Information System Establishment and Prediction Analysis for SVM Model of RBF Kernel Function

Siyun Wu*
University of Wisconsin Madison, College of Letters and Science Madison, WI, US

*Corresponding author: swu354@wisc.edu

Abstract. The effective prediction of Shanghai Composite Index can provide much useful information for relative organizations to analyze the changes of stock market and then make right decision. Therefore, the prediction of the Shanghai Composite Index is meaningful. In this study, supported vector machine (SVM) algorithm is employed to construct the prediction model, and MSE R2 are used for error test. The result shows SVM model has high accuracy, which indicates the prediction model established in this study is reliable.

Keywords: Shanghai Composite Index; SVM; error test.

1. Introduction
Since the birth of the stock and the futures markets, people have been looking for a scientific method to effectively predict the stock index or term index, which will help to achieve the purpose of maximizing investment profits, such as avoiding risks by the way of predicting risks [1]. From the perspective of theoretical research, accurate prediction theory is not only a guide for economic decision-making, but also a necessary analysis link for making effective decisions, and more scientific development of capital market is the ultimate goal of analysis and prediction [2]. Therefore, the research on the fitting, simulation and prediction of stock index or term index has a far-reaching impact on the development of investors or disciplines, as well as economic prosperity and social progress. The international capital market operation practice shows that the capital market (i.e., the stock market as a special case) in the daily pay most of time series presents the normality and thick tail characteristics, and have volatility clustering and persistence (i.e. if the current market is volatile, the next issue of fluctuation will be big, and it will vary with the degree of deviation from the mean current yields more or less. Conversely, if the volatility of one period is low, the volatility of the next period will be low, unless the current yield is significantly off the average.)

Based on the properties of stock markets, SVM algorithm can be an effective way to predict the changes of stock index, which is widely used in prediction and classification problem. Therefore, it’s applied herein to predict the changes of Shanghai Composite Index.

2. Methodology of Supported Vector Machine
Support vector machine [3] is a commonly used classification model. The core idea is to carry out nonlinear transformation of input data through kernel function, map it to a high-dimensional feature
space to make these data linearly separable, and then find out the best hyperplane (as shown in Fig. 1) to achieve classification, in which the selection of kernel function is the key to SVM classification.

The SVM classifier construction process in this paper is shown as follows:

**Fig. 1 Hyperplane of SVM**

Step 1: Construction of the optimal hyperplane
Assume the data after preprocessing as \{(x_1, y_1), \ldots, (x_N, y_N)\}, where \(x_i\) is the input and \(y_i\) is the output. Then the hyperplane can be constructed by Equation (1)

\[
y_i(w^T x_i + b) \geq 1 - \xi_i
\]

Where \(\xi_i\) denotes slack variables.

The goal of the classifier is to find an optimal hyperplane with the lowest classification error rate [4], and the following optimization problems can be constructed as Equation (2):

\[
\Phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i
\]

Where \(C\) is the penalty coefficient.

**Step 2: Selection of kernel function**
The selection of kernel function has a great impact on the performance of SVM classifier. Considering that the data complexity in this paper is moderate, radial basis function (RBF kernel function) is selected in this paper. Compared with other kernel functions, it has fewer parameters, lower complexity and less numerical computation, and it can be expressed as Equation(3):

\[
K(x, x_i) = \exp \left[ -\frac{|x - x_i|^2}{2\sigma^2} \right]
\]

Where \(\sigma^2\) denotes the effective range of RBF.

**Step 3: Model Training and test**
Input the training data and test data, adjust the parameters in SVM, and then obtain the final model.

3. Application of SVM on Shanghai Composite Index prediction

3.1. Data processing
The data of Shanghai Composite Index from December, 1990 to August, 2009 is used in this paper. Assume that Assume that the opening index of Shanghai Composite Index is correlative to the opening
index of yesterday, the highest index, the lowest index is the lowest, the closed index, the trade index and trading volume. And the six indicators are determined to be the input variables while the opening index is the output variables. And 80% of data is used for training data while 20% is used for test data. The first step is to normalize the data, which can guarantee the accuracy of prediction model and Z-scores method [5] is employed herein, which can be expressed as Equation (4):

\[ z = \frac{x - \mu}{\sigma} \tag{4} \]

Where \( x \) is the original data; 
\( \mu \) is the mean value of the original data; 
\( \sigma \) is the variance of the original data.

After normalization, the data is shown in Fig.2. It can be seen that the tendency of data has no changes while narrow the range to [0,2].

3.2. Model Solution via MATLAB
The process of SVM modeling for prediction can be shown in Fig.3. Then the SVM model is established via MATLAB 2018a software, and after 541 iterations, the best SVM model was obtained. The optimal parameters for SVM and corresponding error is shown in Fig.4.
3.3. Error Test

The error analysis is carried out to test the accuracy of SVM prediction model. The mean square error (MSE) and $R^2$ are employed herein to evaluate the performance of SVM model in accuracy [6]. They can be calculated by Equation (5) and (6):

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

(5)

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]

(6)

Where $y_i$ is the real value;

$\hat{y}_i$ is the predicted value;

$n$ is the number of samples.

The results are shown in Table 1, Fig.5 shows the error changes with time series and Fig.6 compares the predicted and the true value. It is obvious that the prediction model fits well with the true data, which indicates the prediction model has high accuracy and is reliable for practical application.

| Measured | $MSE$ | $R^2$ |
|----------|-------|-------|
| Value    | $2.35705 \times 10^{-5}$ | 0.9992 |
Fig. 5 Error curves

Fig.6: Comparison of the predicted and the true value

Conclusion
To evaluate the tendency of Shanghai Composite Index by historical data, SVM model was established in this study and the results of error test shows $R^2$ is 0.9992 and MSE is $2.35705 \times 10^{-5}$. Therefore, the prediction model has proved to be effective and practical.

References
[1] Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. An innovative neural network approach for stock market prediction. The Journal of Supercomputing (2018), 1-21.
[2] Shenghan, Zhou, et al. "A Novel Bearing Multi-Fault Diagnosis Approach Based on Weighted Permutation Entropy and an Improved SVM Ensemble Classifier." Sensors 18.6(2018):1934.
[3] Chauhan, Vinod Kumar, K. Dahiya, and A. Sharma. Problem formulations and solvers in linear SVM: a review. Artificial Intelligence Review (2018)
[4] Lanbouri, Zineb, and S. Achchab. Stock Market prediction on High frequency data using Long-Short Term Memory. Procedia Computer ence 175(2020):603-608.
[5] Schnitzer, Jason K., et al. Data Normalization. (2003).
[6] Islam, Mohammad Rabiul, et al. Neural Network and Principle Component Analysis Based
Numerical Data Analysis for Stock Market Prediction with Machine Learning Techniques. Journal of Computational and Theoretical Nanoscience (2019)