A Hybrid Forecasting Method for Anticipating Stock Market Trends via a Soft-Thresholding De-noise Model and Support Vector Machine (SVM)

Lixuan Zhang, Chang Li, Lee Chen, Don Chen, Zheng Xiang, Bing Pan
Faculty of Computer Science and Information System, Universiti Teknologi MARA (UiTM), Malaysia

ABSTRACT

Stock market time series are inherently noisy. Although support vector machine has the noise-tolerant property, the noised data still affect the accuracy of classification. Compared with other studies only classify the movements of stock market into up-trend and down-trend which does not concern the noised data, this study uses wavelet soft-threshold de-noising model to classify the noised data into stochastic trend. In the experiment, we remove the stochastic trend data from the SSE Composite Index and get de-noised training data for SVM. Then we use the de-noised data to train SVM and to forecast the testing data. The hit ratio is 60.12%. Comparing with 54.25% hit ratio that is forecasted by noisy training data SVM, we enhance the forecasting performance.

KEYWORDS: Soft-thresholding, De-noise, SVM, Stock market, Financial time series, discrimination, adverse selection

1.0 INTRODUCTION

Stock market trend forecasting gives information on the corresponding risk of the investments and it also will influence the trading behavior. Stock market time series are inherent noisy, non-stationary, and deterministically chaotic. It has been shown that data extrapolated from stock markets are almost corrupted by noise and it appears that no useful information can be extracted from such data. Modeling such noisy and non-stationary time series is expected to be a challenging task. In recent years, numerous studies have demonstrated that neural networks are a more effective method in describing the dynamics of non-stationary time series due to their unique non-parametric, non-assumable, noise-tolerant and adaptive properties. However, neural networks still have several limitations. SVM originates from Vapnik's statistical learning theory. Unlike most of the traditional methods which implement the empirical risk minimization principal, SVM implements the structural risk minimization principal which seeks to minimize an upper bound of the generalization error rather than minimize the training error. Many applications of the SVM to forecast financial time series have been reported. Cao and Tay used the theory of SVM in regression to forecast the S&P 500 Daily Index in the Chicago Mercantile. They measured the degree of accuracy and the acceptability of certain forecasts by the estimates' deviations from the observed values. Kim forecasted the direction of the change in daily Korea composite stock price index (KOSPI) with the theory of SVM in classification. The best prediction performance for the holdout data is 57.83% [1-13]. Tony Van Gestel designed the LS- SVM time series model in the evidence framework to predict the daily closing price return of the German DAX30 index (Deutscher Aktien Index). Many of the previous studies have compared the performance of SVM with BP neural network, case-based reasoning (CBR) and so on. All of the results prove that the general performance for SVM is better than the traditional methods. Many studies had selected optimum parameters of SVM when they would enhance the forecasting performance. This study proposes dealing with the noise of the stock market in order to enhance the forecasting performance of SVM. According to the wavelet de-noising model of soft-thresholding, we classify the stock market short- term trend into up-trend, stochastic trend and down-trend. We remove the stochastic trend data from the original Index data and take the rest data which belong to the up- trend and down-trend as the training data. Then we use the trained SVM to forecast the stock market trends. Stock market time series are inherently noisy. Although support vector machine has the noise-tolerant property, the noised data still affect the accuracy of classification. Compared with other studies only classify the movements of stock market into up-trend and down-trend which does not concern the noised data, this study uses wavelet soft-threshold de-noising model to classify the noised data into stochastic trend. In the experiment, we remove the stochastic trend data from the SSE Composite Index and get de-noised training data for SVM. Then we use the de-noised data to train SVM and to forecast the testing data. The hit ratio is 60.12%. Comparing with 54.25% hit ratio that is forecasted by noisy training data SVM, we enhance the forecasting performance [14-28].

2.0 LITERATURE REVIEW

Forecasting trends in the stock market provides valuable information about investment risks and affects trading behavior. However, the inherent noise, non-stationarity, and deterministic chaos of stock market time series pose challenges to accurately modeling them. Although previous studies have shown that neural networks are effective in describing the dynamics of non-stationary time series, they still have limitations. To address these challenges, support vector machines (SVMs) have emerged as a promising method due to their noise-tolerant and adaptive properties. Several studies have applied SVMs to forecast financial time series with promising results. For example, Cao and Tay used SVM regression to forecast the S&P 500 Daily Index, while Kim forecasted the direction of change in the Korea composite stock price index (KOSPI) using SVM classification. Tony Van Gestel designed the LS-SVM time series model to predict the daily closing price return of the German DAX30 index. Compared to traditional methods such as BP neural networks and case-based reasoning, SVMs generally perform better in forecasting stock market trends. To further enhance the forecasting performance of SVM, this study proposes using the wavelet de-noising model of soft-thresholding to classify the stock market short-term trend into up-trend, stochastic trend, and down-trend. By removing the stochastic trend data from the original index data and using the remaining data belonging to the up-trend and down-trend as training data, SVM is trained to forecast the stock market trends. Although SVM has noise-tolerant properties, noisy data can still affect its accuracy. This study addresses this issue by using the wavelet soft-threshold de-noising model to classify the noisy data into stochastic trend. By removing the stochastic trend data from the SSE Composite Index and using the de-noised data to train SVM, the hit ratio of the forecasting performance is increased from 54.25% to 60.12%. In summary, while stock market time series are inherently noisy and challenging to model accurately, SVMs have shown promise in forecasting financial time series. This study proposes a novel approach using wavelet de-noising to enhance the forecasting performance of SVM, and the results demonstrate improved accuracy compared to using noisy data alone [29-34].

Supposing f (i) is the original signal, the polluted image signal is s(i) , and noise signal is e(i) . Then, the model of the noised imaged is where σ denotes a noise level and e(i) is a Gauss white noise.

$$s(i) = f(i) + \sigma e(i)$$

Figure 1 is the block diagram of signal de-noising with wavelet transformation. The three blocks in figure 1 represent the three basic steps of de-nosing respectively.

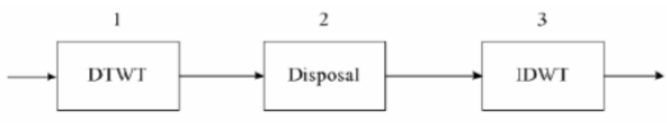

Fig. 1. The block diagram of wavelet de-noising

Wavelet decomposition is the first step: selecting wavelet and decomposition Level, and calculating the coefficients of the transformation from s(i) to the layer J. The second step which is the threshold manipulation step: selecting the threshold and dealing with the coefficients according to the equation as follows [35-42]:

$$d'_{j,k} = \begin{cases} d_{j,k} & |d_{j,k}| \geq t \\ 0 & |d_{j,k}| < t \end{cases}$$

3.0 RESEARCH METHODOLOGY

In our empirical analysis, we set out to examine the five-day moving trend of the Shanghai Stock Exchange (SSE) Composite Index. The original data points cover the time period from 28/04/1997 up to 12/09/2006 which is 2261 data. We select 1920 data from the 2261 data as training data and take the rest 341 data as testing data. As shown in Fig. 2, the 1920 data are illustrated.

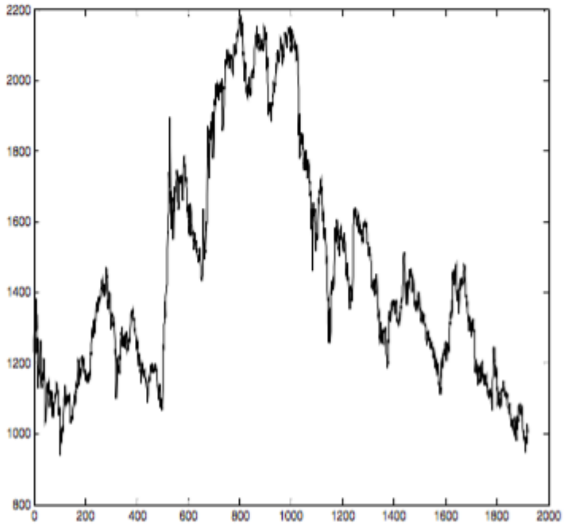

Fig. 2. SSE composite index

Figure 3 illustrate the 1920 smooth SSE Composite Index data which had been de- noised by soft-thresholding.

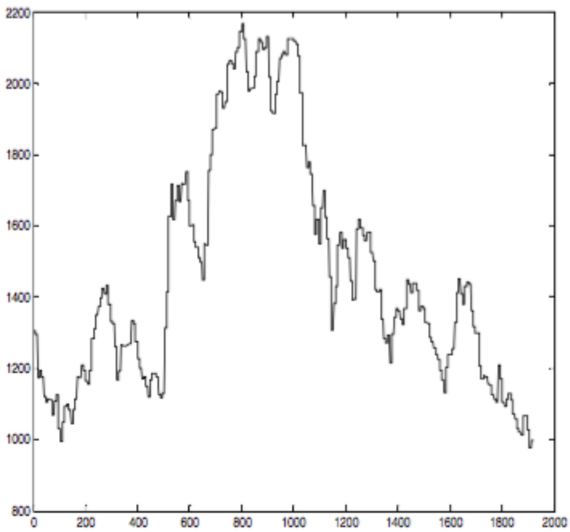

Fig. 3. Smoothed SSE composite index

Based on the soft-threshold which determined by the wavelet de-noise Model, we classify the stock market into up-trend, stochastic trend and down-trend which have 357, 1171,392 data respectively. Figure 4 illustrates the details residual of SSE Composite Index and soft-threshold [13]. We define the noisy data whose five-day SSE Composite Index change value between the up and lower soft-threshold as the stochastic trend. In consequence, the value of data above the upper soft-threshold is defined as the up-trend and the value of data below the lower soft-threshold are defined as the down-trend. Then, we remove the 1171 stochastic trend data from the original SSE Composite Index and take the rest 749 data which belong to the up-trend and down-trend as the training data.

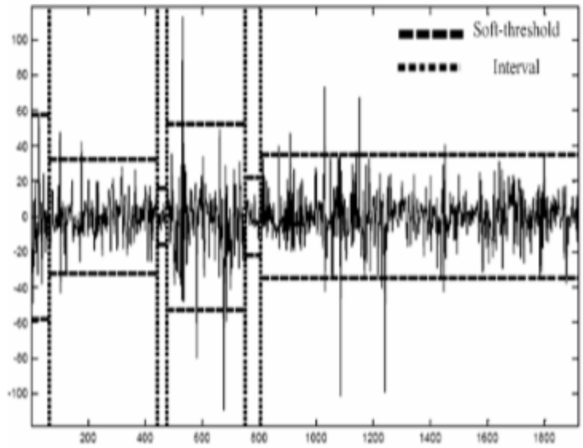

Fig. 4. Residual of SSE composite index and soft-thresholds

4.0 RESULT

In this study, we select the data points covering the time period from 28/04/1997 to 12/09/2006. There are 2261 data points of Shanghai Stock Exchange (SSE) Composite Index. We use the first 1920 data of the 2261 original data as training set and take the rest 341 data as testing data. We use wavelet soft-threshold de-noise Model to de-noise the 1920 training data. As illustrated in Fig.2- Fig.4, we can see the detail process of de-noise. As a result of the soft-threshold de-noising, we get 1171 noised data which are classified as the stochastic trend data. We remove the 1171stochastic trend data from the original SSE Composite Index and take the rest 749 data which belong to the up-trend and down- trend as the training set for SVM. The Gaussian radial basis function is used as the kernel function of the SVM. We conduct the experiment with respect to various kernel parameters and the upper bound C . The range for kernel parameter is between 1 and 100 and the range for C is between 1 and 100. We use the 749 data mentioned above to train the SVM and apply the SVM to classify the 341 test data. For comparison, we also use the 1920 data mentioned above to train SVM and employ SVM to classify the same 341 test data. The forecasting results of two methods are shown in table 2. The results in table 2 show that best hit ratio of the de-noise SVM is 60.12% which are better than the best hit ratio 54.25% of noisy SVM.

Table 2. Best forecasting results of two methods

SVM	Testing/training data	C	σ	Hit ratio
De-noise	Testing data	90	20	60.12%
	Training data	30	10	99.87%
Noisy	Testing data	10	100	54.25%
	Training data	50	10	99.95%

The study utilized 2261 data points of the Shanghai Stock Exchange (SSE) Composite Index from April 28, 1997 to September 12, 2006. Out of the 2261 data points, the first 1920 were selected as the training set while the remaining 341 were chosen as the testing data. The wavelet soft-threshold de-noise Model was used to de-noise the training data, resulting in 1171 data points classified as stochastic trend data. These 1171 data points were removed from the original SSE Composite Index, and the remaining 749 data points belonging to the up-trend and down-trend were used as the training set for the support vector machine (SVM). The SVM employed the Gaussian radial basis function as the kernel function, and various kernel parameters and the upper bound C were tested in the experiment. The SVM was trained using the 749 data points and applied to classify the 341 test data points. In comparison, the SVM was also trained using the first 1920 data points and applied to classify the same 341 test data points. The results showed that the best hit ratio of the de-noise SVM was 60.12%, which was higher than the best hit ratio of 54.25% obtained from the noisy SVM. The detailed process of de-noising can be observed in Figure 2-4.

6.0 CONCLUSION

Many applications of SVM to forecast financial time series have been reported. Most of the researches only paid attention to select optimum parameters of SVM when they want to enhance forecasting performance. However, as SVM has the noise-tolerant property, little study discusses about preprocessing the noisy input data to enhance the forecasting performance. In this study, on the condition of selecting optimum parameters of SVM, we use soft-thresholding to de-noise the training data and get a better optimal hyperplane than the optimal hyperplane learned with noisy training data. Consequently, compared with the 54.25% hit ratio of the noisy SVM, the forecasting performance of the de-noised SVM is 60.12% hit ratio. The hit ratio is also better than Kim's 57.83%, which is the best prediction performance in forecasting the trend of Korea composite stock price index (KOSPI) with SVM [6]. This study proves that de-noising the training data can effectively enhance the forecasting performance of SVM. The paper explores the application of support vector machines (SVM) to forecast financial time series. While previous research has focused on selecting optimum parameters for SVM to improve forecasting performance, little attention has been paid to pre-processing noisy input data to enhance performance. In this study, the authors propose using soft-thresholding to de-noise the training data while also selecting optimum parameters for SVM. By doing so, they were able to obtain an optimal hyperplane that performed better than the hyperplane learned with noisy training data. Specifically, the authors found that the forecasting performance of the de-noised SVM, with a hit ratio of 60.12%, was better than that of the noisy SVM, which had a hit ratio of 54.25%. This improvement in performance was also better than Kim's previously reported hit ratio of 57.83% in forecasting the trend of Korea composite stock price index (KOSPI) with SVM. Overall, the study demonstrates the effectiveness of de-noising training data for enhancing the forecasting performance of SVM in financial time series analysis.

REFERENCES

[1] Richardson, Alan J., Michael Welker, and Ian R. Hutchinson. "Managing capital market reactions to corporate social resposibility." International Journal of Management Reviews 1.1 (1999): 17-43.
[2] Birjandi, Alireza Komeili, Sanaz Dehmolaee, Reza Sheikh, and Shib Sankar Sana. "Analysis and classification of companies on tehran stock exchange with incomplete information." RAIRO-Operations Research 55 (2021): S2709-S2726.
[3] Dionne, Georges, Christian Gouriéroux, and Charles Vanasse. "Testing for evidence of adverse selection in the automobile insurance market: A comment." Journal of Political Economy 109.2 (2001): 444-453.
[4] Nazari Enjedani, Somayeh, and Mandar Khanal. "Development of a Turning Movement Estimator Using CV Data." Future Transportation 3, no. 1 (2023): 349-367.
[5] Edelman, Benjamin, Michael Luca, and Dan Svirsky. "Racial discrimination in the sharing economy: Evidence from a field experiment." American economic journal: applied economics 9.2 (2017): 1-22.
[6] Saadat, MohammadReza, and Benedek Nagy. "Cellular automata approach to mathematical morphology in the triangular grid." Acta Polytechnica Hungarica 15, no. 6 (2018): 45-62.
[7] Hinz, Oliver, II-Horn Hann, and Martin Spann. "Price discrimination in e-commerce? An examination of dynamic pricing in name-your-own price markets." Mis quarterly (2011): 81-98.
[8] Tirole, Jean. "Overcoming adverse selection: How public intervention can restore market functioning." American economic review 102.1 (2012): 29-59.
[9] Amini, Mahyar, and Zavareh Bozorgasl. "A Game Theory Method to Cyber-Threat Information Sharing in Cloud Computing Technology ." *International Journal of Computer Science and Engineering Research* 11.4 (2023): 549-560.
[10] Nazari Enjedani, Somayeh, and Mahyar Amini. "The role of traffic impact effect on transportation planning and sustainable traffic management in metropolitan regions ." International Journal of Smart City Planning Research 12.9 (2023): 688-700
[11] Jahanbakhsh Javidi, Negar, and Mahyar Amini. "Evaluating the effect of supply chain management practice on implementation of halal agroindustry and competitive advantage for small and medium enterprises ." International Journal of Computer Science and Information Technology 15.6 (2023): 8997-9008
[12] Amini, Mahyar, and Negar Jahanbakhsh Javidi. "A Multi-Perspective Framework Established on Diffusion

of Innovation (DOI) Theory and Technology, Organization and Environment (TOE) Framework Toward Supply Chain Management System Based on Cloud Computing Technology for Small and Medium Enterprises ." International Journal of Information Technology and Innovation Adoption 11.8 (2023): 1217-1234
[13] Amini, Mahyar and Ali Rahmani. "Agricultural databases evaluation with machine learning procedure." Australian Journal of Engineering and Applied Science 8.6 (2023): 39-50
[14] Amini, Mahyar, and Ali Rahmani. "Machine learning process evaluating damage classification of composites." International Journal of Science and Advanced Technology 9.12 (2023): 240-250
[15] Amini, Mahyar, Koosha Sharifani, and Ali Rahmani. "Machine Learning Model Towards Evaluating Data gathering methods in Manufacturing and Mechanical Engineering." International Journal of Applied Science and Engineering Research 15.4 (2023): 349-362.
[16] Sharifani, Koosha and Amini, Mahyar and Akbari, Yaser and Aghajanzadeh Godarzi, Javad. "Operating Machine Learning across Natural Language Processing Techniques for Improvement of Fabricated News Model." International Journal of Science and Information System Research 12.9 (2022): 20-44.
[17] Amini, Mahyar, et al. "MAHAMGOSTAR.COM AS A CASE STUDY FOR ADOPTION OF LARAVEL FRAMEWORK AS THE BEST PROGRAMMING TOOLS FOR PHP BASED WEB DEVELOPMENT FOR SMALL AND MEDIUM ENTERPRISES." Journal of Innovation & Knowledge, ISSN (2021): 100-110.
[18] Amini, Mahyar, and Aryati Bakri. "Cloud computing adoption by SMEs in the Malaysia: A multi-perspective framework based on DOI theory and TOE framework." Journal of Information Technology & Information Systems Research (JITISR) 9.2 (2015): 121-135.
[19] Amini, Mahyar, and Nazli Sadat Safavi. "A Dynamic SLA Aware Heuristic Solution For IaaS Cloud Placement Problem Without Migration." International Journal of Computer Science and Information Technologies 6.11 (2014): 25-30.
[20] Amini, Mahyar. "The factors that influence on adoption of cloud computing for small and medium enterprises." (2014).
[21] Amini, Mahyar, et al. "Development of an instrument for assessing the impact of environmental context on adoption of cloud computing for small and medium enterprises." Australian Journal of Basic and Applied Sciences (AJBAS) 8.10 (2014): 129-135.
[22] Amini, Mahyar, et al. "The role of top manager behaviours on adoption of cloud computing for small and medium enterprises." Australian Journal of Basic and Applied Sciences (AJBAS) 8.1 (2014): 490-498.
[23] Amini, Mahyar, and Nazli Sadat Safavi. "A Dynamic SLA Aware Solution For IaaS Cloud Placement Problem Using Simulated Annealing." International Journal of Computer Science and Information Technologies 6.11 (2014): 52-57.
[24] Sadat Safavi, Nazli, Nor Hidayati Zakaria, and Mahyar Amini. "The risk analysis of system selection and business process re-engineering towards the success of enterprise resource planning project for small and medium enterprise." World Applied Sciences Journal (WASJ) 31.9 (2014): 1669-1676.
[25] Sadat Safavi, Nazli, Mahyar Amini, and Seyyed AmirAli Javadinia. "The determinant of adoption of enterprise resource planning for small and medium enterprises in Iran." International Journal of Advanced Research in IT and Engineering (IJARIE) 3.1 (2014): 1-8.
[26] Sadat Safavi, Nazli, et al. "An effective model for evaluating organizational risk and cost in ERP implementation by SME." IOSR Journal of Business and Management (IOSR-JBM) 10.6 (2013): 70-75.
[27] Safavi, Nazli Sadat, et al. "An effective model for evaluating organizational risk and cost in ERP implementation by SME." IOSR Journal of Business and Management (IOSR-JBM) 10.6 (2013): 61-66.
[28] Amini, Mahyar, and Nazli Sadat Safavi. "Critical success factors for ERP implementation." International Journal of Information Technology & Information Systems 5.15 (2013): 1-23.
[29] Amini, Mahyar, et al. "Agricultural development in IRAN base on cloud computing theory." International Journal of Engineering Research & Technology (IJERT) 2.6 (2013): 796-801.
[30] Amini, Mahyar, et al. "Types of cloud computing (public and private) that transform the organization more effectively." International Journal of Engineering Research & Technology (IJERT) 2.5 (2013): 1263-1269.
[31] Amini, Mahyar, and Nazli Sadat Safavi. "Cloud Computing Transform the Way of IT Delivers Services to the Organizations." International Journal of Innovation & Management Science Research 1.61 (2013): 1-5.
[32] Abdollahzadegan, A., Che Hussin, A. R., Moshfegh Gohary, M., & Amini, M. (2013). The organizational critical success factors for adopting cloud computing in SMEs. Journal of Information Systems Research and Innovation (JISRI), 4(1), 67-74.
[33] Khoshraftar, Alireza, et al. "Improving The CRM System In Healthcare Organization." International Journal of Computer Engineering & Sciences (IJCES) 1.2 (2011): 28-35.
[34] Amini, Mahyar and Ali Rahmani. "How Strategic Agility Affects the Competitive Capabilities of Private Banks." International Journal of Basic and Applied Sciences 10.01 (2023): 8397-8406.
[35] Amini, Mahyar and Ali Rahmani. "Achieving Financial Success by Pursuing Environmental and Social Goals: A Comprehensive Literature Review and Research Agenda for Sustainable Investment." World Information Technology and Engineering Journal 10.04 (2023): 1286-1293.
[36] Yun, Chidi, et al. "The Significance of Information Systems in Enhancing Strategic Agility within Supply Chain Context: A Case Study of Telecommunications Industry." International Journal of Engineering and

Applied Sciences 11.02 (2023): 67-74.

[37] Motalo, Kubura, et al. "The Competitive Edge of Strategic Agility in Airlines through Embracing Novel Technologies ."International Journal of Technology and Scientific Research 12.04 (2023): 789-796.

[38] Olutola, Tomiloba, et al. "The Intermediate Function of Sustainable Dynamic Capabilities in the Association between Social Customer Relationship Management and Sustainable Competitive Advantage." Asian Journal of Basic and Applied Sciences 10.04 (2023): 97-104.

[39] Balen, John, et al. "Essential Determinants for Assessing the Strategic Agility Framework in Small and Medium-sized Enterprises (SMEs) ." European Journal of Scientific and Applied Sciences 10.04 (2023): 2124-2129.

[40] Chen, Lee, et al. "Categorization of Surgical Complications using Computer Vision Technique ." American-Eurasian Journal of Scientific Research 11.04 (2023): 5994-6001.

[41] Li, Chang, et al. "Analysis and Categorization of Stock Price Factors via a Novel Framework based on Computer Science Technology ." World Journal of Technology and Scientific Research 12.03 (2023): 361-366.

[42] Xiang, Zheng, et al. "Efficiency of Risk Classification and Insurance Market Regulation through the Integration of Advanced Technology ." World Engineering and Applied Sciences Journal 14.02 (2023): 3986-3992.