ISSN 2616-5902 Vol. 3, Issue 4: 27-30, DOI: 10.25236/AJBM.2021.030407

Review on the Application of Machine Learning in Stock Forecasting

Shengyin Luo

School of Information, Southwest Petroleum University, Nanchong, Sichuan 637001, China

Abstract: *With the improvement of people's awareness of financial management, more and more people put their surplus assets into the stock market. In the dynamic and complex stock system, the future rise and fall of each stock are unknown to a certain extent. This unknown has caused discussion among investors and related scholars, and thus stock forecasting has become a hot topic in the research field. By combing the relevant work of literature in the recent three years, we can find that the research topic of stock forecasting is mainly in the field of machine learning, and its research scope is mainly deep learning and data mining. In the process of applying machine learning to stock forecasting, the influencing factors mainly include stock data collection and preprocessing machine learning model, and machine learning algorithm. After analysis, it is found that the final prediction result of the stock is mainly related to stock data and the machine learning model. Aiming at the deficiency of prediction, the summary and prospect of applying machine learning in stock prediction are discussed at the end of this paper.*

Keywords: *machine learning, stock forecasting, deep learning, data mining*

1. Introduction

Traditionally, two main methods have been proposed to predict the stock price of an organization, which are basic analysis and technical analysis. Nowadays, advanced intelligent technology based on basic analysis and technical analysis has been used to predict stock price. Compared with the past methods, machine learning technology in this field has proved to improve efficiency by 60-86% [1].

In the previous field of machine learning, most of the work used classical algorithms, such as linear regression [2], random walk theory (RWT) [3], moving average convergence/divergence (MACD) [4] Some linear models are also used, such as autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) [5], used to forecast the stock price. Recent work shows that machine learning can be used to enhance stock market forecasting. Support vector machine (SVM). Random forest (RF) [6] Technology such as. Some neural network-based technologies, such as Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recursive Neural Network (RNN), and Deep Neural Network (such as Long-term Short-term Memory (LSTM), also show gratifying results [7].

2. Introduction to Machine Learning

The idea of machine learning is a simulation of the learning process in human life. In the whole process, the selection, collection, and processing of data are the most critical. The core of machine learning is to "use algorithms to analyze data, learn from it, and make decisions or predictions about something in the world." [2].

In stock forecasting, data mining is mainly provided by machine learning through data analysis technology. The results of data mining will be used in stock forecasting decisions. The relationship between machine learning, deep learning, and data mining are shown in the following figure:

ISSN 2616-5902 Vol. 3, Issue 4: 27-30, DOI: 10.25236/AJBM.2021.030407

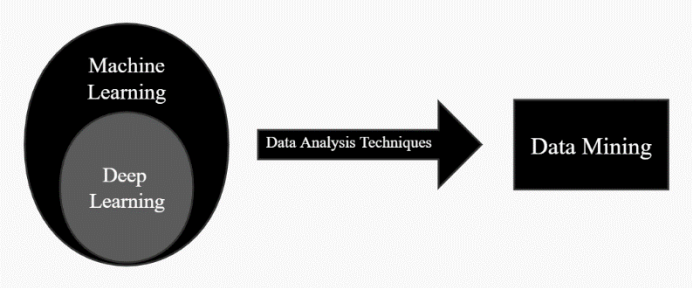

Figure 1: Relationship diagram among machine learning, deep learning, and data mining

3. Application process of machine learning in stock forecasting

According to the application of machine learning, the following application flow chart can be drawn:

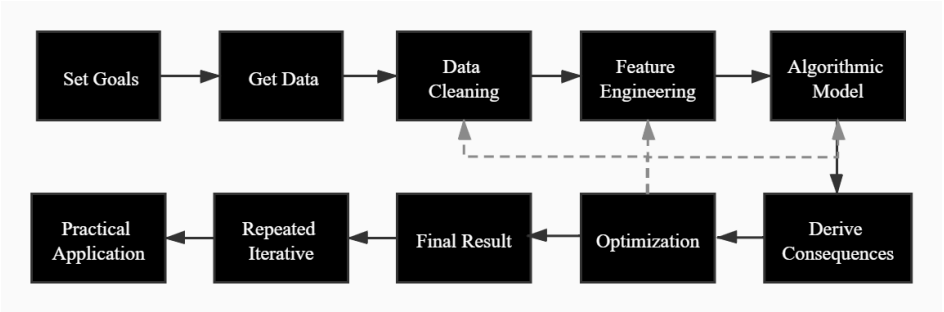

Figure 2: Application flow chart of machine learning

3.1 Stock data collection and preprocessing

In theory, if we use a model based on "deep learning", as long as the amount of input data is enough, using the essence of "deep learning" self-learning and continuous iterative optimization, the data with low correlation and poor quality will be gradually filtered, and the quality of the initial data will be less important. But in machine learning applications, there is a very classic saying "Garbage in. Garbage out." [8] As the name implies, this sentence means that, because the data itself is problematic, the data features will become very irregular, resulting in many missing values, and problems such as the lack of close connection between features and results. Therefore, in the case of low data correlation and poor quality, we should preprocess the data as much as possible to ensure the quality of the data and avoid causing deviation [9].

4. Machine learning model

For machine learning, stock forecasting is applied to its supervised learning part. A summary diagram of machine learning algorithms that can be used for stock forecasting is as follows:

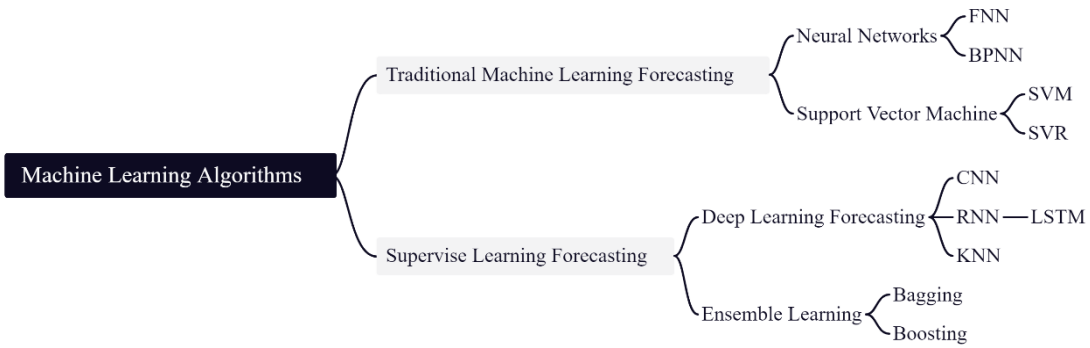

Figure 3: Summary diagram of machine learning model implementation algorithm

ISSN 2616-5902 Vol. 3, Issue 4: 27-30, DOI: 10.25236/AJBM.2021.030407

Based on many research papers, the summary graph is formed by analyzing and integrating research hotspots, mainly including traditional machine learning prediction and supervised learning prediction, both of which are nonlinear structural models.

4.1 Traditional machine learning prediction

4.1.1 Neural network

BP (back propagation) Neural Network [10] In essence, it is a gradient descent method, which will make the algorithm inefficient when the objectives to be optimized are complex; Also, BP algorithm is easy to fall into the local optimal solution, which may lead to abnormal prediction results.

4.1.2 Support vector machine

The classical support vector machine algorithm only gives two kinds of classification algorithms. But in the practical application of data mining, it generally needs to solve the problem of multi-class classification. In stock forecasting, the situation belongs to the latter, which will involve the yield, price-earnings ratio, golden section rate, and other factors, so the support vector machine method can not make the final data reach the upper limit of machine learning.

4.2 Supervised learning prediction

4.2.1 Deep learning prediction

The idea of a feedforward neural network (FNN) is relatively simple by using the FM decomposition model structure, and several full connection layers are directly connected to FM. Using DNN to implicitly cross-feature can reduce the work of Feature Engineering and control the computational time complexity in a reasonable range. But even such FNN will fall into the problem of local optimization.

In deep learning, CNN (convolutional neural network) has been well used in image recognition. In stock research, investors will also use a K-line graph to study and analyze the stock market. It is found that the selection of convolution sample size should not be too large. Otherwise, the accuracy will decline.

After using RNN, there will be the problem that the data obtained during testing are accurate, but the data obtained during practical application are abnormal. This is because RNN had only short-term memory. To overcome the defects of RNN, scholars have developed the long-term and short-term memory network LSTM. It combines short-term memory with long-term memory through subtle control and solves the problem of gradient disappearance to some extent [11]. As far as the current research results are concerned, LSTM analysis of quantitative investment in stock forecasting is the best choice.

4.2.2 Integrated learning

Compared with a traditional single machine learning algorithm, ensemble learning often has higher accuracy. Random forest, XGBoost, and other methods also have the function of feature selection, which is convenient for processing stock data containing multiple features. At the same time, it is highly interpretable and can be displayed visually and easily understood by users. However, this method is mainly applied to the classification of the rise and fall in stock forecasting, and the regression problem of the corresponding stock price cannot be predicted.

4.2.3 Evaluation and Optimization of Machine Learning Algorithms

Each machine learning algorithm has its inherent characteristics, which can solve the prediction problem in different application scenarios to a certain extent. Moreover, the application of machine learning in stock forecasting is also applicable to the forecasting of other complex types of data. When processed data reach the upper limit of machine learning, the results of different machine learning algorithms will approach the upper limit, so data preprocessing is a very important part of machine learning. As far as the final prediction results are concerned, when different algorithms are combined with LSTM, the prediction accuracy is higher.

5. Summary and prospect

As a major problem that puzzles researchers, stock forecasting has made good progress in recent years. Machine learning algorithms, such as support vector machine, deep learning, and ensemble learning, have received extensive attention in this field. However, each model has its unique advantages

ISSN 2616-5902 Vol. 3, Issue 4: 27-30, DOI: 10.25236/AJBM.2021.030407

and disadvantages, such as low prediction accuracy and poor interpretation. Improving the prediction accuracy, increasing the amount of data, fusing models, and enhancing the interpretability of models needed further in-depth study in the future.

References

[1] Mehar V, Deeksha C, Vinay A T, et al. Stock Closing Price Prediction using Machine Learning Techniques [J]. Procedia Computer Science. 2020, 167.
[2] Junze Qu. Constructing stock trading decision-making framework based on machine learning [J]. China High-tech .2019 (7): 54-56.
[3] Bisoi R, Dash PK, Parida A K. Hybrid Variational Mode Decomposition and evolutionary robust kernel extreme learning machine for the stock price and movement prediction on daily basis [J]. APPLIED SOFT COMPUTING. 2019, 74: 652-678.
[4] Yonghao Yang. Stock price trend prediction method based on optimized MACD model [j]. Journal of Guangxi academy of sciences, 33 (1): 65-70.
[5] Haoran Xu, Bo Xu, Kewen Xu. Overview of the application of machine learning in stock forecasting [J]. Computer Engineering and Application .2020, 56 (12): 19-24.
[6] Nana Lin, Jiangtao Qin. A-share stock price forecast based on random forest [J]. Journal of the University of Shanghai for Science and Technology.2018, 40 (03): 267-273.
[7] Fengxin Deng, Hongliang Wang. Application of LSTM neural network in stock price trend prediction-based on the data of individual stocks in us and Hong Kong stock markets [J]. Finance and economics. 2018 (14): 96-98.
[8] Cao H, Lin T, Li Y, et al. Stock Price Pattern Prediction Based on Complex Network and Machine Learning [J]. COMPLEXITY. 2019 (4132485).
[9] Hongsheng Liu, Huawei Feng, Zhang Li, et al. Application of machine learning in MALDI-TOF MS identification of microorganisms [J]. Acta Microbiologica Sinica, 2020(5): 841-855.
[10] Fujia Gao. Review of stock price model based on BP neural network [J]. Economic Outlook around the Bohai Sea. 2019(12): 164-165.
[11] Sheng Lin, Ke Qi, Jiecong Wei, et al. Research Review of Machine Learning in Stock Price Forecasting [J]. Economist .2019 (3): 71-73, 78.