ISSN 2706-6827 Vol. 6, Issue 6: 126-131, DOI: 10.25236/IJFS.2024.060620

Research on Stock Trend Prediction Method in Financial Markets Based on Support Vector Machines

Liu Linfeng, Dong Mei

School of Economics, Qingdao University, Qingdao, Shandong, 266071, China
13159882280@163.com

Abstract: *In the stock market, stock prices are often influenced by macroeconomic indicators, market sentiment, company fundamentals, and other factors, presenting complex nonlinear relationships. However, traditional stock trend prediction methods are often affected by factors such as data noise and nonlinear relationships, resulting in low prediction accuracy. Therefore, a research on financial market stock trend prediction method based on support vector machine is proposed. Firstly, web scraping technology is used to obtain financial market stock data, including historical price data, trading volume data, and technical indicator data, and these data are preprocessed. Then, linear discriminant analysis is used to extract data features, and a support vector machine prediction model is constructed. The extracted features are used as input data to obtain predicted stock trends in the financial market. The experimental results show that the stock prediction trend of the proposed method is consistent with the actual trend, and the mean square error is small, which has practical application value.*

Keywords: *Support Vector Machine; Financial markets; Stocks; Linear discriminant analysis; Trend prediction*

1. Introduction

Stock trend prediction is an important basis for investment decisions. In financial markets, investors often need to formulate investment strategies and decisions based on market trends. By accurately predicting stock trends, investors can better grasp market opportunities, reduce investment risks, and improve investment returns. However, the volatility of the stock market is often influenced by factors such as the overall economic environment, policy changes, and market sentiment. The interaction between these factors makes stock trend prediction a complex and difficult task. Therefore, it is necessary to continuously explore and delve into new prediction technologies in order to better adapt to the changing market trends and enhance the accuracy of predictions. With the development of science and technology, there are more possibilities for stock trend prediction methods. Traditional prediction methods are often based on theories such as statistics or econometrics, which have problems with model assumptions and data limitations, making it difficult to accurately capture real-time dynamic changes in the market. Therefore, exploring a more efficient and accurate prediction method has become an important issue in the financial field.

Currently, many researchers have conducted research on this. Reference [1] proposes an evolutionary game model to deeply analyze the dynamic behavior of stock investors at specific times. By constructing a dynamic replication equation, the trend of investor behavior changes is solved, and the relationship between investor behavior and stock price is characterized by market depth. On this basis, a stock intraday price prediction model based on micro market structure was established, taking into account factors such as financial strength and relevant parameters. The stock market is a highly dynamic and complex environment, influenced by various factors. Methods based on evolutionary game models may not be able to fully capture all changes in the market. Therefore, when dealing with market mutations or emerging trends, the predictive ability of the model may be limited. Reference [2] proposes a stock trend prediction model based on DTW clustering analysis and TCN, which uses feature variables such as opening price for standardization, clusters stocks through DTW, and extracts common features for network training using TCN. The model can predict stock trends and output trends per minute, but DTW processing long time series is time-consuming, which affects the speed of model training and prediction, especially when dealing with large-scale datasets. Reference [3] proposes a new method for predicting stock prices, which utilizes a GRU deep learning network to

ISSN 2706-6827 Vol. 6, Issue 6: 126-131, DOI: 10.25236/IJFS.2024.060620

predict stock prices based on historical stock data, and constructs a prediction model based on the GRU framework. At the same time, the lightweight large model ALBERT is used to extract emotions from media news and obtain emotional features. Finally, input emotional features into the GRU prediction model and adaptively adjust the output of the GRU to obtain the final stock prediction. However, the emotional orientation of news reporting in this method may not always accurately reflect the true emotions of the market, and may also be influenced by the reliability of news sources and the bias of reporting angles, resulting in extracted emotional features deviating from the true emotions of the market.

To solve the problems of the above methods, a research on financial market stock trend prediction method based on support vector machine is proposed. The aim of this study is to provide a more efficient and accurate method for stock trend prediction in the financial market, provide valuable reference information for investors, and promote the continuous innovation and development of financial prediction technology

2. Design of Stock Trend Prediction Methods for Financial Markets

2.1 Financial Market Stock Data Collection and Preprocessing

In the research of financial market stock trend prediction methods based on support vector machines, data collection is an indispensable part. Firstly, it is necessary to clarify the prediction objectives, such as predicting the future price trend of a certain stock and determining the prediction period, such as a daily or weekly chart. Then, select appropriate data sources, including financial data platforms, official securities exchange websites, or obtain data through third-party API interfaces. After determining the data source, further determine the specific content to be collected, which usually includes historical price data, trading volume data, technical indicator data, and fundamental data. These data are crucial for analyzing stock trends and predicting future trends. The historical price data includes opening price, closing price, transaction price, etc. Trading volume reflects market activity. Technical indicator data, such as MACD, RSI, KDJ, etc., can help analyze stock trends and buying and selling signals. Fundamental data, such as company financial statements, industry news, policy changes, etc., have a significant impact on the long-term trend of stocks. To ensure timely and complete stock data, advanced network scraping technology is utilized to accurately collect financial market stock data, and reasonable collection frequency and time are set to ensure real-time updates and comprehensive coverage of data. The obtained financial information, due to its instability and interference factors, can have a negative impact on the accuracy of the model's prediction. Therefore, before applying the model, it is necessary to adjust the data first. Missing and outliers are filled and replaced with their previous normal values. In order to obtain more accurate prediction results, it is necessary to remove noise from financial time series data. This article uses kernel smoothing method to process data noise, which is set as the spatial interval between the center point and its adjacent observation sample points. The formula is as follows:

$$M_h(x)=\frac{\sum_{t=1}^{T}X_tN_h(x-Y_t)}{\sum_{t=1}^{T}N_h(x-Y_t)},N_h(x)=\frac{1}{h\sqrt{2\pi}}e^{-\frac{x^2}{2h^2}} \tag{1}$$

Among them, $N_h(x)$ is the kernel function, X_t is the collected data, Y_t is the coordinates of the sample center point X_t, h is the bandwidth of kernel smoothing, and $M_h(x)$ is the processed data.

Normalize the denoised data using the following formula:

$$X'=\frac{X-X_{\min}}{X_{\max}-X_{\min}} \tag{2}$$

Among them, X is the original financial data, X' is the transformed value, $X_{\min}$ is the

ISSN 2706-6827 Vol. 6, Issue 6: 126-131, DOI: 10.25236/IJFS.2024.060620

minimum value of the original financial data, and $X_{\max}$ is the maximum value of the original financial data.

In summary, complete the collection and preprocessing of stock data in the financial market.

2.2 Feature extraction of stock data in financial markets

After data preprocessing, it is necessary to extract useful features for predicting stock trends from the data. Based on historical experience and theoretical knowledge, basic characteristics such as opening price, closing price, and trading volume can be selected. However, in financial markets, stock data often contains a large number of features, and directly using all features for modeling can result in huge computational complexity and long training time. Therefore, using linear discriminant analysis for feature selection can reduce the dimensionality of the data, reduce computational complexity, and thus accelerate the training and prediction process of the model. The purpose of linear discriminant analysis is to find the best transformation to extract discriminative features of two or more types of objects. The specific process is as follows:

Assuming the stock sample dataset is $X=\{(x_1,y_1),(x_2,y_2),\ldots,(x_m,y_m)\}$, x_i is any data point, y_i is the category, and $y_i\in\{-1,1\}$.In this study, $y=1$ represents a positive class representing an increase in stock price, meaning that the closing price on the following day is higher than the closing price on the previous day; $y=-1$ represents a negative class representing a decline in stock price, meaning that the closing price of the following day is lower than the closing price of the previous day. If N_j is the number of samples of Class j stocks, then the mean vector j of the sample of Class μ_j stocks is:

$$\mu_j=\frac{1}{N_j}\sum_{x\in X_j}x \tag{3}$$

Among them, X_j is the set of sample data points for Class j stocks.

The covariance matrix $\sum_j$ of the j-class stock sample is:

$$\sum_j=\sum_{x\in X_j}\left(x-\mu_j\right)\left(x-\mu_j\right)^T \tag{4}$$

Simplify the stock data points into a line without width, that is, the direction vector ω, so the projection of any sample x_i on vector ω is $\omega^T x_i$; The projection of μ_j on a vector is $\omega^T\mu_j$.

If the intra class divergence matrix is $S_W=\sum_1+\sum_2$ and the inter class divergence matrix is $S_b=\left(\mu_1-\mu_2\right)\left(\mu_1-\mu_2\right)^T$, then the optimization objective is:

$$\arg\max_{\omega} H(\omega)=\frac{\omega^T S_b\omega}{\omega^T S_W\omega} \tag{5}$$

Solve to obtain $\omega=S_W^{-1}\left(\mu_1-\mu_2\right)$, which is the determined optimal projection direction. Project the samples in the training set towards ω to obtain:

$$y=\omega^T X \tag{6}$$

According to the above process, feature extraction of financial market stock data can be completed.

ISSN 2706-6827 Vol. 6, Issue 6: 126-131, DOI: 10.25236/IJFS.2024.060620

By selecting appropriate feature combinations, the predictive ability of SVM models can be improved.

2.3 Stock Trend Prediction in Financial Markets Based on Support Vector Machines

Based on the characteristics of financial market stock data obtained above, construct a support vector machine model for predicting stock trends in the financial market. The main goal of support vector machines is to find a hyperplane that can effectively separate positive and negative samples in the stock dataset. If ω is the normal vector, then the equation for the hyperplane is:

$$\omega^T x+b=0 \tag{7}$$

Among them, b is the displacement term. The distance from any stock sample point x_i to the hyperplane is denoted as:

$$d=\frac{\left|w^T x_i+b\right|}{\|w\|} \tag{8}$$

If the nearest stock sample point to the hyperplane satisfies $\begin{cases}(w\cdot x_i)+b=1, y=1\\(w\cdot x_i)+b=-1, y=-1\end{cases}$, then the distance from that point to the hyperplane is $\frac{\left|\left(w\cdot x_i+b\right)\right|}{\|w\|}=\frac{1}{\|w\|}$, and the distance between the stock sample points is $\frac{2}{\|w\|}$. Therefore, constructing the optimal hyperplane with the maximum classification interval:

$$\min_{w,b}\frac{1}{2}\|w\|^2, s,t.\ y_i(w^T x_i+b)\geq 1, i=1,2,\cdots,m \tag{9}$$

In practical applications, the original stock sample space may not be able to find a hyperplane to accurately separate the two types of samples. This is usually due to the complex distribution of sample data, or the presence of some noise and outliers, making linear classification difficult. Therefore, samples are mapped from low dimensional space to high-dimensional space, and relaxation variables are introduced to allow for a certain degree of classification fault tolerance, thereby balancing the accuracy and fault tolerance of classification.Let $\phi(x)$ represent the feature vector mapped from x, and the decision function corresponding to the optimal classification hyperplane in the feature space is:

$$f(x)=\operatorname{sgn}(w^T\phi(x)+b) \tag{10}$$

Equation (9) can be rephrased as:

$$\min_{w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{m}\xi_i, s,t.\ y_i(w^T\phi(x_i)+b)\geq 1-\xi_i, \xi_i\geq 0, i=1,2,\cdots,m \tag{11}$$

Among them, C is the penalty factor.

Based on the above process, the construction of the support vector machine model is completed. The characteristics of the financial market stock data obtained in sections 1 and 2 are input into the model to identify outliers in the financial market stock data. Finally, the model outputs a prediction result that represents the future trend of the stock.

ISSN 2706-6827 Vol. 6, Issue 6: 126-131, DOI: 10.25236/IJFS.2024.060620

3. Experimental Design

3.1 Experimental Plan

To verify the practical effectiveness of the financial market stock trend prediction method based on support vector machine proposed in this article, this experiment selected stocks from a certain stock exchange as the research object, and selected multiple stocks with a total of 1000 trading day data from 2020 to 2023, of which 860 data were used as the training set and 140 data were used as the testing set. To improve the reliability of the experimental results, the closing price was used as the input variable for the prediction model. The methods described in references [1] and [2] were simultaneously implemented as control experiments, and the prediction accuracy and mean square error were used as evaluation indicators.If $\hat{y}$ is the final predicted value of the model, the average method error formula is:

$$MSE = \frac{\sum_{t=1}^{N}(y-\hat{y})^2}{N} \tag{12}$$

Among them, N is the sample size of stock data, and y is the actual value of the model input data.

3.2 Experimental Results

The predicted results of the second day closing price using three methods are shown in Figure 1.

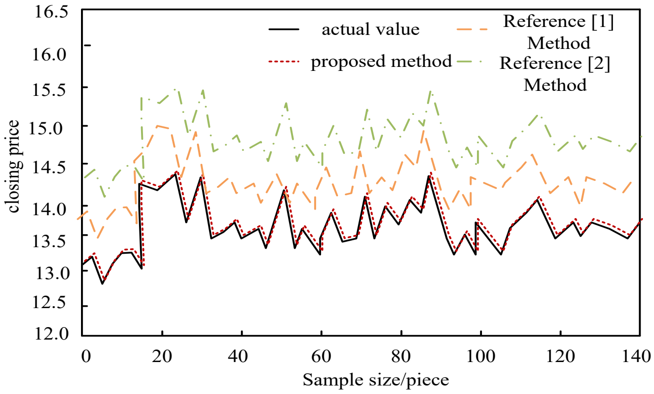

Figure 1: Prediction results of the closing price on the second day using three methods

As shown in Figure 1, the closing price prediction results of our method are consistent with the actual trend, and the predicted values are close to the actual values; The closing price prediction results of reference [1] method deviate from the actual trend, and there is a difference between the predicted value and the actual value; The closing price prediction results of reference [2] are consistent with the actual trend, but the difference between the predicted and actual values is significant. This indicates that the prediction accuracy of the method proposed in this article is higher and has practical application value.

The calculation results of mean square error for stock trend prediction using three methods are shown in Table 1.

Table 1: Calculation results of mean square error for stock trend prediction using three methods

Sample size	Proposed method	Reference [1] Method	Reference [2] Method
20	0.0018	0.0228	0.0201
40	0.0036	0.0294	0.0446
60	0.0912	0.1420	0.0955
80	0.0131	0.1536	0.0895
100	0.1163	0.1687	0.1765

According to Table 1, the maximum mean square error predicted by the method in this article is

ISSN 2706-6827 Vol. 6, Issue 6: 126-131, DOI: 10.25236/IJFS.2024.060620

0.1421, the maximum mean square error predicted by the method in reference [1] is 0.2088, and the maximum mean square error predicted by the method in reference [2] is 0.2143. The mean square error value of the method proposed in this article is consistently lower than that of the methods in reference [1] and [2], indicating that the prediction accuracy of the method proposed in this article is higher.

4. Conclusion

With the increasing prosperity and complexity of global financial markets, the volatility and uncertainty of stock markets are also increasing. In this context, accurately predicting stock trends is crucial for investors. Therefore, a research on financial market stock trend prediction method based on support vector machine is proposed. This study utilized support vector machine technology and crawler technology to obtain financial market stock data, and established an effective stock trend prediction model through preprocessing and feature extraction. The experimental results show that the predicted trend of this method is highly consistent with the actual height, with small mean square error and practical application value. In the future, we will continue to deepen research and explore more advanced technological means to improve prediction accuracy and efficiency, and provide investors with more accurate decision-making assistance.

Acknowledgments

Major Project "Research on Financial Structure Optimization and System Innovation to Promote High-quality Development of Real Economy" (22 & ZD 117); " Shandong Province (18BSJJ03)

References

[1] Yang Longguang. Prediction of intraday stock price trends based on micro market structure [J]. Journal of Southwest Minzu University (Humanities and Social Sciences Edition), 2023,44 (10): 122-130

[2] Sun Ziyu, Ren Ran, Wei Xizhe. Research on Stock Classification and Prediction Based on DTW-TCN [J]. Computer and Modernization, 2023 (08): 31-37

[3] Cui Ting, Huang Feiran. Stock prediction based on sentiment analysis model: a prediction model combining GRU and ALBERT [J]. Dongyue Tribune, 2024, 45 (02): 113-123