Analysis of Stock Market Prediction

Ulugbek Latipov[1], Wei Wang[2]

Zhejiang University of Science and Technology, School of Science, Hangzhou, China

Abstract: *Stock market prediction involves predicting future value of company stock or other financial instrument traded on an exchange. Various types of trading can be done in stock market. It could be short term trading or even long term trading but if someone can predict the value or class of that entity, it can yield very good return for the investment done. Prior to evolution of digital world, predictors continued to use paper work methods like fundamental and technical analysis. Various useful technical indicators such as SMA, EMAand MACD found to be very useful however with the advent of computer technologies and algorithms, prediction moved into technological realm. Analysts started building prediction system using Neural Network, Support Vector Machine, Decision Trees and Hidden Markov Model. Prediction accuracy really improved using algorithmic approach as well as evolutionary data mining techniques used for stock market prediction.*

Keywords: Stock Market, Data mining, Technical indicators, Support vector Machine, Neural Network.

1. Introduction

Stock Market prediction has always had a certain appeal for researchers. While numerous scientific attempts have been made, no method has been discovered to accurately predict stock price movement. The Goal behind making any financial investment is to achieve above average return for invested money while maintaining certain level of involved risks [1] however the difficult of prediction lies in the complexities of modeling market dynamics. Even with a lack of consistent prediction methods, there have been some mild successes. Stock Market research encapsulates two elemental trading philosophies as Fundamental and Technical approaches. In Fundamental analysis stock market price movements are believed to derive from a security's relative data. Fundamentalists use numeric information such as earnings, ratios and management effectiveness to determine future forecasts. In Technical analysis, it is believed that market timing is a key. Technicians utilize charts and modeling techniques to identify trends in price and volume. These later individuals rely on historical data in order to predict future outcomes [2]. One area of limited success in Stock Market prediction comes from textual data. Information from quarterly reports or breaking news stories can dramatically affect the share price of security. Most exciting literature on financial text mining relies on indentifying a predefined set of keywords and machine learning techniques. These methods typically assign weights to keywords in proportion to the movement of a share price. These types of analyses have show a definite, however weak ability to forecast the direction of share prices.

2. Fundamental Analysis

At the company level fundamental analysis includes analysis of financial data, management reports, business concepts and competition. It also explore the relation between financial statement information and fundamental attributes such as revenue rate growth, price to book ratio [3],[4]. At the industry level, there might be an examination of supply and demand forces for the products offered. At economy level fundamental analysis might target economic data to assess the present and future growth of the economy. To predict future stock prices, fundamental analysis combines economic, industry and company analysis [5],[6],[7] to derive a stock's current fair value and forecast value. If fair value is not equal to the current stock price, fundamental analysts believe that the stock is either over or under valued and the market price will ultimately gravitate towards fair value. Fundamentalists do not pay attention to the advice of the random opinions and believe that markets are weak from efficient. By believing that prices do not accurately reflect all available information, fundamental analysts look to capitalize on perceived price discrepancies. Various aspects regarding fundamental analysis are examining business plan, management and financial analysis. Advantage of fundamental analysis is to find out long term trend. It also helps to uncover companies with valuable assets, a strong balance sheet and stable earning.

3. Technical Analysis

Large number of technical indicators is available for technical analysis. They use various statistics generated in the market like closing prices of history, volume traded. Earlier in 1960s and 1970s several researchers studied trading rules based on TIs. Though they did notified them much profitable [8], [9] recent studies [10], [11] show that they are very useful. Commonly used TIs are simple moving averages (SMA), exponential moving averages (EMA), moving average convergence (MACD), exponential moving average (EMA) and relative strength index (RSI).

4. Opinion Mining

With the growth of the web over the last decade, opinions can be found almost everywhere – blogs, social networking sites like Facebook and Twitter, news portals and ecommerce sites. While opinions are meant to be helpful, the vast availability of such opinions becomes overwhelming to users when there is just too much to digest. The simplest form of an opinion summary is the result of sentiment prediction.

The task of sentiment prediction or classification itself has been studied for many years. Beyond such summaries, the newer generation opinion summaries include structured summaries that provide a well-organized breakdown by

aspects, various formats of textual summaries and temporal visualization.

5. Machine Learning Methods

Artificial Neural Network (ANN) – has several advantages but one of the most recognized of these is the fact that it can actually learn from observing data sets. In this way, ANN is known for as a random function approximation tail. These types of tools help estimate the most cost-effective and ideal methods for arriving at solutions while defining computing functions or distributions. ANN uses data samples instead of complete data sets to arrive at solutions, which saves both time and money. ANNs are considered fairly simple mathematical models to improve effectiveness of the available the data analysis technologies. ANNs includes three layers. These layers are connected to each other. The first layer consists of input neurons. Those neurons send data on to the second layer, which in turn sends the output neurons to the third layer. Training an artificial neural network involves choosing from allowed models for which there are several associated algorithms.
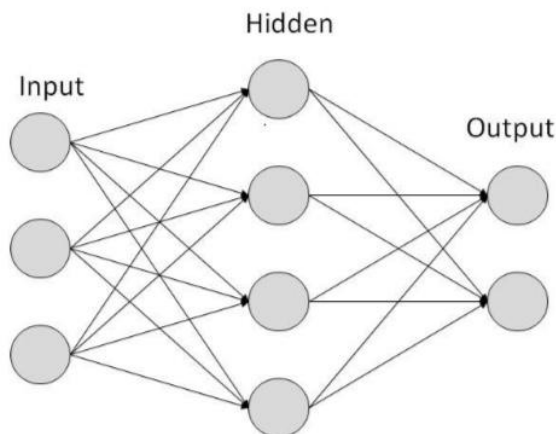

Figure 1: Artificial Neural Network

Phua [12] applied Neural Networks to the financial prediction. It was tested the influence of volume data on Stock price prediction. Khan [13] applied the Neural Networks with different number of hidden layers to analyze the prediction of the Stock prices.

Support Vector Machine (SVM). Although SVM can be applied to various optimizations such as regression, the typical problems are to classify the data. The basic idea is shown in figure 2. The data points are indentified as being positive or negative and the problem is to find a hyper-plane. This plane separates the point data by maximal margin. SVM falls under a category of supervised machine learning algorithm which can be used for both classification and regression challenges. However, researchers mostly use it for classification problems.In this algorithm, we plot each data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.
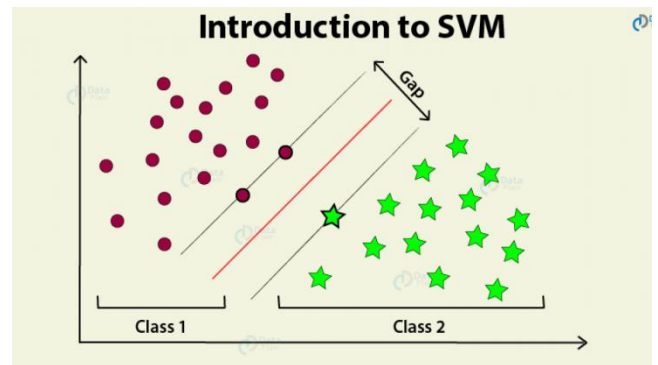

Figure 2: Support Vector Machine

Above the figure shows the 2-dimensional case where the data points are linearly separable SVM approach to linear regression amounts to minimization of ε- insensitive loss and minimization of the norm of linear parameters. This can be formally described by introducing slack variables, to measure the deviation of training samples outside ε - insensitive zone. Pai [14] proposed a hybrid approach with SVM and ARMA model and found it gave promising results.

Hidden Markov Model is a finite state machine. This has some fixed number of states. It gives a probabilistic framework for modeling a time series of multivariate observations. Hidden Markov Models were introduced in the beginning of the 1970's. It is used as a tool in speech recognition. This model which is based on statistical methods has become increasingly popular in the last several years due to its strong mathematical structured basis as it is used a wide range of applications.
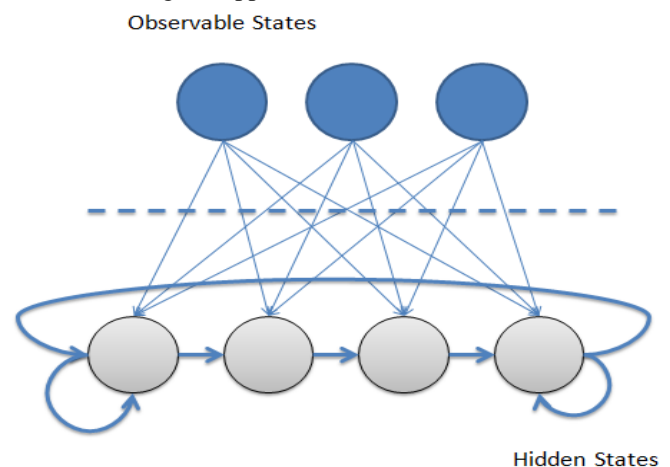

Figure 3: Hidden Markov Model

Recently researchers proposed HMM as a classifier or predictor for speech signal recognition, DNA sequence analysis, hand written characters recognition, natural language domains. It shows that HMM is a very powerful tool for various applications. The advantage of HMM can be summarizes as:
- HMM has strong statistical foundation
- It is able to handle new data robustly
- Computationally efficient to develop and evaluate
- It is able to predict similar patterns efficiently

Decision Trees builds classification or regression models in the form of a tree structure. It breaks down a dataset into

small and then even smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches. Leaf node represents a classification or decision. The top most decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categories and numerical data.
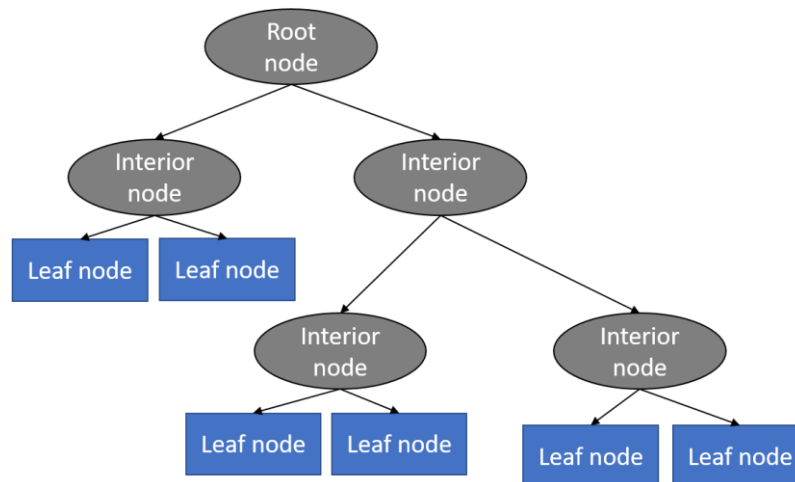

Figure 4: Decision Tree

A decision tree is build top-down from a root node and involves partitioning the data into subsets that contain instance with similar values. ID3 algorithm uses entropy to calculate the homogeneity of sample. If the sample is completely homogeneous the entropy is zero and if the sample is an equally divided it has entropy of one. The information gain is based on the decrease in entropy when a dataset is split on an attribute. Construction a decision tree is all about finding attribute that returns the highest information gain.

6. Conclusion

This paper concludes that though various approaches and techniques are available to increase profit in stock market investment, every methods has its advantages and limitations. Fundamental analysis really helps to find a stock's intrinsic value but it is not much profitable fort term trading. Technical indicators look to predict the future price levels by looking at past patterns and hence useful for long term trading as well as short term trading. SMA smoothen the price movement thus eliminating most fake outs however it also cause a lag in buying and selling signals. EMA reduce the lag by applying more weight to recent prices hence better than SMA in terms of recent movements in the market. Benefit of using RSI is that it immediately indicates the overbought and oversold levels to traders but since the indicator is showing momentum as long as momentum remains strong the indicator can stay in overbought or oversold territory for long periods of time. Therefore, price analysis or some other confirmation is still needed for reversals. The MACD indicator is a so called trend following indicator. With the moving average, a trend can be found with the MACD, the strength of the trend and the possible turning points can be determined but sometimes it is difficult to find reversal in the market using MACD. Because of usefulness and needs from the people, opinion mining became an active research area. As the volume of the opinionated data increases, analyzing and summarizing opinionated data is becoming more important. To satisfy these needs, many kinds of opinion summarization techniques are proposed. Machine learning methods have also their advantages and limitations. Neural network is really an adaptive learning method having well self-organized structure but it sometimes it converges on local minima in optimization problem. Over fitting is another issue with neural network. Over fitting occur in complex decision trees too. Hard concept learning could be difficult in decision trees. Learning model parameter is another constraint in decision trees. HMM uses large number of parameters resulting into large amount of data needed to train it. Although SVMs have good generalization performance, they can be abysmally slow in test phase. Though having limitations with every non algorithmic techniques if properly applied we can predict stock market prices at some extents but use of machine learning algorithm have shown better results. We can predict value as well as trend effectively.

References

[1] Yu-Feng Lin, Chien-Feng Huang, Vincent S. Tseng. (2012). A Novel Methodology for Stock Investment using Episode Mining and Technical Indicators. Conference on Technologies and Application of Artificial Intelligence.

[2] Vivek Rajput, Sarika Bobde. (2016). Stock Market Forecasting Techniques: Literature Survey. International Journal of Computer Science and Mobile Computing. Vol 5, Issue. 6, pp. 500-506.

[3] John Sneed "Estimating earning forecasting model using fundamental analysis: Controlling for different across industries", American business review, pp. 17-24, 1999.

[4] Vastone B., Finnie G. and Tan. C., "Applying fundamental analysis and neural network in the Australian stock market", Proceeding of international conference on Artificial intelligence in Science and

Technology, Hobart, Tsmania, pp. 21-25, November 2004.

[5] E.I. Altman, "Financial ratios, discriminate analysis and the prediction of corporate bankruptcy", Journal of Finance, vol. 23, 1968, pp. 589-609.

[6] S.P. Kothari., "Capital markets research in accounting ", Journal of Accounting and Economics, vol. 31, 2001, pp.105-231.

[7] D. Nissim and S.H. Penman., "Ratio analysis and equity valuation: from research to practice", Review of Accounting Studies, vol. 6, pp. 109-154.

[8] Alexander S. (1961), "Price movements in speculative markets: trends or random walks, Industrial Management Review, pp. 7-26.

[9] Fama E.F. and Blume M.E., (1966), "Filter Rules and Stock-Market Trading", Journal of Business, 39, Part 2: Supplement on Security Pricing, pp 226-241.

[10] Fernandez-Blanco, P., Bodas-Sagi, D. J., Soltero, F. J., and Hidalgo, J.I., 2008, "Technical market indicators optimization using evolutionary algorithm.

[11] Bodas-Sagi, D. J., Fernandez, P., Hidalgo, J.I., Soltero. F.J., (2009), "Multi-objective optimization of technical market indicators.

[12] X. Wang, P.K.H. Phua and W. Lin., "Stock market prediction using neural networks: does trading volume help in short-term prediction", in Neural Networks, 2003. Proceedings of the International Joint Conference on, vol.4 IEEE, 2003, pp. 2438-2442.

[13] A.U. Khan, "Stock rate prediction using back propagation algorithm: Analyzing the prediction accuracy with different number of hidden layers", Glow gift, 2005.

[14] Luigi Troiano, Pravesh Kriplani, "Predict Trend in the Next-Day Market by Hierarchical Hidden Markov Model".