Expert Systems with Applications 39 (2012) 2256–2262

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Grey relational grade in local support vector regression for financial time series prediction

Hui Jiang [a,b,*], Wenwu He [c]

[a] School of Statistics, Renmin University of China, Beijing 100872, China
[b] Department of Mathematical Sciences, Huizhou University Huizhou, Guangdong 516007, China
[c] Department of Mathematics and Physics, Fujian University of Technology, Fuzhou, Fujian 350108, China

ARTICLE INFO

Keywords:
Financial time series prediction
Grey relational grade
Local support vector regression
Cross validation error

ABSTRACT

Support vector regression (SVR) has often been applied in the prediction of financial time series with many characteristics. On account of much time consumption of global SVR, local machines are carried out to accelerate the computation. In this paper, we introduce local grey SVR (LG-SVR) integrated grey relational grade with local SVR for financial time series forecasting. Pattern search method and leave-one-out errors are adopted for model selection. Experimental results of three real financial time series prediction demonstrate that LG-SVR can speed up computing speed and improve prediction accuracy.

1. Introduction

On the grounds that some financial time series contain several specific characteristics: large sample sizes, high noise, non-stationary, non-linearity, and varying associated risk, we are inclined to find some comparatively precise methods to forecast their future values. Namely, given financial time series data set $\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, we usually use a learned model to make accurate predictions of y for future values of x. Recently, due to both the better performance of generalization and the good empirical results (Cristianini & Shawe-Taylor, 2000; Evgeniou, Pontil, & Poggio, 2000; Smola & Schölkopf, 2004; Vapnik, 1998), kernel machines, including support vector regression machines, have been introduced for financial time series forecasting (Kim, 2003; Tay & Cao, 2002; Yang, King, & Chan, 2002). However, the costly computation of global machines becomes an obstacle to apply them to some problems. Despite consideration of model selection, the total computation cost is still large because of repeated SVR algorithm.

In order to solve these problems, some experts introduced lazy learning methods which defer the processing of training data until a query needs to be answered (Kobayashi, Konishi, & Ishigaki, 2007). In addition, experts also proposed some modified local support vector regression models and optimized methods. He and Wang (2007) introduced local kernel machines in which models optimized based on leave-one-out errors, and multiple-kernels are adopted for model selecting and performance improvement. Huang et al. proposed a novel local support vector regression model demonstrated to provide a systematic and automatic scheme to adopt the margin locally and flexibly. This method not only captures the local information in data, but establishes connection with several models (Huang, Yang, King, & Lyu, 2006). Inspired by these literatures, based on the popular grey system theory (Deng, 1982), we introduce a local support vector regression combined with grey relational analysis (GRA).

GRA is an important component of grey system theory. In the uncertain and incomplete system, this method is used to find out not only the key factors affecting a selected object, but also the basic relationship between an influential factor and the selected object. This method compares variation trends of influential factors with that of a selected object. A more similar trend means a closer relationship. Meanwhile, the similarity can be measured in grey relational grade. The greater the grade, the closer the similarity. In addition, GRA does not require strict compliance with certain statistical laws or linear relationships among objects, and is thus widely applied in artificial intelligence (He & Hwang, 2007), hydrology (Wong, Hu, IP, & Xia, 2006), laser technology (Cayda & Hascalik, 2008), material science (Chan & Tong, 2007), mathematics (Xu, Tian, Qian, & Zhang, 2007) and other fields (Liu, Wang, Zhang, & Li, 2009; Moran, Granada, Miguez, & Porteiro, 2006).

In practice, to some financial time series in SVR, training data are seldom distributed evenly in input space regarded as an uncertain and incomplete system because they are affected by many unknown factors. If we only consider the part of training examples close to the test point, the GRA can be adopted here based on that it is a kind of measurement approach that describes the relationship between a test point and the other in an uncertain and incomplete grey space (Deng, 1989). With its great significance in the GRA, the grey relational grade between two points is a measurement of their

* Corresponding author at: Department of Mathematical Sciences, Huizhou University Huizhou, Guangdong 516007, China. Tel.: +86 13433510476.
E-mail address: jianghui200207@163.com (H. Jiang).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.07.100

relationship in a certain data set. Where financial time series data are concerned, we learn the model by choosing a certain fixed number of the nearest former points of the query as the training data, and query point has been matched with every candidate to calculate grey relational grade as the weight of punishment parameter C in SVR. Because the nearest points capture the main information of the query, the accuracy can be improved. In the meantime, the size of training data or computation cost can be changed according to the fixed number. The smaller the number, the faster the computation. In addition, leave-one-out errors and pattern search method can be adopted for model selection. Experimental results demonstrate the good performance of our methods.

The rest of this paper is organized as follows. Local grey support vector regression (LG-SVR) including standard SVR, grey relational grade weighting function (GRF), local support vector regression integrated with GRF and LD-SVR have been provided in Section 2, model optimization methods in Section 3, the experimental results of three real financial time series forecasting in Section 4, conclusion in the last section.

2. Local grey support vector regression (LG-SVR)

2.1. Standard support vector regression (SVR)

Initially developed for solving classification problems (Burges, 1998), support vector machines (SVMs) technology can also be successfully applied in regression problems, especially in financial time series. Let the training data set D be $\{(x_i, y_i)\}_{i=1}^n$, where input $x_i \in X \subseteq R^m$ and output $y_i \in Y \subseteq R$. Our goal is to map the input into a higher dimensional feature space via a nonlinear mapping and find a function f that has an ε-*deviation* from the actually obtained target y_i for all training data at the same time is as flat as possible. Regression can be regarded as the question of how Y could be explained by X, i.e. how to approximate a multivariate function from the observed data. This means how to determine the relationship of the form

$$y_i = f(x_i) + \varepsilon_i, \tag{1}$$

where $f(x_i)$ is the function needs to be estimated and ε_i is a random variable. To estimate the function from the observed data is an ill-posed problem and the effective way to solve it is regularization learning theory (András, 2002; Evgeniou et al., 2000). Based on it, regression problem can be formulated as follows

$$\min_{f \subseteq \mathcal{H}} R[f] = \frac{1}{n}\sum_{i=1}^{n} \Gamma(y_i, f(x_i)) + \lambda \|f\|_\kappa^2, \tag{2}$$

where $\Gamma(y_i,f(x_i))$ is a loss function, $\|f\|_\kappa^2$ is a norm in a Reproducing Kernel Hilbert Space $\mathcal{H}$ defined by the positive define function κ: $X \times X \to R$, n is the size of training data set and λ is the regularization parameter. In the $\varepsilon - SVR$, $\Gamma(\cdot)$ takes following form called ε-*insensitive* loss function

$$\Gamma(x,y) = |x-y|_\varepsilon = \begin{cases} 0, & if |x-y| \leqslant \varepsilon, \\ |x-y| - \varepsilon, & otherwise. \end{cases} \tag{3}$$

Replacing λ with $C(C = \frac{1}{2n\lambda})$, we can re-represent problem (2) as follows

$$\min_{f \subseteq \mathcal{H}} R[f] = C\sum_{i=1}^{n} |y_i - f(x_i)|_\varepsilon + \frac{1}{2}\|f\|_\kappa^2. \tag{4}$$

Considering the soft margin formulation, we have to solve the following problem with slack variables (Chan & Tong, 2007).

$$\min_{\omega,\xi,\xi^*} R[\omega,\xi,\xi^*] = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{5}$$

Subject to

$$\begin{cases} y_i - (\omega \cdot \varphi(x_i) + b) \leqslant \varepsilon + \xi_i, \\ (\omega \cdot \varphi(x_i) + b) - y_i \leqslant \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geqslant 0, \quad i = 1, 2, \ldots, n, \end{cases} \tag{6}$$

where $f(x_i) = \omega \cdot \varphi(x_i) + b$, $\varphi(x)$ is a map function, ξ, ξ^* are the corresponding positive and negative errors at the ith point respectively.

2.2. Grey relational grade weighting function (GRF)

Grey system theory has been applied in a number of fields such as control, pattern recognition and so on (Liu, Dang, & Fang, 2004). As a part of this theory, the grey relational analysis uses information from the grey system to dynamically compare each factor quantitatively. This approach is based on the level of similarity and variability among all patterns to establish their relation. The grey relational analysis is a method to analyze the relational grade for discrete sequences. The grey relational grade, which is a kind of relationship measurement between two patterns in a grey environment, is proposed here.

Let x_0 be reference sequence $x_0 = (x_0(1), x_0(2), \ldots, x_0(m))$, and x_i the compared sequences $x_i = (x_i(1), x_i(2), \ldots, x_i(m))$, $(i = 1, 2, \ldots, k)$. $x_i \in G$, G is an uncertain and incomplete space including k data patterns. The grey relational coefficient between the compared sequence x_i and the reference sequence x_0 for the jth ($j = 1, 2, \ldots, m$) entry is defined as follows

$$\gamma(x_0(j), x_i(j)) = \frac{\min\limits_i \min\limits_j |x_0(j) - x_i(j)| + \delta \max\limits_i \max\limits_j |x_0(j) - x_i(j)|}{|x_0(j) - x_i(j)| + \delta \max\limits_i \max\limits_j |x_0(j) - x_i(j)|}, \tag{7}$$

where $\delta \in [0,1]$ is the distinguishing coefficient used to control the resolution scale, usually taken as 0.5 (Deng, 2002). Then, the grey relational grade between sequence x_0 and x_i can be specified by the coefficient.

$$\gamma(x_0, x_i) = \frac{1}{m}\sum_{j=1}^{m} \gamma(x_0(j), x_i(j)). \tag{8}$$

Here, $\gamma(x_0,x_i)$ denotes the extent to which the compared sequence x_i can affect the reference sequence x_0. Restated, the reference sequence can glean some useful information concerning the variation of data points from other similar sequences. According to the grey relational system theory, the highest relational grade is corresponding to the most important pattern; the larger the grey relational grade $\gamma(x_0,x_i)$ is, the higher the relative importance of x_i to x_0 is.

2.3. Local support vector regression integrated with GRF (LG-SVR)

Different to global algorithm, only part of the training examples (set k) close to the test point are taken into consideration in local learning. In financial time series forecasting, the *k-nearest neighbor* (KNN) method is carried out (Gao et al., 2007). Namely only the nearest k past data points of every test point are considered. In order to realize the locality, we use weighting functions with grey relational grade. Meanwhile, we consider the objective of LG-SVR model is to learn the approximating function in D by making the function locally as involatile as possible while keeping the error as small as possible. Denoting the input of the test point as x_0, we formulate this objective as follows

$$\min_{\omega,\xi,\xi^*} R[\omega,\xi,\xi^*,x_0] = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{k} \gamma(x_0,x_i)\left(\xi_i + \xi_i^*\right) \tag{9}$$