Expert Systems with Applications 40 (2013) 5673–5679

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Effective intervals determined by information granules to improve forecasting in fuzzy time series

Lizhu Wang [a,b,*], Xiaodong Liu [a], Witold Pedrycz [c]

[a] Research Center of Information and Control, Dalian University of Technology, Dalian 116024, China
[b] School of Mathematics and System Science, Shenyang Normal University, Shenyang 110034, China
[c] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada T6R 2G7

ARTICLE INFO

Keywords:
Forecasting
Fuzzy time series
Information granule
Enrollment

ABSTRACT

Partitioning the universe of discourse and determining effective intervals are critical for forecasting in fuzzy time series. Equal length intervals used in most existing literatures are convenient but subjective to partition the universe of discourse. In this paper, we study how to partition the universe of discourse into intervals with unequal length to improve forecasting quality. First, we calculate the prototypes of data using fuzzy clustering, then form some subsets according to the prototypes. An unequal length partitioning method is proposed. We show that these intervals carry well-defined semantics. To verify the suitability and effectiveness of the approach, we apply the proposed method to forecast enrollment of students of Alabama University and Germany's DAX stock index monthly values. Empirical results show that the unequal length partitioning can greatly improve forecast accuracy. Further more, the proposed method is very robust and stable for forecasting in fuzzy time series.

1. Introduction

In recent years, some methods have been presented for handling forecasting problems based on fuzzy time series such as stock index forecasting (Yu, 2005), hydrometeorology forecasting (Wang, Liu, & Yin, 2012), enrollment prediction (Song & Chissom, 1993a), temperature forecasting (Wang & Chen, 2009), etc. The concept of fuzzy time series was first proposed by Song and Chissom (1993b). Using the concept of fuzzy time series, they introduced time-invariant and time-variant fuzzy time series models and used them to forecast the enrollment of students of University of Alabama. Following Song and Chissom, related works mainly focus on either improving the forecast accuracy or reducing the computational complexity.

The forecasting process in fuzzy time series can be summarized as the following four steps: (1) partitioning the universe of discourse, (2) defining fuzzy sets and expressing time series with the use of these fuzzy sets (fuzzification), (3) extracting fuzzy logical relationships from the fuzzy time series, and (4) forecasting and defuzzification of the output of fuzzy time series. Concerning step (1), Huarng (2001) argued that the length of intervals affects significantly forecasting results in fuzzy time series. In the sequel, they proposed distribution and average-based length for the fuzzy time series model; Kuo et al. (2009) presented a method to forecast the enrollment by involving particle swarm optimization; Wang and Chen (2009) proposed a method based on clustering techniques to predict the temperature and the Taiwan futures exchange. There is not too much work dealing with step (2) besides some fuzzy sets theory. Step (3) is deemed to one of the critical phases to influence forecasting result. Chen and Wang (2010) proposed fuzzy forecasting based on fuzzy-trend logical relationship groups; To reduce computational complexity, Chen (1996) presented an efficient forecasting procedure by grouping fuzzy logical relationships into rules and performing simplified arithmetic operations on these groups. Yu (2005), Cheng, Teoh, and Chen (2007), Teoh, Chen, and Cheng (2009) and Lee, Efendi, and Ismail (2009) had also made their efforts to improve it. With regard to phase (4), most of the fuzzy time series models' are the same as that of Song and Chissom.

One of the evident limitations of these models is that they consider ad hoc approaches to the granulation of original numeric data and time variable. Although distribution and average-based length method (Huarng, 2001) was developed to consider a distribution describing of the first differences of data, researchers did not consider the distribution of data itself. Methods such as particle swarm optimization (Kuo et al., 2009), clustering techniques (Chen & Wang, 2010), entropy-based model (Cheng, Chang, & Yeh, 2006), and refined model (Yu, 2005), which utilize heuristics to segment intervals into subintervals, are not supported by underlining semantics.

In this paper, we study an unequal length partitioning. The approach relies on a concept of information granules and fuzzy clustering. The role of the information granules and fuzzy

* Corresponding author. Address: 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 41184709381.
E-mail address: wanglizhu80@126.com (L. Wang).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.04.026

clusters is to determine temporal intervals of unequal length so that the model comes with increased accuracy and enhanced interpretability. The proposed method has been experimentally tested on enrollment and Germany's DAX stock index time series forecasting. The experimental results show that forecasting accuracy is evidently improved when comparing the proposed method with equal length partitioning used in the previous studies.

The paper is organized as follows: in Section 2, we give a brief review of fuzzy time series, fuzzy cluster and information granule. In Section 3, we present our proposed method to partition the universe of discourse into unequal length intervals. The performance of the proposed method on both enrollment and Germany's DAX stock index time series forecasting are examined in Section 4. Some conclusions are presented in Section 5.

2. Related works

In this section, some related background including fuzzy time series, fuzzy clustering and information granules is briefly reviewed.

2.1. Fuzzy time series

We start with a series of pertinent definitions.

Definition 2.1. Let $U = \{u_1, u_2, \ldots, u_n\}$ be the universe of discourse, a fuzzy set A of the universe of discourse U can be defined as follows:

$$A = \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + , \cdots , + \frac{f_A(u_n)}{u_n} \quad (1)$$

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

Definition 2.2. Let $Y(t)(t = \ldots, 0, 1, 2, \ldots)$, which is a subset of R be the universe of discourse in which fuzzy sets $f_i(t)(i = 1, 2, \ldots)$ are defined. Let $F(t)$ be a collection of $f_i(t)(i = 1, 2, \ldots)$. Then, $F(t)$ is called a fuzzy time series on $Y(t)(t = \ldots, 0, 1, 2, \ldots)$.

Definition 2.3. Let $F(t-1) = A_i$ and $F(t) = A_j$. The relationship between two consecutive observations, $F(t-1)$ and $F(t)$, referred to as a fuzzy logical relationship, can be denoted by $A_i \to A_j$, where A_i is called the left-hand side and A_j the right-hand side of the fuzzy logical relationship.

Definition 2.4. Fuzzy logical relationships with the same fuzzy set located in the left-hand side of the relationships can be further grouped into a fuzzy logical relationship group (Huarng, 2001). Suppose there are fuzzy logical relationships such that $A_i \to A_{j1}$, $A_i \to A_{j2}, \cdots$, they can be grouped into a fuzzy logical relationship group $A_i \to A_{j1}, A_{j2}, \ldots$.

In most models, the fuzzy relationship between $F(t)$ and $F(t-1)$ is defined by the fuzzy logical group such as Chen (1996) and Lee et al. (2009). Following Chen's model, the same fuzzy sets can only show up once at the right-hand side of the fuzzy logical relationship group, but in the Lee's model they consider the frequency of occurrence of fuzzy sets.

2.2. Fuzzy clustering

Fuzzy C-Means (FCM), proposed by Bezdek (1981), is one of the commonly used fuzzy clustering algorithms. FCM forms a fuzzy partition of the dataset by minimizing the following objective function with respect to membership grades a_{ij} and the prototype of cluster v_i

$$J_\beta(A, X, V) = \sum_{i=1}^{c}\sum_{j=1}^{n}(a_{ij})^\beta d^2(x_j, v_i) \quad (2)$$

where $\beta > 1$ is the weighting exponent (fuzzification coefficient) used to determine the fuzziness of the resulting clusters, n is the number of feature vectors $x_j, c \geqslant 2$ is the number of clusters, $d(x_j, v_i)$ is the Euclidean distance between x_j and the prototype v_i, and a_{ij} represents the membership degree of x_j belonging to v_i. The minimization of the objective function is realized under the following constraints:

$$0 \leqslant a_{ij} \leqslant 1 \quad \forall\, i, j$$

$$0 < \sum_{j=1}^{n} a_{ij} \leqslant n \quad \forall\, i$$

$$\sum_{i=1}^{c} a_{ij} = 1 \quad \forall\, j$$

The FCM algorithm is an iterative optimization process. The prototypes of the clusters v_i and the membership degrees a_{ij} are updated according to the following expressions:

$$v_i = \frac{\sum_{j=1}^{n}(a_{ij})^\beta x_j}{\sum_{j=1}^{n}(a_{ij})^\beta} \quad 1 \leqslant i \leqslant c$$

$$a_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d(x_j, v_i)}{d(x_j, v_k)}\right)^{2/(\beta-1)}} \quad 1 \leqslant i \leqslant c, \quad 1 \leqslant j \leqslant n$$

A common choice of the weighting exponent is $\beta = 2$ and this value will be used throughout this paper. In addition, the clustering process terminates when the maximum number of iterations reaches 100, or when the objective function improvement between two consecutive iterations is less than 1×10^{-5}. Given the number of clusters, FCM algorithm can calculate the prototypes of the clusters and corresponding membership degrees.

2.3. Information granule

The concept of information granule was first proposed by Zadeh (1979). Pedrycz and Vukovich (2001) introduced a model of generalization and specialization of information granules. Bargiela and Pedrycz (2003) presented recursive information granulation and looked at the aggregation and interpretation issues. Information granule, denoted by Ω, can be expressed in a certain formal framework of granular computing. Here we are concerned with the development of a single information granule based on some experimental evidence (data) coming as a set of a one-dimensional (scalar) numeric data, $D = \{x_1, x_2, \ldots, x_n\}$. The constructed information granule is an interval in R. In its determination, there are two intuitively compelling requirements to be satisfied: justifiable granularity and semantic soundness. With regard to the requirement of justifiable granularity, the more data are included within the bounds of Ω, the better. In this way, the set becomes more legitimate. Justifiable granularity is quantified by counting the number of data falling within the bounds of Ω. We can consider the cardinality of Ω, namely $card\{x_k | x_k \in \Omega\}$ as a suitable description of the legitimacy of information granule. More generally, we may consider an increasing functional of the cardinality, say $f_1(card\{x_k | x_k \in \Omega\})$ where f_1 is an increasing function of its argument. Semantic soundness is quantified by the specificity of Ω. For instance, in case of a numeric interval, $\Omega = [a, b]$, $m(\Omega)$ is a length of this interval. The inverse of the length of this interval, or more generally a decreasing functional of the length f_2, can serve as a sound measure of specificity. The higher the value of $f_2(m(\Omega))$, the more detailed (specific) the