Information Sciences xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

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

Fuzzy forecasting based on automatic clustering and axiomatic fuzzy set classification

Weina Wang [a,b], Xiaodong Liu [a,*]

[a] Research Center of Information and Control, Dalian University of Technology, Dalian 116024, China
[b] Department of Mathematics, Jilin Institute of Chemical Technology, Jilin 132022, China

ARTICLE INFO

Article history:
Received 18 November 2012
Received in revised form 28 August 2014
Accepted 7 September 2014
Available online xxxx

Keywords:
Fuzzy forecasting
Fuzzy time series
Axiomatic fuzzy set (AFS) classification
Automatic clustering
Trend prediction

ABSTRACT

In spite of the impressive diversity of models of fuzzy forecasting, there is still a burning need to arrive at models that are both accurate and highly interpretable. This study proposes a new fuzzy forecasting model designed with the use of the two key techniques, namely clustering and axiomatic fuzzy set (AFS) classification. First, clustering algorithm is utilized to generate clustering-based intervals. Second, the fuzzy trend labeled training data set is constructed based on fuzzy logic relationships and fuzzy trends of historical samples. Then, the AFS classification is exploited to yield the semantic interpretation of each fuzzy trend. The main novelty is that the proposed model not only predicts the value but can also capture the trend prevailing in the time series, and obtain its semantic interpretation. The Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX), inventory demand, and Spanish electricity prices are used in a series of experiments. The results show that the proposed model has both good interpretability and accuracy.

1. Introduction

Investigating the relations within the sequential set of past data and forecasting the future values are the main tasks of the time series. The forecasting problem of time series data plays an important role in various domains, such as finance forecasting, population growth, rainfall prediction, and temperature forecasting [30–32,56]. It deals with forecasting future outcomes from a temporally ordered sequence of past observed data points, whose values are usually real numbers. Traditional time series analysis cannot handle the vagueness and uncertainty inherent in certain data due to inaccuracies in measurements, incomplete sets of observations, or difficulties in obtaining the measurements [43].

Recently, fuzzy logic has been widely recognized as a successful approach for dealing with data uncertainty. For time series, the uncertain values can be modeled as fuzzy variables, resulting in so-called fuzzy time series [43]. Song and Chissom first introduced the theory of fuzzy logic into forecasting time series problems and proposed a new paradigm known as fuzzy time series, capable of dealing with vague and incomplete data represented as linguistic values under uncertain circumstances [50–52]. They established a four-step framework to manipulate the forecasting problem: (1) determine and partition the universe of discourse into intervals; (2) define fuzzy sets from the universe of discourse and fuzzify the time series; (3) derive fuzzy relationships existing in the fuzzified time series; and (4) forecast and defuzzify the forecasting outputs.

* Corresponding author at: Research Center of Information and Control, Dalian University of Technology, No. 2, Linggong Road, Ganjingzi District, Dalian City, Liaoning Province 116024, PR China. Tel.: +86 41184709380; fax: +86 41184707579.
E-mail addresses: wangweina406@sina.com (W. Wang), xdliuros@dlut.edu.cn (X. Liu).

http://dx.doi.org/10.1016/j.ins.2014.09.027
0020-0255/

Since Song and Chissom's pioneering work, a number of related research works have been reported that follow their framework. In the plethora of currently available models of fuzzy time series, their accuracy has been a holy grail of the overall modeling for a long time. With the emergence of more visible need for interpretable models that are easily comprehended by humans, arose an important need to develop models that are not only accurate but transparent as well.

In order to develop a interpretable model, this study exploits axiomatic fuzzy set (AFS) classification which can mimic the human reasoning and in this manner offer a far more transparent and comprehensible way supporting the design of the classifier [35–39]. The proposed classifier consists of AFS fuzzy sets (each class is represented by a fuzzy set) and the degree of the new sample belonging to the class is the membership degree of this sample belonging the fuzzy set representing this class. Thus each class character is linguistically interpretable, comprehensible and similar to the classification schemes exercised by humans.

In this paper, a novel fuzzy forecasting model is presented. The proposed model first utilizes clustering to partition the universe of discourse and generate clustering-based intervals. Then, the model applies fuzzy logical relationships and fuzzy trends of historical samples to construct the fuzzy trend labeled training data set and exploits the AFS classification to yield its semantic interpretation. Finally, the forecasted data can be obtained by integrating the current state with the fuzzy trend. The advantages of the proposed model can be summarized as follows:

- When using fuzzy time series for forecasting, it is obvious that the length of intervals in the universe of discourse is important due to the fact that it can affect the forecasting accuracy. However, most of the existing fuzzy forecasting models based on fuzzy time series used the static length of intervals, i.e., the same length of intervals [1,3,6,8,20,22,25,27,29,33,40,41,45,46,48,49]. The drawback of the static length of intervals is that the historical data are roughly put into the intervals, even if the variance of the historical data is not high. To overcome the drawback, clustering algorithm is applied to generate intervals.
- Fuzzy sets and fuzzy logic were introduced to capture the uncertainty and human perception. Comparing with the most of existing models which employ "black box" structures without a comprehensive explanation facility for their predictions [2,13,19,54,62], the proposed model can yield the semantic interpretation of each fuzzy trend (such as "*upward*", "*unchanged*" and "*downward*").
- The proposed model can predict the fuzzy trend of forecasted data. This trend prediction will help the decision maker to be extra cautious well in advance.

To validate the proposed forecasting model, we conduct three real-world experiments in forecasting the stock indices [63], inventory demand [44] and the Spanish electricity prices time series [64]. The experimental results show the proposed model yields higher average forecasting accuracy rates than the existing models [3,6,7,9,16,18,21,23–25,53,57,59–61].

The paper is organized as follows. Section 2 overviews works related with fuzzy time series. Section 3 briefly reviews basic concepts of fuzzy time series. Section 4 proposes a novel forecasting model based on the automatic clustering and the AFS classification. Section 5 compares the average forecasting accuracy rates of the proposed model with that of existing models, where the historical data of the stock indices, inventory demand and the Spanish electricity prices data sets are used for the experiments. Section 6 concludes the paper.

2. Related works

Forecasting using fuzzy time series has been applied to various problem domains, such as the stock prices forecasting [6,9,15], inventory demand forecasting [21,23,61], and temperature forecasting [8,41]. Traditional time series models are usually highly dependent on historical data, which can be incomplete, imprecise and ambiguous. These uncertainties are likely to be widespread in real-world data and hinder forecasting accuracy, thus limiting the applicability of these models. Unlike traditional time series forecasting approaches, the fuzzy time series is capable of dealing with vague and incomplete time series data under uncertain circumstances.

Based on the fuzzy time series theory, first forecasting model was introduced by Song and Chissom [50–52], 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. The model take a large amount of computation time since max–min operations are used for reasoning. Chen [3] developed first-order fuzzy relationship rules with simplified arithmetic operations rather than the complicated max–min composition operations to forecast the enrollments of the University of Alabama with better accuracy than the Song and Chissom's model. Later, many studies have focused on improving steps of Song and Chissom's framework using Chen's models. The studies can be divided into three categories, (1) those that enhance the accuracy of model by changing interval length in partition step of the universe of discourse, (2) those that establishment of fuzzy relationships with better performance, and (3) those forecasting fuzzy time series by different approaches in forecasting and defuzzication step, separately.

In the first category, Huarng [21] investigated the impact of the interval length on forecasting results and proposed two heuristic approaches, i.e., distribution and average based, to determining the interval length. Chen and Hsu [5] used a two-phase partitioning method with the statistical distributions of historical data. Huarng and Yu [23] presented ratio-based

Q1 Please cite this article in press as: W. Wang, X. Liu, Fuzzy forecasting based on automatic clustering and axiomatic fuzzy set classification, Inform. Sci. (2014), http://dx.doi.org/10.1016/j.ins.2014.09.027