Engineering Applications of Artificial Intelligence 26 (2013) 2443–2457

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

An efficient time series forecasting model based on fuzzy time series

Pritpal Singh [a,b,*], Bhogeswar Borah [b]

[a] School of Mathematics and Computer Applications, Thapar University, Patiala 04, Punjab, India
[b] Department of Computer Science and Engineering, Tezpur University, Tezpur 28, Assam, India

ARTICLE INFO

Article history:
Received 23 January 2013
Received in revised form
2 July 2013
Accepted 8 July 2013
Available online 30 August 2013

Keywords:
Fuzzy time series
Enrollment
Temperature
Stock exchange
Discretization
Defuzzification

ABSTRACT

In this paper, we present a new model to handle four major issues of fuzzy time series forecasting, viz., determination of effective length of intervals, handling of fuzzy logical relationships (FLRs), determination of weight for each FLR, and defuzzification of fuzzified time series values. To resolve the problem associated with the determination of length of intervals, this study suggests a new time series data discretization technique. After generating the intervals, the historical time series data set is fuzzified based on fuzzy time series theory. Each fuzzified time series values are then used to create the FLRs. Most of the existing fuzzy time series models simply ignore the repeated FLRs without any proper justification. Since FLRs represent the patterns of historical events as well as reflect the possibility of appearances of these types of patterns in the future. If we simply discard the repeated FLRs, then there may be a chance of information lost. Therefore, in this model, it is recommended to consider the repeated FLRs during forecasting. It is also suggested to assign weights on the FLRs based on their severity rather than their patterns of occurrences. For this purpose, a new technique is incorporated in the model. This technique determines the weight for each FLR based on the index of the fuzzy set associated with the current state of the FLR. To handle these weighted FLRs and to obtain the forecasted results, this study proposes a new defuzzification technique. The proposed model is verified and validated with three different time series data sets. Empirical analyses signify that the proposed model have the robustness to handle one-factor time series data set very efficiently than the conventional fuzzy time series models. Experimental results show that the proposed model also outperforms over the conventional statistical models.

1. Introduction

Time series data are highly non-stationary and uncertain in nature. Researchers have developed numerous models based on fuzzy time series concept to overcome these limitations. In this paper, we also present a novel forecasting model, which is developed by employing fuzzy time series. The main aim of designing such a model is explained next.

For fuzzification of time series data set, the determination of length of intervals is very important. In most of the fuzzy time series models (Song and Chissom, 1993a, 1994; Chen, 1996; Hwang et al., 1998; Huarng, 2001a), length of the intervals was kept the same. No specific reason is mentioned for using the fix length of intervals. Huarng (2001b) shows that effective length of intervals always affect the results of forecasting. Therefore, for the creation of effective length of intervals of the historical time series data set, a new "Mean-Based Discretization (MBD)" approach is incorporated in the model.

After generating the intervals, the historical time series data sets are fuzzified based on fuzzy time series theory. Each fuzzified time series values are then used to create the FLRs. Still most of the existing fuzzy time series models ignore repeated FLRs. To explain this, consider the following four FLRs at four different time functions, $F(t=1,2,3,4)$ as

$$\begin{aligned} &F(t=4) \quad B_i \rightarrow B_i, \\ &F(t=3) \quad B_k \rightarrow B_j, \\ &F(t=2) \quad B_i \rightarrow B_i, \\ &F(t=1) \quad B_i \rightarrow B_j. \end{aligned} \tag{1}$$

In Eq. (1), three FLRs at functions $F(t=4)$, $F(t=2)$ and $F(t=1)$ have the same fuzzy set, (B_i), in the previous state. Hence, these FLRs can be represented in the following fuzzy logical relationship group (FLRG) as

$$B_i \rightarrow B_i, B_j. \tag{2}$$

Since existing fuzzy time series models do not consider the identical FLRs during forecasting. They simply use the FLR as shown in Eq. (2) by discarding the repeated FLRs in the FLRG. The ignorance of repeated FLRs in the FLRG is not properly justified by these models (e.g., refer to some papers: Chen, 1996; Hwang et al., 1998; Huarng, 2001a; Huarng et al., 2007). Since each FLR represents frequency of

* Corresponding author at: Department of Computer Science and Engineering, Tezpur University, Tezpur 28, Assam, India. Tel.: +91 9957128450.
E-mail addresses: pritpal@tezu.ernet.in, pritpal.singh@thapar.edu (P. Singh), bgb@tezu.ernet.in (B. Borah).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.07.012

occurrence of the corresponding event in the past. For example, in Eq. (1), $B_i \rightarrow B_i$ occurs three times and $B_i \rightarrow B_j$ appears only once. These occurrences represent the patterns of historical events as well as reflect the possibility of appearances of these types of patterns in the future. If we simply discard the repeated FLRs, then there is a chance of information lost. This can impact on robustness as well as effectiveness of the model. Hence, to utilize more information, we propose the following approach to represent the FLRs of Eq. (2) in FLRG as

$$B_i \rightarrow B_i, B_j, B_i, B_j. \tag{3}$$

Hence, to check the performance of the model, it is evaluated with the repeated FLRs. Empirical analyses signify that the proposed model outperforms over existing competing models that ignore the repeated FLRs.

In existing fuzzy time series models, each FLR is given equal importance, which is not effective to solve real time problems. Because, each fuzzy set in the FLR represents various uncertainties involved in the domain. According to Yu (2005), there are two possible ways to assign weights, i.e., (i) assign weights based on human interpretation, and (ii) assign weights based on their chronological order. Assignment of weights based on human-knowledge is not an acceptable solution for real world problems as human-interpretation varies from one to another. Moreover, human-interpretation is still an issue which is not understood by the computational scientists. Therefore, Yu (2005) considered the second way, where all the FLRs are given importance based on their chronological order. In this scheme, weight for each FLR is determined based on their sequence of occurrence. That is, Yu (2005) gives more importance in recently occurred events than previously occurred events. However, this scheme of assigning weight is not justifiable, because it does not consider the severity of the events, which are represented by the fuzzy sets. To resolve this problem, we introduce a new "Index-based weights technique (IBWT)" in the model. In this approach, weight for each FLR is determined by using index (i) of the fuzzy set (B_i) associated with current state of the FLR. To explain this, consider the following FLRs as

$$\begin{aligned} &B_i \rightarrow B_i \quad \text{with weight } i,\\ &B_i \rightarrow B_j \quad \text{with weight } j,\\ &B_i \rightarrow B_k \quad \text{with weight } k,\\ &B_i \rightarrow B_l \quad \text{with weight } l. \end{aligned} \tag{4}$$

In Eq. (4), each FLR is assigned a weight i, j, k, and l, based on indices of the current state of the FLRs, which are B_i, B_j, and B_k, and B_l, respectively. The advantage of using such approach is that the model can capture more persuasive weights of the FLRs based on the severity of the events during forecasting.

Song and Chissom (1993a) adopted the following method to forecast enrollments of the University of Alabama:

$$Y(t) = Y(t-1) \circ R, \tag{5}$$

where $Y(t-1)$ is the fuzzified enrollment of year $(t-1)$, $Y(t)$ is the forecasted enrollment of year "t" represented by fuzzy set, "$\circ$" is the max–min composition operator, and "R" is the union of fuzzy relations. This method takes much time to compute the union of fuzzy relations R, especially when the number of fuzzy relations is more in (5) (Chen and Hwang, 2000; Huarng et al., 2007). In 1996, Chen (1996) used simplified arithmetic operations for defuzzification operation by avoiding this complicated max–min operations and their method produced better results than Song and Chissom (1993a,b, 1994) models. Most of the existing fuzzy time series models use Chen's (1996) defuzzification method in order to obtain the forecasting results. However, forecasting accuracy of these models are not good enough. Therefore, to resolve the problem of data defuzzification, we introduce a new "Index-based defuzzification technique (IBDT)" in this paper.

Hence, the contributions of this paper are fourfold: (1) we incorporate a new data discretization approach "MBD" in fuzzy time series forecasting model, which is an improvement over the original works presented by Song and Chissom (1993a), and later modification by Chen (1996); (2) we consider the repeated FLRs, which is the limitation of previous existing fuzzy time series models. Later, these FLRs are used to obtain the forecasted values; (3) we suggest to define weight for each FLR using the IBWT; (4) we recommend to employ the IBDT for defuzzification operation.

The proposed model is trained and tested with two major key parameters, viz., number of intervals generated by the proposed "MBD" approach and high-order FLRs. The rationale behind choosing these two parameters is that both these parameters significantly affect the result of forecasting as shown by Singh and Borah (2013a). The application of the proposed model is demonstrated by using three real-world data sets: (a) university enrollments data set of Alabama (Song and Chissom, 1993b), (b) daily average temperature data set of Taipei (Chen and Hwang, 2000), and (c) daily stock exchange price data set of State Bank of India (SBI) (Finance, 2012). However, the university enrollments data set of Alabama is used for the model verification, whereas the daily average temperature data set of Taipei and the daily stock exchange price data set of SBI is used for the model validation.

This paper is organized as follows. In Section 2, we present related works for fuzzy time series models. In Section 3, some basic concepts of fuzzy set theory has been explained with an overview of fuzzy time series. Section 4 demonstrates the application of the proposed approach to find the effective length of intervals in the universe of discourse. The proposed model has been introduced in Section 5. In Section 6, we discuss some statistical parameters, which are used to evaluate the performance of the model. The performance of the model has been assessed and the results are presented in Section 7. The proposed model has been validated in Section 8 using the data sets of daily average temperature of Taipei and daily stock exchange price of SBI. Finally, discussions and concluding remarks are presented in Section 9.

2. Related works

Forecasting using fuzzy time series is applied in several areas including forecasting university enrollments, sales, road accidents and financial forecasting. In a conventional time series models, the recorded values of a special dynamic process are represented by crisp numerical values. But, in a fuzzy time series model, the recorded values of a special dynamic process are represented by linguistic values.

Based on the fuzzy time series theory, first forecasting model was introduced by Song and Chissom (1993a,b, 1994), which were used to forecast the time series values based on linguistic values. They presented the fuzzy time series model by means of fuzzy relational equations involving max–min composition operation, and applied the model for forecasting the enrollments in the University of Alabama. In 1996, Chen (1996) developed fuzzy time series model based on first-order FLRs, and obtained the forecasted results with simplified arithmetic operations rather than complicated max–min composition operations. Chen (1996) forecasted the enrollments in the University of Alabama with better accuracy than the models proposed by Song and Chissom (1993a, b, 1994). Later, many studies provided some improvements in Chen (1996) method in terms of following issues:

- effective lengths of intervals (Li and Chen, 2004; Cheng et al., 2006, 2008; Huarng and Yu, 2006; Kai et al., 2010; Liu and Wei, 2010; Singh and Borah, 2011; Chen and Tanuwijaya, 2011a; Huang et al., 2011),