Stock prediction in machine learning: Comparison and analysis

Shengmao Bai

High School Affiliated to Shanghai Jiao Tong University, Shanghai, China

baishengmao@jdfzib.org

Abstract. In recent years, the rapid development of machine learning has made it go into the field of stocks prediction. There have already been a few mature algorithms in the world that could help analyze the futural movements of the stocks market. In this paper, the author aims to exploit the effectiveness of machine learning methods in stock prediction. The utilized algorithms contain SOM(Self-organizing Maps) and SVR(Support vector regression) SVM, RNN, LSTM as well as work on several papers from others to estimate the probable advantages or disadvantages of the algorithms. After these experiments and testing, the author finds out some conclusions about the algorithms and realizes that the SOM algorithm has the advantage of exploiting the neighbor relationship on the cluster centroids is beneficial to the interpretation of the clustering results while having the disadvantage that users must select parameters, neighbor function, grid type and number of centroids.

Keywords: Stock prediction, machine learning, support vector machine.

1. Introduction

With the emerging development of machine learning based method[1,2,3,4], many methods with novel architecture are proposed, such as AlexNet[5], VGG[6], ResNet[7], DenseNet[8] and inception[9] network.Stock forecasting refers to the behavior of securities analysts who have a deep understanding of the stock market to predict the future development direction and the degree of fluctuation of the stock market according to the development of the stock market. This forecasting behavior is only based on assumed factors and established preconditions. The importance of stock prediction is that the financial market has certain rules. If it is completely chaotic and irregular, it is like buying a lottery ticket, and predictions are meaningless. We know that there are random tickets to buy lottery tickets. A lot of times, random selection is better than thinking about it yourself. That's because it's completely random and unpredictable in the lottery. So is this the case in the stock market? We can say with certainty that this is not the case, the stock market has certain rules. Various historical trends of various indicators are useless if there is no pattern at all. Rather than research, draw lots to decide which stock to pick. Why stocks are worth studying is precisely because they are based on historical representations.

Only when history repeats itself can stocks have the value of research and the basis and conditions for prediction. Since the stock market will not simply repeat itself, there will be some changes that analysts want to study and that individual investors lack. It is because history can repeat itself that we have prophecies. Prediction does not mean blindly relying on luck and so-called intuition, but a combination of various comprehensive methods to study the future direction of the stock market. Through this direction, you can buy stocks and do stocks to see the general trend. Real analysts make

forecasts, because forecasting is the only way to test their true skills in the stock market. So, how can we make effective stock market forecasts? Using some machine learning algorithms can help us achieve this goal quickly. A variety of different machine learning algorithms can make predictions about future stock markets to some extent from past data paradigms and nascent data. And a good machine learning algorithm can obviously make it easier for us to get accurate prediction results. Therefore, the purpose of writing this paper is to integrate and summarize the existing stock market forecasting computer algorithms, analyze their own advantages and disadvantages through their forecasting results, and provide ideas for the evolution trend of future stock market forecasting algorithms.

2. Machine learning algorithms for stock prediction

Principal Component Analysis is a data dimensionality reduction algorithm. It transforms high-dimensional data into lower-dimensional data, thus maintaining the main trends in the data and eliminating abnormal data and unimportant features. This method increases the speed of data processing, but also faces some data loss. This method reduces computational costs and gives more obvious results.

In this section, 3 papers are reviewed based on stock prediction. First, in this paper[10], the authors argue that machine learning (ML) has become an integral part of the stock market forecasting process. Various ML algorithms, in their own unique ways, have helped make stock market predictions more accurate and have made fortunes for many individual investors. Therefore, they obtained the advantages and disadvantages of several computer algorithms through the calculation of a large amount of data and the analysis of the existing prediction results. Among them, Recurrent Neural Network(RNN) has a strong ability to display time, but it has the problem of difficult training. SOM（Self-organizing Maps） is very robust in parameter selection, but its sensitivity to specific data is a big problem. Supported Vector Machine(SVM) can provide the optimal global solution, and the prediction results are also very accurate, but it also has the problem of being very sensitive to outliers. SVR(Support Vector Regression) is more suitable for processing multiple inputs, and the prediction is more accurate, but it is very sensitive to user-defined free parameters. In summary, this paper summarizes most of popular machine learning algorithms for stock prediction and provided a comprehensive review on them.

In reference[11], the author retains that if the model is accurate enough, predictions of stock prices can become very accurate. So, he reaches a conclusion that in order to get a close prediction to the stock market, they should use the appropriate algorithm. The author's experimental result gives us important insights into stock prices prediction. The author obtained very accurate predictions of stock price movements in the short term by using LSTM, which can be a great help in solving problems about price movements. The architecture of the proposed network is shown in Figure 1.
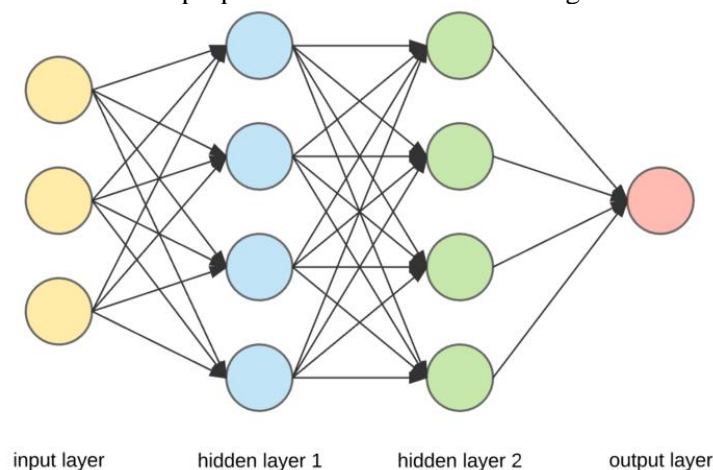

Figure 1. Architecture of the proposed method[11].

The author also tries to predict the stock market through several major popular algorithms like SOM\SVR\SVM\LSTM\RNN. The results show that SVR has strong predictive ability, and the forecast accuracy is also greatly improved in periods of low volatility. Using random walk models for some blue-

chip and small-cap stocks in the three countries studied, the authors find that SVR yields poor predictions relative to random walk models for nearly all the stocks studied.

In the paper[12], the author mentions that with the rapid development of the economy, more and more people begin to invest in the stock market. Although the stock price will be affected by many factors, using computer algorithms to predict can still effectively help people understand the actual situation of the stock market and choose the right stock to invest. This article proposes an algorithm called CNN to predict the closing price of the stock market the next day. It is mentioned in the article that CNN is mainly composed of three parts: conversion layer, pool layer and full connection. The experimental results show that the sharing of CNN algorithm can greatly reduce the pressure of parameter model, thus improving the learning efficiency model. Figure 2 shows the visualization results of [12].

Figure 2. Visualization result of the proposed method[12].

3. Experimental results

We also conduct comparison on different algorithms for stock prediction, which is shown in Table 1. As shown in the table, all the methods have great potential on handling stock prediction task. Furthermore, the SVC and SVM get the top performance, which achieve 100% precision. All other algorithms have satisfying performances.

Table 1. Comparison results of different methods on public dataset.

Method	Precision	Recall	F1-score
SVC	100%	100%	100%
CNN	99%	99%	99%
SVM	100%	-	-
Naïve Bayes	98.61%	-	-

4. Conclusion

In this article, the author mainly summarized the effects, advantages and disadvantages of five computer algorithms in several papers, namely SOM, SVR, SVM, RNN and LSTM, on stock forecasting. Among them, SOM has the advantage of visualization, and the network is self-stable, but the number of clusters and the initial network structure are fixed, so users need to specify the number of clusters and the initial weight matrix in advance. The advantage of SVR is that the computational complexity does not depend on the dimensions of the input space. In addition, it has excellent generalization ability, high prediction accuracy and support vector regression. However, it is not suitable for a large number of dataset operations. The advantage of SVM is that it can map to high-dimensional space and solve nonlinear classification problems by using kernel functions. However, it will be difficult to solve the multi

classification problem. The advantage of RNN is that it can apply the previously obtained information to the current task, and the output results at each moment are not independent of each other. However, it cannot solve the problem of long-term dependence. LSTM algorithm can easily add more computing cells, but it is difficult to add more storage bandwidth. Finally, the author put forward the prospect of computer algorithms in the field of stock forecasting. The current public data may be less, which is not conducive to the learning of deep neural networks. More databases should be extended, and use a large number of data to obtain the advantages and disadvantages of various existing computer algorithms and integrate them, so as to obtain new and more effective computer forecasting algorithms through a large number of calculations.

References

[1] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

[2] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[3] Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., ... & Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. ACM Computing Surveys (CSUR), 51(5), 1-36.

[4] Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. IEEE Access, 7, 53040-53065.

[5] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

[6] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[7] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[8] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[9] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[10] Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., & Wills, G. (2020). A survey on machine learning for stock price prediction: algorithms and techniques.

[11] Sen, J., & Chaudhuri, T. D. (2018, December). Stock price prediction using machine learning and deep learning frameworks. In Proceedings of the 6th International Conference on Business Analytics and Intelligence, Bangalore, India (pp. 20-22).

[12] Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. The Journal of finance and data science, 4(3), 183-201.