Computers & Industrial Engineering 99 (2016) 153–161

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Taiwanese export trade forecasting using firefly algorithm based K-means algorithm and SVR with wavelet transform

R.J. Kuo [a,*], P.S. Li [b]

[a] Department of Industrial Management, National Taiwan University of Science and Technology, No. 43, Section 4, Kee-Lung Road, Taipei 106, Taiwan
[b] AU Optronics Corporation, No. 1 Jhong-Ke Road, Central Taiwan Science Park, Taichung 407, Taiwan

ARTICLE INFO

Article history:
Received 8 March 2016
Received in revised form 9 July 2016
Accepted 12 July 2016
Available online 14 July 2016

Keywords:
Cluster analysis
Firefly algorithm
Support vector regression
Forecasting
Wavelet transform

ABSTRACT

In order to develop a prediction system for export trade value, this study proposes a three-stage forecasting model which integrates wavelet transform, firefly algorithm-based *K*-means algorithms and firefly algorithm-based support vector regression (SVR). First, wavelet transform is utilized to reduce the noise in data preprocessing. Then, the firefly algorithm-based *K*-means algorithm is employed for cluster analysis. Finally, a forecasting model is built for each cluster individually. For evaluation, this study compares methods with and without clustering. In addition, both non-wavelet transform and wavelet transform for data preprocessing are investigated. The experimental results indicate that the forecasting algorithm with both wavelet transform and clustering has better performance. Besides, firefly algorithm-based SVR outperforms the other algorithms.

1. Introduction

Because of Taiwan's relatively small domestic market and the nation's lack of resources, Taiwan's economic development counts heavily on international trade. Thus, being able to make forecasts about export trade is a very important issue. In recent years, because of the innovation of forecasting techniques and improvement in forecasting accuracy, forecasting methodology is necessary for enhancing the decision-making process in both industry and government. However, according to previous researches on export trade forecasting, most forecasting methods usually construct a traditional model or fuzzy time series model as a tool for forecasting future value and data analysis related to export trade (Tu, Hsien-Lun, & Chi-Chen, 2009; Wang, 2011; Wong, Tu, & Wang, 2010). Therefore, other intelligent methods may be employed in order to obtain greater accuracy in forecasting.

Currently, the support vector regression (SVR) model is being widely used in many fields to analyze time series forecasting, for example tourism demand (Chen, 2011; Pai, Hung, & Lin, 2014; Shahrabi, Hadavandi, & Asadi, 2013), traffic flow (Castro-Neto, Jeong, Jeong, & Han, 2009; Hong, Dong, Chen, & Wei, 2011; Hong, Dong, Zheng, & Lai, 2011; Li, Hong, & Kang, 2013) and demand forecasting (Guanghui, 2012; Lu & Wang, 2010; Wu, 2010). Besides, SVR usually derives more precise results than the other methods. However, determining the SVR parameters is very important for establishing the SVR model. Some metaheuristics have been employed to determine SVR parameters, like genetic algorithm (GA) (Cai, Sheng, & Xiao-bin, 2009; Chen & Wang, 2007; Fang, Wang, Qi, & Zheng, 2008; Ju & Wu, 2010), particle swarm optimization algorithm (PSO) (Chen & Liu, 2013; Jiansheng & Enhong, 2010; Siamak, Bahrami Jovein, & Ramezanianpour, 2012), differential algorithm (DE) (Cai, Qu, & Li, 2013; Li & Cai, 2008; Pan, Cheng, & Ding, 2013; Wang, Li, Niu, & Tan, 2012) and firefly algorithm (FA) (Kavousi-Fard, Samet, & Marzbani, 2014; Kazem, Sharifi, Hussain, Saberi, & Hussain, 2013; Xiong, Bao, & Hu, 2014). In addition, a two-stage forecasting model was developed based on SVM and metaheuristic which brings a good predictive value than other machine learning model (Olatomiwa et al., 2015).

This study intends to present a novel three-stage forecasting model which integrates wavelet transform, FA-based *K*-means algorithms and the FA-based SVR model. Wavelet transform is first applied to reduce the noise in data preprocessing. The FA-based *K*-means algorithm is then utilized for cluster analysis. Finally, the FA- based SVR model is used to develop the forecasting model for each cluster individually.

The remainder of this study is organized as follows. Section 2 presents the SVR model, while the proposed three-stage forecasting model is proposed in Section 3. Section 4 shows the experimental results for export forecasting. Finally, the concluding remakes are offered in Section 5.

* Corresponding author.
E-mail addresses: rjkuo@mail.ntust.edu.tw (R.J. Kuo), karenli127@gmail.com (P.S. Li).

http://dx.doi.org/10.1016/j.cie.2016.07.012
0360-8352/

2. Support vector regression

Support vector machine (SVM) is an artificial intelligence method which was first developed by Vapnik (1995). SVM is based on structural risk minimization (SRM) principle which aims to minimize an upper bound of generation error; it consists of the sum of training error and a confidence interval (Guo, Sun, Li, & Wang, 2008). In Vapnik (1999, 1995) promoted the SVM method to support vector regression (SVR) by using a new type of loss function called ε-insensitive loss function which is used to penalize errors as long as they are greater than ε; it is assumed that ε is known beforehand (Guo et al., 2008). SVR is a non-linear kernel-based regression method which seeks to locate the best regression hyperplane with the smallest risk in high dimensional feature space (Yeh, Huang, & Lee, 2011).

The SVM model consists of both classification and regression tasks, it has many advantages, including: a global optimal solution can be found, the result is a general solution avoiding overtraining, and nonlinear solutions can be calculated efficiently due to the usage of inner products (Thissen, van Brakel, de Weijer, Melssen, & Buydens, 2003). The SVM technique has been used in a range of applications, including financial stock market prediction (Ding, Song, & Zen, 2008; Huang, Nakamori, & Wang, 2005; Kim, 2003; Tay & Cao, 2001, 2002), electric load and price (Che & Wang, 2010; Pai & Hong, 2005; Yan & Chowdhury, 2014) and sales forecasting (Wu, 2009).

It is assumed that a set of training data which consists of ℓ data points $\{(x_1, y_1), \ldots, (x_\ell, y_\ell)\}$, i = 1,2,...,$\ell$ and x_i denotes the space of the input pattern, and y_i is the corresponding target value. The input is first mapped onto an n-dimension feature space by $\varphi(x)$, which is a kernel function, to transform the non-linear input into a linear mode in a high dimensional feature space (Lu, 2014). The goal of SVR is to estimate a function $f(x)$ that is as accurate as possible to get the target value y_i for every x_i (Guo et al., 2008). The function f is represented using a linear function, as in Eq. (1):

$$f(x) = (w \cdot \varphi(x)) + b, \tag{1}$$

where $f(x)$ denotes the forecasting value, w is a vector of weight coefficients, b is a bias constant, $\varphi(x)$ is a nonlinear mapping from the input space to the feature space and $(w \cdot \varphi(x))$ describes the dot production in the feature space. In SVR, the problem of nonlinear regression in the lower dimension input space (x) is transformed into a linear regression problem in a high dimension feature space (Fig. 1(a) and (b)) (Lu & Wang, 2010).

The robust ε-sensitive loss function (L_ε) (Fig. 1(c)) given below is the most commonly used equation:

$$L_\varepsilon(f(x), y) = \begin{cases} |f(x) - y| - \varepsilon & \text{if } |f(x) - y| \geqslant \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where ε is a precision parameter representing the radius of the tube located around the regression function $f(x)$. The region enclosed by the tube is called the ε-sensitive zone since the loss function assumes a zero value in this region and the prediction errors with a value smaller than ε are not penalized (Lu & Wang, 2010).

Furthermore, w and b are estimated by minimizing the following optimization problem:

$$\text{Minimize } \frac{1}{2}\|w\|^2 \tag{3}$$

$$\text{Subject to } \begin{cases} y_i - (w, \varphi(x_i) + b) \leqslant \varepsilon \\ (w, \varphi(x_i) + b) - y_i \leqslant \varepsilon \end{cases}$$

In addition, to deal with feasibility issues and to make the method stronger, points forming the ε-sensitive band are not eliminated. Instead, we penalize these points by introducing slack variables ξ_i and ξ_i^* (Kazem et al., 2013). The SVR minimizes the overall errors as follows:

$$\text{Minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}(\xi_i + \xi_i^*) \tag{4}$$

$$\text{Subjected to } \begin{cases} y_i - (w, \varphi(x_i) + b) \leqslant \varepsilon + \xi_i \\ (w, \varphi(x_i) + b) - y_i \leqslant \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geqslant 0 \text{ for } i = 1, \ldots, l \end{cases}$$

The first term of Eq. (4), having the concept of maximizing the distance of two separated training data, is used to regularize weight sizes, penalize large weights and maintain regression function flatness. The second term is to penalize training errors of $f(x)$ and y by using the ε-sensitive loss function (Hong, Dong, Chen, et al., 2011; Hong, Dong, Zheng, et al., 2011). C is a modifying coefficient representing the trade-off between empirical risk and structure risk. The optimum value of each parameter can be solved by Lagrange with an appropriate modifying coefficient C, band area width ε and kernel function K (Kao, Chiu, Lu, & Chang, 2013). The general form of the SVR-based regression function can be written as Eq. (5):

$$f(x, w) = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)K(x_i, x) + b, \tag{5}$$

where α_i, α_i^* are nonzero Lagrangian multipliers and the solution for the dual problem. Any function that meets Mercer's condition can be used as the kernel function. Although several options for the kernel function are available, the most widely used kernel function is the radial basis function (RBF) which is defined as follows:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \tag{6}$$

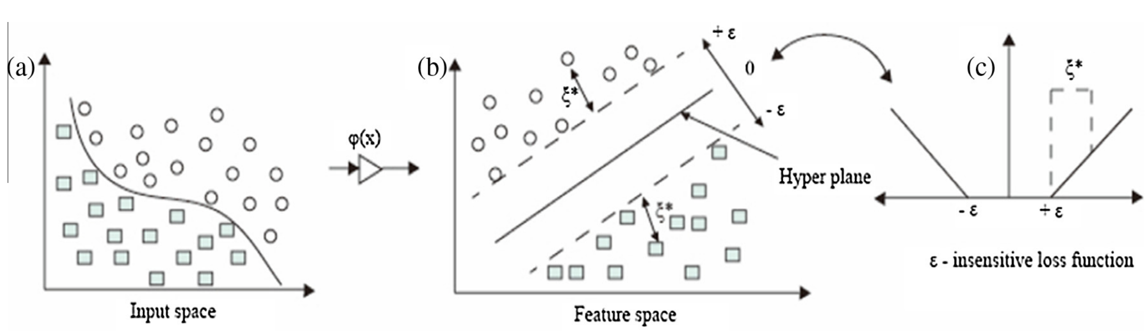

Fig. 1. Transformation process illustration of an SVR model.