu E, Aladag CH, Yolcu U, Uslu VR, Basaran MA (2010) Finding an optimal interval length in high order fuzzy time series. *Expert Syst Appl* 37:5052–5055
27. Egrioglu E, Aladag CH, Basaran MA, Uslu VR, Yolcu U (2011a) A new approach based on the optimization of the length of intervals in fuzzy time series. *J Intell Fuzzy Syst* 22:15–19
28. Chen SM, Chung NY (2006) Forecasting enrolments using high order fuzzy time series and genetic algorithms. *Int J Intell Syst* 21:485–501
29. Lee LW, Wang LH, Chen SM, Leu YH (2006) Handling forecasting problems based on two factor high-order fuzzy time series. *IEEE Trans Fuzzy Syst* 14(3):468–477
30. Uslu VR, Bas E, Yolcu U, Egrioglu E (2013), A fuzzy time series approach based on weights determined by the number of recurrences of fuzzy relations. *Swarm Evol Comput*. doi:10.1016/j.swevo.2013.10.004
31. Fu FP, Chi K, Che WG, Zhao QJ (2010) High-order difference heuristic model of fuzzy time series based on particle swarm optimization and information entropy for stock markets. In: International conference on computer design and applications
32. Huang YL, Horng SJ, Kao TW, Run RS, Lai JL, Chen RJ, Kuo IH, Khan MK (2011) An improved forecasting model based on the weighted fuzzy relationship matrix combined with a PSO adaptation for enrollments. *Int J Innov Comput Inform Control* 7(7):4027–4046
33. Kuo I-H, Horng S-J, Kao T-W, Lin T-L, Lee C-L, Pan Y (2009) An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. *Expert Syst Appl* 36:6108–6117
34. Kuo I-H, Horng S-J, Chen Y-H, Run R-S, Kao T-W, Chen R-J, Lai J-L, Lin T-L (2010) Forecasting TAIEX based on fuzzy time series and particle swarm optimization. *Expert Syst Appl* 37:1494–1502
35. Davari S, Zarandi MHF, Turksen IB (2009) An improved fuzzy time series forecasting model based on particle swarm intervalization. In: The 28th North American fuzzy information processing society annual conferences (NAFIPS 2009). Cincinnati, Ohio, USA, June 14–17
36. Park J-I, Lee D-J, Song C-K, Chun M-G (2010) TAIEX and KOSPI 200 forecasting based on two factors high order fuzzy time series and particle swarm optimization. *Expert Syst Appl* 37:959–967
37. Hsu L-Y, Horng S-J, Kao T-W, Chen Y-H, Run R-S, Chen R-J, Lai J-L, Kuo I-H (2010) Temperature prediction and TAIEX forecasting based on fuzzy relationships and MTPSO techniques. *Expert Syst Appl* 37:2756–2770
38. Cheng CH, Chen TL, Teoh HJ, Chiang CH (2008) Fuzzy time series based on adaptive expectation model for TAIEX forecasting. *Expert Syst Appl* 34:1126–1132
39. Li ST, Cheng YC, Lin SY (2008) FCM-based deterministic forecasting model for fuzzy time series. *Comput Math Appl* 56:3052–3063
40. Alpaslan F, Cagcag O, Aladag CH, Yolcu U, Egrioglu E (2012) A novel seasonal fuzzy time series method. *Haceteppe J Math Stat* 4(3):375–385
41. Egrioglu E, Aladag CH, Yolcu U, Uslu VR, Erilli NA (2011b) Fuzzy time series forecasting method based on Gustafson–Kessel fuzzy clustering. *Expert Syst Appl* 38:10355–10357
42. Alpaslan F, Cagcag O (2012) A seasonal fuzzy time series forecasting method based on Gustafson-Kessel fuzzy clustering. *J Soc Econ Stat* 1:1–13
43. Huarng K, Yu HK (2006) The application of neural networks to forecast fuzzy time series. *Phys A* 363:481–491
44. Aladag CH, Basaran MA, Egrioglu E, Yolcu U, Uslu VR (2009) Forecasting in high order fuzzy time series by using neural networks to define fuzzy relations. *Expert Syst Appl* 36:4228–4231
45. Egrioglu E, Aladag CH, Yolcu U, Uslu VR, Basaran MA (2009a) A new approach based on artificial neural networks for high order multivariate fuzzy time series. *Expert Syst Appl* 36:10589–10594
46. Egrioglu E, Aladag CH, Yolcu U, Basaran MA, Uslu VR (2009b) A new hybrid approach based on Sarima and partial high order bivariate fuzzy time series forecasting model. *Expert Syst Appl* 36:7424–7434
47. Egrioglu E, Uslu VR, Yolcu U, Basaran MA, Aladag CH (2009c) A new approach based on artificial neural networks for high order bivariate fuzzy time series. In: Mehnen J, et al. (eds) Applications of soft computing AISC 58. Springer-Verlag, Berlin Heidelberg, pp 265–273
48. Sullivan J, Woodall WH (1994) A comparison of fuzzy forecasting and Markov modeling. *Fuzzy Sets Syst* 64(3):279–293
49. Yolcu U, Aladag CH, Egrioglu E, Uslu VR (2013) Time series forecasting with a novel fuzzy time series approach: an example for stanbul stock market. *J Comput Stat Simul* 83:597–610
50. Aladag CH, Yolcu U, Egrioglu E (2010) A high order fuzzy time series forecasting model based on adaptive expectation and artificial neural networks. *Math Comput Simul* 81:875–882
51. Huarng K, Yu THK (2006) Ratio-based lengths of intervals to improve fuzzy time series forecasting. *IEEE Trans Syst Man Cybern-Part B: Cybern* 36:328–340
52. Ringwood JV (1997) Optimization of fuzzy electricity forecasting models using genetic algorithms. Proceedings of the 5th European congress on intelligent techniques and soft computing Aachen Germany, vol. 3, pp 2457–2464
53. Kang HII (2005) A fuzzy time series prediction method using the evolutionary algorithm. *Adv Intell Comput Lect Notes Comput Sci* 3654:530–537
54. Lee LW, Wang LH, Chen SM (2007) Temperature prediction and TAIEX forecasting based on fuzzy logical relationships and genetic algorithms. *Expert Syst Appl* 33:539–550
55. Egrioglu E (2012) A new time invariant fuzzy time series forecasting method based on genetic algorithm. *Advances in Fuzzy Systems*, Article ID 785709, p 6
56. Holland JH (1975) *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor
57. Goldberg DE (1985) *Optimal initial population size for binary-coded genetic algorithms*. Department of Engineering Mechanics, University of Alabama, Alabama
58. Storn R, Price K (1995) *Differential evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces*. Technical Report TR-95-012, International Computer Science Institute, Berkeley
59. Cavicchio DJ (1970) *Adaptive search using simulated evolution*. Thesis (PhD). University of Michigan
60. De Jong K (1975) *An analysis of the behavior of a class of genetic adaptive systems*. Thesis (PhD). University of Michigan

61. Wagner S, Affenzeller M, Beham A, Kronberger GK, Winkler SM (2010) Mutation effects in genetic algorithms with offspring selection applied to combinatorial optimization problems. In: Proceeding of 22nd European modeling and simulation symposium EMSS
62. Singh SR (2007a) A simple method of forecasting based on fuzzy time series. *Appl Math Comput* 186:330–339
63. Singh SR (2007b) A robust method of forecasting based on fuzzy time series. *Appl Math Comput* 188:472–484
64. Cheng CH, Chang RJ, Yeh CA (2006) Entropy-based and trapezoid fuzzification-based fuzzy time series approach for forecasting IT project cost. *Technol Forecast Soc Chang* 73:524–542
65. Tsaur RC, Yang JCO, Wang HF (2005) Fuzzy relation analysis in fuzzy time series model. *Comput Math Appl* 49:539–548
66. Lee HS, Chou MT (2004) Fuzzy forecasting based on fuzzy time series. *Int J Comput Math* 817:781–789
67. Aladag CH, Yolcu U, Egrioglu E, Dalar AZ (2012) A new time invariant fuzzy time series forecasting method based on particle swarm optimization. *Appl Soft Comput* 12:3291–3299
68. Huarng K (2001b) Heuristic models of fuzzy time series for forecasting. *Fuzzy Sets Syst* 123:369–386
69. Cheng C-H, Cheng G-W, Wang J-W (2008) Multi-attribute fuzzy time series method based on fuzzy clustering. *Expert Syst Appl* 34:1235–1242
70. Aladag CH (2013) Using multiplicative neuron model to establish fuzzy logic relationships. *Expert Syst Appl* 40:850–853
71. Jilani TA, Burney SMA, Ardil C (2007) Multivariate high order fuzzy time series forecasting for car road accident. *World Acad Sci Eng Technol* 25:288–293
72. Jilani TA, Burney SMA (2008) Multivariate stochastic fuzzy forecasting models. *Expert Syst Appl* 353:691–700