# Financial Data Processing Nethod Based on SVM

**ZhuoHan Liu[a], YuHan Li[b], MingLei Han[c,*], YuXin Li[d], KaiYue Wu[e]**

Harbin University of Commerce, Harbin, 150028 Heilongjiang, China

*Corresponding Author

**Keywords:** SVM, Financial data, Processing method

**Abstract:** Under the background of economic and financial globalization, the economic and financial risks faced by countries will continue to intensify, resulting in more violent and changeable foreign exchange fluctuations. The financial market is easily influenced by information, and the Internet, with its real-time, richness and coverage, has gradually replaced the traditional media as the main channel for people to obtain information. How to deal with financial data quickly and effectively has become a hot issue that people urgently need to pay attention to. Therefore, this paper studies the financial data processing method based on SVM, and uses certain data mining methods. In the experiment, we use SVM to mine the financial information on the Internet and the fluctuation of the financial market to find out the quantitative nonlinear relationship between them, thus providing some help for the prediction of the financial market. Using a certain data mining method, we use SVM in the experiment to mine the fluctuation of financial information and financial market on the Internet to find out the nonlinear relationship between them, thus providing some help to the prediction of financial market.

## 1. Introduction

Since the reform and opening up, the number of financial institutions and practitioners in China has increased significantly, the financial scale has expanded significantly, and various banking institutions of different nature have spread all over the country, assuming the functions of absorbing deposits and issuing loans. Insurance institutions have grown from small to large, and securities institutions have grown from scratch, showing a rapid development momentum, and a relatively complete financial system of banking, securities, insurance and other functions has initially formed. In the context of economic and financial globalization, the economic and financial risks faced by countries will continue to intensify, resulting in more volatile foreign exchange fluctuations. The financial market is very vulnerable to the impact of information, and the Internet has gradually replaced the traditional media as the main channel for people to obtain information by virtue of its real-time, rich and coverage characteristics[1]. As the exchange rate is the hub and closely related to each country's economy, regional financial and economic crisis will spread to other regions through financial globalization and exchange rate transmission, which will expand into a larger regional financial and economic crisis, leading to global economic recession and social unrest[2].

The continuous development of the financial industry has played an important role in optimizing the allocation of resources, supporting economic development and maintaining social stability. However, with the rapid development of the financial market, a large number of complex financial data are also emerging. In terms of mining the relationship between financial information and the volatility of the financial market, the current research is mainly based on the relationship between the amount of news information and volatility, that is, the number of news on the Internet is the main research object, while the research on the specific content and commendation of information has not yet been carried out. How to process financial data quickly and effectively has become a hot issue that people urgently need to pay attention to[3]. In the financial market, Internet financial information plays an important role. Compared with traditional media, financial information obtained from the Internet will be the most real-time and comprehensive. The global financial market is dynamic, complex and unpredictable, and the exchange rate is the bridge and link

between the various elements of the economic system and between the system elements. It has evolved into a complex system, which increases the difficulty of exchange rate prediction in the foreign exchange market[4-5]. Therefore, it is necessary to adopt a new exploration mode for exchange rate prediction, such as the prediction method using nonlinear approximation. Therefore, the mining of Internet financial information is of great practical value.

Based on the above analysis, this paper applies SVM technology to financial data processing methods to model financial market data. As we all know, the reason why the financial market is difficult to be grasped by people is because of its randomness. Through the analysis and processing of financial data by SVM method, it is verified that the sum of squares of the residuals obtained is larger than the least squares regression, and the tolerance to sick data is also stronger than the least squares method, which is more suitable for analyzing financial data in practical work.

## 2. Reasons for Using SVM to Process Financial Data

### 2.1 The Needs of the Financial Industry's Own Development

How to accurately draw effective conclusions from the analysis and processing of financial data is the focus of people's attention, because the analysis of financial data can bring a lot of valuable information to financial enterprises or financial investors. The data stream mining algorithm of integrated SVM proposed in this paper is an integrated classifier composed of multiple SVM base classifiers. Among them, the data stream detection framework is based on CDSMM algorithm; The weight of the base classifier is obtained by calculating the classification accuracy of the current data block; The concept drift is detected by hypothesis testing, and the base classifier is dynamically adjusted to adapt to the concept change[6]. This paper proposes an event detection framework, which can be used to detect potential corporate financial events from Weibo. The framework is divided into several parts, such as predefined corporate financial event types, trigger dictionary construction, document vectorization and event type classification detection, as shown in Figure 1.

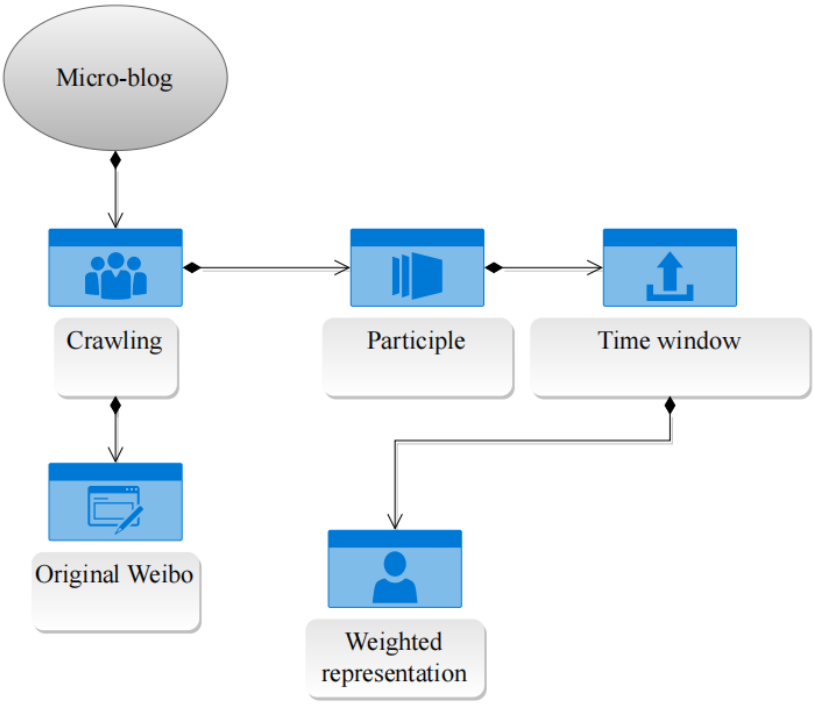


Fig.1 Event Detection Framework

The analysis of financial data is not only conducive to financial enterprises or institutions to understand their current operation, obtain valuable business information, and prevent financial risks, but also conducive to financial investors to understand the nature of the financial market, so as to better conduct financial investment. SVM is a supervised machine learning method that can be used for data analysis and pattern recognition when dealing with financial data. It adopts the structural

risk minimization criterion, which can reduce the upper bound of the model generalization error while minimizing the sample point error, and has good generalization ability. Support vector machine and its regression extension model support vector regression have been widely used in simulating the stock market[7-8]. There are two main types of information related to finance on the Internet. One is the news of the financial sector of major websites, and the other is the replies to the news and the posts in various forums. These two types of information have different characteristics.

The news in the financial sector of the website is not very obvious, but its content is true and there is something in it. The new training set is reconstructed by using the error classification and the replaced base classifier training set to train the new base classifier and replace the lower accuracy base classifier to improve the overall classification accuracy. Using certain data mining methods, in the experiment, we use SVM to mine the financial information on the Internet and the volatility of the financial market to find out the non-linear relationship between the two in quantity, thus providing certain help for the prediction of the financial market.

### 2.2 Limitations of Traditional Algorithms for Financial Data

In the financial field, a large number of financial data often exist in the form of time series, such as stock price data and futures price data. The financial derivatives market is influenced by various factors, which makes a lot of noise in financial data, especially in financial time series data, resulting in the characteristics of time series instability, which greatly affects its analysis and processing[9]. In the traditional regression analysis method, the design and selection of linear regression model is closely related to the analysis of financial data. If we can predict the volatility of financial markets, it is of great significance for investors to make decisions. In view of the above situation, denoising for financial time series is particularly important. However, financial time series itself has the characteristics of non-stationarity, nonlinearity and high signal-to-noise ratio, and there are often many defects in traditional denoising methods. We know that the volatility of financial market can be expressed as a time series, and its current value is closely related to its previous values and external variables. Based on the existing literature and the characteristics of information in corporate Weibo, the types of corporate financial events are defined as seven categories, and the specific event types and descriptions are shown in Table 1.

Table 1 Types of Corporate Financial Events

| Serial number | Event type | Event elaboration |
|---|---|---|
| 1 | Policy events | Macro-policy influences such as government. |
| 2 | Profitability | Stock price changes, profits and liabilities |
| 3 | Social events | Clarify warning category |
| 4 | Strategic activities | Strategic layout class |

The traditional algorithm of financial data is often difficult to solve the nonlinear problem in the financial time series. Later, with the development of nonlinear science, the neural network method was proposed for the prediction research of financial time series data. However, the neural network method uses the empirical risk minimization criterion to minimize the sample error in the training set, which will lead to the over-fitting problem, and the generalization ability of the model is poor[10]. In the financial time series, the occurrence of singular values is usually caused by unexpected events, which often indicates the end of the trend or the strengthening of the trend. Therefore, singular value is not only noise, but also the focus of research.

The traditional algorithm of financial data fully embodies the value of singular value. Due to the particularity of financial time series prediction, the traditional denoising method is not applicable. Each company will predict the volatility of the next window, record whether the predicted volatility will rise or fall compared with the previous window, and observe whether the real volatility will rise or fall compared with the previous window. If the two conclusions are consistent, the company's prediction of the trend is correct. In the actual operation analysis, it is also very rare that the explanatory variables are completely uncorrelated, and most variables have a certain degree of collinearity.

## 3. Experimental Results and Discussion

With the development of the stock market, more and more investors find that there are a lot of investment opportunities in the stock market. Financial news is the first-hand news that the public investors can get, and investors will make investment decisions based on the news. Therefore, it is of great practical significance for investors to find a way to predict the impact of news on the stock market. We use the regression function of SVM to explore the relationship between the value of financial news and the stock price volatility of financial markets. Both public opinion value and stock price volatility are divided into different time windows. Because of the large amount of financial data and its complexity, the actual data sequence is bound to be filled with a lot of noise, and a large number of assumptions that the ordinary least square method must meet are often difficult to realize in the actual financial data processing, so this paper uses SVM method to process financial data.

As a biased estimation method specially used for collinearity data analysis, SVM is more in line with the actual regression process, although it gives up the unbiased least square and partial accuracy, and the regression effect is slightly worse. In order to further verify the algorithm, we statistically found that the prediction accuracy of stock price volatility reached 64.25% and the square value of complex correlation coefficient reached 75.48%. Figure 2 describes the effect of MDT in stock price prediction of one of the companies. Fig. 2 describes the change of VTFA value in the experimental results.

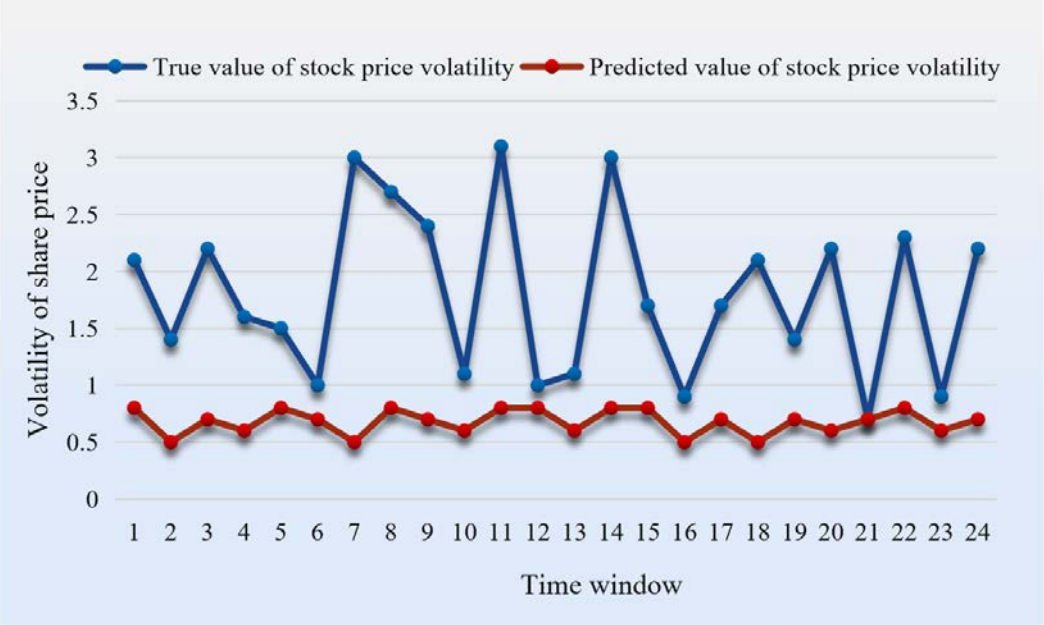


Fig.2 Real Value and Predicted Value of Stock Price Volatility in Each Time Window

The data is divided into training set, test set and verification set by SVM, and the prediction effect of different model verification sets is compared. The result shows that the prediction effect of support vector machine model is the best. Most of the above documents compare the prediction effect of support vector machine method with other neural network methods. The empirical results show that support vector machine has more advantages. Support vector machine model has been favored by many scholars because of its many advantages.

## 4. Conclusions

This paper analyzes and processes financial data by SVM method, effectively avoids the influence of multicollinearity on financial data, and overcomes many problems, such as model distortion and the result running counter to the reality caused by the traditional least square method for regression analysis in the past. The regression analysis is carried out by using training sets, test sets and model comparison. The results show that the model between the two is constructed by

using the text tendentiousness analysis in natural language processing and combining the characteristics of the financial market itself. Through the research on the data in the stock market, it is proved that there is a relatively close relationship between the two. The SVM method is far more conducive to people's application in actual analysis and research than the traditional method that must meet many assumptions. Because the transaction data is divided by windows, the calculation of volatility is based on these divided windows. In the experiment, we mainly study the volatility of stock prices. Moreover, the sum of squares of the residuals obtained by SVM is larger than that of the least squares regression, and its tolerance to sick data is far stronger than that of the least squares method, which is more suitable for the actual financial data analysis. Using certain data mining methods, in the experiment, we use SVM to mine the financial information on the Internet and the volatility of the financial market to find out the non-linear relationship between the two in quantity, thus providing certain help for the prediction of the financial market.

## References


[1] Kang Q. Information supervisory model for financial risk prevention and control based on twin-SVM[J]. International Journal of Information Technology and Management, 2022, 21(17):62-77.

[2] Wang Y. Empirical analysis on the financial governance risk based on C-SVM data mining[J]. Basic & clinical pharmacology & toxicology. 2019, 49(21):124-198.

[3] Mishra K A. Indian stock market prediction using artificial neural networks on tick data[J]. Financial Innovation (English), 2020, 5(1):267-278.

[4] Liang M. Optimization of Quantitative Financial Data Analysis System Based on Deep Learning[J]. Complexity, 2021, 21(1):1-11.

[5] Harikrishnan R, Jebakumar R, Kumar S G, et al. Insurance Customer Authentication Using SVM and Financial Time Series Analysis for Mobile Applications[J]. Webology, 2020, 17(2):945-956.

[6] Lu Y, Li J, Yang Z, et al. Data mining and social networks processing method based on support vector machine and k-nearest neighbor[J]. Journal of Computational Methods in Sciences and Engineering, 2020, 55(1):1-13.

[7] Lu S, Caishigang. Study on papermaking enterprise financial accounting data analysis model based on decision tree and optimized SVM algorithm[J]. Paper Asia, 2022, 34(6):57-62.

[8] Luo C, Jiang Z, Zheng Y. A novel reconstructed training-set SVM with roulette cooperative coevolution for financial time series classification[J]. Expert Systems with Application, 2019, 123(58):283-298.

[9] Peng W, Xun H. Intelligent Early Warning for Multifractal Risk of Financial Market Based on Twin-SVM[J]. Statistical Research, 2020, 45(11):36-52.

[10] Jie S A, Hfbc D, Yz A, et al. Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods[J]. Information Sciences, 2021, 56(11):17-26.