2020 International Conference on Computer Engineering and Intelligent Control (ICCEIC) | 978-1-7281-8573-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICCEIC51584.2020.00032

Stock Price Prediction Model Based on RBF-SVM Algorithm

Zixuan Liu
School Of Control And Computer Engineering
North China Electric Power University
Beijing, China
zixuanliu@ncepu.edu.cn

Ziyuan Dang
School Of Control And Computer Engineering
North China Electric Power University
Beijing, China
Ziyuan_Dang@ncepu.edu.cn

Jie Yu
School Of Control And Computer Engineering
North China Electric Power University
Beijing, China
jieyu@ncepu.edu.cn

***Abstract*—At this stage, China's economic development continues to progress, and the emergence of various emerging industries has caused the stock market to show strong volatility. However, people's research and prediction on the stock market have never stopped. In order to improve the accuracy of stock prediction, this paper studies the network model based on the improved support vector machine (SVM) algorithm to realize the correct judgment of the stock price trend, so as to achieve the purpose of accurate stock price prediction, and improve the accuracy of the prediction while ensuring the speed of the model. Experiments show that the proposed prediction model can approximate the short-term price trend of the stock market, and provide a more reliable data basis for the accurate prediction of stock prices, which benefit the high-tech development and progress of the stock market.**

Keywords-support vector machine; stock price; prediction model

I. INTRODUCTION

In the past hundred years, the world economy has developed rapidly, which has brought great development opportunities to the financial industry. At the same time, the impact of capital market on the quality and speed of economic development is also expanding, so the role of the stock market is becoming more and more important. Stock market is not only one of the important ways of financing for listed companies, but also with the improvement of people's living standards, buying stocks has become a very common way of investment in people's life.

Investors earn the price difference by buying and selling stocks. Predicting the rise and fall of stock price is an important basis for predicting the stock market and making buying and selling operations. The rise and fall of stock price determines the loss and profit, so there is a certain risk in investing in stocks. People expect to reduce the risk and increase the return in the stock investment. If people only rely on subjective judgment and estimation to predict the stock price and operate, it will not only increase the risk, but also greatly reduce the return. Stock selection, timing and trading strategy are the three main research directions to reduce the risk and increase the return of stock investment. Among them, quantitative investment and portfolio investment strategy are the two most commonly used methods. The formulation of these investment strategies generally needs to be combined with the prediction of the trend of stock price rise and fall. Therefore, how to improve the accuracy of stock price prediction has always been a hot issue.

Stocks are derived from the long-term development of the market economy, and the issuance and trading of stocks are directly related to the development trend of the market economy. In the field of stock investment, people use more scientific methods to predict the trend of stock market prices, and the accuracy of the results directly determines the level of profit and the level of investment risk [1]. From the current development trend of our country, the prediction and research methods of stock market price trends are in a state of continuous updating. The more common methods include time series method, trend extrapolation method, and exponential smoothing method. The most classic prediction method is the time series prediction method, which mainly converts the stock price or the index form of the stock price into a time series, and predicts the stock price through the establishment of a time model, and then grasps the fluctuation state of the stock market price.

In recent years, the development and application of emerging science and technology have become more and more popular. Some international scholars have paid attention to the application value of artificial neural network [2], support vector machine (SVM) [3] and other methods in the processing of securities transaction data, especially the processing of stock prices. They did a lot of experiments and research, trying to realize the prediction of the fluctuation trend of the stock market. Among them, SVM, as a new technology in data mining, avoids excessive dependence on experience and can obtain the global optimal solution. In addition, the excellent generalization ability of SVM can effectively avoid the influence of local extreme values on prediction [4]. In order to realize the accurate prediction of stock prices, researchers at home and abroad have explored a large number of prediction methods and models, which have strong practical applicability. Junyoung Heo and Jin Yong Yang propose a prediction model based on SVM, and use

978-1-7281-8573-6/20/$31.00 ©2020 IEEE
DOI 10.1109/ICCEIC51584.2020.00032

financial statements as the input of the model to evaluate the predictability of the model to stock prices [5]. In order to improve the accuracy of stock trend prediction, Li Hui et al. propose a two-layer feature selection and prediction method based on the DFS-BPSO-SVM prediction model, which shows a good effect in the prediction of stock price [6]. Jihong Xiao et al. propose a combination prediction model based on the fusion of SSA and SVM for the characteristics of stock price trends, market fluctuations and noise, which open a new door to the research of combination models [7]. Li Tang et al. use SVM to establish a stock price prediction model to predict stock prices, and find a certain positive correlation between the stock market and social network information [8]. Hong Li et al. combine fractal interpolation and SVM model to build a combined model, which has strong applicability for the prediction of high-frequency index data [9]. Chenglin Xia et al. propose a prediction model based on ARI-MA-LS-SVM, and verify the actual prediction effect of the model on stock prices through simulation [10].

However, there are many factors that affect stock prediction in the actual stock prediction process, which will affect the authenticity of the prediction results. Therefore, this paper proposes a model based on the RBF-SVM algorithm to further optimize the stock price prediction, accurately judge the short-term trend of stock price changes, and provide more reliable guidance for stock market research and stock investors' investment decisions.

II. Non-Linear RBF-SVM Prediction Model

A. Basic principle of SVM algorithm

SVM is a kind of machine learning method, which is derived from the basic theory of statistical learning proposed by Vapnik. SVM has a strong advantage in minimizing structural risks, which can reduce the error of sample points to the greatest extent. In addition, SVM can ensure the global optimal results in solving and processing optimization problems. Therefore, it can be said that SVM is the optimal algorithm for solving convex quadratic programming [11].

In this paper, we suppose that the input sample data set is:

$$Q=\{(x_1,y_1),(x_2,y_2),\ldots,(x_j,y_j),\ldots,(x_n,y_n)\} \quad (1)$$

Where $x_j \in R^m$ and $y_j \in \{-1,1\}, j=1,2,\ldots,n$.

The basic idea of SVM learning is to solve the separation hyperplane that can correctly divide the training data set and have the largest geometric interval, which is shown in Figure 1. $\omega \cdot x+b=0$ is the separated hyperplane, where ω is the normal vector of the separated hyperplane. For a linearly separable data set, there are infinitely many such hyperplanes, but the separating hyperplane with the largest geometric interval is unique [12].

The basic SVM model assumes that all data points are linearly separable, that is, there is a linear equation separating the two categories of points on both sides. Because the stock price sample data is linear and inseparable, it is necessary to use SVM to perform non-linear mapping on the price data, so that the data is mapped to a feature space more in line with the classification requirements, and the price data is linearly separable.

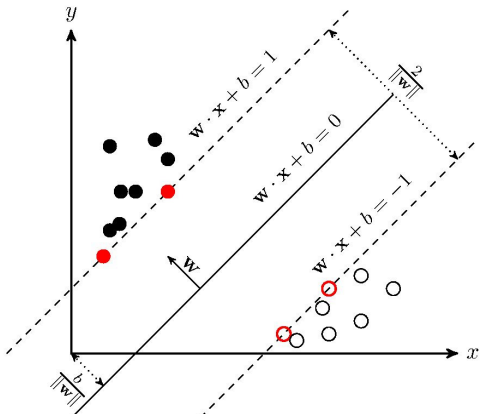

Figure 1. Schematic diagram of separated hyperplane.

B. Principle of Non-Linear SVM

In the basic linear SVM, the discriminant function is:

$$y(x)=\mathrm{sgn}(\omega \cdot x+b) \quad (2)$$

The basic situation of nonlinear and linear SVM is similar, the main difference lies in the linear mapping function $\varphi(x)$. The discriminant function of nonlinear SVM is:

$$y(x)=\mathrm{sgn}\left(\sum_{j=1}^{k}\alpha_j y_j\left(\varphi(x_j)\cdot\varphi(x)\right)+b\right) \quad (3)$$

where, $\alpha \geq 0$ represents the Lagrange multiplier, $\varphi(x_j)\cdot\varphi(x)$ is the operation form of inner product in high-dimensional space. Using kernel function $f(x_j,x)$ to represent $\varphi(x_j)\cdot\varphi(x)$, the dual problem in quadratic programming can be transformed into:

$$\max\left\{\sum_{j=1}^{k}\alpha_j-\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{k}\alpha_j\alpha_i y_j y_i f(x_j,x_i)\right\} \quad 0\leq\alpha_j\leq C \quad (4)$$

Among them, $\sum_{j=1}^{k}\alpha_j y_j=0$, the final discriminant function is:

$$y(x)=\mathrm{sgn}\left(\sum_{x_j\in SV}\alpha_j y_j f(x_j,x)+b\right) \quad (5)$$

Commonly used kernel functions include linear kernel functions, polynomial kernel functions, gaussian kernel functions, and sigmoid kernel functions. The more commonly used ones are polynomial kernel functions and gaussian kernel functions. The gaussian kernel function is usually called the RBF kernel function. The data segmentation effect of RBF kernel function is shown in Figure 2.

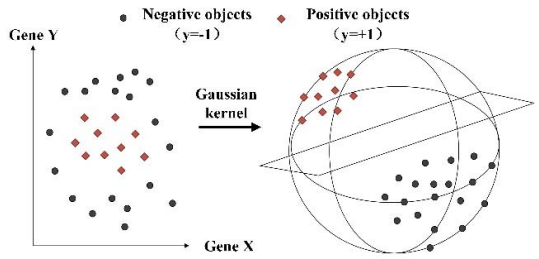

Figure 2. The data segmentation effect of RBF kernel function.

The RBF kernel function has the advantage of fewer parameters, which can simplify the complexity of the calculation process and improve the efficiency of the entire model. In this paper, we use RBF kernel function to build the non-linear model based on SVM. The expression of the RBF kernel function is:

$$f(x_j, x_i) = \exp\left\{-\left\|x_j - x_i\right\|^2 / \sigma^2\right\} \tag{6}$$

In this paper, SVM mainly analyzes the characteristics of sample data, and divides the prediction target data into two basic ranges: up and down, so as to achieve the prediction of stock price fluctuations. This paper uses the seven basic data information of stock prices as the basis for the selection of SVM algorithm training feature values. By training the relationship between the feature value and the 'rising' and 'falling' of the stock, the stock price prediction based on the characteristic data is realized.

III. Stock Price Prediction Based on SVM

A. Data set establishment

The experimental data in this paper mainly comes from the Yahoo Finance website. Through the analysis of the stock market in recent years, 10 representative companies are finally selected as the data acquisition objects established by the stock data set. The company's stock codes are AAPL, AMZN, MSFT, GOOG, IBM, ORCL, INTC, HPQ, LNVGY and NDAQ, respectively.

The download of data is mainly realized by using the pandas_datareader download tool in python. The data content includes the stock data of various companies between 2000/01/01 and 2016/12/31. Each piece of information includes seven basic attributes, namely time (Data), highest price (High), lowest price (Low), opening price (Open), closing price (Close), trading volume (Volume), the adjusted closing price (Adj.Close). The downloaded raw data is named according to the stock code, and after data classification and integration, a data file named Input_Dataset1.CSV is obtained. Part of the data in file are shown in Figure 3.

Close	Change	Stock_Price_Volatility	Stock_Momentum	Index_Volatility	Index_Momentum	Sector_Momentum
30.01571465	0	0.006990741	0.4	0.005945019	0.6	0.422222222
130.3099976	0	0.018429452	0.4	0.005945019	0.6	0.422222222
299.4326477	0	0.013546929	0.2	0.005945019	0.6	0.422222222
23.81017303	0	0.004878907	0.6	0.005945019	0.6	0.422222222
129.4799957	0	0.00850847	0.2	0.005945019	0.6	0.422222222
24.69000053	1	0.006338556	0.4	0.005945019	0.6	0.422222222

Figure 3. Part of the data in file Input_Dataset1.CSV.

It is very obvious that the stock price information in the stock market is in a state of volatility, and there are often obvious characteristics of price differences. Therefore, the characteristics of the data must be fully considered during the experiment to avoid too long learning time, too slow training speed, and minimize the error of the prediction result. In this paper, we have carried out a unified preprocessing of the data set, and standardized the data in Input_Dataset1.CSV in a normalized form. In addition, we construct a variable matrix and predicted values, and name the processed data SVM_Target.CSV, in which, a rise is recorded as 1 and a fall is recorded as 0.

B. Experimental platform construction and parameter settings

The system configuration mainly includes: the hardware device model is HP Z8 G4 (Z3Z16AV-SC001), the CPU type is Intel Xeon Bronze 3104, the memory size is 16GB, the operating system is Ubuntu 16.04 64bit, the experimental platform is PyCharm 3.1, and the programming language is Python language. We set C as 0.7, gamma as 1, and n_epoch as 30.

C. Model prediction results

Based on the SVM method proposed in this paper, a stock price prediction model is constructed on the collected stock test data. Actual value and predictive value of the prediction are shown in Figure 4.

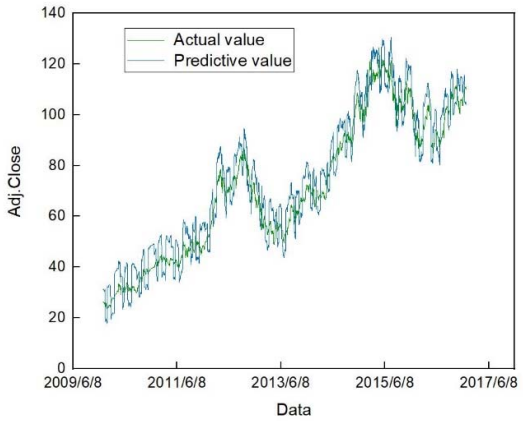

Figure 4. Actual value and predictive value of the prediction.

In addition, in order to evaluate the accuracy of the prediction model more scientifically, we conduct a comprehensive analysis of the accuracy, standard deviation, and running time of the model, and the results shown in Table I can be obtained.

TABLE I. Data Analysis of Prediction Results

Algorithm	EVALUATION CRITERIA		
	Mean_Accuracy	*Standard_Deviation*	*Total_Time*
RBF-SVM	65.64%	0.2338	145.9
PCA-SVM	62.17%	0.3645	178.5
GA-SVM	64.67%	0.2648	162.3
DFS-BPSO	61.31%	0.3784	174.6

It can be seen intuitively that the fluctuation range of the predictive value in the figure is basically consistent with the fluctuation range of the actual value. According to the performance evaluation indicators in Table I, although there are certain errors at some points on the curve, the magnitude of the

errors does not exceed the acceptable range. Compared with the original prediction algorithm, the prediction model based on SVM has higher accuracy and better prediction effect. On the whole, the final predictive value is basically consistent with the volatility of the actual value, which means that the SVM-based prediction model proposed in this paper has a very accurate prediction effect.

IV. Conclusion

The current stock market prices in China are showing relatively obvious volatility, which poses a greater challenge to the prediction of stock prices. In view of the difficulties in the current stage of stock market price prediction, this paper opens up a new research idea. Based on the traditional SVM algorithm, this paper optimizes and improves the model according to the characteristics of the overall trend of stock prices, realizes a stock price prediction model with higher accuracy, and provides a reliable basis for applications in stock market investment and stock analysis. However, in order to realize the practical value of prediction models in quantitative investment, the stability and applicability of prediction need to be further strengthened, and the accuracy of forecasting results also need to be improved to a certain extent.

Acknowledgment

This work was supported by Catholic University of Leuven.

References

[1] A. Kong, H. Zhu, "Predicting Trend of High Frequency CSI 300 Index Using Adaptive Input Selection and Machine Learning Techniques," Journal of Systems Science and Information, vol. 6, pp. 120–133, 2018.

[2] A. K. Jain, J. Mao, K. M. Mohiuddin, "Artificial Neural Networks: A Tutorial," Computer, vol. 29, pp. 31–44, 2015.

[3] M. Santoso, R. Sutjiadi, R. Lim, et al, "Indonesian Stock Prediction Using Support Vector Machine (SVM)," Matec Web of Conferences, pp. 164, 2018.

[4] L. I. Hong, Y. W. Hong, "Prediction of Stock Index Based on Fractal Interpolation and SVM," Journal of Jishou University (Natural Sciences Edition), vol. 3, pp. 19–24, 2018.

[5] J. Y. Heo, J. Y. Yang, et al, "Stock Price Prediction Based on Financial Statements Using SVM," International Journal of Hybrid Information Technology, vol. 9, pp. 57–66, 2016.

[6] H. Li, Y. H. Zhao, "Stock Trend Predicting Method Based on DFS-BPSO-SVM Predicting Model," Software Guide, vol. 16, pp. 147–151, 2017.

[7] J. H. Xiao, X. H. Zhu, C. X. Huang, et al, "A New Approach for Stock Price Analysis and Prediction Based on SSA and SVM," International Journal of Information Technology & Decision Making, pp. 1–17, 2018.

[8] L. Tang, S. Zhang, L. He, et al, "Research on Stock Prediction in China based on Social Network and SVM Algorithm," 2nd International Conference on Economic Development and Education Management (ICEDEM 2018), pp. 435-438, 2018.

[9] H. Y. Wang, H. Li, J. Y. Shen, et al, "A Novel Hybrid Fractal Interpolation-SVM Model for Forecasting Stock Price Indexes," Fractals, pp. 1-20, 2019.

[10] C. Xiao, W. Xia, J. Jiang, et al, "Stock Price Forecast Based on Combined Model of ARI-MA-LS-SVM," Neural Computing and Applications, pp. 1-10, 2020.

[11] Y. Zhang, B. Li, H. Lu, et al, "Sample-Specific SVM Learning for Person Re-identification," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[12] S. M. Erfani, S. Rajasegarar, S. Karunasekera, et al, "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," Pattern Recognition, pp. 121-134, 2016.