Research on the Application of Machine Learning Technology in Intelligent Stock Selection

Hongtao Chen

Jinan University, Guangzhou 510000, China

samchen0501@163.com

Abstract. With the development of big data era and artificial intelligence, intelligent investment has gradually replaced artificial investment consultants for consulting and financial services. Robo-advisor can also be called robot investment, intelligent financial management, automated financial management and so on. The traditional quantitative investment model mostly depends on the indicators designed by the financial knowledge and experience of the researchers and cooperates with the economic model or strategy. Financial knowledge and experience play a major role in the construction of the quantitative investment model, so the efficiency of model development and updating is low and the human influence factors are large. The quantitative model is established through big data and intelligent algorithms, and the market is judged according to the risk preference of investors, intelligent allocation and investment of assets are carried out, and automatic strategic trading services are implemented. Based on machine learning theory, this paper studies the two types of investment in quantitative investment: quantitative timing and quantitative stock selection, and puts forward corresponding quantitative investment models. Through testing the models, it is verified that they have good profitability and risk control ability, which is a quantitative investment model that can provide guidance for investors.

Keywords: Machine Learning Technology; Smart Stock Selection; Application Research.

1. Introduction

In the 1870s, Barclays Investment Company of the United States issued the world's first index fund, which marked the birth of quantitative investment. At the same time, with the rapid development of computer technology, computer programs can process massive data more efficiently [1]. Investors have also gradually realized that if the characteristics of computers are integrated into securities investment, it is bound to produce better stocks and continuously optimize investment returns. Therefore, based on the massive historical data in the financial market, the quantitative investment method of building models with mathematical and statistical knowledge, realizing models and completing investment strategies with computer software tools has attracted the attention of more and more investors [2]. The so-called quantitative stock selection refers to the selection of high-quality stocks by selecting appropriate stock selection index system and using quantitative statistical analysis tools. It is an important content of quantitative investment, and its essence is the classification problem in the field of data mining [3].

Both quantitative investment and traditional qualitative investment are based on the theoretical basis of market inefficiency or weak efficiency, and establish investment strategies that beat the market and produce average returns in excess market. However, traditional qualitative investment mainly relies on people's feelings and experience to manage assets, while quantitative investment uses computers to adjust model parameters by fitting a large amount of data. When the performance of the model in historical data meets certain conditions, it will be applied to real trading [4]. In the early days of the development of computer technology, investors used computers instead of manpower to complete some mechanized steps of stock trading. For example, investors can set trading orders according to investment strategies, and realize position control and fund allocation by computers [5]. At that time, observing the laws existing in the financial attributes of stocks and designing trading strategies according to the laws can only be realized manually. With the further development of technology, whether it can increase the autonomy of computers and further assist stock trading through computer analysis has aroused the thinking of scholars and investors [6].

Energy investment advisers measure the risk preferences of different users and calculate the return of the investment portfolio by building a dynamic algorithm model, so as to match the appropriate investment scheme for users. Previous researchers usually considered single transaction data or single financial data in stock selection technology. With the advent of the era of big data and the improvement of computing power in recent years, some researchers began to combine transaction indicators with financial indicators for empirical simulation, but still ignored other important measurement indicators [7]. The corresponding research methods also show the trend of a hundred flowers blooming and a hundred schools of thought contending. In addition to the researchers in the traditional financial field, researchers in physics, cognitive science, artificial intelligence, computer science, biology and other fields have flocked to create many different quantitative investment strategies based on the ideas and theories of their respective fields and the powerful data processing capabilities of computers, this has led to various branches of quantitative investment [8].

This paper attempts to start from the quantitative investment strategy, analyze the relevant theories of quantitative investment, and try to build a stock selection and timing model suitable for China's capital market on the basis of theoretical research, design corresponding products according to their own model, and carry out effectiveness tests to test whether the designed products can defeat the market.

2. Multi-indicator Intelligent Investment Advisor Stock Selection Strategy Design

2.1 Strategic Design Ideas

With the progress of Internet and computer technology, quantitative investment based on mathematical model and computer technology has gradually entered the field of vision of researchers and investors. Quantitative investment should be traced back to portfolio theory. The mean variance model established by Macovei shows the importance of risk while pursuing returns in portfolio. On the basis of this theory, the capital asset pricing models proposed by Sharp, Lint and Mohsen respectively illustrate the quantitative relationship between risk and return [9]. Quantitative investment is a concept to promote the use of quantitative tools to realize investment. It mainly invests by combining the methods of various disciplines, and at the same time, it will have its own set of investment system. This investment method mainly uses quantitative investment tools to mine the laws of securities business, thus corresponding investment model mode. In fact, quantitative investment is not new and has been developed abroad for 30 to 40 years. Due to the stable investment performance, quantitative investment has always been favored by investors [10].

After entering the 21st century, quantitative investment has developed rapidly. In addition, although some research has been carried out on the application of machine learning algorithms in China's quantitative investment, most of them are limited to the discussion of algorithms and the accuracy test of algorithm models, or the simplification of the back-test system. It is not very convincing, and there is not a complete dynamic quantitative method system for investment. A typical stock selection model, such as the value stock selection model of Royal Bank of Canada, uses six stock selection index systems, including PE excluding economic cycle fluctuations, PE based on consensus expectations, PE excluding non-recurring profit and loss, PB index, three-stage dividend discount model and dividend yield. The model has achieved excellent returns in the United States and Canada markets. However, foreign stock selection models are mainly based on the experience of foreign investment managers, and there is still a process of localization in the application of Chinese market. The ultimate goal of smart investment advisor's stock selection strategy is to select the top stocks. The basic stock selection strategy is: to predict the best first few stocks of the day, it is necessary to train the historical data indicators of the days before the day to obtain the best model parameters, so as to achieve the best stock prediction of the day. As shown in Figure 1.

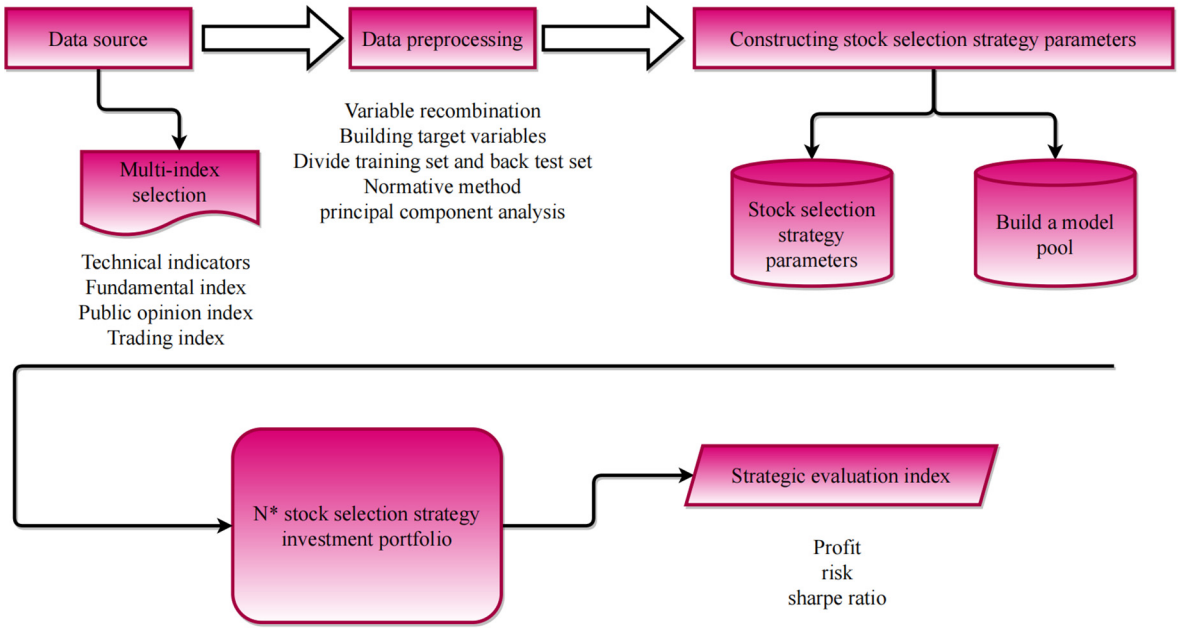


Figure 1. Schematic diagram of strategy design

Figure 1 is a schematic diagram of the strategy design idea. The strategy design logic of this study is to obtain the data source through multi-index selection, construct the stock selection strategy parameters through data preprocessing steps such as variable reorganization and normalization, and obtain several stock selection strategy portfolios. Then, these stock selection strategy portfolios are tested through data set training, and the stock selection strategy is evaluated in terms of income, risk and Sharp ratio, and finally several groups of optimal stock selection strategy portfolios are selected.

2.2 Quantitative Investment and Machine Learning

The rapid development of machine learning has not only changed the computer field, but also brought impact to other fields. Many scholars try to apply machine learning to the financial field. Because the stock market is widely concerned in the financial field, and the neural network is suitable for batch processing of data and needs a large amount of data, and has high adjustability and flexibility, scholars try to use neural network in stock trading. According to the theorem, nonlinear projection of complex pattern classification problems into high-dimensional feature space may be linearly separable. Therefore, as long as the dimension of the feature space is high enough, the original pattern space can be transformed into a new high-dimensional feature space, so that the pattern in the feature space is linearly separable with a high probability. The change of the national macro strategy and the introduction of relevant policies may not change the financial indicators in the enterprise report in a short time, but the future trend of the company can be quickly judged from the social public sentiment. Generally speaking, the recent company news events of a stock can be reflected by the number of short-term discussions, and the increase of short-term attention can explain the popularity of this stock and the possibility of more buying in the future. The cumulative number of people concerned can roughly judge the long-term industry background trend behind the stock, whether it belongs to key sectors and other information. Since the fluctuation of all the observed stock price data is the result of market state fluctuation and individual stocks' own factors, the key of quantitative stock selection model lies in mining the driving factors behind all the stock price data, and then selecting the stocks to be invested by analyzing and forecasting the internal factors of each stock.

According to this idea, this paper also designs the Alpha collaborative Kalman filter model used to mine the internal factor variables of stock price changes, and combines the support vector regression model to carry out regression prediction of the excess return factors, so as to build a

quantitative stock selection strategy. The strategy is rolling update, which is different from previous studies. Most of the previous studies used single factor and fixed time period, and then used to train data. The difference is that we use the method of updating data every day. The selected training set is rolling and updated. We use the dynamic rolling data to train the model and get more accurate output results. It is that the selected stock pool is more predictive. At the same time, the parameters of our training set and model are constantly updated with the training data, so as to obtain the optimal model. At the same time, the matching and optimization between the data and the model make our stock pool more accurate and achieve good performance on the trading day. It is the outstanding characteristics of the strategy realization to obtain an effective quantitative strategy.

3. Application of Machine Learning Technology in Intelligent Stock Selection

3.1 Use Correlation Analysis to Construct Stock Selection Index

After the initial index system is established, in order to make the model have a good theoretical basis and ensure the correctness of the quantitative model, this paper intends to introduce the correlation theory of indicators and analyze the correlation between the initial influencing factors and the response factors, so as to realize the construction of the quantitative stock selection index system. Correlation analysis is mainly to analyze the direction and degree of the relationship between phenomena. Determine whether there is a correlation between phenomena and the manifestation of the correlation. In the process of quantitative stock selection, the impact factor is used to estimate the response factor, and the reliability of its estimation can be measured by the correlation coefficient between the impact factor and the response factor. If the correlation coefficient between the impact factor and the response factor is larger, it means that the impact factor is beneficial to explain the response factor and should be added to the model.

When using support vector machines for stock selection, we use the improved algorithm as a means of stock selection, as a concept of machine learning, so as to obtain a stock pool that meets our requirements from the input of indicator data. We use data to train model parameters, and finally use relevant rules to output data for model prediction. In this process, redundant and useless interference data will be considered, and only necessary data will be left for data screening and prediction. Finally, we will get the desired results, select the stock pool that meets our requirements, and carry out subsequent strategy operation and empirical research. As shown in Figure 2.

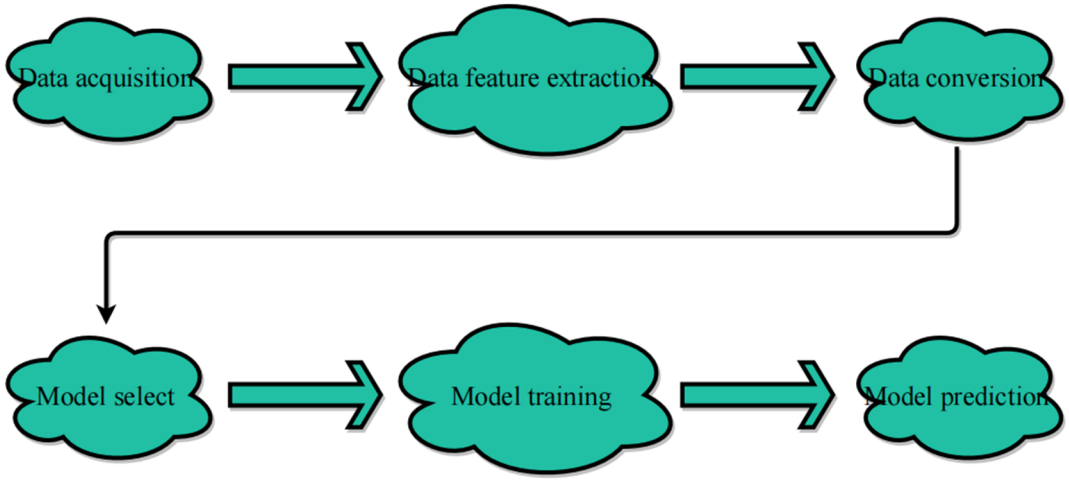


Figure 2. Model processing flow chart of stock selection stage

Although the deep learning method has been able to capture the characteristics of the input data, it often needs model integration to further improve the performance of the model. Model integration can be divided into data level integration and model level integration. Data level integration includes data expansion, simple integration and other integration methods, mainly through data expansion,

downsampling and other means. Model level integration includes single model integration and multiple model integration. For deep learning, single model integration is the fusion of multi-layer features. Multi-model integration includes the integration of the same model and different models. For the integration of the same model, multiple models with different initial parameters are trained and the prediction results are integrated. The model is used to mine internal factor variables on stock price data to verify its rationality and effectiveness. Then, a new stock selection model is designed by combining SVR and dynamic updating mechanism, and the model is back-tested during the testing period. The advantages of the alpha collaborative Kalman filter model are verified through the revenue process, and the profitability and risk control ability of the quantitative stock selection model based on alpha collaborative Kalman filter proposed in this paper are also demonstrated.

3.2 Empirical Analysis

Because the eight indexes in the stock selection index system are continuous variables, the random forest algorithm needs to discretize the continuous variables to improve the classification accuracy of the algorithm. We analyze the results of the strategy, mainly including two aspects, the first is the asset allocation of the industry sector, and then the factors of stock selection in the industry. We know that good industries perform well, which brings high yield performance to stocks. On the contrary, poor industries and sectors will not bring performance beyond the market. And the choice of stocks can also bring excess returns, so we use attribution to analyze the influence of these two aspects on the strategic results. In order to measure the effectiveness of the stock selection model, we should set an overall average return of the market as the benchmark, and take the difference between the return of the stock selection model and the benchmark return as the standard to measure the quality of a stock selection model. The red line here is the benchmark set in this paper, which is the yield curve generated by investing equal amount of funds in all sample tickets in the Shanghai and Shenzhen 300 Index.

From a comprehensive analysis, the main reason is that our stock selection benchmark is HS300 component stocks. These stocks are characterized by large-cap blue-chip stocks, which are significantly superior to other stocks in terms of value and liquidity, and other factors are the same. The overall analysis of the style of the strategy has not changed much in the past three years, which proves that the strategy is still very stable. Only blue-chip stocks with high value and good liquidity are preferred. Based on the above analysis, we know that this strategy can achieve better returns by increasing the proportion of high-quality stock selection sectors, which is also the direction to further improve the strategy, so we can carry out some follow-up thinking from this aspect. Due to the limited ability, this paper only puts forward relevant ideas and does not give relevant results and methods.

4. Conclusion

China is gradually promoting the development of intelligent investment research, and the current development momentum is good. With the breakthrough of artificial intelligence technology, the advent of the era of large capital management, the innovation of financial products and the continuous improvement of the needs of investors, smart investment will surely usher in a larger market with great potential. Comparing with quantitative investment strategies, we find that both of them have their own advantages and disadvantages. First, quantitative investment overcomes the lack of calculation of analysts' investment. As long as the rules of the strategy are formulated, quantitative investment can perfectly implement the rules and will not be affected by emotional fluctuations. This paper explores the application of machine learning theory in quantitative investment, mainly focusing on the two types of investment strategies of quantitative timing and quantitative stock selection in quantitative investment, and strives to use machine learning algorithm to mine the internal laws of financial data operation and provide investors with quantitative investment guidance. At the same time, it verifies the advantages of the Alpha collaborative Kalman filter model compared with the traditional linear regression method in mining the intrinsic factor variables of the stock price. Finally,

aiming at the "failure" phenomenon caused by the long running time of the model, the dynamic update model mechanism is introduced to optimize.

References

[1] Li, H. M. Zhang, Q. P. Zheng, Z. L. APPLICATION OF MACHINE LEARNING IN INTELLIGENT KNOWLEDGE SEARCH ALGORITHM[J]. Latin American Applied Research, 2018, 48(4):12-16.

[2] Yu G. Research on computer network information security based on improved machine learning[J]. Journal of Intelligent and Fuzzy Systems, 2020, 40(3):1-12.

[3] Lv X, Li M. Application and Research of the Intelligent Management System Based on Internet of Things Technology in the Era of Big Data[J]. Mobile Information Systems, 2021, 2021(16):1-6.

[4] Gao J, Guo H, Xu X. Multifactor Stock Selection Strategy Based on Machine Learning: Evidence from China[J]. Complexity, 2022, 2022(3):15-18.

[5] Schneider N. Industrial condition monitoring with smart sensors using automated feature extraction and selection[J]. Measurement Science & Technology, 2018, 29(9):30-34.

[6] Muhammad, Shafiq, Xiangzhan, et al. A machine learning approach for feature selection traffic classification using security analysis[J]. Journal of Supercomputing, 2018,55(4):26-30.

[7] Paraschos P D, Xanthopoulos A S, Koulinas G K, et al. Machine learning integrated design and operation management for resilient circular manufacturing systems[J]. Computers & Industrial Engineering, 2022, 167:10-17.

[8] Liu H, Xiao Y, Yang P, et al. Transmit Antenna Selection for Full-Duplex Spatial Modulation Based on Machine Learning[J]. IEEE Transactions on Vehicular Technology, 2021,20(4):11-13.

[9] Yadav R K, Sivakkumar M, Kshirsagar P. Design Framework of Stock Price Forecasting Using Cascaded Machine Learning and Swarm Intelligence[J]. Solid State Technology, 2021, 64(1):24-38.

[10] Yan M, Feng G, Zhou J, et al. Smart Multi-RAT Access based on Multi-Agent Reinforcement Learning[J]. IEEE Transactions on Vehicular Technology, 2018,36(12):1-1.