COMP7705 Project

Detailed Project Proposal

Project Title: Stock Price Analysis and Prediction

Supervisor: Professor S.M.Yiu

Student 1 (Leader) Chen Zixuan

Student 2 Chen Pengxiu

Student 3 Peng Wanlan

Student 4 Yan Jun

Student 5 Ye Siqi

Aim

In this project, we aim to analyze the price of several stocks in Hang Seng Index (HSI) In the past decade and seek the features to help us predict its trend.

Generally, people do stock price prediction with time series analysis. In the procedure of time series analysis, we always use several kinds of Moving Average (MA), variance and standard deviation to grasp the features of the price. Then, we can apply certain index to chase the trend, such as: Moving Average Convergence/Divergence (MACD), Relative Strength Index (RSI) and Boll Line. These methods can really help us traditionally.

However, with the development of Artificial Intelligence, we want to apply the methods of deep learning in stock price analysis.

In the first step, we use historical data only. After the pre-processing of data, we want to use RNN neural network and add LSTM layer to it. Thus, after training, we can verify the output and compare the results with traditional methods.

In the second step, we want to add more message and features in the data like real-time public opinions and compare the result with the former work. In the third step, we would like to try some other networks and add certain layer to optimize the results.

Brief Literature Review

Stock market data are highly complex and difficult to predict, even for human experts, due to a number of external factors, e.g., politics, global economy, and trader expectation. The trends in stock market data tend to be nonlinear, uncertain, and non-stationary. In most of the cases, proposed approaches have been mainly based on statistical models used to analyze time series of interest. such as: Moving Average Convergence/Divergence (MACD), Relative Strength Index (RSI) and Boll Line.

However, the problem of time-series is that stock prediction is often viewed as a single channel problem, which explains the difficulties to produce accurate prediction systems, since stocks depend on a myriad of other factors, and arguably not at all on past values of the stock itself.

On the other hand, with the development of deep learning, this area has aroused the interest of academia and practitioners, especially in the field of finance. Deep learning methods offer better representation and classification on a multitude of time-series problems compared to shallow approaches when configured and trained properly. As a result, different types of models have been considered about solving the problem of predicting asset price movements and the behaviors in time of more structured financial instruments. In previous works, better predicting results have been obtained using linear classifiers as the logistic regression one, which has been used to predict the Indian Stock market, see [1], or with respect to the S&P500 index, see [2]. More complicated techniques, as large-margin classifier or Support Vector Machine (SVM) were also been tried.

After the rise of neural network. Based on the type of application, various types of deep neural network architectures are used in financial areas. These include multi-layer perceptions (MLP), Recursive Neural Networks (RNN), Long Short-Term Memory (LSTM), CNN (Convolutional Neural Network) etc. [3]. A recurrent neural network (RRN) is any artificial neural network whose neurons send feedback signals to each other. The idea behind RNNs is to make use of sequential information. In a traditional neural network, we assume that all inputs (and outputs) are independent of each other. But for many tasks this is not the better idea. In particular, if one wants to predict the next word in a sentence, then it is better if he knows which words came before it. RNNs are called recurrent because they perform the same task for every element of a sequence, with the output being depended on the previous computations. With the introduction of LSTM [4], the analysis of time dependent data become more efficient. These types of networks have the capability of holding past information. They have been used in stock price prediction by [5], [6] .

However, most approaches of RNN applied to stock prediction have given unsatisfactory results [9]. This is because the price of a stock is not only related to its own historical trend, but also to a large extent will be affected by emergencies and public opinions. A major earthquake in Chile, for example, would shake public opinion and would inevitably affect the share prices of Chile's proprietary energy.

In summary, it can be concluded that there is still room to improve existing techniques for making safe and accurate stock prediction systems. If additional information from sources that affect the stock market can be measured and obtained, such as general public opinions from social media[7], trading volume[8], market specific domain knowledge, and political and economic factors, it can be combined together with the stock price data to achieve higher stock price predictions[9].

[1] S.-H.Chen, Genetic Algorithms and Genetic Programming in Computational Finance, Springer, 2002

[2] A. Dutta, G. Bandopadhyay and S. Sengupta, Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression, *Inter- national Journal of Business and Information*, 2012

[3] J. Heaton, N. Polson, and J. Witte, "Deep learning in finance," arXiv preprint arXiv:1602.06561, 2016.

[4] H. Jia, "Investigation into the effectiveness of long short term memory networks for stock price prediction," arXiv preprint arXiv:1603.07893, 2016.

[5] Y. Bengio, I. J. Goodfellow, and A. Courville, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[7]Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. Journal of Computational Science 2, 1-8

[8]Zhu, X., Wang, H., Xu, L., Li, H., 2008. Predicting stock index increments by neural networks: The role of trading volume under different horizons. Expert Systems with Applications 34, 3043-3054.

[9]Agrawal, J.G., Chourasia, V.S., Mittra, A.K., 2013. State-of-the-art in stock prediction techniques. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 2, 1360-1366.

Proposed Methodology

The method we used in this project mainly contains two parts:
1. statistical models: Moving Average Convergence/Divergence (MACD), Relative Strength Index (RSI) and Boll Line
2. machine learning method and neural network: multi-layer perceptions (MLP), Recursive Neural Networks (RNN), Long Short-Term Memory (LSTM), CNN (Convolutional Neural Network)

We want to compare the result of different methods and optimize the neural network to get better result.

Milestones

	Tasks	***Estimated completion time***	***Estimated number of learning hours***
1	Project Webpage	June 1, 2020	30 hours
2	Statistical methods analysis	June 1, 2020	75 hours
3	Basic neural network analysis	June 15, 2020	75 hours
4	Optimization of machine learning methods	July 1, 2020	75 hours
5	Poster design	July 15, 2020	15 hours
6	Final report and conclusion	August 1, 2020	30 hours
7			
8			
9			
10			
			Total: 300

Deliverables

Items	
1	A detailed report between traditional methods and machine learning methods on stock price analysis
2	Optimization of machine learning method
3	Grasping more information and features on this field
4	
5	
6	
7	
8	
9	
10	