Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2022.030203
ISSN 2790-0983 Vol. 3, Issue 2: 24-31

Implementation of Price Prediction Model Based on Support Vector Machine Algorithm

Shulan Li*
Information Engineering, Jingdezhen University, Jingdezhen 334000, Jiangxi, China
lsljdz@126.com
**corresponding author*

Keywords: Support Vector Machine, Prediction Model, Price Prediction, Kernel Function

Abstract: In recent years, people decide the investment trading behavior by quantitative means, through the combination of different stocks and trading time to select the optimal strategy for trading, therefore, the analysis of stocks and how to accurately predict the price of stocks has become the key to make investment decisions. This paper mainly carries on the support vector machine (SVM) algorithm price forecast model research and implementation. In this paper, the SVM model is used to model the stock price changes and conduct the corresponding strategy analysis, so as to play its advantages in small samples and nonlinearity. From the experimental results, it can be seen that the mean square error of the prediction result of SVM is far less than that of the time series model. This shows that the performance of SVM in stock price prediction is very excellent.

1. Introduction

The securities market is an important part of the market economy, a concentrated embodiment of the modern enterprise system, and an important place to realize the effective allocation of resources. The issue of the price prediction and selection of securities is a hot topic in the current research of financial statistics technology [1-2]. At present, the stock market has occupied a unique position in the economy of developed countries. Because of the inadequacy of the management system of the securities market in China, the policy-based factors have a great influence on the securities market. Basically, every policy issue will cause some time fluctuation to the securities market. There's also the international impact. The global economy has been hit hard by COVID-19, and the U.S. stock market has triggered circuit breakers four times in one week, which has never happened in history. At the same time, China's A - share securities market also experienced sharp fluctuations. Under the influence of globalization economy, any changes in international environment will have certain influence on our securities market [3]. In order to further improve and supervise the market, market analysis and prediction is the basis of market management and development, because only by discovering the fluctuation rules of the market, can we establish a better management system, prevent the occurrence of systemic financial risks, and realize the effective allocation of resources. With the development of science and technology, researchers are more inclined to study the development direction of the security market with the help of artificial intelligence [4-5]. Therefore, this paper tries to study the prediction and selection of securities market with the help of artificial intelligence technology.

The prediction of the security market has always been the focus of financial statistics. How to improve the prediction accuracy of the security market has become the focus of researchers. Because successful stock price prediction can help managers make judgments in advance in the face of market fluctuations, and can also bring considerable profits to investors [6]. However, due to the unstable, jumping and chaotic data of the stock market, the prediction of its price has become a major challenge at present. Therefore, it is becoming more and more challenging to predict the profit of investors based on the investment plan [7]. Many statistical models have been developed to predict the trend of the stock market. Meanwhile, stock market prediction based on fuzzy time series data is based on time series [8]. Such methods are widely used to predict nonlinear and dynamic data sets in the changing field, so they are well adapted to the volatility of stock data [9]. With the development of technology, many intelligent technologies, namely soft computing algorithm, neural network, back propagation algorithm and genetic algorithm, are used to predict stock market returns [10]. The use of artificial intelligence technology to forecast the stock market can be roughly divided into two categories, namely, the technology used for trend prediction and the technology used for classification [11]. Techniques used for prediction include artificial neural network, convolutional neural network, recurrent neural network (RNN), decision support system, hidden Markov model Naive Bayes, Support vector regression (SVR), SVM and other techniques [12]. The techniques used for classification include filtering, blurring, K-means and optimization [13].

SVM algorithm was used to study the price of a security at the method innovation not only can further understand the method of artificial intelligence application in the field of finance, strengthen the mutual fusion of different disciplines indivisibility, more important is to optimize the stock price prediction model, provide the basis for investors to invest and make contributions to the healthy development of China's stock market.

2. Stock Price Prediction Based on SVM

2.1. Stock Price Impact Factors

The financial rules in the stock market are complex, and there are various factors affecting the stock price. The impact of various factors will eventually be reflected in the rise and fall of stock prices. The following is a brief introduction to the three main factors affecting stock price volatility.

(1) Policy factors. National policy adjustment, new policies or major events at home and abroad act on the stock market and profoundly affect the volatility of stock prices.

(2) Internal market factors. When a large number of investors are bullish on a stock and buy it in droves, it usually leads to an increase in the price of the stock. Stock prices, similar to commodity prices, are affected by the relationship between supply and demand in the market.

(3) Internal factors of the company, mainly referring to the operation status, development strategy and economic basis of the listed company. Only when the issuing company maintains good operating conditions and steady profitability can its stock price keep rising steadily. On the contrary, often cause stock turbulence.

2.2. Principle of SVM

Based on the principle of minimizing structural risk, SVM avoids overfitting and improves the generalization ability of the model. It performs well in solving machine learning problems such as classification and regression. As a new technology applied to pattern recognition in the field of data mining, the model has few parameters and good practical application effect, which has been widely used in pattern recognition, gene classification, time series prediction and other fields [14-15].

(1) Classification algorithm

How do you get the optimal hyperplane? It is necessary to move the hyperplane to both sides in parallel and stop when the first data point in the data sets on both sides is encountered. At this time, the straight line distance between the two hyperplanes after moving is called margin. The standard for the best hyperplane is the hyperplane that can obtain the maximum margin [16-17].

In this paper, we can define a hyperplane in the following form;

$$w^* \cdot x + b^* = 0 \tag{1}$$

And the corresponding decision classification function, also known as decision function:

$$f(x) = sign(w^* \cdot x + b^*) \tag{2}$$

Under this definition, the hyperplane's margin can be calculated from the expression as m=wT(xi-xj). Where xi and xj represent the first points touched by the hyperplane as it moves sideways. By adjusting the values of w and b such that yi(fxi) is greater than o for any i. Finding the hyperplane that produces the largest margin between the two classes of training points is equivalent to solving an optimization problem. So the optimal solution is w*, b*.

In the process of data linear inseparability, the method of maximizing the margin is still applicable, but some points are allowed not to meet the optimal conditions. And then we can define a relaxation variable.

The optimal problem becomes:

$$\min_{w,b} \frac{1}{2} \| w \|^2 + C\sum_{i=1}^{N} \xi_i \tag{3}$$

Where C is the penalty parameter, which determines how much you value the cost of outliers. The larger c is, the relaxation variable is expected to be close to 0, that is, the penalty for misclassification increases, and the training set tends to be fully divided into pairs. The consequence of this is that the prediction accuracy is relatively high, but the generalization ability becomes weak. When C becomes smaller and smaller, the penalty for misclassification decreases, the possibility of allowing errors increases, and the generalization ability becomes stronger accordingly.

SVM (SVM) proposes the concept of kernel function to deal with this problem. That is, the original two-dimensional data is mapped to higher dimensions through a kernel function, and the transformed data is linear, and the principle introduced above can be applied for classification [18]. This makes SVM more effective for solving high-dimensional problems, and avoids the "dimensional disaster" problem often encountered in traditional machine learning methods.

(2) Principle of regression

In order to prevent some deviation points from leading to no solution of the optimal problem, we define an ε-insensitive loss function, which means that when training the training set samples, the maximum deviation between the regression function obtained in the training process and the actual value is allowed to be s.

In this case, the objective function is:

$$\min \frac{1}{2}\| w\|^2 + C\sum_{i=1}^{N}| y_i - f(x_i)|_{\varepsilon} \tag{4}$$

The expression of the optimal hyperplane, also known as the regression function, can be obtained by solving the above optimal problem.

2.3. SVM Prediction Model Construction

For the stock price data studied in this paper, by constructing a nonlinear mapping function φ, the time series data is mapped to a high-dimensional feature space, and then the data is analyzed, processed and predicted in the high-dimensional space.

Given the learning sample D={(x1, y1), (x2, y2),... , (xi, yi)}, I = 1, 2,... ,n,xi represents the sample input, and yi represents the sample expected output. Make f(x) as close as possible to y, ω and b are undetermined model parameters:

$$f(x) = \omega^T x + b \tag{5}$$

Unlike traditional models, SVR assumes that we can tolerate at most an þdeviation between the two. The optimization function is used to optimize the target value, namely:

$$\min \frac{1}{2}\| \omega\|^2 + C\sum_{i=1}^{m}(\theta_i + \hat{\theta}_l) \tag{6}$$

Kernel function is mainly used to solve the dimensional disaster that may occur in the process of SVM model building. It reduces the probability of dimensional disaster by simplifying the calculation of inner product of eigenvectors in high-dimensional space. When using kernel functions, you usually don't need to worry about the computation, just the end result.

RBF function is used as kernel function in this paper. Therefore, the stock price prediction model based on RBF function SVM is:

$$f(x) = \sum_{i=1}^{l}\alpha_i \exp(-\frac{\| x_i - x\|^2}{2\sigma^2}) + b \tag{7}$$

3. Example Analysis

3.1. Data Sources

The data used in this document is the stock data obtained from the network stock interface, and 480 days of daily commercial data are selected. The specific data used is the highest price, lowest price, opening price, closing price and trading volume of the stock. The close value is the dependent variable, and the other variables are independent variables. Regression is happening. Because forecasting the future price trend must use the information of the past to predict the future, this document uses today's closing price sequence to retreat with yesterday's higher price, lower price, opening price and trading volume. In this way, adjust the model to today's data, and you can get tomorrow's closing price, thus completing the forecast.

3.2. Experimental Setting

This paper constructs SVM model in R language. The input variables are the highest price, the lowest price, the opening price and the trading volume. The data of 400 days are selected as the training set, and the data of the remaining 80 days are used as the test set. The e1071 package was used in R language software, and the svm function was called to fit the model. Since the fitting speed was faster after data standardization, the parameter scale=T was set, and the default parameter Settings were used, namely, cost=1, gamma=0.25, epsilon=0.1 to fit the model.

4. Analysis of Experimental Results

In order to compare with the traditional univariate time series model, only the closing price series is fitted first, as shown in FIG. 1. The fitted line is very similar to the unary regression, which cannot fit the fine trend and has a poor fitting effect.

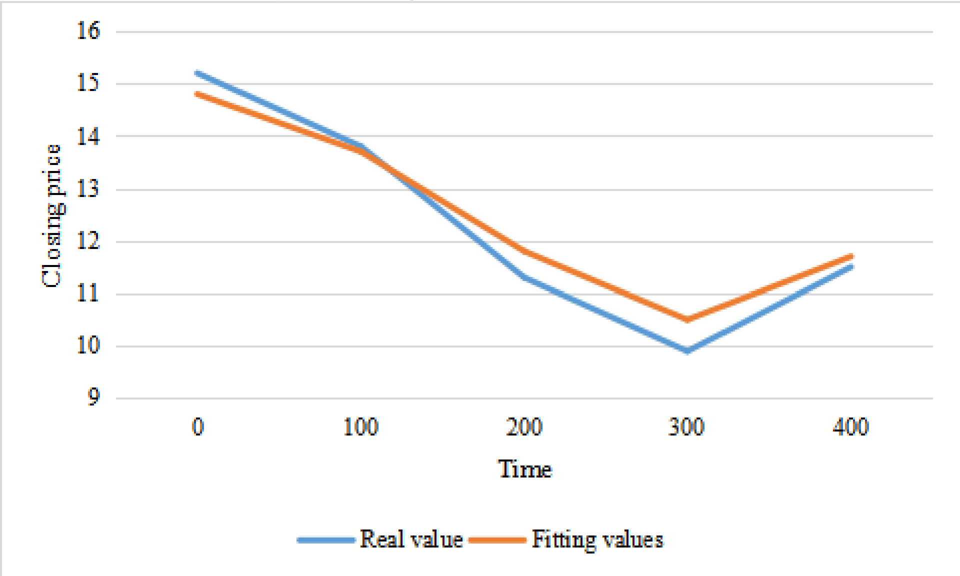

Figure 1. Closing price SVM regression

It can be seen from FIG. 2 that most of the fitted values are close to the real values, but the fitting of the data in some intervals is not very sufficient, especially in the places where the data changes greatly, the results are not very good, and the data fluctuation trend is not fully reflected.

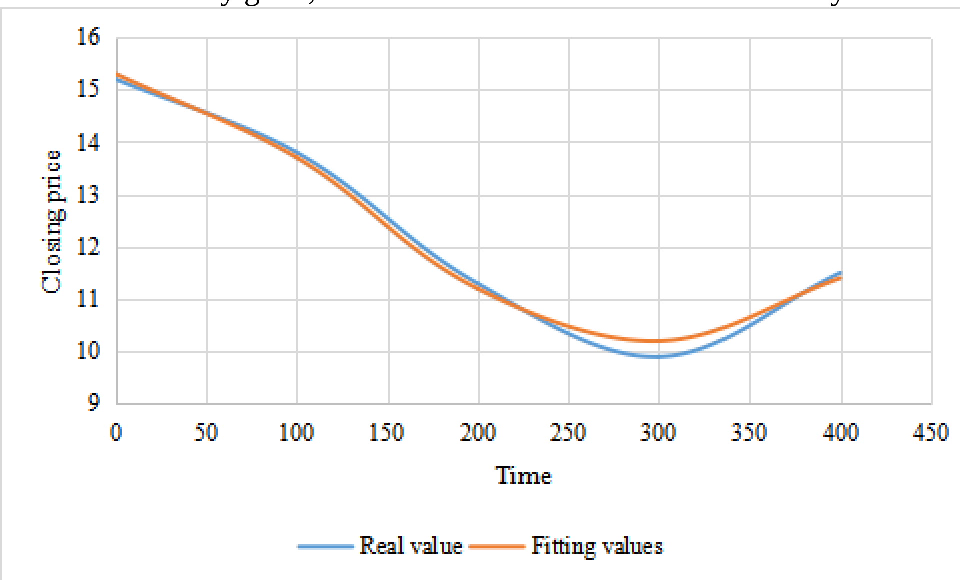

Figure 2. Time sequence diagram of fitted values

Next, the prediction of the model is tested on the test set, as shown in Figure 3.

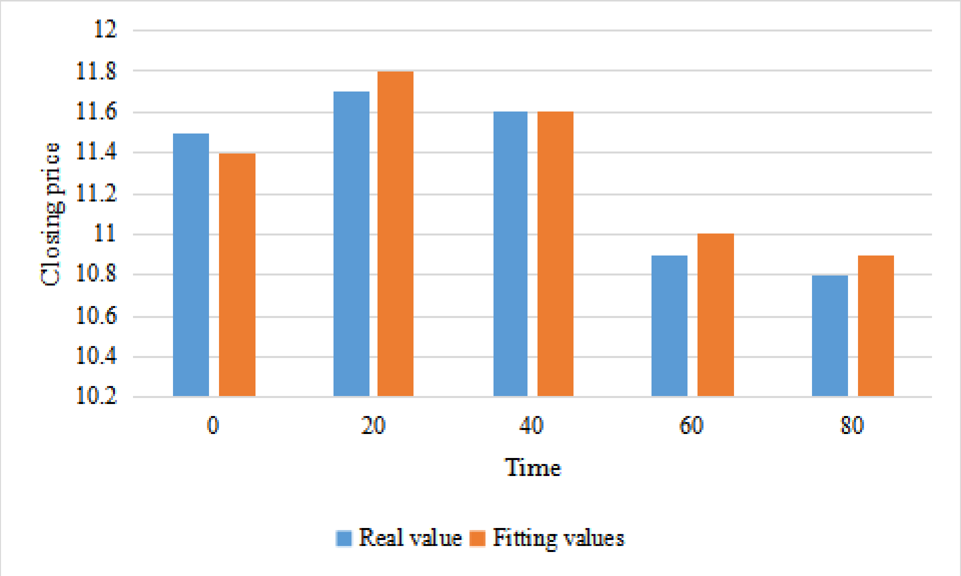

Figure 3. Predicted value timing diagram

It is obvious that the predicted value is consistent with the overall trend of the real value.

As shown in Table 1, this paper lists the comparison between the predicted values of ARIMA model and SVM model.

Table 1. ARIMA model and SVM model predicted value

	1	2	3	4	5
Real Value	11.45	11.63	11.58	12.45	12.17
ARIMA	11.41	11.32	11.35	12.32	11.98
SVM	11.43	11.54	11.55	12.41	12.05

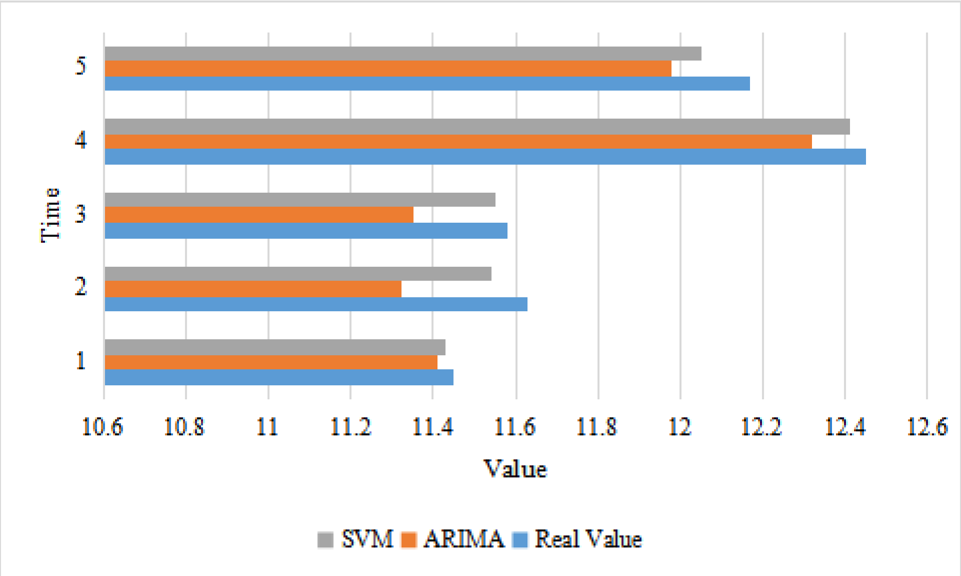

Figure 4. Comparison of ARIMA model and SVM model prediction results

As shown in Figure 4, the prediction trend of the model is relatively accurate, and the gap between the predicted value and the real value is also close.

5. Conclusion

Based on the understanding of vector support machine and time series model, this document proposes a stock price prediction model based on vector support engine regression, and establishes a traditional time series model for comparison. Because the SVM fits the nonlinear problem well, and uses multiple variables to carry on the regression, it can get the better prediction result. Time series analysis emphasizes the role of time factor in forecasting and does not consider the influence of external factors. Therefore, the prediction error is defective. When significant changes take place in the outside world, large deviations will occur. The effect of forecast time series on short- and medium-term forecasts is better than long-term forecasts. Financial time series data, such as stock prices, have very complex factors that affect their changes, so it is not accurate to predict the future development only from the perspective of time based on the past trend of the market.

Funding

Science and Technology Project of Jiangxi Provincial Department of Education, Project No.: GJJ2202418 Project Name: Research on intelligent Engineering Audit System of communication operators.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Ren R, Wu D D, Liu T. Forecasting stock market movement direction using sentiment analysis and SVM. IEEE Systems Journal, 2018, 13(1): 760-770. https://doi.org/10.1109/JSYST.2018.2794462

[2] Sedighi M, Jahangirnia H, Gharakhani M, et al. A novel hybrid model for stock price forecasting based on metaheuristics and SVM. Data, 2019, 4(2): 75. https://doi.org/10.3390/data4020075

[3] Gowthul Alam M M, Baulkani S. Local and global characteristics-based kernel hybridization to increase optimal SVM performance for stock market prediction. Knowledge and Information Systems, 2019, 60(2): 971-1000. https://doi.org/10.1007/s10115-018-1263-1

[4] Nti I K, Adekoya A F, Weyori B A. Efficient stock-market prediction using ensemble SVM. Open Computer Science, 2020, 10(1): 153-163. https://doi.org/10.1515/comp-2020-0199

[5] Reddy V K S. Stock market prediction using machine learning. International Research Journal of Engineering and Technology (IRJET), 2018, 5(10): 1033-1035.

[6] Pant M, Kumar S. Fuzzy time series forecasting based on hesitant fuzzy sets, particle swarm optimization and SVM-based hybrid method. Granular Computing, 2022, 7(4): 861-879. https://doi.org/10.1007/s41066-021-00300-3

[7] Das S P, Padhy S. A novel hybrid model using teaching–learning-based optimization and a SVM for commodity futures index forecasting. International Journal of Machine Learning and Cybernetics, 2018, 9(1): 97-111. https://doi.org/10.1007/s13042-015-0359-0

[8] Gururaj V, Shriya V R, Ashwini K. Stock market prediction using linear regression and SVMs. Int J Appl Eng Res, 2019, 14(8): 1931-1934.

[9] Henrique B M, Sobreiro V A, Kimura H. Stock price prediction using support vector regression on daily and up to the minute prices. The Journal of finance and data science, 2018, 4(3): 183-201. https://doi.org/10.1016/j.jfds.2018.04.003

[10] Zhang J, Teng Y F, Chen W. Support vector regression with modified firefly algorithm for stock price forecasting. Applied Intelligence, 2019, 49(5): 1658-1674. https://doi.org/10.1007/s10489-018-1351-7

[11] Pattanayak R M, Panigrahi S, Behera H S. High-order fuzzy time series forecasting by using membership values along with data and SVM. Arabian Journal for Science and Engineering, 2020, 45(12): 10311-10325. https://doi.org/10.1007/s13369-020-04721-1

[12] Parray I R, Khurana S S, Kumar M, et al. Time series data analysis of stock price movement using machine learning techniques. Soft Computing, 2020, 24(21): 16509-16517. https://doi.org/10.1007/s00500-020-04957-x

[13] Gupta D, Pratama M, Ma Z, et al. Financial time series forecasting using twin support vector regression. PloS one, 2019, 14(3): e0211402. https://doi.org/10.1371/journal.pone.0211402

[14] Anand C. Comparison of stock price prediction models using pre-trained neural networks. Journal of Ubiquitous Computing and Communication Technologies (UCCT), 2021, 3(02): 122-134. https://doi.org/10.36548/jucct.2021.2.005

[15] Srivastava P R, Zhang Z J, Eachempati P. Deep neural network and time series approach for finance systems: predicting the movement of the Indian stock market. Journal of Organizational and End User Computing (JOEUC), 2021, 33(5): 204-226. https://doi.org/10.4018/JOEUC.20210901.oa10

[16] Zahariev A, Zveryakov M, Prodanov S, et al. Debt management evaluation through SVMs: on the example of Italy and Greece. Entrepreneurship and sustainability issues, 2020, 7(3): 2382. https://doi.org/10.9770/jesi.2020.7.3(61)

[17] Vilela L F S, Leme R C, Pinheiro C A M, et al. Forecasting financial series using clustering methods and support vector regression. Artificial Intelligence Review, 2019, 52(2): 743-773. https://doi.org/10.1007/s10462-018-9663-x

[18] Horak J, Vrbka J, Suler P. SVM methods and artificial neural networks used for the development of bankruptcy prediction models and their comparison. Journal of Risk and Financial Management, 2020, 13(3): 60. https://doi.org/10.3390/jrfm13030060