Information Sciences 220 (2013) 196–209

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

Revenue forecasting using a least-squares support vector regression model in a fuzzy environment

Kuo-Ping Lin [a], Ping-Feng Pai [b,*], Yu-Ming Lu [a], Ping-Teng Chang [c]

[a] Department of Information Management, Lunghwa University of Science and Technology, Taoyuan 333, Taiwan
[b] Department of Information Management, National Chi Nan University, 1 University Rd., Puli, Nantou 545, Taiwan
[c] Department of Industrial Engineering and Enterprise Information, Box 985, Tunghai University, Taichung 407, Taiwan

A R T I C L E I N F O

Article history:
Available online 22 September 2011

Keywords:
Least-squares support vector regression
Membership function
Genetic algorithms
Revenue forecasting

A B S T R A C T

Revenue forecasting is difficult but essential for companies that want to create high-quality revenue budgets, especially in an uncertain economic environment with changing government policies. Under these conditions, the subjective judgment of decision makers is a crucial factor in making accurate forecasts. This investigation develops a fuzzy least-squares support vector regression model with genetic algorithms (FLSSVRGA) to forecast seasonal revenues. The FLSSVRGA uses the *H*-level to control the possibility distribution range yielded by the fuzzy model and to provide the fuzzy prediction interval. Depending on various factors, such as the global economy and government policies, a decision maker can elect a different level for *H* using the FLSSVRGA. The proposed FLSSVRGA model is a rolling forecasting model with time series data updated monthly that predicts revenue for the coming month. Four other forecasting models: the seasonal autoregressive integrated moving average (SARIMA), generalized regression neural networks (GRNN), support vector regression with genetic algorithms (SVRGA) and least-squares support vector regression with genetic algorithms (LSSVRGA), are employed to forecast the same data sets. The experimental results indicate that the FLSSVRGA model outperforms all four models in terms of forecasting accuracy. Thus, the FLSSVRGA model is a useful alternative for forecasting seasonal time series data in an uncertain environment; it can provide a user-defined fuzzy prediction interval for decision makers.

1. Introduction

Seasonal time series prediction has been widely investigated and used in many fields. The seasonal time series is a sequence of seasonal data points recorded sequentially and chronologically. In the literature, trend and seasonality have been modeled using different approaches. Over the past several decades, many studies have focused on developing and improving seasonal time series forecasting models. One of the most popular and best-known approaches is the SARIMA model [1]. The SARIMA model has been successfully utilized in many fields of forecasting, such as a soil dryness index [43], tourism demand [23,35] and municipal solid waste management [55]. However, the seasonal time series is a complex and nonlinear problem. The artificial neural network (ANN) model provides an alternative for forecasting seasonal data patterns [91]. Some literatures [25,54,57,77,85] indicate that the ANN can obtain desirable results in seasonal and trend forecasting. However, in recent years numerous authors have proposed a hybrid approach method to improve seasonal prediction accuracy [15,19,81].

* Corresponding author. Tel.: +886 49 2910960.
E-mail address: paipf@ncnu.edu.tw (P.-F. Pai).

0020-0255/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2011.09.003

Recently, support vector regression (SVR) has emerged as a prediction method for nonlinear time series data [83]. With the introduction of Vapnik's ε-insensitive loss function, SVR schemes have been extended to cope with forecasting problems in various areas, including control systems [66,86], thickness strains [69], bioinformation [2], electric loads [10,20,21], electricity consumption [39], customer demand [42], finance [36,40,78], intermittent demand [34], tourism demand [13,31,57], air quality [48,49,56], wind speed [52,65], rainfall [30,33], prices for the electricity market [14,22], and flood control [50,89]. Pai et al. [59] were the first to propose seasonal decomposition using SVR. The experimental results indicated that seasonal SVR could perform better than the traditional SVRGA model. Van Gestel et al. [82] combined the Bayesian evidence framework with LS-SVR for financial time series forecasting. In predicting the DAX30 index, LS-SVR achieved better performance than the ARIMA and nonparametric models did. Li et al. [45] proposed a consensus LS-SVR method for calibrating the near-infrared (NIR) spectra. By comparing LS-SVR with the partial least squares (PLS) model, they showed that LS-SVR generates more stable and accurate prediction results. Li et al. [44] also used LS-SVR to forecast the parts weight in a plastic injection model. The results indicated that LS-SVR is a very effective forecasting method.

There have been other extensions of the SVR approach. Goodarzi et al. [24] developed a genetic algorithm/least square support vector regression (GA-LSSVR) to predict pH indicators. The prediction ability of this model was found to be superior to classical regression tree and SVR methods. Lin et al. [46] used least square support vector regression with a genetic algorithm (LS-SVRGA) to forecast Radio-Wave Path-Loss. Compared with various other methods, including SVRGA, GRNN and the traditional Radio-Wave Path-Loss approach, the LS-SVRGA was superior in terms of forecasting accuracy. Qin et al. [62] employed SVR and LS-SVR to approach the hormetic dose–response curves (DRC). The results showed that the SVR and LS-SVR can accurately fit the DRC. Quan et al. [63] developed a weighted least squares support vector machine (WLS-SVM) for Henon time series, Lorenz time series and neuronal data prediction. It was reported that the WLS-SVM local region can keep a stable forecast in a range of training sample size and has superior performance to the least squares support vector machine. Finally, Yang et al. [87] adopted GA-LSSVR to forecast the conductivity and tensile strength performance of carbon fiber/ABS composite material. The GA-LSSVR model can be effectively used for predicting and optimizing the conductivity of carbon fiber/ABS composites; moreover, it outperformed the principle component analysis-genetic back propagation neural networks model (PCA-GABPNN).

While the foregoing results are promising, some time series data contain uncertain or unpredictable factors. Hence, this study employs fuzzy regression (FR) [74] to deal with unpredictable factors or uncertainties in seasonal time series data. The fuzzy regression method is based on fuzzy set theory, which provides a possibility distribution for imprecise or vague phenomenon. This distribution is expressed by fuzzy parameters or coefficients. Fuzzy regression represents the data without losing their original meaning; hence, it can analyze both trends of variability and the mean of data. A number of investigations have proposed different FR methods (e.g., see [3,5,7–9,17,18,38,41,47,64,75,76,84], among others). One that stands out is the study of Chang [6], who has proposed seasonal fuzzy regression. Here, fuzzy seasonality is defined by realizing the membership grades of the seasons to the fuzzy regression model. Unfortunately, fuzzy seasonality analysis methods have not been widely studied. This research is the first to discuss the phenomenon of fuzzy seasonality, which it applies to the prediction of milk product sales. Moreover, approaches dealing with fuzzy time series problem were firstly introduced by Song and Chissom [70–72]. Since then, fuzzy time series models have been applied in various fields [1,12,11,29,37,53,67,68,79,80]. Pedrycz and Vasilakos [60] used the fuzzy clustering approach to handle uncertain data. The proposed approach generated meaningful linguistic labels and adopted the linguistic labels to design a fuzzy system. For time series problems, the developed fuzzy approach can obtain better performance than traditional models. Furthermore, some studies combining the fuzzy set theory with SVR to handle uncertain problems have been reported [4,26,27,32,41,88].

Combining different forecasting models can be an effective way of improving forecasting performance [58,92]. Therefore, this study develops an FLSSVRGA model that exploits the unique strength of decomposition techniques and fuzzy regression to forecast monthly revenue data. This paper is organized as follows. The following section (Section 2) introduces the LSSVR approach. Section 3 presents the FLSSVRGA model. In Section 4, numerical examples are utilized to demonstrate performances of different forecasting models, and Section 5 concludes the study.

2. Least-squares support vector regression

The LSSVR [73,61] approach is to approximate an unknown function by using a training data set $\{(x_i, y_i), i = 1, \ldots, N\}$. The regression function can be formulated as follows:

$$y = f(x) = w^T \varphi(x_i) + b, \tag{1}$$

where $\varphi(\cdot)$ denotes the feature of the inputs, and w and b indicate coefficients. LSSVR introduces a least squares version to SVR by formulating the regression problem as:

$$\begin{aligned} &Min && J_1(w, b, e) = \tfrac{1}{2}\|w\|^2 + \tfrac{1}{2} C \sum_{i=1}^{N} e_i^2 \\ &subjective\ to && y_i = w^T \varphi(x_i) + b + e_i, \quad i = 1, 2, \ldots, N. \end{aligned} \tag{2}$$

Introduce the Lagrangian as