

# Research on Investment Strategy and Benefit of A-share Traditional Chinese Medicine Industry Value Quantification Based on Machine Learning

Jiatong Hu

Shanghai University of Finance and Economics, Shanghai, China

---

**Abstract:** Before the development of quantitative investment, investors only used qualitative investment to participate in the market. With the great development of information technology and computer technology, investment institutions can use emerging technologies to analyze and process data, apply mathematical models to the market, and execute them by computer instructions. This paper summarizes the research progress of applying machine learning technology to quantitative strategy, especially the influence of BP neural network on quantitative investment after stock selection.

**Keywords:** Machine learning, Quantitative investment, BP neural network, Deep learning.

---

## 1. Introduction

In recent years, the Artificial Neural Network, (ANN) model has made great achievements in natural language processing, image recognition and other fields[1]. With the development of GPU technology, the improvement of computer computing power and the proposal of neural network optimization algorithms, such as Backpropagation, BP), the problem of model training speed that has long plagued the development of deep learning has been solved[2]. The application of neural network in financial field has a long history[3]. At present, an important application of neural network and deep learning algorithm in financial field is quantitative investment. Quantitative investment refers to the realization of securities analysis and prediction with the help of computers, and then the formulation of feasible quantitative strategies for trading in the securities market.

The traditional quantitative investment strategy is mainly based on statistical learning, but the traditional statistical learning method has great limitations in model fitting and price prediction[4]. In recent years, machine learning algorithms, which are developing rapidly, are different from statistical learning methods. They have no assumption of distribution for input data, and can fit a variety of relationships[5]. Using machine learning algorithms including neural network models to formulate quantitative investment timing strategies is a very popular direction in quantitative investment strategy research in recent years, and is also the main application field in the field of quantitative investment[6]. It can be predicted that more machine learning methods will be used in the future financial quantitative research, and how to find a suitable application mode of machine learning has become one of the most promising research directions in the current quantitative research field[7]. At present, China's research in the field of quantitative investment is still in the exploratory stage, so promoting the vigorous development of China's quantitative investment is the requirement of the times under the current international environment, and is also the significance of this study.

## 2. Overview of Mass Production Investment and Machine Learning and Related Theories

### 2.1. Overview of Mass Production Investment and Machine Learning Theory

The so-called machine learning, in fact, is to use computers to simulate and realize the thinking of human brain; The so-called quantitative investment is actually the method of using computer to realize the choice of investment targets, that is, analyzing the investment objects, extracting transaction information and making a reasonable investment plan[8]. Quantitative investment emerged in 1970s, and in the following decades, this quantitative investment method has been developed by leaps and bounds, and quantitative investment has gradually been widely studied and used[9]. With the great development of information technology and computer technology, investment institutions can use emerging technologies to analyze and process data, apply mathematical models to the market, and execute computer instructions programmatically[10]. Different from traditional qualitative investment, quantitative investment uses a set computer model and completes market transactions according to certain rules[11]. In fact, it is a practical application of methodology. The application process is more combined with mathematics, statistics, information technology and other disciplines. In recent years, it has incorporated many emerging computer technologies such as machine learning, artificial intelligence, and combined with classical statistical theory to complete data prediction and estimation, The emergence and development of quantitative investment has expanded the asset allocation method and promoted the development and perfection of asset allocation theory.

In the stock market, the application of quantitative investment is the quantitative stock selection, among which the multi-factor quantitative stock selection strategy is one of the stock selection strategies which is widely used and highly accepted by the market. The factors used in this strategy mostly come from the fundamental data of enterprises and market transaction data, and the stock selection in the strategy

also has a strong logical support. Good factors mean long-term stable income, and more and more investors realize multi-factor strategies through different channels and ways in order to obtain excess returns from the market. Therefore, the nonlinear dimension can be considered when constructing the stock return prediction model. Various algorithms in machine learning can simulate nonlinear relations well, among which neural network can simulate any function in theory. The machine learning methods represented by neural network have strong self-learning ability, and certain robustness and fault tolerance.

## 2.2. The impact of machine learning on quantitative investment

Machine learning algorithm is a kind of data analysis method that can learn from historical data. It learns and perfects existing patterns from feature data sets, and is used to accomplish common tasks such as data classification, regression and unsupervised fitting. The general flow chart of machine learning is shown in Figure 1.

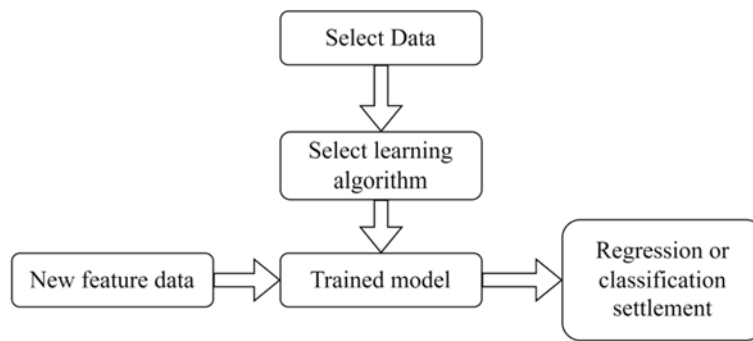


Figure 1. Machine Learning Flow Chart

Learning is a branch of artificial intelligence, and its core advantage is learning ability, which is also the difference between it and traditional algorithms. The machine can learn historical data and find its own rules. Secondly, compared with traditional programming methods, machine learning has the advantages that firstly, it can simplify the code and improve the execution ability of the code through machine learning algorithms; secondly, it can find a solution to the problems that cannot be solved by traditional methods through machine learning technology. Furthermore, machine learning can adapt to the changes of new data and realize system personalization according to different environments.

Learn the historical data of the machine, and self-explore the rules. Secondly, compared with traditional programming methods, machine learning has the advantages that firstly, it can simplify the code and improve the execution ability of the code through machine learning algorithms; secondly, it can find a solution to the problems that cannot be solved by traditional methods through machine learning technology. Furthermore, machine learning can adapt to the changes of new data and realize system personalization according to different environments. Applying machine learning to the field of financial quantification can help investors find the laws between financial data that cannot be found by manpower and reproduce them. This is based on machine learning, which can analyze massive and high-dimensional transaction data, synthesize various information, and discover potential correlation factors among financial data. Moreover, its distributed operation can greatly shorten the operation time and reduce the transaction timeliness risk.

If you want to build a successful quantitative investment model based on machine learning, you need to carefully select samples and preprocess data; The selection of machine

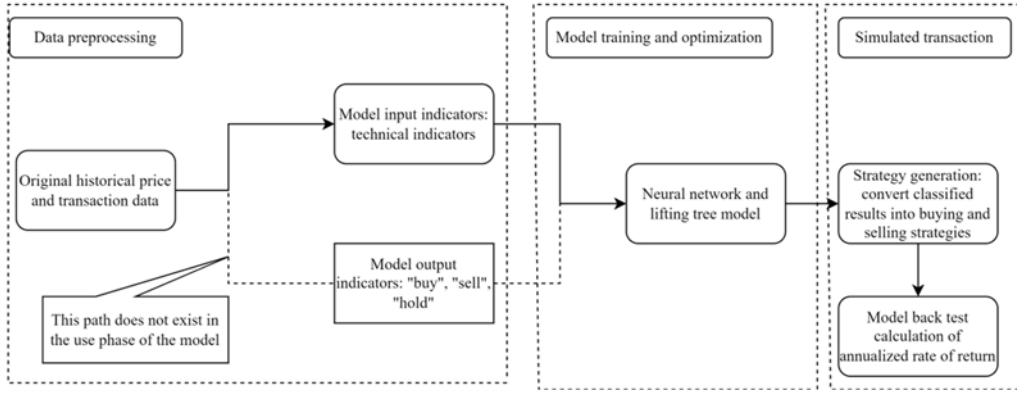
learning model, the selection of model algorithm and the setting of parameters are all complicated preparatory work; After running the machine learning model, it is necessary to modify the parameters repeatedly according to the results, test the model again and again, and optimize the model several times to obtain an optimal learning result.

## 3. Make Quantitative Investment Through Machine Learning Stock Selection

### 3.1. Machine Learning and Quantitative Investment

The ultimate goal of quantitative research is to predict the future trend of securities, and to make a profit by "buying low and selling high" according to the prediction. Quantitative timing research is to judge whether the current time point is the low point or the high point of the price only by judging the price trend of a security, and to buy at the low point of the price and sell at the high point to realize the profit.

The data processing flow of this paper is shown in Figure 2, which is mainly divided into three parts: data preprocessing, model calculation and simulated income. What needs to be noted here is that in the data preprocessing stage, the article takes the market open data source -- the price and trading volume data of historical stock transactions -- as the original input, and obtains the input indicators and output indicators of the model respectively. The "future" data is used in the calculation of output indicators, so it can be obtained in the model training and optimization stage, but in the use stage of the model, the output indicators should be given by model prediction, as shown in Figure 2.



**Figure 2.** Quantification based on historical data is the strategy research framework

In the work of the paper, we use the model with artificial features as input, which has better effect in the case of less data. The work of this paper explores the possibility of combining neural network and machine learning model to predict the stock market and make profits from it with the manually extracted features as input. It is hoped that the process of extracting features using models can be replaced by manually extracted features, so as to achieve better prediction results using simpler neural networks or machine learning models.

In order to obtain a better quantitative investment effect of large-scale assets, this paper systematically selects six machine learning models with strong classification and prediction ability to predict the income direction of large-scale assets, which plays a role in screening superior assets to assist the subsequent asset allocation. The six selected machine learning models include support vector machine, XGBoost, random forest, GBDT, BP neural network and LSTM. Here, the theoretical method of one of the models, namely BP neural model, and the basic idea of its classification algorithm are introduced.

### 3.2. Use BP algorithm for stock selection

The learning ability of multi-layer neural network is far stronger than that of single-layer neural network, and it trains multi-layer neural network through more powerful algorithms. The BP neural network model used in this paper is a neural network model based on error back propagation algorithm (abbreviated as BP algorithm). BP algorithm is a widely used neural network algorithm, which can be used to train not only multilayer feedforward neural networks, but also other neural networks.

The specific BP algorithm workflow is as follows:

(1) Input training set  $D = \{(x_k, y_k)\}_{k=1}^m$  and determine learning rate  $\eta$ .

(2) According to the input training set and learning rate, randomly initialize the connection and threshold of the neural network between (0,1)

(3) According to the set parameters and formula  $\hat{y}_j^k = f(\beta_j - \theta_j)$ , the output value  $\hat{y}_j^k$  of the training sample is obtained.

(4) According to the formula  $g_j = \hat{y}_j^k (1 - y_j^k) (y_j^k - y_j^k)$ , the gradient term  $g_j$  of the output neuron is obtained.

(5) According to the formula  $\Delta \gamma_h = -\eta e_h$ , the gradient term  $e_h$  of hidden neurons is obtained

(6) According to the updated formula  $\Delta \omega_{hj} = \eta g_j b_h$ ,  $\gamma_i = +1$ ,  $\Delta v_{ih} = \eta e_h x_i$ ,  $\Delta \gamma = -\eta e_h$ , get the updated connection weights  $\omega_{hj}$  and  $v_{ih}$ , and the thresholds  $\theta_j$  and  $\gamma_h$ .

(7) From this cycle, the stopping condition of the neural network is guided, and finally the neural network whose connection weight and threshold are determined is output.

In this paper, the financial statements of 3000 Chinese medicine listed companies in A stock market from 2017 to 2021 are selected as samples. The establishment of stock selection model consists of two parts: training samples and testing samples. To determine the sample first, this paper starts from

700 stocks from 230 stocks were selected as samples for training test. According to the following criteria, it is classified as good company and bad company: companies with negative net profit, negative main business income and negative net profit can be judged as bad companies. After a series of tests, the parameters of the BP neural network stock selection model are 20 neurons in the hidden layer, the threshold is 0.005, and the learning rate is 0.1. Through BP neural network, we can find out those companies with excellent performance and good growth in listed companies and make quantitative investment. Considering the factors of the company's growth, BP neural network is used to establish the stock selection model, and these indexes are trained and studied by BP neural network, so as to obtain a stock value feature set based on value investment. Through repeated training and testing by cross-validation, 656 stocks with investment value are finally selected. It can be seen that it is a very prescient and useful method to select stocks and invest through BP neural network.

## 4. Conclusion

As a computer language, the thinking mode of machine learning algorithm is still fixed and inflexible, and the changes brought by unexpected events in the real world are still unpredictable by computers. In the future, we need to invest more efforts and energy to contribute our own strength to the research of quantitative investment. This paper shows you the research on quantitative investment strategy of A-share market based on machine learning, especially the model

built by BP neural network. The experiment proves that it is a very prescient and useful method to select stocks and invest by BP neural network.

## References

- [1] Huang Qing Research on the application of support vector machine in China's A-share market quantification strategy -- based on Fama French three factor model [J] Times Finance, 2017 (11): 3.
- [2] Chen Siwen Stock Investment and Prediction Based on Machine Learning [J] Digital Design. CG WORLD, 2020, 009 (009): P.122-123.
- [3] Wei early morning Research on the Application of Machine Learning in the Analysis of Securities Investment [J] Information Weekly, 2019 (52): 1.
- [4] Yuan Junfeng When machine learning is encountered in the field of financial investment [J] AI, 2018 (5): 6.
- [5] Huo Yan Empirical Research on Credit Rating and Credit Debt Default -- Based on Machine Learning Algorithm [J] Financial Regulation Research, 2022 (3): 21
- [6] Su Bingzhi Exploration of portfolio investment based on machine learning algorithm [J] Knowledge base, 2017 (15): 2.
- [7] Li Bin, Shao Xinyue, Li Yueyang Research on Fundamentals Quantitative Investment Driven by Machine Learning [J] China Industrial Economy, 2019 (8): 19.
- [8] Wang Yunkai, Lan Jinhui ML-FFA: Quantitative Investment Strategy Based on Machine Learning and Fundamental Factor Analysis [J] Times Finance, 2018 (32): 3.
- [9] Lin Yaohu, Liu Shancun, Yang Haijun Research on a Stock Market Investment Strategy Based on Machine Learning and Candle Chart [J] Journal of Econometrics, 2022, 2 (1): 15.
- [10] Yin Xingguang Research on Portfolio Selection Based on Machine Learning [J] Information System Engineering, 2021 (12): 4.
- [11] Bao Junying, Shi Chengxiang Research on the Application of Machine Learning in China's High Yield Debt Investment [J] Journal of Chongqing Second Normal University, 2021, 34 (3): 7.