Expert Systems with Applications 42 (2015) 3868–3874

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Time series forecasting based on wavelet filtering

Tae Woo Joo [1], Seoung Bum Kim *

School of Industrial Management Engineering, Korea University, Anam-Dong, Seongbuk-Gu, Seoul 136-713, Republic of Korea

ARTICLE INFO

Article history:
Available online 20 January 2015

Keywords:
ARIMA
Forecasting
Time series
Wavelet transforms

ABSTRACT

Forecasting time series data is one of the most important issues involved in numerous applications in real life. Time series data have been analyzed in either the time or frequency domains. The objective of this study is to propose a forecasting method based on wavelet filtering. The proposed method decomposes the original time series into the trend and variation parts and constructs a separate model for each part. Simulation and real case studies were conducted to examine the properties of the proposed method under various scenarios and compare its performance with time series forecasting models without wavelet filtering. The results from both simulated and real data showed that the proposed method based on wavelet filtering yielded more accurate results than the models without wavelet filtering in terms of mean absolute percentage error criterion.

1. Introduction

A time series is a set of observations collected over time (Lai et al., 2010). The analysis of time series data is one of the most important areas in statistics in both theory and application. The main objective of time series analysis is to obtain inherent structural characteristics, such as autocorrelations, trends, and seasonal variations, and to use this information to formulate an appropriate mathematical model for analysis and prediction (Anderson, 1971). In general, time series data can be analyzed in either the time or frequency domains. The most widely used methods in analyzing the time domain include time series regression, decomposition methods, exponential smoothing, and the Box–Jenkins autoregressive integrated moving average (Li, Gai, Kang, Wu, & Wang, 2014).

Analyses in the frequency domain are usually conducted for periodic and cyclical observations. The methodologies of the frequency domain are based on Fourier transforms that allow us to determine the number of frequency components and detect the dominant cyclic frequency, all of which are embedded in the time domain. However, Fourier transforms-based methods have limitations in that they require an assumption of stationarity and produce no information associated with time (Croarkin & Tobias, 2006; Popoola, 2007; Tabak and Feitosa, 2010). Further, Fourier transforms do not work well in large data, and thus, they can be implemented only on an interval between 0 and 2π (Tran, 2006).

To address the limitations of Fourier transforms, wavelet transforms that are localized in both the time and frequency domains (Mallat, 1989) have been proposed (Morettin, 1996, 1997; Mousa, Munib, & Moussa, 2005; Percival & Walden, 1999; Priestley, 1996). More specifically, because wavelet basis functions exist over a finite time limit and are typically irregular and asymmetric, they are better suited for those time series analyses that exhibit sharp discontinuities and local behavior (Graps, 1995).

Many prediction approaches based on wavelet transforms have been developed recently (Chen, 2014; Nguyen, Khosravi, Creighton, & Nahavandi, 2015). Nguyen et al. (2015) presents a combination of wavelet features with fuzzy standard additive model for medical diagnosis. In general, if a time series is nonstationary, it is difficult to determine a relevant global model. For example, Hydrological time series forecasting is a difficult task because of its complicated nonlinear, non-stationary and multiscale characteristics (Di, Yang, & Wang, 2014). To overcome this problem, local models based on wavelet transforms have been proposed (Weigend & Mangeas, 1995; Zhang, Coggins, Jabri, Dersch, & Flower, 2001). In addition, it is known that wavelet transforms have the potential to increase the accuracy of time series predictions (Ramsey, 1999; Schlüter & Deuschle, 2010). Once forecasting models using wavelet transforms succeeded in eliminating noise before preceding to construct a model (Alrumaih & Al-Fawzan, 2002), improvement in overall forecasting performance followed (Chou, 2014; Ferbar, Čreslovnik, Mojškerc, & Rajgelj, 2009).

Other approaches that use wavelet transforms estimate components in a structural time series model (Arino, 1995; Sang, 2013; Wong, Wai-Cheung, Zhongjie, & Lui, 2003; Zhang et al., 2001). Arino (1995) proposed a methodology that used wavelet

* Corresponding author. Tel.: +82 2 3290 3397; fax: +82 2 929 5888.
E-mail addresses: mrblond002@gmail.com (T.W. Joo), sbkim1@korea.ac.kr (S.B. Kim).
[1] Tel.: +82 2 3290 3769.

http://dx.doi.org/10.1016/j.eswa.2015.01.026
0957-4174/

transforms to decompose time series into long-term trends and seasonal components. Specifically, let $\mathbf{y} = \{y_t: t = 1, \ldots, T\}$ be a time series set. $\mathbf{y}$ is decomposed into two sets $\mathbf{z}^1 = \{z_t^1: t = 1, \ldots, T\}$ and $\mathbf{z}^2 = \{z_t^2: t = 1, \ldots, T\}$, that is, $\mathbf{y} = \mathbf{z}^1 + \mathbf{z}^2$, where $\mathbf{z}^1$ and $\mathbf{z}^2$, respectively, represent long-term trends and seasonal behavior. Each component is used to construct an appropriate model, and total prediction accuracy can be obtained by aggregating the prediction results from each component model. However, the generalizability of this method is questionable because Arino's method considered only limited cases such as long-term trends and seasonal variations.

Moreover, some methods applied linear or nonlinear predictive models to wavelet coefficients (approximation and detail coefficients) at each level, and the final prediction was obtained by inverting the predicted wavelet coefficients. Renaud, Starck, and Murtagh (2002) introduced the multiscale autoregressive models that use Haar wavelets and scale coefficients during decomposition. A method that combines nonlinear models with wavelet coefficients has also been proposed (Chen, Qian, & Meng, 2013; Hadaś-Dyduch, 2014; Rocha, Paredes, Carvalho, Henriques, & Harris, 2010; Soltani, 2002; Wang, 2014). Chen, Nicolis, and Vidakovic (2010) classified the wavelet coefficients into trend, seasonal, and high frequency components and used them to construct forecasting models based on exponential smoothing, harmonic regression, and autoregressive moving average with exogenous input (ARMAX) model. However, these methods involve a high computational load because they considered each level of a series. Moreover, considering each level of series could cause an overfitting problem.

In this paper, we propose a forecasting method that uses wavelet filtering. Through wavelet transforms, the series is partitioned into two parts (trends and variations). We then construct a separate model for each part. By this partitioning, the proposed method is especially useful for the time series with a large amount of noise. The potential overfitting problem of this approach caused by using two separate models is addressed by adjusting the decomposition levels.

The rest of this paper is organized as follows. In Section 2, we briefly introduce the basic concept of wavelet transforms. In Section 3, we present our proposed forecasting approach based on wavelet filtering. Section 4 presents a simulation study to examine the properties of the proposed method and evaluate its performance under various scenarios. Section 4.2 presents the experimental results from real-life problems. Finally, Section 5 contains our concluding remarks.

2. Wavelet transforms

Wavelets have the advantages of the locality of the analysis and their ability to handle multiscale information efficiently. Numerous studies of wavelets have been conducted in the fields of signal/image processing (Meerwald & Uhl, 2001; Prasad & Iyengar, 1997; Rao & Bopardikar, 1998; Subasi, 2007; Avci & Derya, 2008), statistics (Abramovich, Bailey, & Sapatinas, 2000; Antoniadis, 1999; Vidakovic, 1999), and manufacturing processes (Guo, Linyan, Gang, & Song, 2008; Jeong, Lu, & Wang, 2006; Jin & Shi, 2001; Lada, Lu, & Wilson, 2002; Saravanan & Ramachandran, 2009). Although Fourier transforms give only the frequency information of given data, wavelet transforms can simultaneously deliver both the time and frequency localizations (Mallat, 1989). Wavelet transforms are conducted from wavelet basis functions that consist of a scale wavelet part $\phi(t)$ and a detail part $\psi(t)$.

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k),$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \tag{1}$$

$$\int \phi(t)dx = 1 \int \psi(t)dx = 0 \quad \phi(t), \quad \psi(t) \in L^2(\mathbb{R}),$$

where j and k denote, respectively, the scaling parameter and translation index. $L^2(\mathbb{R})$ is the space of square integrable real function defined on the real line $\mathbb{R}$. The data can be decomposed into the following wavelet series form:

$$y(t) = \sum_{k=1}^{n} c_{j,k}\,\phi_{j,k}(t) + \sum_{j=1}^{J}\sum_{k=1}^{n} d_{j,k}\,\psi_{j,k}(t), \tag{2}$$

where $c_{j,k}$ represents the approximation coefficient at scale j and location k; J is the decomposition level; $d_{j,k}$ represents the detail coefficient at scale j and location k; and n is the size of the time series data. Reconstruction can also be done through Mallet's pyramid algorithm (Burt & Adelson, 1983). Fig. 1 shows an example of the overall process of decomposition and reconstruction by wavelet transforms.

In Fig. 1, $c_{j,k}$ represents the coarse approximation (trend and seasonality) and $d_{j,k}$ represents detailed information (noise or random fluctuation). The difference between the first level approximation coefficients $c_{1,k}$ and original series y yields the detail coefficients of the first level $d_{1,k}$. To obtain $c_{2,k}$, $c_{1,k}$ is approximated by a set of basis functions. As can be seen from Fig. 1, when the level becomes higher, $c_{j,k}$ represents the overall pattern of the original series.

3. Proposed forecasting method

3.1. Data decomposition through wavelet filtering

Our proposed method consists of three steps: (1) wavelet transforms to decompose the data in the trend part (TP) and variation part (VP), (2) determination of the optimal decomposition level, and (3) construction of forecasting models. First, wavelet transforms are conducted to the original time series data to obtain the same number of coefficients as the size of the data. Further, denoising that shrinks the empirical detail wavelet coefficients toward zero is performed to remove noise. We used soft thresholding in this study (Donoho & Johnstone, 1994). A wavelet filtering method is used to partition the data into the TP and VP (Chang and Yadama, 2010). More precisely, the TP can be obtained by setting all the detail coefficients at all levels $d_{j,k}(j = 1, \ldots, J)$ to zero while maintaining all coarse approximation coefficients, $c_{j,k}(j = 1, \ldots, J)$. Similarly, the VP can be obtained by setting all the coarse approximation coefficients at the last level, $c_{J,k}$ to zero while maintaining all detail coefficients. The TP and VP can be considered the overall trend and variation information in a time series. Fig. 2 shows how a series can be decomposed into the TP and VP by wavelet filtering. Note that in this paper, we used Daubechies wavelets as a basis function whose scale and detail wavelets are shown in Fig. 3.

3.2. Choosing the optimal decomposition level

For modeling, the dataset is divided into training (80%), validation (10%), and testing (10%) sets. The training set is used to create a model, and the testing set is used to evaluate the model. The validation set is used to select the optimal parameters for the decomposition level of wavelet transforms. To determine an adequate decomposition level, we varied the number of levels for both the TP and VP from one to six until the model delivered the best accuracy in terms of the minimum mean absolute percentage error (MAPE). We believe a potential overfitting problem by using two separate models (TP and VP) can be addressed by adjusting the decomposition levels in the wavelet transforms.