

Application of Machine Learning Technology in Financial Market

Mengtian Wang, Pan Xu, and Jun Liu*

School of Management Science and Engineering, Nanjing University of Finance and Economics, No.3
Wenyuan Road, Nanjing, China

Keywords: Machine learning; Quantitative trading; Financial market

Abstract: With the development of economy, people are paying more attention to the financial investment. The prediction of the Stock/Future prices could have significant influence on the financial market. But it is hard to predict the prices for they are time series, dynamic and non-linear. To have valid data, machine learning methods are adopted, because they are good at dealing with those problems and have good performance on the financial market. Therefore, this paper presents a detailed review of 30 research papers about the application of machine learning technology in financial market suggested methodologies like Support Vector Machine (SVM), Artificial Neural Networks (ANN), Decision Tree (Random Forest) and so on. We use publication years, datasets, methods and research direction to analyze the works. Most papers reviewed here are published in recent years, and SVM and ANN are the most common methods used in the quantitative trading. According to the research direction, Stock Selection and Timing are the main targets. Based on lots of research results, the paper finds that many disadvantages exist in the current quantitative trading. From this survey, it can be concluded that machine learning technology used in financial market is not very efficient. And a better machine learning method should be found and more factors should be considered for quantitative trading as well.

1. Introduction

As the economy develops, it has witnessed a rapid increase of residents' wealth and distributable assets. A problem they are faced with is how to better allocate their assets. The forming of a large-scale and multi-level capital market pattern and the deepening of structural reform on the financial supply side in China bring vigor to the Chinese capital market, which is full of realistic driving forces now. Meanwhile, the real economic capacity of financial services has been correspondingly enhanced. Asset allocation, an important channel for financial services to the real economy, could inject new economic strength into the real economy.

In 2017, Artificial Intelligence (AI) was first written into the government report and the development of AI was raised to the national development strategy. Artificial Intelligence, Block Chain and 5G technology have led to the rapid development of financial technology. As an important means of financial technology, quantitative trading has greatly improved the income of asset allocation and reduced risks. It is gradually becoming an important part of the domestic financial market. Some financial institutions, such as Brokers, Banks, Trusts and Insurance are also paying more and more attention to quantitative trading. In the big data environment, artificial intelligence technology, represented by Machine Learning and Deep-learning, plays an increasingly important role in the financial market and has gradually attracted the interest and extensive attention of the financial industry and academia.

In recent decades, quantitative investment has become a hot spot in the development of capital markets in Europe and the United States. Harry Markowitz [1] (1952) proposed a mean variance model based on the mathematical statistics theory. In his model, the fluctuation of expected return rate of assets is adopted to represent financial market risks and this principle is applied to securities investment portfolio, opening a new way for quantitative trading. The market scale of quantitative investment has once reached 70% of the U.S. investment market. It has become a new method for investors. In recent years, the trading strategy of quantitative investment has performed in a stable way. With the market scale and share expanding, it has become one of the three mainstream

methods together with Fundamental analysis and Technical index analysis. The concept of quantitative investment originated in the United States. Investors used computer technology, statistics and mathematical knowledge to build models to realize investment concepts and strategies. They trained and summarized market rules in a large number of historical transaction data to obtain excess market returns. The development of Chinese financial market is not very mature for the irrational pricing mechanism of stock price, the seriousness of the market speculation atmosphere and weak effectiveness of the market in China. Therefore, there is huge potential for finding excess returns and quantitative trading is feasible in Chinese market.

The main types of quantitative investment are trend trading, market neutral strategy and high frequency trading. Among them, market neutral investment measurement is currently the most mainstream and largest strategy in the market, including quantitative statistical arbitrage, quantitative hedging, quantitative Stock Selection and quantitative Timing strategy. Financial data, as a time series, are dynamic, non-stationary, non-linear and unpredictable. Machine learning enjoys great advantages in non-linear computation. It covers probability theory knowledge, statistics knowledge and mathematical algorithm knowledge. Computer is used as a tool to simulate human learning methods, and the existing content is divided into knowledge structures to effectively improve learning efficiency. Machine learning has experienced a tortuous development since the 1950s, the era of big data which makes it attracted much attention. It can be divided into supervised learning, unsupervised learning and semi-supervised learning. The main difference among the three learning methods lies in the difference of data sample labels. Supervised learning requires labels of samples, such as the rise and fall of stocks price, while unsupervised learning requires no label value and Semi-supervised learning some labels. The literature on the application of machine learning technology in financial markets is not very rich. Unlike some papers review the machine learning methods used in stock price predicting, this paper is to introduce the main methods of machine learning used in all financial market, classify the dozens of papers read according to the types of quantitative investments, briefly introduce and summarize them, and finally point out the shortcomings of current machine learning technology applied in financial market so as to give suggestions for future research directions.

Literature review is a method of investigating research topics. The next section will offer a brief introduction of the main machine learning technologies designed in the selected articles. Then, it will analyze, classify and summarize the selected papers in the third section. Finally, it will review all the works and draw a conclusion.

2. Machine Learning Technology

Machine Learning is multi-disciplinary and cross-disciplinary. After learning a large amount of historical data, the model itself can have good self-learning ability, thus achieving Artificial Intelligence. Machine learning usually has two parts. In the first part, appropriate models and corresponding super variables are selected to learn the pre-segmented training set, and the models are verified and optimized. In the second part, the optimized model is applied to the untested data, and the prediction ability of the model is judged according to the corresponding indexes. The machine learning techniques used in the literature selected in this paper include Support Vector Machine (SVM), Decision Tree (Random Forest), Artificial Neural Network (ANN).

Support Vector Machine (SVM) is a two-classification model. Its basic idea is to transform the original dimension space into the high dimension space through kernel function, thus solving the separation hyperplane that can correctly divide the training data set and maximize its geometric interval. SVM is used to classify data according to the input variables in the model. The final optimization problem is:

$$\min \frac{1}{2} W^T W \quad (1)$$

$$s. t. \quad y_i (W x_i + b) \geq 1 \quad (2)$$

where W is the vector and (x_i, y_i) is the data set. Jiang Peng [2] chose 39 factors as input, used

SVM to predict the price of PingAn Bank, and also compared the performance of used Principal Component Analysis (PCA) to reduce dimension.

Random Forest (RF) enjoys a wide application prospect as an efficient and flexible machine learning algorithm, and financial markets have been frequently used in recent years. For example, Wu Weixing [3], Tan Zheng [4] both used RF to carry out multi-factors stock selection. Lai Tian [5] applied the RF to commodity futures to classify the market volatility and improve the profitability of investment strategies in different time periods. RF takes decision tree (DT) as the basic unit and integrates many decision trees to form a random forest. The basic working principle of random forest is to select m features from the data set, with a total of n features, and then build a decision tree according to the m features, and repeat it many times. After each random combination of the features, multiple decision trees are built, the results of each decision tree are stored, and finally the result with the largest number of votes is taken as the predicted result.

Artificial Neural Network (ANN) is currently the hottest foundation for deep learning. ANN is a machine learning technology that simulates the neural network of human brain. It is estimated that there are more than 100 billion neurons in the adult brain. Jiao Licheng [6] reviewed the development of neural network in recent 70 years. Therefore, ANN is more complex than SVM and RF. It consists of input layer, hidden layer and output layer. These layers are connected with each other. Each link is given weights at the beginning. These weights will be adjusted continuously during training and learning process to minimize errors. Finally, the output layer sums all signals of the previous layer into one output signal. Hu Wenpeng [7] compared the performance of different Neural network methods include RNN, CNN, LSTM used in Stock index Futures. Chen Cang [8] use neural network to predict the trends of Shanghai-Shenzhen index, select the best time to buy or sell. ANN are generally asymptotical to common functions or algorithms, and do not obtain accurate values. At the same time, ANN may also be an expression of a logic structure strategy.

Three cobblers with their wits are equal to Zhuge Liang. Scholars try to combined SVM, RF, ANN to achieve much better results, and it called ensemble Algorithm in the filed of machine learning. Li Xiang [9] used more than 300 factors to select stocks based on XGBoost, the quarterly return as the standard. Jiang Minqi [10] used an improved stacking framework to predict the trend of stock index by leveraging tree-based ensemble models and deep learning algorithms. All of ensemble methods listed above have good performance in the financial market.

3. Analysis

With the help of CNKI database and Elsevier database, a total of 30 papers [2-31] in recent years have been selected using the combination of quantitative trading, stock, futures, machine learning and other terms. The article is in line with the research topic, and the research fields are finance, economics, management science and engineering, etc. The specific content of each article is shown in table 1 below. This paper analyzes the publication year, usage method, research object, performance index, etc. From the analysis, it can be seen that the most articles published in 2019, and according to the understanding, there are fewer articles published before 2017. In terms of research objects, most of them are stock markets, but few are Future. The commonly used performance metrics are Mean Absolute Percent Error (MAPE), Accuracy, Root-Mean-Square Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), and Precision. Other metrics such as F-measure, Information Gain, F1-Score. The most used performance metric for analyzing the performance of stock market techniques is MAE and Accuracy.

Table 1 Brief introduction of the selected papers

Publication years	Author	Papers Title	Methods	Classification
2019	Jiang Peng	Research based on Support Vector Machine (SVM)of Stock Timing in Quantitative Trading	SVM	Timing
2019	Xie Chenyu	Research on Factor Selection Strategy based on LSTM Neural Network Model	ANN	Stock Selection
2019	Zhang Xinyu	Multi-factor Quantitative Stock Selection Scheme Planning Based on Random Forest Algorithm	DT	Stock Selection
2019	Chen Yang	Research on forecasting and investment strategy of Shanghai-Shenzhen 300 Index based on Support Vector Machine	SVM	Timing
2019	Liu Guangyuan	Research on Quantitative Investment Strategy based on CSI300 Stocks	SVM	Stock Selection
2018	Wu Yuhai	Research on multi-factor Stock Picking Strategy in A-share market based on Neural Network models	ANN	Stock Selection
2018	Lai Tian	Research on quantitative investment strategy of commodity Futures based on Random Forest	DT	Timing
2018	Zhang Meng	SVM Multi-factor Stock Selection Model Based on Industry Rotation Strategy and Empirical Analysis of Investment Effect	SVM	Stock Selection
2018	Wu Weixing	Application of Random Forest in Quantitative Stocks selection of Technical Indicators	DT	Stock Selection
2018	Liu Jianxia	Research on the Improvement of Pairs Trading Strategy of Stock Based on PCA-SVM model - A case study of Bank Shares	SVM	Stock Selection
2018	Hu Wenpeng	Research and Application of Stock Index Futures Price Forecast Based on Neural Network	ANN	Timing
2018	Wang Mengzhu	Research on Stock Quantitative Trading Strategy Based on SVM	SVM	Timing
2017	Lan Qiangtai	An Empirical Study on Comprehensive Stock selection based on Principal Component Analysis and BP neural network	ANN	Timing
2017	Chen Cang	A study of quantitative Timing Strategy based on deep learning	ANN	Timing
2017	Song Wenda	The quantitative timing-choose Strategy based on SVM and its Empirical Research	SVM	Timing
2017	Zhou Jian	An empirical research on multi-factor stock selection model based on SVM	ANN	Stock Selection
2017	Chen Lili	The research for quantitative analysis of stock price predict based on BP neural network theory	ANN	Timing
2017	Li Xiang	Multi-factor quantitative stock option planning based on XGBoost Algorithm	DT	Stock Selection
2016	Sun Hongquan	Research on quantitative trading	SVM	Timing
2016	Li Hanshen	Quantitative Trading Strategy of Stock Index Futures based on SVM	SVM	Timing
2020	Chen Zheshi, Li Chunhong	Bitcoin price prediction using machine learning: An approach to sample dimension engineering	DT+SVM+ANN	Timing
2019	Lee T K, Cho J H, Kwon D S	Global stock market investment strategies based on financial network indicators using machine learning techniques	SVM	Timing
2019	Long Wen Lu Zhichen	Deep learning-based feature engineering for stock price movement prediction	ANN	Timing
2019	Keywan Christian Rasekhschaffe	Machine learning for stock selection	ANN+SVM	Stock Selection
2018	Sotirios P.Chatzis	Forecasting stock market crisis events using deep and statistical machine learning techniques	DT	Timing
2017	Guo Bin	The five-factor asset pricing model tests for the Chinese stock market	Regression	Stock Selection
2019	Jiang Minqi; Liu Jiapeng	An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms	DT+ANN	Timing
2018	Felipe Dias Paiva	Decision-making for financial trading: A fusion approach of machine learning and portfolio selection	SVM	Stock Selection
2019	Tan Zheng, Yan Ziqin	Stock selection with random forest: An exploitation of excess return in the Chinese stock market	DT	Stock Selection
2019	Dattatray P. Gandhmal, K. Kumar	Systematic analysis and review of stock market prediction techniques		review

3.1 Analysis based on the publication years

This part is based on statistical analysis of the publication years of the selected articles. Figure 1 illustrates the number of articles published in 2016-2020. Among these 30 papers, 11 papers were published in 2019 and 10 papers were published in 2018.

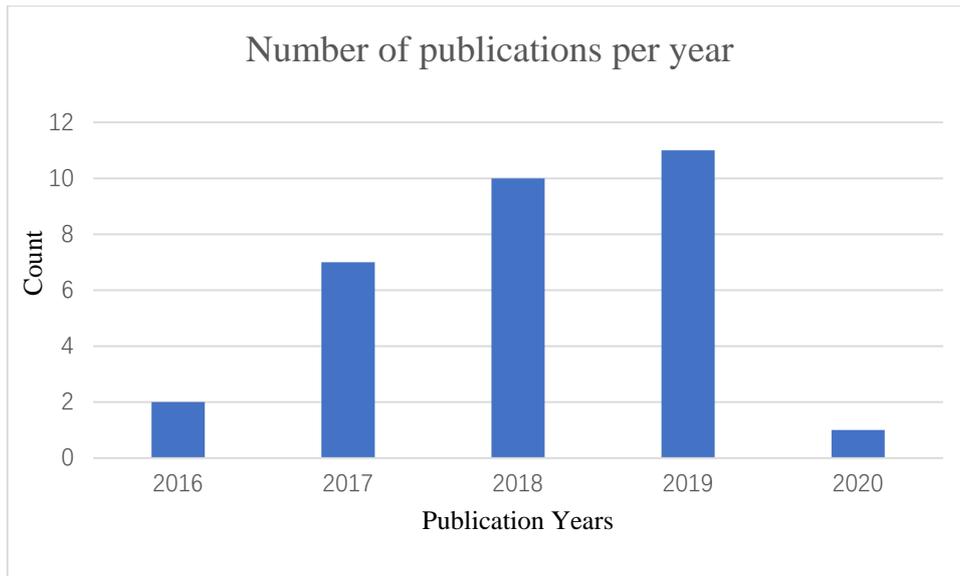


Figure 1 Numbers of publication of the articles considered in the survey

3.2 Analysis based on used methods

According to the second section, the papers selected in this paper are all based on support vector machines, Decision Trees (Random Forests) and Neural Networks. Figure 2 describes the respective proportions of the three methods used in the paper. From Figure 2, it can be seen that 38% of the papers use SVM, 28% of the papers use Neural Networks, and 21% of the papers use Neural Networks, and SVM, Neural Networks and Decision Trees in combination account for 3% of the total. Using the traditional quantitative trading regression, Guo Bin [11] tested the classic five-factor asset pricing model in Chinese Stock market and proved that was effective. Therefore, SVM and Neural Network are the most commonly used methods.

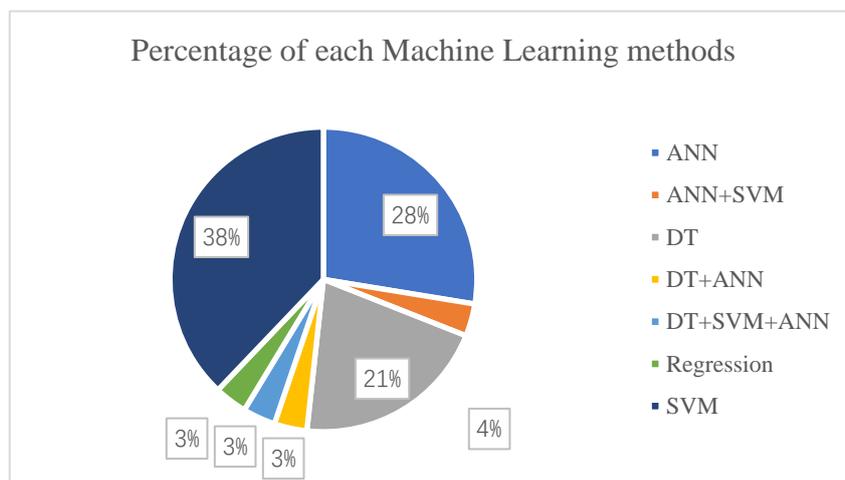


Figure 2 Frequency of machine learning methods employed in the research

3.3 Analysis based on classification

According to the selected papers, the paper can be divided into two categories after read. One is Stock Selection strategy in stock market, which selects multiple factors according to the selected financial factors or technical factors, while another is Timing, predicting the trends. In the Stock and Futures markets, some financial data, index data and momentum data are used to predict the price of a stock. Some articles also predict stock indexes such as Shanghai and Shenzhen 300 Index. As can be seen from figure 3, there are 16 papers on quantitative Stock selection about 53% percent, 13 papers on quantitative Timing, 44% to the all papers. Dattatray P. Gandhmal [12] pointed out that stock market trend prediction is very important. He selected 50 papers, introduced the machine learning method used in the article, and analyzed the methods used, model evaluation indexes and

other aspects. Finally, he pointed out that the selection of characteristics in prediction is very important, and the accuracy and effectiveness need to be further improved. Therefore, it can be concluded that the main direction of quantitative trading is quantitative stock selection and quantitative timing, and there are also some parts such as matching trading and arbitrage trading. Zhang Xun [33] designed a statistical arbitrage strategy for commodity futures based on fundamental analysis, and applied it to copper futures arbitrage with good results.

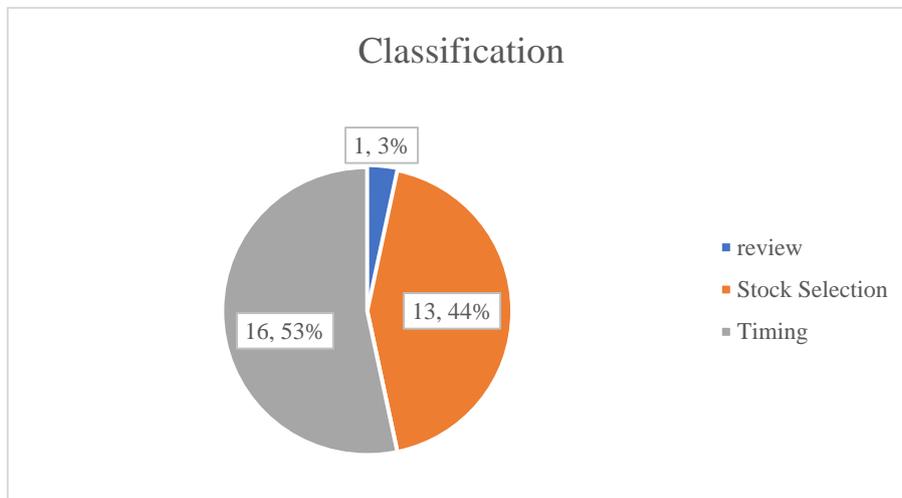


Figure 3 Classification to the papers.

4. Conclusion

This article summarizes the application of machine learning in financial markets, selects 30 articles and classifies the articles according to the year of publication, the method adopted and the research direction (Stock Selection and Timing). There are few papers on quantitative transactions before 2016, and the number of papers grows rapidly in 2018 and 2019, presumably benefiting the development of big data and artificial intelligence technology. Machine learning methods commonly used in financial markets include ANN, SVM, SVR, DT (RF), LR, K-means, etc. However, ANN and SVM are applied more frequently in quantitative trading and have relatively better results. At present, the main problem of quantitative trading is that the accuracy of some models has reached a bottleneck, the winning rate is relatively low, and it is also affected by some other factors, including government policies and investment sentiment. Therefore, we should seek better data mining methods and more comprehensive data to improve the performance of quantitative trading models.

Acknowledgements

This work was supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province: KYCX19_1430.

References

- [1] Markowitz H. M. Portfolio Selection [J]. Journal of Finance, 1952, 7: 77-91.
- [2] Jiang Peng. Research based on Support Vector Machine (SVM) of Stock Timing in Quantitative Trading [D]. Nanchang. Jiangxi University of Finance and Economics, 2019.
- [3] Wu Weixing. Application of Random Forest in Quantitative Stocks selection of Technical Indicators [D]. Chengdu. University of Electronic Science and Technology of China, 2018.
- [4] Tan Zheng, Yan Ziqin and Zhu Guangwei. Stock selection with random forest: An exploitation of excess return in the Chinese stock market [J]. Heliyon, 2019, 5(8): e02310.

- [5] Lai Tian. Research on Quantitative Investment Strategy of Commodity Futures Based on Random Forests [D]. Shanghai. East China Normal University, 2018.
- [6] Jiao Licheng, Yang Shuyuan, Liu Fang, Wang Shigang and Feng Zhixi. Seventy Years of Neural Networks: Retrospect and Prospect [J]. Chinese journal of computers, 2016, 39(08): 1697-1716.
- [7] Hu Wenpeng. Research and application of stock index futures price forecast based on neural network [D]. Hunan. Hunan University, 2018.
- [8] Chen Cang. Research on Quantitative Timing Strategies Based on Depth Learning [D]. Lanzhou, University of Finance and Economics, 2017.
- [9] Li Xiang. Multi-factor Quantitative Stock Selection Scheme Planning Based on XGBoost Algorithm [D]. Shanghai. Shanghai Normal University, 2017.
- [10] Jiang Minqi, Liu Jiapeng and Zhang Lu. An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms [J]. Physical A: Statistical Mechanics and its Applications, 2020, 541, 122272.
- [11] Li Hanshen. Quantitative Trading Strategy of Stock Index Futures based on SVM [D]. Zhejiang University, Hangzhou, 2016.
- [12] Guo Bin, Zhang Wei and Zhang Yongjie. The five-factor asset pricing model tests for the Chinese stock market [J]. Pacific-Basin Finance Journal, 2017, 43: 84-106.
- [13] Dattatray P. Gandhmal, K. Kumar. Systematic analysis and review of stock market prediction techniques [J]. Computer Science Review, 2019, 34, 100190.
- [14] Sun Hongquan. Research on quantitative trading [D]. University of Technology, Beijing, 2016.
- [15] Chen Zheshi, Li Chunhong and Sun Wenjun. Bitcoin price prediction using machine learning: An approach to sample dimension engineering [J]. Journal of Computational and Applied Mathematics, 2020, 365, 112395.
- [16] Lee T K, Cho J H, and Kwon D S. Global stock market investment strategies based on financial network indicators using machine learning techniques [J]. Expert Systems with Applications, 2019, 117, 228-242.
- [17] Liu Jianxia. Research on the Improvement of Pairs Trading Strategy of Stock Based on PCA-SVM model - A case study of Bank Shares [D]. Harbin Engineering University, Harbin, 2018.
- [18] Wang Mengzhu. Research on Stock Quantization Trading Strategy Based on Support Vector Machine [D]. Chongqing University, Chongqing, 2018.
- [19] Wu Yuhai. Research on Multi-factor Stock Selection Strategy of A-share Market Based on Neural Network Method [D]. University of Electronic Science and Technology of China, Chengdu, 2018.
- [20] Chen Lili. The research for quantitative analysis of stock price predict based on BP neural network theory [D]. Zhejiang University of Technology, Hangzhou, 2017.
- [21] Zhang Xinyu. Multi-factor Quantitative Stock Selection Scheme Planning Based on Random Forest Algorithm [D]. Shanghai Normal University, Shanghai, 2019.
- [22] Chen Yang. Research on forecasting and investment strategy of Shanghai-Shenzhen 300 Index based on Support Vector Machine [D]. Northwest University, Xi'an, 2019.
- [23] Liu Guangyuan. Research on Quantitative Investment Strategies Based on Shanghai and Shenzhen 300 Component Stocks [D]. Guizhou University, Guizhou, 2019.
- [24] Xie Chenyu. Research on factor stock selection strategy based on LSTM neural network model [D]. Liaoning University, Shenyang, 2019.

- [25] Zhang Meng. SVM Multi-factor Stock Selection Model Based on Industry Rotation Strategy and Empirical Analysis of Investment Effect [D]. Shanghai Normal University, Shanghai, 2018.
- [26] Zhou Jian. An Empirical Study on Multi-factor Stock Selection Model Based on SVM Algorithm [D]. Zhejiang University of Technology and Industry, Hangzhou, 2017.
- [27] Song Wenda. The quantitative timing-choose Strategy based on SVM and its Empirical Research [D]. University of technology, Xi'an, 2017.
- [28] Lan Qiangtai. An Empirical Study on Comprehensive Stock Selection Based on Principal Component Analysis and BP Neural Network Algorithm [D]. Jinan University, Guangzhou, 2017.
- [29] Long Wen, Lu Zhichen and Cui Lingxiao. Deep learning-based feature engineering for stock price movement prediction [J]. Knowledge-Based Systems, 2019, 164: 163-173.
- [30] Sotirios P. Chatzis, Vassilis Siakoulis, Anastasios Petropoulos, Evangelos Stavroulakis and Nikos Vlachogiannakis. Forecasting stock market crisis events using deep and statistical machine learning techniques [J]. Expert Systems with Applications, 2018, 112: 353-371.
- [31] Felipe Dias Paiva, Rodrigo Tomás Nogueira Cardoso, Gustavo Peixoto Hanaoka and Wendel Moreira Duarte. Decision-making for financial trading: A fusion approach of machine learning and portfolio selection [J]. Expert Systems with Applications, 2019, 115: 635-655.
- [32] Keywan Christian Rasekhschaffe, Robert C. Jones. Machine Learning for Stock Selection [J]. Financial Analysts Journal, 2019, 75: 70-88.
- [33] Zhang Xun. Design and Application of Quantitative Arbitrage Strategy for Commodity Futures Based on Fundamental Analysis [D]. Beijing University of Technology and Industry, Beijing, 2018.