In this paper, we focus on the topic stock prediction in e-finance area. We proposed a hybrid model to predict stock price change. We classified the stock price change as two classes, up and down. Also, the training data set in our model is not universal. We applied opinion mining technique to get the opinion on current stock market condition. Different stock market condition has the different type of training data set, bullish, bearish or neutral. We run the experiment on technical indicator, RSI. The results showed the limitation of traditional technical analysis model and the effectiveness of our model.

KEYWORDS: Stock market prediction, sentimental analysis, classification

INTRODUCTION

In recent years, stock prediction has become one of the hottest topics in e-finance area. There has been a rapid growing interest in applying artificial intelligence technique to predict the stock price. The data sources are usually from two types, fundamental analysis and technical analysis. Most of the studies focus on building the model based on the technical indicators, such as SMA (Simple Moving Average), OBV (On Balance Volume), RSI (Relative Strength Index). Recently, the role of web sentiment on stock prediction also attracts the researchers’ notice. In this paper, we proposed a new hybrid model on stock prediction. This hybrid model covers not only technical analysis, but also the fundamental analysis, and furthermore, the sentimental analysis. We define the sentimental analysis as the third type of data source on stock prediction, paralleled with fundamental analysis and technical analysis. The sentimental analysis focused on the sentiment from investors. The source of sentiment could be from news websites, personal blogs, user forum, etc.

The stock price change could be treated as two classes, up and down. Based on this idea, we use classification technique to predict the stock price change. The novelty of this model is also from its training data set. We use the combined analysis (fundamental, technical and sentimental) to decide the market trend, bullish, bearish or neutral. Under different market trend, there will be different training data set. This idea will definitely increase the accuracy on stock prediction.

The structure of this paper is as the following, section 2 is the related work in stock prediction area. Our proposed model will be in section 3. We will run the experiment in section 4 to show the limitations of traditional model. The last section is our conclusion.
Many researchers have been working on stock prediction area. In this section, we will pick up some studies in this area.

Phua and Lin [1] provided a study of the Stock Exchanges of Singapore (SES). The assumption of their study was that there is a positive correlation with the movement of Straits Times Index (STI). They believed the prediction of STI will be very helpful for trading stocks in SES. They used genetically evolved neural networks to predict the stock price and got 81% accuracy on market direction.

Kwon and Moon [2] proposed a hybrid neurogenetic system, a recurrent neural network, for stock trading. The input was the technical indicators used by financial experts. Genetic algorithm was used to optimize the parameters. The experiments were based on 36 companies in NYSE and NASDAQ for 13 years from 1992 to 2004. Their method showed notable improvement on the average on the buy-and-hold strategy.

Ren et.al [3] proposed a model based on C4.5 decision tree for stock prediction based on fundamental stock data. The results showed that the generated rules have exceptional predictive performance.

Schierholt and Dagli [4] proposed a method on the prediction on index. They built a multiplayer perception architecture and probabilistic neural network are used to predict the incline, decline and steadiness of the S&P index. The results showed that the probabilistic neural network performed slightly better than the multi-layer perception.

Thawornwong et al. [5] presented a study on the performance of neural network on the stock market prediction. They input the technical indicators into the neural network to get the more accurate stock trend predictions.

Lee [6] proposed an intelligent agent-based stock prediction system—iJADE stock advisor, to predict the stock performance. iJADE used a hybrid radial basis-function recurrent network (HRBFN). The data source is from 10 year stock information (1990-1999), with 33 companies on the HONG KONG market. The results showed that iJADE achieved promising performance on efficiency and accuracy.

Sapena [7] et al. talked about the stock prediction on “pool” company, which is made of several distribution warehouses in different locations. They used artificial neural networks to get the accurate prediction. Their method could help the pool company reduce material in storage.

Han and Chen [8] built a model based on SVM method and selected Gaussian radial basis function (RBF) as the kernel function. The results showed that their method meets both the accuracy requirements and stockholder’s expection.

Klassen [9] provided an investigation of some technical indexes, such as RSI, for predicting the stock price. The method used was Levenberg-Marquardt algorithm.

The above studies are based on the technical analysis. All the input data are technical indicators. The following case is based on sentimental analysis. This is a different way on stock prediction area.

Sehgal and Song [10] proposed a model on stock prediction by using not technical indicator, but using web sentiment. They calculate the trust value of each publishers on web and finally to predict the stock value.

**PROPOSED MODEL**

As we know, the quality of trading data set is very important in artificial intelligence study.
One big limitation stock prediction model based on technical analysis is from the data source. Supposed that the data source of training data set on your model is from 1990-2004, if you use this model to predict a technical stock, you will get the inaccurate prediction. In 1999, there is an IT fever; while in 2001, there is IT recession. How could you predict a stock based on the data set with such a big different background? You must find the appropriate data set as your training data set, which means your stock market condition of training data set must fit the current market condition. Otherwise, your prediction will be in the wrong way.

Based on this idea, we build this hybrid model

![Figure 1. Hybrid Model](image)

We use opinion mining technique to decide the market condition. We collect the opinions from news websites, personal blogs, user forum, etc. The customers can make decision on the collected opinions. And then, they can pick the type of training data set, like bullish, bearish or neutral. Also, the customers can select the technical indicators which they thought important to build the model. Finally, the system will pop up the decision on stock price change.

EXPERIMENT

In this section, we run an experiment to show the limitation of prediction based technical analysis.
We picked C (Citi group) as the test security. The period of data source is from Sep, 2008 to Apr, 2009. In this period, the financial crisis happened. The sentiment of the market is bearish. So we use bearish dataset as training dataset. The fundamental analysis, like P/E ratio, ROI, cash flow, insider transactions, etc., will help us determine if this company is in good track, so decide if continue work on the TA select part. The technical indicator we picked is RSI (Relative Strength Index). Relative Strength Index (RSI) is an extremely useful and popular momentum oscillator. The RSI compares the magnitude of a stock's recent gains to the magnitude of its recent losses and turns that information into a number that ranges from 0 to 100. It takes a single parameter (usually 14), the number of time periods to use in the calculation [11]. RSI can indicate the buy or sell signal. Generally, if the RSI rises above 30 it is considered bullish for the underlying stock. Conversely, if the RSI falls below 70, it is a bearish signal. Based on this statement, we build the rules,

If RSI (14) < 30, buy;
If RSI (14)> 70, Sell;

The following graph is the result

![Figure 2. Experiment Results](image)

As we can see, in some cases, the technical analysis gave us a good indication. In each buy point, we can earn some profit in a short time. But in the Oct, 2007, when the financial crisis happened, there is no alert from the technical part. This will be a big risk on customers' investment.

From this case, it is clear that the technical analysis is not enough. We also need the sentiment analysis and fundamental analysis to protect our investment.

By adding sentiment analysis and fundamental analysis, we can get more comprehensive system. By collecting the opinions from the news, web forum, etc. The sentiment we retrieved shows up as “overheat”, “alert”. From fundamental analysis, the P/E ratio, ROI, profit margin can also give you a different opinion.
This study is still in progress. In next step, we will collect enough data to build three types of training data set and test the performance of this model.

CONCLUSION

In this paper, we discussed the current study in stock prediction based on artificial intelligence technique. We talked about the limitation of traditional technical analysis. A hybrid model was proposed. This model combined the sentimental analysis and technical analysis. Also, we used different training data set, which meted the current stock condition, to get more accurate prediction. This study is still in progress, we believe this study will be a good contribution on stock prediction area.

REFERENCES


