International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Long-term forecasting of time series based on linear fuzzy information granules and fuzzy inference system

Xiyang Yang [a], Fusheng Yu [b,*], Witold Pedrycz [c]

[a] Key laboratory of Intelligent Computing and Information Processing of Fujian Province, Quanzhou Normal University, Quanzhou, China
[b] School of Mathematic Science, Beijing Normal University, Beijing, China
[c] Department of Electrical and Computer Engineering, University of Alberta, Edmonton T6R 2V4 AB, Canada

ARTICLE INFO

Article history:
Received 8 April 2016
Received in revised form 14 August 2016
Accepted 29 October 2016
Available online xxxx

Keywords:
Linear fuzzy information granule
Gaussian fuzzy number
Fuzzy inference system
Long-term forecasting
Time series

ABSTRACT

Long-term time series forecasting is a challenging problem both in theory and in practice. Although the idea of information granulation has been shown to be an essential concept and algorithmic pursuit in time series prediction, there is still an acute need for developing a sound conceptual framework for time series prediction so that information granulation can capture the essence of collections of data better, including average and trend information. In this paper, a novel type of fuzzy information granule involving a time-dependent (non-stationary) membership function is proposed to structure numerical time series into granular time series. We show that the underlying arithmetic along with the concept of distance for this type of information granules can be expressed in a simple way, which facilitates the ensuing processing of information granules. With this regard, distances between observation granules and antecedent granules presented in fuzzy rules can be easily determined. The design of long-term prediction method based on fuzzy inference system is then realized through interpolation completed with the aid of fuzzy rules. Experiments involving chaotic Mackey–Glass time series and real-world time series demonstrate that the proposed model produces better long-term forecasting than some existing numeric models such as Autoregressive (AR) models, nonlinear autoregressive (NAR) neural networks, Support Vector Regression (SVR) and fuzzy inference systems involving triangular and interval information granules.

1. Introduction

Time series refers to a sequence of data, which are collected at time intervals. Time series commonly appear in numerous areas including econometrics, finance, environment, ecology and many others. The past decades have witnessed intensive developments in applications of time series analysis to various fields owing to the progresses in network hardware and the algorithmic developments. Systems generate large amounts of temporal data, from which one can discover knowledge and learn about essential relationships existing in data [1]. In recent years, time series forecasting arises among important issues in time series analysis and as such has received significant attention.

* Corresponding author.
E-mail address: yufusheng@bnu.edu.cn (F. Yu).

http://dx.doi.org/10.1016/j.ijar.2016.10.010
0888-613X/

1.1. Time series forecasting and related methods: a brief review

Autoregressive model (AR), which forms a special case of the more general ARMA model [2] is one of the most commonly used time series models. For example, Duran et al. proposed autoregressive with exogenous variable (ARX) models to carry out wind power prediction [3]. Pappas et al. used an autoregressive moving average model to predict electricity demand loads [4]. Erdem et al. considered an ARMA model to predict tuples of wind speed and direction [5]. These traditional time series assume that a value for any given time moment depends linearly on the values reported in some previous time instants and stochastic (probabilistic) terms [6,7]. Although these time series methods are widely used, the data to be analyzed in this way require to satisfy some assumptions, such as stationary and normality [8], which may not always hold for some real world data encountered in economics, finance, and environmental sciences.

Artificial Neural Networks (ANNs) form another type of prediction models with main characteristics involving learning from examples, generalization, nonlinearity, which make them attractive to forecast time series in many areas like water end-use demand [9], flood forecasting [10], electrical energy consumption [11]. For example, Benmouiza et al. used k-means and nonlinear autoregressive (NAR) neural networks to forecast hourly global horizontal solar radiation [12], Bennett et al. applied NAR neural networks to predict a residential water end-use demand [9]. Although ANNs come with the abilities to capture the nonlinear relationship between input and output variables, the prediction mechanism is difficult for humans to comprehend due to their low interpretability, because the trained ANNs are basically data-driven black-box models [13].

Support Vector Regression (SVR) forms a certain type of supervised learning models based on statistical learning theories. SVR comes with good generalization capabilities as it utilizes the structural risk minimization, rather than the empirical risk minimization to derive learning algorithms for regression analysis [14]. SVR has been widely used in the area of time series prediction. For example, Lu et al. integrated independent component analysis (ICA) with SVMs to predict financial time series [15]. Ö. Baydaroglu and K. Koçak [16] applied SVR to estimate evaporation from free water surface for the prediction of water amount in reservoirs. Wen et al. [17] used SSA and SVM for stock price prediction. Tay and Cao [18] proposed a C-ascending support vector machine by modifying the cost parameter C of SVR to model non-stationary financial time series. Kaneda et al. [19] combine SVR and a data extraction method to build a smart greenhouse environmental control system. Other forecasting applications based on SVR include credit risk [20], business default forecasting [21], customer churn prediction [22], warranty claim forecasting [23], and image enhancement [24].

As mentioned in [15] and [25], SVR has some drawbacks. In particular, the accuracy of SVMs may be strongly influenced by noisy and/or imbalanced data and the SVMs' efficiency is directly affected by the dimensionality of input data. The generalization ability may be limited if the parameters of the SVR, type of kernel functions and parameters in such kernels are not properly chosen. However, these parameters are hard to decide without enough prior knowledge. Considering the complex parameters selection, a large number of evolutionary algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Direct Search (DS) have been employed. For example, a real-value GA is used to optimize the SVM for predicting bankruptcy [26], Liu et al. proposed a network traffic forecast model by SVR algorithm optimized by global artificial fish swarm algorithm [27]. Nieto et al. put forward a PSO-SVM model to predict the remaining life of aircraft engines [28]. Wang et al. applied Direct Search algorithm to optimize the parameters in a prediction method based on a combination of vector AR and SVM regression [29].

The methods mentioned above usually involve a single point ahead prediction, which may perform poorly in situation that requires predicting a sequence of values over a long time horizon. One can note that there is no unique definition for long-term prediction. Usually, short-term forecasts refer to a single step prediction; and long-term forecasts often refer to 1–10 year predictions [30,31]. In this study, given the data used here, long-term forecasts refer to multi-step predictions [32,33], and the forecasting horizon is at least 3 months for data collected on a monthly basis.

Fuzzy information granulation [34,35] constitutes an important tool to provide appropriate solutions in predicting long-term future values. To granulate a large-scale time series, the series is first broken down into successive pieces of simpler subseries and each subseries is then represented by a fuzzy set, referred to as fuzzy information granule (FIG). The main advantages of applying FIG to time series forecasting problem include: (1) FIG based forecasting is useful for long-term prediction, because a predicted granule usually involves many predicted values positioned in the future time window; (2) and a predicted granule is meaningful, which makes it easier and more transparent for human to interpret changes in this window; (3) furthermore, analysis of this time series is realized at the level of newly constructed granules, instead of the original numeric time series. Consequently, the dimensionality of the problem becomes greatly reduced and the computation overhead is reduced as well. FIG can be sought as an emerging tool in time series predicting. For example, Ruan et al. [1] integrated FIG and SVM to develop a fast interval prediction method for large-scale, nonlinear time series with noisy data. Wang et al. [36,37] considered the construction method of information granules with unequal temporal intervals lengths to improve accuracy of forecasting. Lu et al. [38] proposed a new modeling approach to realize interval prediction via fuzzy granules that are built by exploiting the principle of justifiable granularity. Hryniewicz et al. [39] provided a conceptual framework for the Bayesian time series forecasting within the realm of granule computing. Lu et al. extended a Sugeno-type model to its granular counterpart in forecasting of time series [40]. Lu et al. proposed a process of granulating to partition the universe of discourse to improve forecasting in fuzzy time series [41]. Chen et al. proposed a partition method for the construction of granules for fuzzy time series to forecast stock market prices [42].

Please cite this article in press as: X. Yang et al., Long-term forecasting of time series based on linear fuzzy information granules and fuzzy inference system, Int. J. Approx. Reason. (2016), http://dx.doi.org/10.1016/j.ijar.2016.10.010