

Stock Prediction and Analysis Based on Support Vector Machine

Menglei Hu, Zili Tang, Xiaoxian Xie and Min Jiang

Anhui University of Finance and Economics, Bengbu, 233030, China

Abstract: The changes of the stock market are closely related to the market development and economic research trends of the whole country. Correctly predicting the stock price trend is not only beneficial for investors to make correct investment management decisions, but also plays an important role in promoting the effective allocation of resources and improving market effectiveness. To this end, the daily closing prices of Shanghai Hengrui Medicine and Baosteel Co., Ltd. are used as the basic data, and the support vector machine of python software is used to empirically predict and analyze the development of my country's stock market. The accuracy rate is as high as about 90%; then the parameter optimization is carried out on the basis of the model again, and it is also concluded that the accuracy of stock prediction is as high as about 90%, indicating that the stock prediction accuracy rate based on support vector machine is ideal and has meaningful.

Keywords: Support vector machine, Stock forecast, Parameter optimization.

1. Research Background and Purpose

The changes of the stock market are inextricably linked with the development of the national market, and the stock market has a very important impact on the continuous growth of the national economy. The future stock price trend has always been the core issue for investors. Correct stock price trend judgment is not only beneficial for investors to make correct investment decisions, but also is of great significance for promoting the effective allocation of resources and enhancing the effectiveness of the market. This article predicts and analyzes my country's stock market by drawing on the research experience of domestic and foreign scholars on the development of securities investment at the national level and combining some basic stock indicators. The daily closing price is selected as the indicator of stock market forecast to measure the development trend of my country's stocks, and through the analysis of the status quo and the results of market measurement, solutions to the development trend of my country's stocks are put forward (Chen Fangfang [1], 2017).

2. Review of Domestic and Foreign Stock Forecast Literature

Liu Qingxia [2] (2017) verified that the improved BP network based on principal component analysis can be well adapted to stock data technology through learning and training, and has a better prediction effect. Shen Jinrong [3] (2017) took financial indicators as the analysis object, and used the improved CART decision tree and stepwise regression to measure, and concluded that the stepwise regression model based on the decision tree can reduce the financial indicators that affect the target variable and improve the model prediction accuracy. Li Dan [4] (2018) studied the stock forecasting problem from the perspective and conducted empirical analysis, and analyzed the stock forecasting optimal network structure, forecasting results and experimental results of SVFD-BPNN, MVFDIF-BPNN and MVFDIL-BPNN. Hu Di and Huang Wei [5] (2019) based on SVM combination algorithm and clustering stock prediction

algorithm neighbor propagation clustering to empirically analyze stock correlation, and verify that the combination of AP algorithm and other algorithms improves the accuracy of stock prediction. Zhang Jinghua and Gan Yujian [6] (2019) proposed that the deep learning support vector machine optimizes the configuration of the model parameters, and uses the model to conduct simulation experiments. The results show that the deep learning SVM has significantly improved prediction accuracy compared with the existing SVM. There are more foreign studies on stock forecasting than domestic ones. Charles Dow [7] (1902) wrote a review of the market view. Sam Nelson developed his views on the basis of his comments on the view of the market, eventually forming the Dow Theory. W.D. Gann[8] studied the importance of time and put forward the concept of "price-time equivalence". Frank Rosenblatt (1957) invented linear classifiers called perceptrons. In the mid-1990s, Corinna Cortes and Vapnik proposed a statistical learning-based SVM with many unique advantages in the face of nonlinear, small-sample, high-dimensional pattern recognition problems. Lerner and Vapnik [9] (1963) introduced the maximum interval classification algorithm. Soft margin classifiers were introduced by Cortes and Vapnik (1995), the same year SVM was extended to regression models. Gavrishchaka [10] et al. (2006) studied the volatility and risk of the stock market. Compared with the existing mainstream models, the use of SVM to build a volatility evaluation framework can effectively deal with high-dimensional data, and can conduct longer-term and larger volatility analysis. Scale evaluation, the effect is better than other mainstream evaluation models. Funatsu and Kaneko [11] (2013) proposed an online support vector machine based on time series to study adaptive software-aware prediction models. In addition, the window size and appropriate hyperparameter settings are investigated, resulting in regression reliability predictions.

To sum up, although great progress has been made in stock market forecasting, there are still many places to explore the depth and scope of application of the theory. At present, it is not only an important period for the development of the national economy, but also the development of changes in China's securities investment. important period. Therefore, on

the basis of the advanced theories put forward by scholars at home and abroad, how to find a method that can effectively improve the current defect becomes the key. Based on the above considerations, this paper chooses SVC in SVM with strong generalization ability as the core model for predicting stock prices.

3. Theoretical Basis

3.1. The Concept of Support Vector Machine

Support Vector Machine (SVM) is a machine learning method based on VC dimension theory and structural risk minimization theory proposed by Vapnic et al. At the same time, it can also analyze nonlinear classification through kernel method research, which breaks through the phenomenon of small-scale data overfitting that is easy to occur in traditional machine learning based on empirical risk minimization theory.

(1) Introduction to the kernel function of the support vector machine

In the feature space, we want the samples to be linearly separable. But without knowing these feature maps, we cannot know for ourselves which kernel function is appropriate. Therefore, the selection of the kernel function is very important for the quality of the support vector machine model. The following are several common kernel functions. Linear kernel function (linear) is the simplest kind of kernel function, and the calculation method is: $K(X_i, X_j) = X_i^T X_j$. The polynomial kernel function (poly) is a non-standard kernel function, which is very suitable for orthonormalized data sets. The calculation method is: $K(X_i, X_j) = (X_i^T X_j)^d, d \geq 1$. The Gaussian kernel function (rbf) has better anti-interference ability in dealing with the noise of the data. The calculation method is: $K(X_i, X_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\delta^2}\right), \delta > 0$.

(2) Introduction to the parameters of the support vector machine

The correct choice of SVM parameters has a great influence on the classification management effect. In general, the parameters that need to be optimized are the C penalty parameter and the σ kernel parameter, respectively. But at present, there is no good theory to optimize the guiding parameters. The commonly used methods are experiment, grid, gradient descent method, etc. In this paper, the grid method is used to manage and optimize C, which simplifies the operation process of parameter selection and improves the classification performance of SVM based on the selected parameters.

4. Sample Selection

In the empirical analysis of the stock forecast in this paper, it is considered that the stock market is a very unstable dynamic process, and its future development trend is also affected by the government's macro-control, and the impact of the epidemic in 2020 may have a greater impact on medicine. Therefore, in the data selection At this time, Hengrui Medicine, which has been affected more, and Baosteel Co., Ltd., whose price is stable, are selected as the research objects, in order to compare the prediction results of the two types of stocks to verify the credibility of SVM.

5. Empirical Analysis of Stock Prediction Based on Support Vector Classifier

5.1. Data Preprocessing

This paper selects the data of Baosteel Co., Ltd. and Hengrui Medicine from 2018.01.01 to 2020.03.01, a total of 523 data. At the same time, in order to test the data training of python, this paper established a large sample and a small sample data training set for Baosteel Co., Ltd. and Hengrui Medicine. The large sample data adopts the full sample data, and the small sample data adopts the 2019.06. Data, the data comes from the online crawling of python crawler.

5.2. Operation Process

Use python to grab the historical data of two stocks online from 2018.01.01 to 2020.03.01, and preliminarily organize the data of the two stocks. Specific implementation: Use value (today's closing price minus yesterday's closing price) to indicate ups and downs. If the difference is greater than 0, it will be assigned a value of 1 for an increase, and if the difference is less than 0, it will be assigned a value of 0 for a decrease.

For further label classification, the first 80% of the data is taken as the training set, and the last 20% of the data is used as the test set, and then the sample data is normalized. Start to use the kernel function for periodic prediction, predict a value forward each time, select 'ploy', 'linear', and 'rbf' to classify the predicted values, and finally calculate the correct rate in the test set to obtain the actual value of the output value and The results of the predicted values are as follows:

Large sample: 'ploy': Baosteel Correct=91.43%, Hengrui Medicine Correct=93.33%; 'linear': Baosteel Correct=96.19%, Hengrui Medicine Correct=96.19%; 'rbf': Baosteel Correct=90.48 %, Hengrui Medicine Correct=93.33%.

Small sample: 'ploy': Baosteel Correct=93.10%, Hengrui Medicine Correct=96.55%; 'linear': Baosteel Correct=93.10%, Hengrui Medicine Correct=93.10%; 'rbf': Baosteel Correct=89.66 %, Hengrui Medicine Correct=96.55%.

The above is the prediction analysis based on the parameters by default. It can be seen that in the three kernel functions, whether it is a large sample or a small sample, the accuracy rate is about 90%. It can be seen that the SVM prediction effect is ideal. However, since the SVM parameters have an important influence on the prediction effect of the model, and the accuracy of rbf is relatively low, this paper selects a large sample of rbf for parameter optimization:

The SVM parameter is selected to be $C=1000000.0$, so this paper changes the C parameter in rbf from the default value of 1.0 to the optimal parameter of 1000000.0, and the accuracy rate is Correct=98.10%, which is much higher than the previous 90.48%. It can be seen that the parameters have been modified Afterwards, it has a positive effect on the prediction effect. Printing the predicted grade confusion matrix yields:

```
[[38 16]
 [9 42]]
```

The confusion matrix is the error matrix that we can use to evaluate the performance of supervised learning algorithms. In the confusion matrix, the more values that appear in the second and fourth quadrants, the better; conversely, the fewer

values that appear in the first and third quadrants, the better. It can be seen from the above output results that the values appearing in the second and fourth quadrants are 38 and 42, respectively, which are much larger than 16 and 9. It can be seen that the prediction effect of the model is considerable.

Since the statistics of the confusion matrix are only numbers, in the face of a large amount of data, it is difficult to measure the pros and cons of the model only with numbers. Therefore, several indicators are extended on the basis of the

basic statistical results: accuracy, which refers to all the results that the model prediction is more positive than the model prediction category; sensitivity refers to all the results that the model prediction is more positive than the actual category; F1 Score is the output result that combines the precision rate and the recall rate. Its value ranges from 0 to 1. 1 represents the best prediction model, 0 represents the worst prediction model; Support refers to the number of original data categories. Print the predicted grades report to:

Table 1. Print Predicted Score Report

	precision	recall	f1-score	support
0.0	0.81	0.70	0.75	54
1.0	0.72	0.82	0.77	51
accuracy			0.76	105
macro avg	0.77	0.76	0.76	105
weighted avg	0.77	0.76	0.76	105

It can be seen from Table 1 that the data of each indicator are more than 70%, and the position of the distance 1 is relatively close. It can be concluded that the prediction effect of the model is ideal. It is also ideal to perform the same operation on Hengrui Medicine to obtain its model prediction effect.

6. Main Conclusions and Policy Recommendations

6.1. Main Conclusions

This paper applies SVM to the prediction of my country's stock market. Use SVM to select the kernel function and optimize the parameters, and then find the optimal model to measure the stock trend. The main conclusions are as follows:

(1) Based on the fast convergence speed and high precision of the SVM model, the SVM model can predict the stock data very well, so that the prediction results are very close to the actual value.

(2) Kernel function selection and kernel parameter selection have a very important impact on the learning and prediction performance of SVM. Different kernel functions and kernel parameters are directly related to the accuracy of the operation results.

(3) SVM has good accuracy in predicting stock prices, and provides a very meaningful analysis tool for investors.

6.2. Policy Recommendations

Stock prices fluctuate frequently, and often quite dramatically. To make the stock market develop more stably, it is necessary to reduce the uncertainty of investors and fundraisers, so that stock forecasting can be used more flexibly between investors and fund-raisers, so as to achieve the optimal resource allocation market. This paper proposes to improve my country's stock market roughly from the following aspects, in order to reduce the uncertainty of the stock market.

First, through extensive social research and discussion, determine the stage target of my country's stock market economic development. Second, build a dynamic monitoring system for stock market quality as soon as possible, in order

to achieve timely and accurate assessment and grasp of stock market quality. Third, on the basis of the above, regulatory authorities should pay attention to changes in market quality, in order to achieve the stability of the stock market and reduce the risk of investors and fundraisers.

7. Conclusion

The Chinese stock market is a policy market, and once bad news or good news leaks out, it will have a big impact on the stock market. In addition, some dealers and institutions also operate the stock market, making it difficult to accurately predict stock prices. Therefore, we should refer to some policy indicators or information in the actual operation process, and then make relevant investments according to the actual predicted trend. It is only a single use of SVM to predict the stock trend, and further research and improvement are needed. In this article, Baosteel and Hengrui Medicine are selected for stock forecasting. But if you choose more volatile stocks, is there still an advantage in the forecasting effect of SVM? Can you further research and explore stock picking? Does the 9-dimensional input feature space contain all the information needed for model training and prediction? If all the information is not included, the prediction effect and accuracy will be greatly reduced, so how to find the most representative prediction information attributes needs further research. But in general, stock market forecasting is very challenging, but its practical significance is obvious.

References

- [1] Chen Fangfang. Research on stock price prediction based on support vector machine [D]. Chongqing: Chongqing University of Technology, 2016.
- [2] Liu Qingxia. Stock price prediction based on principal component analysis and BP neural network [D]. Suzhou: Soochow University, 2017.
- [3] Shen Jinrong. Stepwise regression algorithm based on decision tree and its application in stock forecasting [D]. Guangzhou: Guangdong University of Technology, 2017.
- [4] Li Dan. Research on stock prediction based on multi-view feature data [D]. Xi'an: Northwest University, .2018.

- [5] Hu Di, Huang Wei. Stock Price Prediction Based on AP_SVM Portfolio Model [J]. Journal of Wuhan Engineering University, 2019, (6): 297-301.
- [6] Zhang Jinghua, Gan Yujian. Prediction of Shanghai Stock Exchange Index Based on Deep Learning Support Vector Machine [J]. Statistics and Decision, 2019, (2): 178.
- [7] Li Yue. Stock market forecast based on support vector machine [D]. Shanghai: Fudan University, 2014.
- [8] Jiang Peng. Research on Stock Timing Strategy Based on Support Vector Machine SVM [D]. Nanchang: Jiangxi University of Finance and Economics, 2019.
- [9] C Cortes, V Vapnik. Support-Vector Networks. [J]. Machine Learning, 1995, 20(3): 273-297.
- [10] Sun Ruoyu. Stock Price Prediction Based on Support Vector Machines [D]. Dalian: Dalian University of Technology, 2019.
- [11] Hiromasa Kaneko, Kimito Funatsu. Adaptive Soft Sensor Model Using Online support Vector Regression with Time Variable and Discussion of Appropriate Hyperparameter Settings and Window Size[J]. Computers and Chemical Engineering, 2013, 58(11): 288-297.