Applied Soft Computing 28 (2015) 422–430

Contents lists available at ScienceDirect

Applied Soft Computing

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

A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand

Riswan Efendi [a], Zuhaimy Ismail [a], Mustafa Mat Deris [b,*]

[a] Mathematics Department, Faculty of Science, Universiti Teknologi Malaysia, Skudai, Johor Bahru 81310, Johor, Malaysia
[b] Faculty of Computer Science, Universiti Tun Hussein Onn Malaysia, Batu Pahat 86400, Johor, Malaysia

ARTICLE INFO

Article history:
Received 17 August 2012
Received in revised form 7 December 2012
Accepted 30 November 2014
Available online 12 December 2014

Keywords:
Fuzzy time series
Index number
Weight
Electricity load demand
Linguistic time series
Out-sample forecast

ABSTRACT

The fuzzy logical relationships and the midpoints of interval have been used to determine the numerical in-out-samples forecast in the fuzzy time series modeling. However, the absolute percentage error is still yet significantly improved. This can be done where the linguistics time series values should be forecasted in the beginning before the numerical forecasted values obtained. This paper introduces the new approach in determining the linguistic out-sample forecast by using the index numbers of linguistics approach. Moreover, the weights of fuzzy logical relationships are also suggested to compensate the presence of bias in the forecasting. The daily load data from National Electricity Board (TNB) of Malaysia is used as an empirical study and the reliability of the proposed approach is compared with the approach proposed by Yu. The result indicates that the mean absolute percentage error (MAPE) of the proposed approach is smaller than that as proposed by Yu. By using this approach the linguistics time series forecasting and the numerical time series forecasting can be resolved.

1. Introduction

In the power management, the forecasting of load demand is the main problem especially in obtaining the high accuracy level. In this decade, the short time load forecasting (STLF) is frequently developed by researchers. Many approaches have been proposed to STLF demand ranging from the statistical to the artificial intelligence approaches [15]. For example, regression method [10,25,36,41], time series [1,9,12,35,54], neural network [2,27,37,38,44], similar day approach [18], expert system [20,40], fuzzy logic [26,49], data mining [16,17], wavelets [33], and evolutionary algorithms [19,21,50]. For the TNB (the largest electricity utility company in Malaysia) data, some previous studies investigated the STLF by using the time series Box–Jenkins method [35,54]. However, these studies do not provide sufficient justification on the results as compared with other approaches. Whereby, this situation encourages us to investigate the STLF for the TNB data by using one artificial intelligence, namely, the fuzzy time series (FTS).

FTS is a new approach which has been developed by Song and Chissom [45,46] in resolving the linguistic time series data problems. This approach is a combination between the fuzzy logic and the time series analysis. Furthermore, its application can be found in some domain problems, such as, enrollment [3,5,11,24,28,30,31,34,39,42,43,45–47,55], the stock index [6–8,11,22,24,48,52,53] temperature [4] and financial prediction [29,32]. The most important thing in the fuzzy time series forecasting is the assumption regarding data that are not needed, which is the main difference from the statistical approaches. In general, the model has been established by using fuzzification, fuzzy logical relationship (FLRs), fuzzy logical group (FLG), and defuzzification.

Many different models have been proposed on the fuzzy time series forecasting by researchers. Huarng [22] initiated a study on heuristic models of the fuzzy time series for forecasting by using the stock index data. In addition, Huarng et al. [23] also continued to analyze a multivariate heuristic model for the fuzzy time series forecasting. On the other hand, Yu [52] enhanced the weighted fuzzy time series models for the Taiwan Stock Index (TAIEX) forecasting. It is assigned by the recurrent FLRs in the FLG. Furthermore, Cheng et al. [7] presented the trend-weighted fuzzy time series model for the TAIEX forecasting. Qiu et al. [39] also presented a generalized approach in the forecasting by using the fuzzy weights. Yu [52] proposed the final forecast was equal to the product of midpoints matrix and the transpose of the weight matrix. However, the approach proposed by Yu is yet to be improved to resolve the linguistic out-sample forecast and also the numerical out-sample forecast.

* Corresponding author. Tel.: +60 7 453 3723.
E-mail address: mmustafa@uthm.edu.my (M.M. Deris).

http://dx.doi.org/10.1016/j.asoc.2014.11.043
1568-4946/

In this study, the discussion will be enhanced to design the new linguistic out-sample approach and to determine the weight of FLRs. In the forecasting phase, two types of forecasting phase are used: the linguistics time series forecasting and the numerical time series forecasting. The linguistics time series forecasting is applied to predict the linguistics values by using the index numbers using Box Jenkins procedure. The weights and midpoint intervals will be applied for the numerical time series forecasting. The weights can also be assigned by using index numbers of close relationship in the FLG. In addition, the length of interval and partition number in the universe of discourse will be considered using Sturges and the two powers of p (p is a number interval or class) rules. At the end of this procedure, both of the mean absolute percentage errors (MAPE) from the proposed approach will be compared with the MAPE approach proposed by Yu.

The remainder of the paper is organized as follows: Section 2 presents the basic theory of fuzzy set and fuzzy time series. Section 3 describes the importance of weighted fuzzy time series with few examples. Section 4 proposes weight and forecasting procedures for FLRs. Section 5 describes the empirical analysis using the daily TNB data from 01/01/2006 to 31/08/2006. The final section gives some conclusions of the study.

2. The fundamental theories in the fuzzy time series

This section describes the fuzzy set, fuzzy time series, and some related definitions that can be used in this paper.

2.1. Fuzzy set definition

Let U be the universe of discourse. A fuzzy subset A on the universe of discourse U can be defined as follows:

$$A = \{(u_i, \mu_i(u_i)) | u_i \in U\}$$

where μ_A is the membership function of A, $\mu_A : U \to [0, 1]$, and $\mu_A(u_A)$ is the degree of membership of the element u_i in the fuzzy set A [23]. If U be finite and infinite sets, then fuzzy set A can be expressed as follows:

$$A = \sum \frac{\mu_A(u_i)}{u_i} = \frac{\mu_A(u_1)}{u_1} + \frac{\mu_A(u_2)}{u_2} + \cdots + \frac{\mu_A(u_n)}{u_n}$$

and

$$A = \int \frac{\mu_A(u_i)}{u_i} du, \quad \forall u_i \in U$$

2.2. Fuzzy time series definitions

Song and Chissom [45,46] presented some definitions for fuzzy time series as follows:

Definition 1. Let $Y(t)$ ($t = 0, 1, 2, \ldots$), a subset of real numbers, be the universe of discourse on which fuzzy sets $f_i(t)$ ($i = 1, 2, \ldots$) are defined in the universe of discourse $Y(t)$ and $F(t)$ is a collection of $f_i(t)$ ($i = 1, 2, \ldots$). Then $F(t)$ is called a fuzzy time series defined on $Y(t)$ ($t = 0, 1, 2, \ldots$). Therefore, $F(t)$ can be understood as a linguistics time series variable, where $f_i(t)$ ($i = 1, 2, \ldots$), are possible linguistics values of $F(t)$.

Definition 2. Suppose $F(t)$ is caused by $F(t-1)$ denoted by $F(t-1) \to F(t)$, then this relationship can be represented as:

$$F(t) = F(t-1)^{\circ} R(t, t-1)$$

where "$\circ$" represents an operator, $R(t, t-1)$ is a fuzzy relationship between $F(t)$ and $F(t-1)$ and is called the first-order model of $F(t)$.

The other definitions also presented by Yu [52] and Huarng et al. [23] regarding the FLRs and FLG are as follows.

Definition 3. Let $F(t-1) = A_i$ and $F(t) = A_j$. The relationship between two consecutive data (called a fuzzy logical relationship, FLR), i.e., $F(t)$ and $F(t-1)$, can be denoted as $A_i \to A_j$, $i, j = 1, 2, \ldots, p$ (where p is interval or subinterval number) is called the left-hand side (LHS), and A_j is the right-hand side (RHS) of the FLR.

Definition 4. Let $A_i \to A_j$, $A_i \to A_k$, $\ldots$, $A_i \to A_p$ are FLRs with the same LHS which can be grouped into an ordered FLG (called a fuzzy logical group) by putting all their RHS together as on the RHS of the FLG. It can be written as bellow:

$$A_i \to A_j, A_i \to A_k, \ldots, A_i \to A_p; \quad i, j, k, \ldots, p = 1, 2, \ \ldots, \ n (n \in N)$$

2.3. Fuzzy time series forecasting

The forecasting procedure was developed by Song and Chissom (1993) into several steps as follows:

- Define the universe of discourse (U) and divide it into several equal length intervals.
- Fuzzify each interval into linguistics time series values (A_i, $i = 1, 2, \ldots, p$, p is partition number).
- Establish fuzzy logical relationships among linguistics time series values ($A_i \to A_j$, $i, j = 1, 2, \ldots, p$).
- Establish forecasting rule.
- Determine the forecast value.

3. The importance of weight in the fuzzy time series forecasting

In fuzzy time series, the forecasting model uses fuzzy relationships among the linguistic time series values. Two fuzzy types of relationships are (i) the same-fuzzy logical relationship and (ii) the different-fuzzy logical relationship. Both types of relationships may occur either recurrently or frequently. The occurrence of a particular fuzzy relationship explains the number of its appearances in the past. Some of the reasons for establishing the weight factor are:

i. To compensate for the presence of bias especially when the events are frequently occurred [14].
ii. To raise the influence of the more accurate input data, and to reduce the influence of the less accurate ones [13].

Fundamentally, these are the reasons for finding the weights in the fuzzy relationships. Similar to Yu [52] and Cheng et al. [7] findings, the weight factors are denoted within the weight matrix given in the following definition.

Definition 5. Yager's OWA operator of dimension n is a mapping

$$\emptyset : \mathbb{R}^n \to \mathbb{R}$$

which has an associated of weights $\mathbf{W} = (w_1 \quad w_2 \quad w_3 \cdots w_n)^T$ or can be written as:

$$\mathbf{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

such that