Applied Soft Computing 40 (2016) 132–149

Contents lists available at ScienceDirect

Applied Soft Computing

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

A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting

Hossein Javedani Sadaei [a], Rasul Enayatifar [b], Muhammad Hisyam Lee [c,*], Maqsood Mahmud [d]

[a] Graduate Program in Electrical Engineering, Federal University of Minas Gerais, Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG, Brazil
[b] Department of Computer Science, Islamic Azad University, Firookooh Branch, Iran
[c] Department of Mathematical Sciences, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
[d] Department of Computer Sciences, Al Mareefa College, Riyadh, Saudi Arabia

ARTICLE INFO

Article history:
Received 10 May 2015
Received in revised form 8 September 2015
Accepted 9 November 2015
Available online 27 November 2015

Keywords:
Fuzzy time series
Differential fuzzy set
Stock forecasting
Trend
Imperialist competitive algorithm

ABSTRACT

In this study, a new kind of fuzzy set in fuzzy time series' field is introduced. It works as a trend estimator to be appropriate for fuzzy time series forecasting by reconnoitering trend of data appropriately. First, the historical data are fuzzified into differential fuzzy sets, and then differential fuzzy relationships are calculated. Second, differential fuzzy logic groups are established by grouping differential fuzzy relationships. Finally, in the defuzzification step, the forecasts are calculated. However, for increasing the accuracy of the models, an evolutionary algorithm, namely imperialist competitive algorithm is injected, to train the model. A massive stock data from four main stock databases have been selected for model validation. The final project, has shown that outperformed its counterparts in term of accuracy.

1. Introduction

Accurate stock index forecasting is one of the most profitable and challenging prediction for stock investors. Therefore, there is a growing demand and motivation to conduct new researches in this field. Subsequently, a major novel studies have been conducted to promote more accurate forecasting [1–7]. However, among these models, recently, researchers have shown an increased interest in using Fuzzy Time Series (FTS) models for stock forecasting [8–11]. The concept of FTS, first, introduced to the world by Song and Chissom [12,13]. Their proposed model employed fuzzy relation equations and approximate reasoning. Subsequently, Chen (1996) proposed a first-order FTS, by basic arithmetic operations and fuzzy logical relationship groups to make Song and Chissom model's more applicable [14]. Huarng reviewed Chen's method by integrating problem-specific heuristic knowledge [15]. In addition, in 2005, Yu introduced weighted FTS models to resolve two fuzzy relationship problems, recurrence and weighting [8]. The model was for the Taiwan Capitalization Weighted Stock Index (TAIEX) forecasting. Lately, other fields have applied FTS models to forecast, including university enrollment [16–18], stock index predicting [8,19,20], temperature forecast [21–23] and short term load forecasting [24,25].

In stock data time series, naturally, there are some increasing and decreasing tendencies which classified into short, medium and long term trends. Among these groups, the short term trends which has received more concern for investors and traders, for making business decisions, are more difficult to investigate as well. In FTS literature, exist certain studies about exploring the trend of data, partially. For instance, Cheng in 2006, proposed a methodology which combined trend-weighting into the fuzzy time-series, by employing actual trading data of Taiwan Stock Index (TAIEX) and the enrolment data of the University of Alabama [26]. In 2012, Chen and Chen presented a method that calculated the value of the variable between the subscripts of adjacent fuzzy sets appearing in the antecedents of high order fuzzy logical relationship. They used the method to forecast TAIEX, the enrollments of the University of Alabama and the inventory demand [27]. Although those methods are slightly successful for the trend prediction inside stock data, they can be more improved if trend of the stock also been completely explored in a better way. In fact, applicants and people in the mission area who are involved in business, investing or other relevant fields except new forecasting methods. These methods must produce precise results and promote quick-outcome and include straightforward concepts to understand. From the results of the current study, it concludes that these goals can be in a more proper way achieved by the proposed models.

In this research, the initial aim is to propose a new forecasting FTS model which its basis is predicts the trends of data. It promotes better forecast for data having various trends. Especially, the proposed method has been designed to be applied on stock index data. This project is called Proposed 1 method. However the Proposed 2 method, which also is a second concern of this study, is obtained by combing optimization concept with Proposed 1 to increase the accuracy of forecasts. To do so, a novel evolutionary algorithm (EA), namely, imperialist competitive algorithm (ICA), is deployed on Proposed 1, to train the model for optimizing the required parameters. In fact, ICA, which was presented by [28] is a computational method that is applied to resolve optimization problems of in different applications [24,29]. Such as most of EAs, ICA does not need the gradient of the function in its optimization procedure.

The rest of this paper is organized as follows; Next section demonstrates related works and preliminary concepts of FTS; Section 3 introduces the methodology. In Section 4, the proposed method is explained by certain examples. Section 5 introduces the benchmark stock datasets to readers. Subsequent section, provides the discussions and results, finally Section 7 concludes the paper.

* Corresponding author. Tel.: +60 7553 4236; fax: +60 7556 6162.
E-mail addresses: jshossein@ufmg.br (H.J. Sadaei), r.enayatifar@gmail.com (R. Enayatifar), mhl@utm.my (M.H. Lee), maqsood@mcst.edu.sa (M. Mahmud).

http://dx.doi.org/10.1016/j.asoc.2015.11.026
1568-4946/

2. Preliminaries

In this section the prerequisite concepts of FTS and ICA, which are relevant to this paper demonstrated. The first part of this section is about the fundamentals of FTS, the second part is Chen's 1996 algorithm, and in last section, the ICA algorithm is illustrated.

2.1. Fundamentals and definitions

Song and Chissom first presented the concepts of FTS [12,13], where the values in a FTS are presented by fuzzy sets [30]. Let U be the universe of discourse, where $U = \{u_1, u_2, \ldots, u_b\}$. A fuzzy set of U is defined as follows:

$$A_i = \frac{f_{A_i}(u_1)}{u_1} + \frac{f_{A_i}(u_2)}{u_2} + \cdots + \frac{f_{A_i}(u_b)}{u_b} \tag{1}$$

where f_{A_i} is the membership function of the fuzzy set A_i, $f_{A_i} : U \to [0, 1]$; u_a is a generic element of fuzzy set $A_i : f_{A_i}(u_a)$ is the degree of belongingness of u_a to $A_i : f_{A_i}(u_a)$ and $1 \le a \le b$.

Definition 1 *(Song and Chissom [12]).* Let $Y(t) : t = \ldots, 0, 1, 2, \ldots$, a subset of real numbers $\mathbb{R}$ be the universe of discourse by which fuzzy sets $f_j(t)$ are defined. If $F(t)$ is a collection of $f_1(t), f_2(t), f_3(t), \ldots$ then $F(t)$ is called a FTS defined on $Y(t)$.

Definition 2 *(Song and Chissom [12]).* If there exists a fuzzy relationship $R(t-1, t)$, such that $F(t) = F(t-1) \circ R(t-1, t)$ where $\circ$ is an arithmetic operator, then $F(t)$ is said to be caused by $F(t-1)$. The relationship between $F(t)$ and $F(t-1)$ can be denoted by $F(t) \to F(t-1)$.

Definition 3 *(Song and Chissom [12]).* Suppose $F(t)$ is calculated only by $F(t-1)$ and $F(t) = F(t-1) \circ R(t-1, t)$. For any t, if $R(t-1, t)$ is independent of t, then $F(t)$ is considered a time-invariant FTS. Otherwise, $F(t)$ is time-variant. Assuming $F(t-1) = A_i$ and $F(t) = A_j$, a fuzzy logical relationship can be defined as $A_i \to A_j$, where A_i and $A_j$are called the left-hand side (LHS) and right-hand side (RHS) of the fuzzy logical relationship, respectively.

2.2. The algorithm of Chen's first-order model

Chen (1996) proposed a model by applying simplified arithmetic operations in forecasting algorithms, replacing the complicated max-min composition operation presented by Song and Chissom [12]. Since it will be refined for proposing first method in this paper, it is explained as follows:

Step 1: *Define the universe of discourse.* Based on the min and max values in the dataset, define the D_{min} and D_{max} variables. Then, choose two arbitrary positive numbers, D_1 and D_2 to portion universe of discourse, i.e., $U = \{D_{min} - D_1, D_{max} + D_2\}$, into equal intervals in length, i.e., $u_1, u_2, \ldots, u_m$.

Step 2: *Define fuzzy sets.* Based on the defined sub intervals in former step, define fuzzy sets as follows:

$$A_i = \frac{0}{u_1} + \frac{0}{u_2} + \cdots + \frac{0.5}{u_{i-1}} + \frac{1}{u_i} + \frac{0.5}{u_{i+1}} + \cdots + \frac{0}{u_m}, \quad i = 1, 2, 3, \ldots, m. \tag{2}$$

Step 3: *Fuzzify observed rules.* In this step, each datum in time series is fuzzified into its corresponding fuzzy set using Eq. (2).

Step 4: *Establish fuzzy logic relationships.* In this step, fuzzy logic relationships (FLRs) and fuzzy logic relationships groups (FLRGs) are established. Establish FLRs and group them based on the current states of the data.

Step 5: *Calculate the forecasted outputs.* Let $F(t) = A_i$. In this step three cases may occur: Situation 1: There is only one fuzzy logical relationship in the sequence. If $A_i \to A_j$, then $F(t+1)$, the forecast output, equals A_j. Situation 2: If $A_i \to A_1, A_2, A_3, \ldots, A_k$, then $F(t+1)$, the forecast output, equals $A_1, A_2, A_3, \ldots, A_k$. Situation 3: There are no fuzzy relationship groups, i.e., $A_i \to \#$, then, the forecast output, equals A_i.

Step 6: *Defuzzify.* If $F(t+1) = A_{i1}, A_{i2}, A_{i3}, \ldots, A_{ik}$, the forecasted values at time $t+1$ is calculated as

$$\text{Forecasted-value}(t+1) = \frac{M_{i1} + M_{i2} + M_{i3} + \cdots + M_{ik}}{k}$$

where $M_{i1}, M_{i2}, M_{i3}, \ldots, M_{ik}$ are defuzzified values of $A_{i1}, A_{i2}, A_{i3}, \ldots, A_{ik}$, respectively.

2.3. Imperialist competitive algorithm (ICA)

Recently, ICA, which has been presented as an innovative algorithm, has been generally utilized to seek for solutions in myriads of ranges [1]. It differs from most of the meta-heuristic algorithms in that it is not generated from a natural event. In fact, it uses an imperialism and imperialist competition process which is both social and political (see Fig. 1).

3. Methodology

This section consists of three parts. Firstly, new required fundamentals and new definitions related to proposed methods, are stated. Secondly, the Proposed 1 algorithm is demonstrated and finally, Proposed 2 is explained.

3.1. Definitions and algorithm for proposed models

Suppose the universe of discourse for the forecasting problem is defined as $U = \{u_1, u_2, \ldots, u_m\}$. If the fuzzy sets on U define as follows:

$$\begin{aligned} A_1 &= 1/u_1 + 0.5/u_2 + \cdots + 0/u_m \\ A_2 &= 0.5/u_1 + 1/u_2 + \cdots + 0/u_m \\ &\vdots \\ A_m &= 0/u_1 + 0/u_2 + \cdots + 1/u_m. \end{aligned} \tag{3}$$

then by considering m which come from Eq. (3), we define a new universe of discourse namely

$$U' = [-(m-1), (m-1)] \tag{4}$$

and portion it, into following equally-length of intervals:

$$\begin{aligned} u'_{-(m-1)} &= [-(m-1) - 0.5, -(m-1) + 0.5] \\ u'_{-(m-2)} &= [-(m-2) - 0.5, -(m-2) + 0.5] \\ &\vdots \\ u'_{r-2} &= [(r-1) - 0.5, (r-1) + 0.5] \\ u'_{r-1} &= [(r-2) - 0.5, (r-2) + 0.5] \\ u'_{r} &= [r - 0.5, r + 0.5] \\ u'_{r+1} &= [(r+1) - 0.5, (r+1) + 0.5] \\ u'_{r+2} &= [(r+2) - 0.5, (r+2) + 0.5] \\ &\vdots \\ u'_{(m-2)} &= [(m-2) - 0.5, (m-2) + 0.5] \\ u'_{(m-1)} &= [(m-1) - 0.5, -(m-1) + 0.5] \end{aligned} \tag{5}$$