



A multiple-kernel support vector regression approach for stock market price forecasting [☆]

Chi-Yuan Yeh, Chi-Wei Huang, Shie-Jue Lee ^{*}

Department of Electrical Engineering, National Sun Yat-Sen University, Kaohsiung 804, Taiwan

ARTICLE INFO

Keywords:

Stock market forecasting
Support vector regression
Multiple-kernel learning
SMO
Gradient projection

ABSTRACT

Support vector regression has been applied to stock market forecasting problems. However, it is usually needed to tune manually the hyperparameters of the kernel functions. Multiple-kernel learning was developed to deal with this problem, by which the kernel matrix weights and Lagrange multipliers can be simultaneously derived through semidefinite programming. However, the amount of time and space required is very demanding. We develop a two-stage multiple-kernel learning algorithm by incorporating sequential minimal optimization and the gradient projection method. By this algorithm, advantages from different hyperparameter settings can be combined and overall system performance can be improved. Besides, the user need not specify the hyperparameter settings in advance, and trial-and-error for determining appropriate hyperparameter settings can then be avoided. Experimental results, obtained by running on datasets taken from Taiwan Capitalization Weighted Stock Index, show that our method performs better than other methods.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Accurate forecasting of stock prices is an appealing yet difficult activity in the modern business world. Many factors influence the behavior of the stock market, including both economic and non-economic. Therefore, stock market forecasting is regarded as one of the most challenging topics in business. In the past, methods based on statistics were proposed for tackling this problem, such as the autoregressive (AR) model (Champernowne, 1948), the autoregressive moving average (ARMA) model (Box & Jenkins, 1994), and the autoregressive integrated moving average (ARIMA) model (Box & Jenkins, 1994). These are linear models which are, more than often, inadequate for stock market forecasting, since stock time series are inherently noisy and non-stationary. Recently, nonlinear approaches have been proposed, such as autoregressive conditional heteroskedasticity (ARCH) (Engle, 1982), generalized autoregressive conditional heteroskedasticity (GARCH) (Bollerslev, 1986), artificial neural networks (ANN) (Hansen & Nelson, 1997; Kim & Han, 2008; Kwon & Moon, 2007; Qi & Zhang, 2008; Zhang & Zhou, 2004), fuzzy neural networks (FNN) (Chang & Liu, 2008; Oh, Pedrycz, & Park, 2006; Zarandi, Rezaee, Turksen, & Neshat, 2009), and support vector regression (SVR) (Cao & Tay, 2001, 2003; Fernando, Julio, & Javier, 2003; Gestel et al., 2001; Pai &

Lin, 2005; Tay & Cao, 2001; Valeriy & Supriya, 2006; Yang, Chan, & King, 2002).

ANN has been widely used for modeling stock market time series due to its universal approximation property (Kecman, 2001). Previous researchers indicated that ANN, which implements the empirical risk minimization principle, outperforms traditional statistical models (Hansen & Nelson, 1997). However, ANN suffers from local minimum traps and difficulty in determining the hidden layer size and learning rate. On the contrary, SVR, proposed by Vapnik and his co-workers, has a global optimum and exhibits better prediction accuracy due to its implementation of the structural risk minimization principle which considers both the training error and the capacity of the regression model (Cristianini & Shawe-Taylor, 2000; Vapnik, 1995). However, the practitioner has to determine in advance the type of kernel function and the associated kernel hyperparameters for SVR. Unsuitably chosen kernel functions or hyperparameter settings may lead to significantly poor performance (Chapelle, Vapnik, Bousquet, & Mukherjee, 2002; Duan, Keerthi, & Poo, 2003; Kwok, 2000). Most researchers use trial-and-error to choose proper values for the hyperparameters, which obviously takes a lot of efforts. In addition, using a single kernel may not be sufficient to solve a complex problem satisfactorily, especially for stock market forecasting problems. Several researchers have adopted multiple-kernels to deal with these problems (Bach, Lanckriet, & Jordan, 2004; Bennett, Momma, & Embrechts, 2002; Crammer, Keshet, & Singer, 2003; Gönen et al., 2008; Lanckriet, Cristianini, Bartlett, Ghaoui, & Jordan, 2004; Ong, Smola, & Williamson, 2005; Rakotomamonjy, Bach, Canu, &

[☆] This work was supported by the National Science Council under the grant NSC 95-2221-E-110-055-MY2.

^{*} Corresponding author.

E-mail address: leesj@mail.ee.nsysu.edu.tw (S.-J. Lee).

Grandvalet, 2007, 2008; Sonnenburg, Ratsch, Schäfer, & Schölkopf, 2006; Szafranski, Grandvalet, & Rakotomamonjy, 2008; Tsang & Kwok, 2006; Wang, Chen, & Sun, 2008).

The simplest way to combine multiple-kernels is by averaging them. But each kernel having the same weight may not be appropriate for the decision process, and therefore the main issue concerning multiple-kernel combination is to determine optimal weights for participating kernels. Lanckriet et al. (2004) used a linear combination of matrices to combine multiple-kernels. They transformed the optimization problem into a semidefinite programming (SDP) problem, which, being convex, has a global optimum. However, the amount of time and space required by this method is demanding. Other multiple-kernel learning algorithms include Bach et al. (2004), Sonnenburg et al. (2006), Rakotomamonjy et al. (2007), Rakotomamonjy, Bach, Canu, and Grandvalet (2008), Szafranski et al. (2008) and Gönen et al. (2008). These approaches deal with large-scale problems by iteratively using the sequential minimal optimization (SMO) algorithm (Platt, 1999) to update Lagrange multipliers and kernel weights in turn, i.e., Lagrange multipliers are updated with fixed kernel weights and kernel weights are updated with fixed Lagrange multipliers alternatively. Although these methods are faster than SDP, they are likely to suffer from local minimum traps. Multiple-kernel learning based on hyperkernels has also been studied (Ong et al., 2005; Tsang & Kwok, 2006). Tsang and Kwok (2006) reformulated the problem as a second-order cone programming (SOCP) form. Crammer et al. (2003) and Bennett et al. (2002) used boosting methods to combine heterogeneous kernel matrices.

We propose a regression model, which integrates multiple-kernel learning and SVR, to deal with the stock price forecasting problem. A two-stage multiple-kernel learning algorithm is developed to optimally combine multiple-kernel matrices for SVR. This learning algorithm applies SMO (Platt, 1999) and the gradient projection method (Bertsekas, 1999) iteratively to obtain Lagrange multipliers and optimal kernel weights. By this algorithm, advantages from different hyperparameter settings can be combined and overall system performance can be improved. Besides, the user need not specify the hyperparameter settings in advance, and trial-and-error for determining appropriate hyperparameter settings can then be avoided. Experimental results, obtained by running on datasets taken from Taiwan Capitalization Weighted Stock Index (TAIEX), which is a stock market index for companies traded on the Taiwan Stock Exchange, show that our method performs better than other methods.

The rest of this paper is organized as follows. Section 2 presents basic concepts about support vector regression. Section 3 describes our proposed multiple-kernel support vector regression approach for stock price forecasting. Experimental results are presented in Section 4. Finally, a conclusion is given in Section 5.

2. Support vector regression (SVR)

In a regression problem, we are given a set of training patterns $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)$, where $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1, \dots, l$, and $y_i \in \mathbb{R}$. Each y_i is the desired target, or output, value for the input vector \mathbf{x}_i . A regression model is learned from these patterns and used to predict the target values of unseen input vectors. SVR is a nonlinear kernel-based regression method which tries to locate a regression hyperplane with small risk in high-dimensional feature space. It possesses good function approximation and generalization capabilities.

Among the various types of support vector regression, the most commonly used is ε -SVR which finds a regression hyperplane with an ε -insensitive band (Cristianini & Shawe-Taylor, 2000; Vapnik, 1995). To make the method more robust, the image of the input data does not need to lie strictly on or inside the ε -insensitive band.

Instead, the images which lie outside the ε -insensitive band are penalized and slack variables are introduced to account for these images. For convenience, in the sequel, the term SVR is used to stand for ε -SVR. The objective function and constraints for SVR are

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle + C \sum_{i=1}^l (\xi_i + \hat{\xi}_i), \\ \text{s.t.} \quad & (\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b) - y_i \leq \varepsilon + \xi_i, \\ & y_i - (\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b) \leq \varepsilon + \hat{\xi}_i, \\ & \xi_i, \hat{\xi}_i \geq 0, \quad i = 1, \dots, l, \end{aligned} \quad (1)$$

where l is the number of training patterns, C is a parameter which gives a tradeoff between model complexity and training error, ξ_i and $\hat{\xi}_i$ are slack variables for exceeding the target value by more than ε and for being below the target value by more than ε , respectively. Note that $\phi : X \rightarrow F$ is a possibly nonlinear mapping function from the input space to a feature space F . Also, $\langle \cdot, \cdot \rangle$ indicates the inner product of the involved arguments. The regression hyperplane to be derived is

$$f(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle + b, \quad (2)$$

where \mathbf{w} and b are weight vector and offset, respectively.

To solve Eq. (1), one can introduce the Lagrangian, take partial derivatives with respect to the primal variables and set the resulting derivatives to zero, and turn the Lagrangian into the following Wolfe dual form

$$\begin{aligned} \max_{\alpha, \hat{\alpha}} \quad & \sum_{i=1}^l y_i (\hat{\alpha}_i - \alpha_i) - \varepsilon \sum_{i=1}^l (\hat{\alpha}_i + \alpha_i) \\ & - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\hat{\alpha}_i - \alpha_i) (\hat{\alpha}_j - \alpha_j) K(\mathbf{x}_i, \mathbf{x}_j), \\ \text{s.t.} \quad & \sum_{i=1}^l (\hat{\alpha}_i - \alpha_i) = 0, \\ & C \geq \alpha_i, \hat{\alpha}_i \geq 0, \quad i = 1, \dots, l, \end{aligned} \quad (3)$$

where α_i and $\hat{\alpha}_i$, $i = 1, \dots, l$, are Lagrange multipliers, and $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_l]$ and $\hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_l]$. Note that $K(\mathbf{x}_i, \mathbf{x}_j)$ is a kernel function which represents the inner product $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$. The most widely adopted kernel function is the radial basis function (RBF) which is defined as

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}_j) &= \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle, \\ &= \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \end{aligned} \quad (4)$$

where γ is the width parameter of the RBF kernel. Now, Eq. (3) can be solved by SMO (Platt, 1999). Suppose α_i^* and $\hat{\alpha}_i^*$, $i = 1, \dots, l$, are the optimal values obtained. The regression hyperplane for the underlying regression problem is then given by:

$$f(\mathbf{x}) = \sum_{i=1}^l (\hat{\alpha}_i^* - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b^*, \quad (5)$$

where $b^* = y_k + \varepsilon - \sum_{i=1}^l (\hat{\alpha}_i^* - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}_k)$ is obtained from any α_k^* with $0 < \alpha_k^* < C$.

3. Proposed method

In this section, the idea of multiple-kernel support vector regression is formulated. Then a two-stage multiple-kernel learning algorithm for deriving optimal kernel weights and Lagrange multipliers is described.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات