The SVR Parameters Optimization For Stock'S Closing Price Forecast

Tian Ye[*]

* Wuhan University of Technology, Wuhan, China
+86-13508628699
Tianmaomao701@163.com

Abstract

Stock's closing price determines the price of the stock. So getting good prediction effect of stock's closing price is helpful to master the direction of the stock in some degree. In order to get better forecast effect in the stock's closing price to make use of SVR training and regression, fully considering the influence of parameters on the result of the experiment, mainly adopts the grid search parameters optimization, genetic algorithm and particle swarm optimization algorithm, three kinds of optimization algorithm of SVR parameters optimization. And 2275 rows historical data from April 9, 2007 to 2007 on August 10 are used for the simulation experiment. The experimental results show that the use of different optimization algorithms has much effect on the experimental results, and the regression results used the optimization algorithm are more accurate..

1 Introduction

As the rapid development of China's economy and financial industry, under the premise of global economic integration, the price of stock is affected not only by the domestic economic situation, but also by the economic situation of the world. Although the development of stock market has gone through less than 30 years, more and more people have turned their attention to the stock market. Therefore, it is very important to forecast the stock price and trend to the development of contemporary economic investment.

2 Background and Significance of Stock Forecasting

The traditional stock forecasting methods are portfolio investment analysis, time series forecasting, non-linear prediction, combination forecasting and so on [1]. Due to the relatively large number of factors affecting the stock price, the measurement index is also a little complicated, and the limitation of these traditional methods makes them difficult to meet the demand of forecasting precision and speed. The main problems are：

a. The non-linearity of stock market leads to the difficulty of establishing unified model.

b.The price of stock is influenced by many complicated factors, and there are many uncontrollable factors.

c.The noise in the stock price data is large, and the traditional prediction analysis method is not perfect enough to deal with the noise data.

At present, the technology and research of extracting information from massive historical data are developing rapidly. The characteristics of non-linearity, noise, influence and restriction of stock price attract many experts and scholars to study it.

Compared with the traditional learning methods, the modern machine learning algorithms have a stronger ability of data mining, and they are good at extracting effective information from historical data and have significant advantages in forecasting results.

3 Support Vector Machines

3.1 Support Vector Regression

Support Vector Machines (SVM) is a learning method based on VC dimension and structural risk minimization [2]. It solves a convex quadratic programming problem by mapping the feature dimension to the high-dimensional feature space and constructing the hyperplane as the decision function of the original space in the high-dimensional space [3]. The idea of locating the solution space in the high-dimensional space effectively solves the problem that the decision function is nonlinear and has good generalization ability, and the complexity has nothing to do with the number of sample dimensions, only with the number of support vectors.

SVM itself is proposed for the classic binary classification problem. Support Vector Regression (SVR) is the application of SVM in regression.

For a given training data set, $D = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, $y_i \in R$, x is mapped to the high dimensional feature by the nonlinear mapping $\phi(x)$, and establish a regression function:

$$f(x, \omega) = \omega \cdot \phi(x) + b \qquad (1)$$

Where, ω is the normal vector, determine the direction of the hyperplane; b is the displacement, determine the distance between the hyperplane and the origin.

In contrast to support vector machine classification, the optimal hyperplane sought by SVR is not to separate the samples from different classes as much as possible, but to make the sample points within the allowable deviation range ε, that is, calculating the loss when the deviation is larger than ε [4].

3.2 Kernel Function Parameter

The performance of SVM depends on the selection of kernel functions and parameters [5]. However, there is no uniform selection criterion in the industry. Because for different areas and different types of problems, different parameters of the kernel function will have its advantages and disadvantages. So it is not possible to determine which is optimal. Commonly used kernel functions are:

a. linear kernel function:

$$k(x, y) = x^T y + c \qquad (2)$$

b. polynomial kernel function:

$$k(x, y) = (ax^T y + c)^d \qquad (3)$$

c. radial basis kernel function:

$$k(x, y) = \exp(-\gamma\|x - y\|^2) \qquad (4)$$

d. sigmoid kernel function:

$$k(x, y) = \exp(-\frac{\|x - y\|^2}{2\sigma^2}) \qquad (5)$$

Different kernel functions have their own characteristics. These different features determine the difference of the effect of the kernel function in different fields. Among them, RBF radial basis function's characteristics that not appearing too much deviation, good tracking performance, no memory, suitable for static modeling determine the problem for the stock closing price forecast.

And for the selection of parameters, generally only through a variety of optimization algorithms or constantly adjust parameters and experiments, to get the optimal parameters of specific problems.

In the case of using the RBF kernel function, the parameters of regression estimation of the stock closing price using SVR are: c and g. C is the penalty coefficient in SVR. If c is too small, prone to over-fitting; and if c is too large, the results of fitting and actual results are too different; G is the coefficient of RBF, the value of g directly affects the ability to find the optimal hyperplane. We can see that the choice of c and g on the experimental results have a great impact, so optimization of c and g is necessary [6].

4　Parameter optimization method

With the deepening of intelligent algorithm research, more and more optimization algorithms have emerged, and many have been widely used in various fields. At present the most common optimization algorithms are grid search, genetic algorithm and particle swarm optimization.

4.1 Grid search method

The mesh method is a kind of optimization method based on blind path finding [7]. The basic idea is to take a set of points regularly in the range of the variables, select the point with the smallest objective function value as the best point, and then separate the smaller grid in a certain range near the optimal point. Find the optimal point, and iterate until it meets the requirement of precision, then the optimal point will be taken as the optimal result.

4.2 Genetic Algorithm (GA)

Genetic algorithm (GA), also known as evolutionary algorithm, is inspired by Darwinian evolution theory and is a heuristic search algorithm based on biological evolutionary process [8]. Genetic algorithm will be the problem of analogy to a biological evolution of the process, through replication, crossover, mutation and a series of processes to produce the next generation, the problem corresponds to a series of solutions. And then through the fitness function, out of the lower part of the solution function, the solution retains the high value of the solution, such a process as an evolution. After evolution n, it is highly probable that individuals with high fitness function values can be evolved, that is, the optimal solution for the problem.
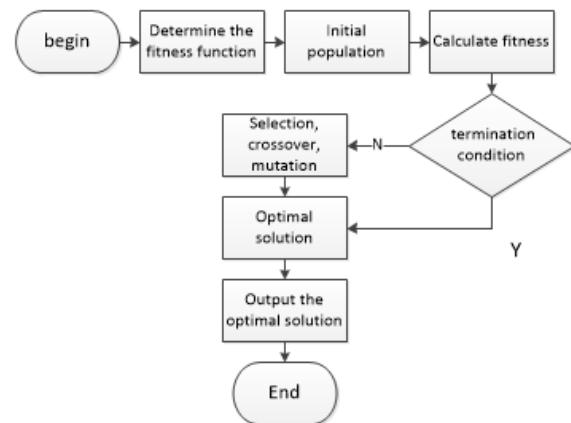

Figure 1. The process of GA

This algorithm has the characteristics of heuristic search and group search. It can effectively solve complex problems quickly and is not easy to fall into the local optimum. Its ability to parallel computing also greatly guarantees the probability of finding the global optimal solution. At the same time, the scalability of genetic algorithm is good, which provides a good prerequisite for its wide application.

4.3 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is an optimization algorithm proposed by Eberhart and Kennedy for predicting bird behavior [9]. The algorithm compares all possible solutions to a particle in the search space. Each particle has its own velocity. The velocity determines the direction and distance of the motion of the particle. At the same time, each particle has a fitness function value. The optimization process evolves into a particle. Following the current fitness function, the optimal particle motion is searched in the solution space.
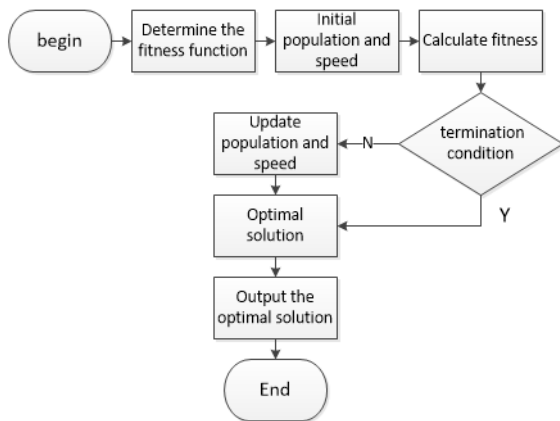
Figure 2. The process of PSO

Particle Swarm Optimization (PSO), like genetic algorithm, is a kind of evolutionary algorithm, which finds the optimal solution from the iterative iteration of random solution, but also takes the fitness function as the good or bad judgment of a solution. But compared with the genetic algorithm, particle swarm optimization algorithm does not need to debug a lot of parameters and convergence faster.

5 Experiment

Closing price as the last transaction price of the day defines the ups and downs of the stock. This section will use the Shanghai Composite Index from April 9, 2007 to August 10, 2016 of the 2275 historical data to experiment. Using the opening price, the highest index, the lowest index, the closing price, the volume of the day and day trading volume as the dependent variable to do regression forecast for daily closing prices. Using grid search, genetic algorithm and particle swarm algorithm to find the optimal combination of c and g, which makes the prediction accuracy of the closing price the highest.

5.1 Experimental Tools

LIBSVM Toolkit is a software package developed by Prof. Chih-Jen Lin's of Taiwan, whose main purpose is to make SVM's pattern recognition and regression easier. It not only provides the implementation of the document, but also open source code for us to learn and improve. LIBSVM package is small, flexible, small number of parameters, open source and easy to expand, and these advantages make most people use it as the best choice when using SVM.

5.2 Evaluation index

In this paper, the mean-square error, MSE, is used to be the evaluation index:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$$

(6)

The smaller the MSE, the higher the accuracy.

5.3 Experimental Results and Analysis

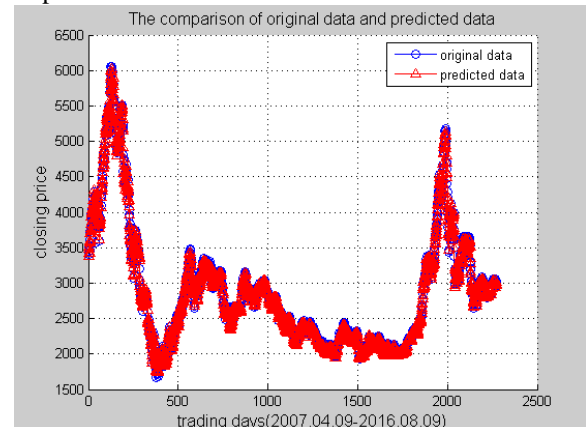The experimental results are summarized in Table 1：

Figure 3. The comparison of original data and predicted data based on grid search method
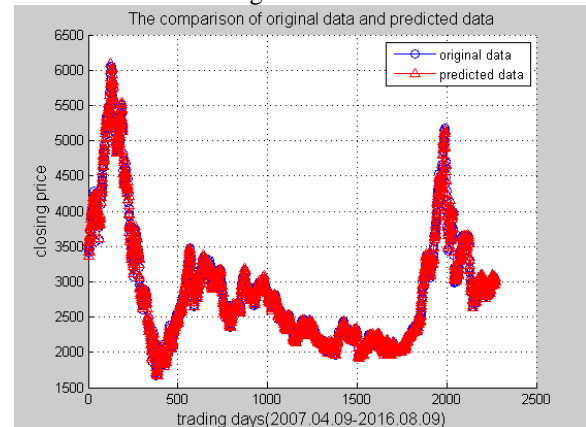

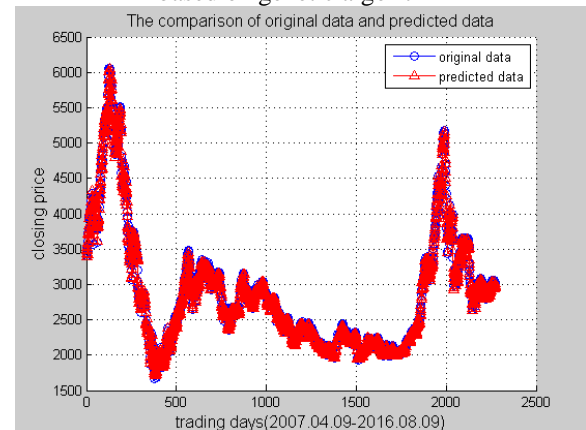Figure 4. The comparison of original data and predicted data based on genetic algorithm

Figure 5. The comparison of original data and predicted data based on PSO

Table 1 Summary of experimental results

optimization	Grid Search	GA	PSO
best c	0.176777	14.192	2.55698
best g	2	69.6662	0.01
Time/s	39.736633	559.210169	505.467248
MSE	0.000197408	0.00008082	0.000217061
Correlation coefficient R	99.47%	99.78%	99.42%

From the data and charts can be seen：

（1）The difference between the optimal solutions c and g obtained by different optimization algorithms is quite different. The source of this difference may be the limitations of the algorithm itself.

（2）We know that the MSE is smaller, the higher the accuracy rate. As can be seen from Table 1, the genetic algorithm with respect to grid search and PSO MSE minimum.

（3）Compared with particle swarm optimization and particle swarm optimization algorithm, it can be seen that grid search is superior to particle swarm optimization in terms of accuracy and time..

（4）Although the genetic algorithm has the highest accuracy, its time is the longest. This experiment uses 2275 data from 2007 to 2016, the genetic algorithm takes nearly nine minutes, and the grid search takes less than one minute.

6 Summary

It can be seen from the above experimental results that, for the problem of stock closing price forecast in this paper, the effect of the three optimization algorithms of grid search, genetic algorithm and particle swarm optimization is very different.

In general, the accuracy of PSO is the lowest and the time is not too short, which indicates that the PSO is not ideal for SVR stock price optimization problem.

Grid search method costs the least time, but its accuracy rate is not very good. So grid search method is not the best choice. From the accuracy rate can be seen, the genetic algorithm is most suitable for SVR parameters optimization. However, the genetic algorithm also cost the longest time. So how to ensure accuracy under the premise of improving efficiency will also be the next question to think about

References

[1] Wang Sha. Application of BP Neural Network in Stock Forecasting Research [D], Central South University. 2008.

[2] Ding Shi-fei. Qi Bing-juan. and Tan Hong-yan. An Overview on Theory and Algorithm of Support Vector Machines [J]. Journal of University of Electronic Science and Technology of China. 2011, 40(1):2-10.

[3] Song Zhaoqing. Cui He. Hu Yunan. Research and development of support vector machine theory[J]. Journal of Naval Aeronautical and Aeronautical University. 2008, 02:143-148+15.

[4] Tian Ying-jie. Support vector regression and its application [D]. College of Economics & Management, China Agricultural University. 2005.

[5] Guo Lijuan. Sun Shiyu. Duan Xiusheng. Research for Support Vector Machine and Kernel Function [J]. Sxience Technology and Engineering. 2008, 8(2):487-490.

[6] Lin Shengliang. Liu Zhi. Parameter selection in SVM with RBF kernel function [J]. Journal of Zhejiang University of Technology. 2007, 35(2):163.

[7] Wang Jianfeng, Zhang Lei, Chen Guoxing, He Xuewen. A parameter optimization method for an SVM based on improved grid search algorithm [J]. Applied Science and Technology. 2012, 39(3) : 28-31.

[8] Liu Haiyue. Genetic algorithm and Neural Network in Stock Forecasting Analysis [D]. North University of China. 2011.

[9] Ai Yongguan. Particle Swarm Optimization Neural Network Modeling and Application in the Stock Market Forecast [D]. HeFei University of Technology. 2009