Applied Soft Computing 14 (2014) 156–166

Contents lists available at ScienceDirect

Applied Soft Computing

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

Online fuzzy time series analysis based on entropy discretization and a Fast Fourier Transform

Mu-Yen Chen*, Bo-Tsuen Chen

Department of Information Management, National Taichung University of Science and Technology, Taichung 404, Taiwan, ROC

ARTICLE INFO

Article history:
Received 22 August 2012
Received in revised form 28 July 2013
Accepted 29 July 2013
Available online 29 August 2013

Keywords:
Fuzzy time series
Entropy-based discretization
Fast Fourier Transform algorithm

ABSTRACT

Fuzzy time series analysis has been used successfully for forecasting in various domains including stock performance, academic enrollment, temperature, and traffic patterns. Research in this field has concentrated primarily on two issues: the reasonable partition of discourse, and defuzzification methods for discrete datasets. Both issues have a huge impact on the prediction performance of forecasting models. This paper integrates the entropy discretization technique with a Fast Fourier Transform (FFT) algorithm to develop a novel fuzzy time series forecasting model to resolve these issues. The Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) and Dow-Jones Industrial Average (DJIA) financial datasets were used to evaluate the model's performance. The results demonstrate that the presented model is a major improvement over previous fuzzy time series models produced by Chen (1996), Yu (2005), Chang et al. (2011), and Hsieh et al. (2011), and five other conventional time series models. The proposed model is implemented using the bootstrapping method, after which it incrementally updates its prediction capability. Results show that the proposed model's incremental learning mechanism allows it to effectively handle large online financial datasets.

1. Introduction

Time series analysis refers to the prediction of future trends based on historical records. Conventional forecasting methods are based on statistical or mathematical concepts, including the autoregressive moving average model [4,7], factor analysis [44], linear multiple discriminant approaches (MDA) [1], and univariate approaches [3]. These methods assume the time series to be stationary around a time-invariant mean. In finance, the efficient markets hypothesis (EMH) claims that current stock prices fully reflect firm value, and assumes that the efficiency of the stock market prevents the earning of excess profits. In other words, EMH assumes that profits cannot be made from predicting price movements. However, this hypothesis has been refuted by considerable empirical evidence which shows that markets are highly nonlinear and not truly efficient. Nevertheless, mistakes in the data collection process may create errors or ambiguity in time series data. Thus, time series data should always be represented by fuzzy sets with linguistic expressions. However, conventional time series methods have difficulty processing time series forecasting puzzles in a linguistic terms setting. In fact, fuzzy time series are commonly used to produce stock price forecasts because they can be applied to linguistic value datasets and model the fuzzy relationships among observations.

A literature survey on fuzzy time series and financial forecasting reveals three major drawbacks: (1) a lack of reliable interval lengths [11], (2) an excess of linguistic values [46], and (3) intervals being set too short – a condition which can result in some null sets among the fuzzy logical relationships (FLRs). The fuzzy relationship between two consecutive observations can be referred to as an FLR [38,40], a detailed definition of which is given in Section 2.2. This paper integrates the entropy discretization technique with a Fast Fourier Transform (FFT) algorithm to implement a novel fuzzy time series forecasting model for the accurate prediction of stock prices. The proposed model considers several factors critical for prediction [11] such as a reasonable linguistic value for the fuzzy logical relationship, a reasonable universe of discourse, and reliable lengths of intervals. Experimental analysis is performed on datasets covering a seven-year period of both the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) and the Dow-Jones Industrial Average (DJIA). The analysis results are compared to those produced by models suggested by Chen [10], Yu [45], Chang et al. [8], and Hsieh et al. [19]. The results are also compared to those from five conventional time series models. Experimental results indicate that the prediction accuracy of the proposed model is higher than that of the listed conventional and hybrid fuzzy time series models.

The remainder of this paper is organized as follows. A survey of fuzzy time series models and the FFT algorithm is provide in Section 2. Section 3 presents a brief description of the research

* Corresponding author. Tel.: +886 4 22196314.
E-mail addresses: mychen@nutc.edu.tw, s1899b109@nutc.edu.tw (M.-Y. Chen).

1568-4946/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2013.07.024

methodology. Section 4 details the empirical analysis of the forecasting results from the TAIEX and DJIA datasets. Finally, our conclusions and recommendations for future studies are drawn in Section 5.

2. Related works

This research applies entropy-based discretization and a Fast Fourier Transform algorithm to implement fuzzy time series forecasting. This section is a brief introduction to fuzzy time series and the Fast Fourier Transform algorithm.

2.1. Fuzzy time series models

Internet usage continues to rise, and the popularity of the now-ubiquitous computing devices rises along with it. The availability of many large volume and distributed data flows has furthered the evolution of system designs; thus, in the future, learning algorithms and stream mining will be the core approach [17]. Learning algorithms can be categorized primary into online, offline, incremental, supervised, and unsupervised approaches [25]. These approaches can be used to solve different kinds of problems in various models of evolving intelligent systems (EIS) [2], including computational intelligence-evolving simple connectionist systems, evolving rule-based and fuzzy systems, evolving Takagi-Sugeno fuzzy systems, and many different hybrid models. Among these evolving intelligent system techniques, the evolving fuzzy system (EFS) is an ideal model for various applications emerging in dynamically evolving environments, especially considering its capability for self-adaptive tuning of parameters, and its online, real-time and incremental learning mechanisms for model revisions [6,31]. These learning mechanisms save retraining time, achieve high accuracy, and prevent memory constraint limitations with incoming time-intensive datasets [5,35]. Therefore, the capabilities of fuzzy theory have become more suitable and attractive for time series or online data stream models.

Stock prediction research generally uses time series analysis [26] and multiple regression models to deal with large financial datasets. Conventional time series methods such as ARIMA (Autocorrelation Regressive Integrated Moving Average) [7], ARCH (Autoregressive Conditional Heteroskedasticity) [15], and GARCH (Generalized ARCH) [4] are widely used to deal with stationary and non-stationary time series problems, but are unable to predict future trends with linguistic historical records, which are often used to record daily observations. The fuzzy time series was developed to manage such linguistic problems.

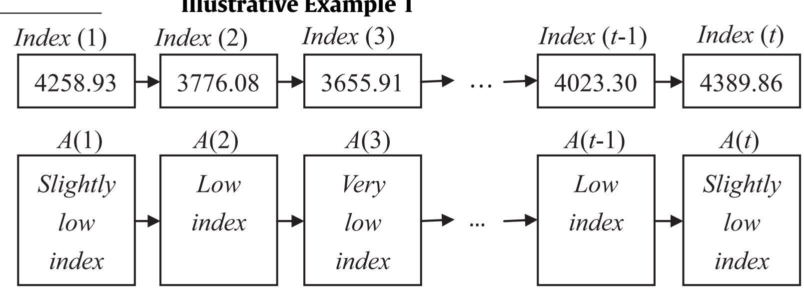
Fuzzy theory was originally proposed by Zadeh [48] to handle human linguistic expressions; it subsequently performed well in both academic and practical applications. However, fuzzy theory failed to address dynamic processes and observations with linguistic values. Song and Chissom [39] first presented a fuzzy time series model to solve this kind of dynamic process in 1993, but the max–min composition operations in this model required a highly complicated computational process. Chen [10] proposed simplified arithmetic operations to deal with this, but Chen's model suffered from a lack of proper weighting for the FLRs. Since then, most studies have focused on the lengths of the linguistic intervals [32] and the problem of proper weighting. For instance, Huarng (2001) presented distribution-based and average-based approaches to decide the reasonable interval length and discussed the influence of different interval lengths to enhance prediction performance [21]. Yu (2005) incorporated a weighted mechanism into a fuzzy time series model and improved the model's performance in predicting TAIEX stock prices [45]. Subsequent studies developed many modifications including first-order models [22,23], high-order models [29,33,37], the deterministic vector long-term forecasting (DVL) model [30], ratio-based lengths of intervals [20], evolution computation models [28], adaptive expectation models [11,27], and hybrid models [8,12,19,47]. Unfortunately, the lack of reliable interval lengths requires all of the above models to set linguistic intervals based on distributions and averages, resulting in negligible improvements in stock price forecasting accuracy.

2.2. Fuzzy time series definitions and algorithm

Song and Chissom [38–40] applied fuzzy set concepts into time series datasets to solve the human linguistic terms concerns. The significant difference between conventional time series and fuzzy time series is that the former refers to real numbers, while the latter is constructed from fuzzy sets [48]. Recently, many researchers proposed their improved fuzzy time series models and applied them to the prediction of course enrollments [10,21,39,41], temperatures [43], stock indices [9,11,45,46] and other domains [24,42]. The related fuzzy time series definitions based on Song and Chissom [40] are discussed in the following section.

Definition 1. Let U, a subset of R^1, be the universe of discourse, where $U = \{u_1, u_2, \ldots, u_n\}$, in which the possible linguistic values $f_{A(T)}$ of fuzzy sets $A(t)$ are defined, where $f_{A(T)}$ denotes the membership function of the fuzzy set $A(t)$, $f_{A(T)}: U \to [0, 1]$, and the $A(t)$ can be understood as a linguistic variable which is a collection of $f_{A(T)}(u_1), f_{A(T)}(u_2), \ldots, f_{A(T)}(u_n)$. Then $F = \{A(t), t = 1, 2, \ldots\}$ is called a fuzzy time series defined on U.

Illustrative Example 1



Assume the stock index on the first day is 4258.93, denoted as *Index* (1) = 4258.93, while the stock price on the second day is 3776.08, denoted as *Index* (2) = 3776.08, and the stock price on day t is 4389.86, denoted as *Index* (t) = 4389.86. These stock prices would then each obtain a linguistic value following the fuzzification process; for instance, $A(1)$ = "*Slightly low index*," $A(2)$ = "*Low index*", $A(3)$ = "*Very low index*," $A(t\text{-}1)$ = "*Low index*," $A(t)$ = "*Slightly low index*." From Definition 1 above, the F is a fuzzy time series definition on U, and F = {"*Slightly low index*," "*Low index*," "*Very low index*," . . ., "*Low index*," "*Slightly low index*"}.

Definition 2. Assume that $A(t)$ is caused by $A(t-1)$ only, denoted as $A(t-1) \to A(t)$; then this fuzzy relationship can be expressed as $A(t) = A(t-1) \circ R$, where $A(t-1) \circ R$ is called the first-order model of $A(t)$, R is the fuzzy relationship between $A(t-1)$ and $A(t)$ in the time independent case, and "$\circ$" is the max–min composition operator.