Application of Data Mining Technology in Portfolio Optimization

GuoRong Hu1,*, Hui Liu1
1Huzhou Vocational & Tech. Coll., Huzhou Zhejiang, China

*Corresponding author e-mail: balanwu@163.com

Abstract. With the development of the times and the progress of society, the level of human science and technology is constantly improving. A large number of scientific and technological forces help us to work and study in our daily production and life. In this era of big data, data mining technology is a hot technology in the current era, and it can be found in many industries in many fields. People analyze and manage the information in our daily life through big data technology. But in this era of rapid development, the traditional big data information analysis and management ability can not meet the new needs of the new era, and the emergence of data mining technology can fully meet the needs of information data processing in the new era. In order to better study the portfolio optimization problem in this era of big data, this paper proposes a method of integrating data mining technology with portfolio optimization problem, and makes full use of data mining technology. In a large number of portfolio optimization problem data, the relevant data are accurately analyzed and efficiently managed, so as to reasonably optimize the portfolio Research. It is found that the method proposed in this paper is of great significance to the application of data mining technology in portfolio optimization.

Keywords: Science and Technology, Data Mining Technology, Big Data Technology, Portfolio Optimization

1. Introduction

With the popularization and application of Internet technology, information technology has entered a new stage of development, greatly improving people's production efficiency and living standards. All walks of life have gradually realized the road of information development, greatly improving the management level and economic benefits of all walks of life. However, with the development of market economy, information systems[1-2] in all walks of life have accumulated a large amount of data information. How to make statistics, analysis, utilization and decision-making of these massive data
information has become an urgent problem for all walks of life. At the same time, massive data mining technology emerges as the times require, which is widely used in information management, scientific research, financial decision-making, processing and retail industry, medical and pharmaceutical. It is because of the emergence of data mining technology\textsuperscript{[3-5]} and data warehouse technology to solve the effective use of massive data, and further promote the development of information technology.

Index investment is a passive investment strategy. Its original idea comes from the academic debate about "holding market portfolio". It is a portfolio that replicates and tracks the benchmark index to achieve average market returns. It not only does not have the excessive risk of active fund management\textsuperscript{[6-8]}, but also has no time limit. Therefore, with the continuous development of China's securities market, index construction method as the main form of index investment research has been paid more and more attention by theoretical and practical circles. Generally speaking, index construction methods can be divided into three types: complete replication method, optimized replication method and sampling replication method. Due to the reasonable and effective construction principle of optimal replication method, the optimization of index combination constructed by optimal replication method is the focus of this paper.

This paper mainly studies the application of data mining technology in the portfolio optimization problem\textsuperscript{[9-10]}. In the current era of rapid development of information, the competition of all walks of life is becoming increasingly fierce. In order to have a suitable development path in the new era, this paper puts forward the method of integrating data mining technology and portfolio optimization problem, and uses data mining technology to solve a large number of investment portfolio optimization problems. Through the analysis of portfolio optimization data, a new path suitable for the development of portfolio optimization is proposed.

2. Application of Data Mining Technology in Portfolio Optimization

2.1. Data Mining Technology

The main purpose of data mining is to extract, sort and classify massive data information, and to mine the hidden useful information, so as to provide decision support information for the survival and development of all walks of life. Concepts similar to data mining include knowledge discovery (KDD), data analysis, etc. In a word, data mining is a technology that uses various data analysis tools to mine the relationship between model data information and model in massive data information. In the first mock exam, the president analyzed and understood the corresponding relationship between them through the understanding and understanding of this model, so as to guide the production development of all walks of life and provide support for major decisions.

2.2. Portfolio Theory

Portfolio is determined by securities and their weights. The expected return of the portfolio is the weighted average of the expected return of each security. The risk of a portfolio is defined by the standard variance of the rate of return. These statistics describe the extent to which yields vary around their average. If the volatility is serious, it shows that the yield has greater uncertainty, that is, high risk. The covariance between the two combinations reflects the interaction. If the correlation
coefficient is negative, the smaller the covariance, the smaller the overall risk of the portfolio. On the contrary, if the correlation coefficient is positive, the overall risk of the portfolio will be greater. If the correlation coefficient is positive, it means that the two securities are independent and independent of each other. Therefore, it is better to choose negatively correlated securities to construct portfolio.

3. Experimental Correlation Analysis

3.1. Experimental Background

According to the technical analysis method, the historical price information of securities includes the future price information of securities. We use the logarithmic method to calculate the stock return from 9 days to 9 months in 2018.

3.2. Experimental Design

The sample data model can be used to solve the problem of portfolio optimization. For the sake of simplicity, we only select 12 stocks in the Shanghai Shenzhen 300 index with high liquidity, large scale and strong stability as the research object. The names and stock codes of the 12 stocks are shown in Table 1 below:

| Stock name | Stock code | Stock name | Stock code |
|------------|------------|------------|------------|
| Minsheng Bank | 600016 | Baosteel Co., Ltd | 600019 |
| SINOPEC (China Petrochemical Corporation) | 600028 | Wuliangye | 000858 |
| Yangtze Power | 600900 | Zhenhua heavy industry | 600320 |
| Vanke Group | A000002 | Shanghai Airport | 600009 |
| China Unicom | 600050 | Bank of China | 601988 |
| Maotai, Guizhou | 600519 | ZTE | 000063 |

4. Discussion

4.1. Analysis on the Application of Data Mining Technology in Portfolio Optimization

The model is an empirical analysis of the results, mainly from two aspects of income and risk. In terms of income, this paper uses a comparative test with benchmark market indicators. Benchmark rate of return is the benchmark value of the market when evaluating the performance of investment strategy. This benchmark should be able to objectively reflect the overall changes of each stock in the market and capital operation. The model is based on the CSI 300 index, so the CSI 300 index is also
selected as the benchmark.

The interval of model back testing is from December 25, 2013 to December 25, 2019. The post adjustment period is monthly. The weight of each equity is calculated at the end of the month, and the opening operation is conducted according to the average price on the closing day of the next month. Regardless of timing, the warehouse weight is 100%. Shanghai Shenzhen 300 index was selected as the control. The total return curve of the model is shown in Figure 1:

![Return rate curve of back test interval](image)

**Figure 1.** Return rate curve of back test interval

It can be seen from Figure 1 that the yield in recent years has basically shown an upward trend, and the return of CSI 300 has also maintained an upward trend in the same period. According to the numerical proportion, it can be seen that the model yield is about five times of the benchmark yield, and the excess return is very obvious. This shows that the application of model return rate in data mining of excess return is very effective.

Exit degree is a common risk evaluation index in the field of quantitative investment, represents the maximum difference between the net investment value at the lowest point of the model at any time point in the investment period and the return at the previous time point. The maximum pullback reflects the worst possible return on the portfolio at any point in time. Maximum correction is an important risk index, more important than volatility. The annual yield of the model and the relevant data of benchmark CSI 300 index are shown in Figure 2:
Figure 2. The relative excess return rate of Shanghai and Shenzhen 300 in recent years

As shown in Figure 2, the annual average abnormal return of the model portfolio reaches 15.725%, and the excess return rate of other years is more than 10%, except for the decrease in 2014 and 2017. It shows that the division of excess earnings is relatively stable, stable returns can be obtained in bull market, bear market and adjustment period, and the applicability is strong.

4.2. Suggestions on the Application and Development of Data Mining Technology in Portfolio Optimization

At present, data mining technology with information characteristics can be better integrated into the current social environment. Considering the social economy, science and technology, culture and other aspects, it can more reasonably cooperate with the production and operation activities of enterprises, explain the changes of the nature and value of materials in the production and operation activities, and find out the change characteristics and laws of the relevant data of materials, so that enterprises can objectively and reasonably adjust the direction and mode of production and operation activities and improve the effect of production and operation activities, Create more economic benefits for enterprises. From this point of view, the application of data mining technology expands the functional scope of data information resources, better serves enterprise business, and reaches the technical level of enterprise business level.

From the current application of data mining technology in various fields, data mining technology plays an important role in marketing, scientific research, telecommunications, education and other fields. Based on the current application trend of data mining technology, in the future, with the continuous optimization, innovation and improvement of data mining technology, the technology will be extended to more fields, such as biopharmaceutical, criminal investigation, aerospace and other fields, and play a greater role. Therefore, the development and application of data mining technology is very important.

The essence of portfolio selection is an online decision-making problem. The rapid development of online learning algorithm in the field of theoretical computer science provides a new research direction and technical means for online decision-making. At the same time, the in-depth study of market abnormal behavior financial theory also provides theoretical support for online portfolio
selection. In particular, as an emerging securities market, China has both the characteristics of western mature markets and distinctive characteristics different from other emerging markets. Therefore, it is necessary to study the online portfolio strategies and algorithms in different financial markets at home and abroad. This paper breaks through the limitations of the competition analysis framework established by cover et al. Start with the worst in the market. In order to maximize the return, the algorithm is simple and easy to operate. Based on the legal information of market anomalies, the correlation relationship and online learning technology are used to construct the portfolio model. By analyzing the characteristics of online algorithm, online portfolio strategy and market volatility, the algorithm is applied to test the performance of different securities markets at home and abroad.

There are also other market anomalies in the financial market, such as scale effect, weekend effect, and other investment strategies, such as small cap stock strategy. Therefore, according to the characteristics of market anomalies, other forms of loss function are constructed to describe the market characteristics more accurately, which is more in line with the actual situation of the market, so as to establish a better return investment strategy. Therefore, the degree of investor's aversion to loss varies with the degree of investor's aversion to loss.

This paper studies the optimization problem of static portfolio, that is, assuming that the investor's investment strategy is buy and hold strategy, the investor does not change the asset allocation state in the subsequent holding period. Under this strategy, the portfolio is completely exposed to market risk, and its advantage lies in lower transaction cost and management cost, but at the same time, it gives up making profits from market changes. Therefore, when the market environment and investor preferences change slightly, or the cost of changing the asset allocation ratio is greater than the income, investors are suitable to take the buy and hold strategy. But in the actual market, many investments are long-term investments, which require investors or asset managers to dynamically adjust asset allocation according to the capital market environment and economic conditions. In the future, we can do further research on the dynamic risk measurement model of portfolio and the optimization of portfolio based on dynamic risk measurement, which is of great significance. Once a reasonable dynamic risk measurement model is proposed, it will be a major breakthrough in the field of risk measurement.

The cash management in the process of index investment needs further discussion, including the control of tracking error and the daily maintenance of index investment. Cash management is an aspect of daily maintenance in liquidity management of index investment. Because of the existence of cash, it will cause cash drag and increase the tracking error of index investment. In order to achieve the target of index tracking, what kind of optimal cash management strategy should be adopted by indicator tracking managers, which should not only meet the liquidity demand, but also track the performance of target indicators well, which is the direction of further research in the future. As the research on index investment in this paper is still at the basic level, many frontier issues such as derivative securities index investment, moral responsibility index investment and special customized index investment are not involved. Therefore, further exploration is needed in the follow-up study.

At present, the rapid development and application of data mining and its technology have made great achievements, which can provide important decision support and development industry for production, especially for the work of colleges and universities, and help to establish a harmonious campus. I believe that in the near future, data mining technology will be further developed, which is
of great significance to human production and life.

5. Conclusion

This paper mainly introduces the application method of data mining technology in portfolio optimization. In the current information age, the development of our science and technology has reached a peak, and all walks of life are facing the impact of the new era. In order to meet the new requirements of the new era for portfolio optimization, this paper puts forward the combination of data mining technology and investment. The method of combining portfolio optimization problems, through the rational use of data mining technology, will be a large number of portfolio optimization data for accurate analysis, for the solution and development of portfolio optimization problems to provide new ideas. It is found that the method proposed in this paper is of great significance to the application of data mining technology in portfolio optimization.

References

[1] Schomburg I, Jeske L, Ulbrich M, et al. The BRENDA enzyme information system–From a database to an expert system[J]. Journal of Biotechnology, 2017:194-206.

[2] Fang S, Xu L, Zhu Y, et al. An integrated information system for snowmelt flood early-warning based on internet of things[J]. Information Systems Frontiers, 2015, 17(2):321-335.

[3] Edwards M, Rashid A, Rayson P. A Systematic Survey of Online Data Mining Technology Intended for Law Enforcement[J]. ACM Computing Surveys, 2015, 48(1):1-54.

[4] Yan L, Hu P, Li C, et al. The performance prediction of ground source heat pump system based on monitoring data and data mining technology[J]. Energy & Buildings, 2016, 127(sep.):1085-1095.

[5] Liu, Kaixuan, Wang, Jianping, Hong, Yan. Wearing comfort analysis from aspect of numerical garment pressure using 3D virtual-reality and data mining technology[J]. International Journal of Clothing ence & Technology, 2017, 29(2):166-179.

[6] Taffler R J, Spence C, Eshraghi A. Emotional economic man: Calculation and anxiety in fund management[J]. Accounting, Organizations and Society, 2017:53-67.

[7] Duyvené de Wit, Jean-Jacques, Polakow D. How disruptive can Smart Beta be to the South African Active Fund Management Fraternity?[J]. Investment Analysts Journal, 2017, 46(1):17-31.

[8] Shih H S. Policy analysis on recycling fund management for E-waste in Taiwan under uncertainty[J]. Journal of Cleaner Production, 2017, 143(FEB.1):345-355.

[9] Low R K Y, Yao Y, Faff R. Diamonds vs. precious metals: What shines brightest in your investment portfolio?[J]. International Review of Financial Analysis, 2016, 43(Jan.):1-14.

[10] Jin Y, Roca E, Li B, et al. Sprinkle your investment portfolio with water![J]. International
Journal of Water, 2015, 9(1):43-59.