Information Sciences 294 (2015) 227–241

Contents lists available at ScienceDirect

Information Sciences

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

A hybrid fuzzy time series model based on granular computing for stock price forecasting

Mu-Yen Chen *, Bo-Tsuen Chen

Department of Information Management, National Taichung University of Science and Technology, Taiwan

ARTICLE INFO

Article history:
Received 9 December 2012
Received in revised form 13 September 2014
Accepted 24 September 2014
Available online 13 October 2014

Keywords:
Granular computing
Binning-based partition
Entropy-based discretization
Time series forecasting

ABSTRACT

Given the high potential benefits and impacts of accurate stock market predictions, considerable research attention has been devoted to time series forecasting for stock markets. Over long periods, the accuracy of fuzzy time series model forecasting is invariably affected by interval length, and formulating effective interval partitioning methods can be very difficult. Previous studies largely relied on distance partitioning, but this approach neglects the distribution of datasets and can only handle scalar forecasting. But the magnitude of stock price movements is often severe and difficult to predict. Thus, the distribution of stock price datasets is always skewed and the straightforward partitioning method is not well suited to these types of time series datasets. In this research, a novel fuzzy time series model is used to forecast stock market prices. The proposed model is based on the granular computing approach with binning-based partition and entropy-based discretization methods. The proposed model is verified using experimental datasets from the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX), Dow-Jones Industrial Average (DJIA), S&P 500 and IBOVESPA stock indexes, and results are compared against existing fuzzy time series models, three different SVM models, and three modern economic models - GARCH, GJR-GARCH, and Fuzzy GARCH. Compared to other current forecasting methods, the proposed models provide improved prediction accuracy and the results are verified by paired two-tailed *t*-tests. The experimental results clearly provide improvements for obtaining optimized linguistic intervals and ensuring the accuracy of the proposed model.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The ongoing global financial crisis underlines the importance of the time series concept, and methods are urgently needed to analyze time series data from daily closing prices with the aim of accurately predicting stock price movements. Currently, financial forecasting relies mainly on mathematical and statistical methods [1,2,20,43,57], and time series models [3,4]. The 2007 financial crisis increased the importance of financial forecasting among general investors and researchers, prompting the development and application of many theories and techniques in fundamental and technical analysis.

A time series is defined as a set of sequential observations which can be either *continuous* or *discrete*. Time series analysis is widely used by researchers studying methods of stock market forecasting (originally proposed by Kendall and Ord [34]) and logistical regression models based on traditional statistical assumptions. Financial forecasting problems are usually

* Corresponding author at: 129 Sec. 3, San-Min Road, Taichung 404, Taiwan. Tel.: +886 4 22196310; fax: +886 4 22196311.
E-mail address: mychen@nutc.edu.tw (M.-Y. Chen).

http://dx.doi.org/10.1016/j.ins.2014.09.038
0020-0255/© 2014 Elsevier Inc. All rights reserved.

handled using traditional time series methods such as autoregressive moving average (ARMA) model [4], autoregressive conditional heteroscedastic (ARCH) model [19], and Generalized ARCH (GARCH) model [3], but these methods require extensive historical data and assumptions such as normality postulates [33]. Unlike conventional time series which deal with real numbers, fuzzy time series are architecture by fuzzy sets [62]. Fuzzy time series are frequently applied into producing stock price predictions due to the handling capability of linguistic value datasets to produce accurate forecasting results. Nowadays, it has been widely and successfully utilized to forecast nonlinear and dynamics datasets in such widely varying domains including course enrollment [9,25,50,52], temperature [56], traffic accidents [33], tourism demand [55] and stock markets [13,14,28,59,60].

First proposed by Zadeh (1965), fuzzy set theory is still widely used in a broad range of applications. Song and Chissom [50] developed a fuzzy time series model to predict university enrollment levels. However, this model used max–min composition operations which significantly complicate the computational process. Chen [7] presented one simplified calculation process to solve this drawback, but this model lacked a suitable weight mechanism for the fuzzy logical relationships (FLRs). The model has since been widely extended and adapted, with researchers seeking to improve forecasting accuracy by adjusting the length of linguistic intervals or by changing the weighting approach. Huarng (2001) determined interval lengths using distribution-based and average-based lengths [25]. Yu (2005) applied different weighted technologies on fuzzy time series models to strength forecasting accuracy [59]. Huarng and Yu (2006) added the ratio-based interval length to the fuzzy time series model [24] while Cheng et al. (2006) integrated a fuzzy time series model with a trend-weighting mechanism to predict actual trading data of stock prices and university enrollment [12]. They later predicted hospital outpatient traffic adopting a revised fuzzy time series which integrated a weighted-transitional matrix [15]. These methods provide lower forecasting error rates did the methods proposed by Chen and Chung [9] and Yu [59].

These aforementioned studies all focus on linguistic interval length for fuzzy time-series models, which raises several problems: (1) while distance partitioning is the most widely used method for interval partitioning [9,37,45,49–52,65], it is unable to accurately reflect the distribution of authentic data and (2) the difficulty of defining and selecting reliable interval lengths leaves existing methods unable to provide adequate prediction accuracy [10,13]. These two drawbacks created problems for fuzzy time series forecasting models. First, distance partitioning sets each linguistic interval range to a uniform width, but this can easily result in either excessive linguistic values [60] or excessively short intervals which can lead to the generation of null sets among the FLRs. In fact, the magnitude of stock price movements is often severe and difficult to predict. Thus, the distribution of stock price datasets is always skewed, and is thus better suited to the frequency partitioning method. Secondly, reliable interval lengths are hard to obtain from dynamic and complex time series models (e.g., daily stock prices) [10,13]. However, the entropy-based discretization method has better results with datasets with continuous attributes than unsupervised discretization methods [17,18].

Granular computing (GrC) techniques can provide appropriate solutions to the limitations of fuzzy time series. GrC applies concepts including value discretization, group aggregation, classification or universe clustering to solve a range of problems [63]. To sum up, GrC is suddenly becoming an emerging technology and can be viewed as a superset of fuzzy information granulation, rough sets and interval computations [46]. From this perspective, the main research issues of granular computing for data mining can be categorized as belonging to three types: rule representation, rule mining, and soft computing integration [58]. Rule representation means one fuzzy granule can be determined to generalize constraints and can be represented in natural language. Rule mining refers to the capacity of GrC to obtain more general rules by grouping attributes into granules. It can also be used to investigate semantic relationships among these attributes through establishing a hierarchy of granules. Finally, soft computing integration refers to combining GrC with evolutionary computation or bio-inspired computation to significantly improve forecasting performance. Here, this research proposed one novel fuzzy time series method to forecast the stock market prices through binning-based partitioning and entropy-based discretization methods to obtain reliable interval lengths from skewed datasets. In addition, the entropy-based discretization method can also adjust the length of the individual intervals for each iteration and obtain reliable interval lengths. The proposed model was applied to experimental datasets taken from the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX), Dow-Jones Industrial Average (DJIA), S&P 500, and IBOVESPA stock indexes. Results were then compared against the models suggested by Chen (1996) [7], Yu (2005) [59], Chang et al. (2011) [5], Hsieh et al. (2011) [23], Chen and Chen (2011) [8], Chen et al. (2011) [11], Cheng et al. (2013) [16], Chen and Kao (2013) [10], and three support vector regression (SVR) models with different kernel functions [53,54]. This research also compared our model to current economic models including GARCH [3], GJR-GARCH (Glosten–Jagannathan–Runkle GARCH) [21], and Fuzzy GJR-GARCH [31].

The structure of this research is illustrated as follows. Section 2 investigates relevant studies regarding fuzzy time series definitions, fuzzy time series models, and entropy concepts. The research methodology is described in Section 3. In Section 4, the experimental design and experimental results are discussed. Finally, Section 5 summarizes the conclusions and potential issues in the future researches.

2. Literature review

Solutions to strategic decision problems rely heavily on accurate forecasting to reduce losses from poor decisions and to increase firm competitiveness. Modern firms are increasingly reliant on fuzzy time series for forecasting, and this research reviews theories related to fuzzy time series and related interval computing methods.

Download English Version:

https://daneshyari.com/en/article/392246

Download Persian Version:

https://daneshyari.com/article/392246

Daneshyari.com