© ECTI-CARD 2009, May'09, Bangkok, Thailand. ISBN: 978-974-8285-62-7

Relative Movement to Support Stock Forecasting of the Thai Market using the Neural Network

Pongsak Srithongnopawong
Computer Science and Information Management
Asian Institute of Technology
pongsak.srithongnopawong@ait.ac.th

Vatcharaporn Esichaikul
Computer Science and Information Management
Asian Institute of Technology
vatchara@ait.ac.th

ABSTRACT

Nowadays, stock analysis is a challenging task in stock prediction. Stock analysis methods including fundamental analysis and technical analysis are commonly used among financial professionals to help them on investment decisions. In recent days, AI-based system becomes a common tool to predict stock price. Among AI-based models for stock forecasting, Artificial Neural Network (ANN) is the most popular and accurate model. This research proposes data preprocessing using relative movement to improve performance of ANN-based stock forecasting. Both fundamental and technical indicators are chosen as input to the system. The common preprocessing including Principal Component Analysis (PCA) and Z-Scaling is also applied. The evaluation metrics include Root Mean Squared Error (RMSE), hit ratio, and total return. The k-fold cross validation is used to utilize the dataset of stocks in banking sector. The significance of those three metrics is determined through t-test over cross validation. The experiments show that the proposed model outperforms a traditional model, random walk model, and buy & hold strategy for all evaluation metrics.

Keywords

Stock forecasting, Artificial Neural Network, Relative Movement, Fundamental indicator, Technical indicator

1. INTRODUCTION

Stock analysis is an important task in stock market. However, it is not simple to predict stock price. Over many years, many analyses and researches have been conducted to model the stock from its historical data. Stock price as time series data is commonly analyzed by statistical and mathematical models. AI-based model is a new generation of methodology used to forecast stock price. Artificial Neural Network (ANN) is the most popular and accurate model among AI-based models in recent days.

In general, there are two main methodologies in stock analysis including 1) fundamental analysis and 2) technical analysis. Fundamental analysis is to analyze the stock from financial statements, competitive advantages, competitors, and markets. In general, state of the economy, of politics, and of industry in macro and micro terms are analyzed in fundamental analysis. These are in the forms of interest rate, gold price, and oil price.

Technical analysis is known as a technique that forecasts stock price particularly in short-term by studying price and volume movement. Technical analyst believes that price and volume are the most important factors on investor's decision whether to buy or sell stocks. Specifically, technical analysis is widely used among traders and financial professionals. Therefore, technical analysis directly reflects the change on the stock market.

Many researches on stock forecasting have been conducted by using mixed indicators between fundamental and technical indicators. Given the fact that stock price movement is quite random-walk, all indicators that improve the error rate are necessary to be applied to the stock forecasting model. In most studies on stock prediction, the predictor receives its historical price and volume together with several fundamental and/or technical indicators to predict the close price in the next day.

The data preprocessing is generally used to improve ANN performance. In this research, PCA (Principle Component Analysis) and Z-Scaling normalization are applied. This research uses ANN model as the predicting model. ANN receives the inputs from the preprocessing and then generates the signal to use in finding buying or selling signals at some threshold.

In this research, relative movement or daily price change is proposed as an additional preprocessing to stock predicting system. The proposed model predicts the change in current price from that of the previous day while the exact close price is predicted in a traditional model.

2. RELATED WORK

There have been several researches on stock forecasting models using the Neural Network. In 2003, Egeli's research [1] has proved that the prediction models based on ANN were more accurate than the ones based on MA (Moving Average). Egeli used both MLP (Multi Layer Perceptron) and GFF (Generalized feed-forward) networks, and trained them with the backpropagation learning method. Egeli compared MLP and GFF models, and the result has shown that GFF was more appropriate for the stock prediction than MLP. In the same study, it was also found that using 1 hidden layer gave less error rate than using more than one layer for both MLP and GFF.

In 1998, Mizuno trained feed-forward network with backpropagation and equalized learning method to predict the buying/selling signals of stocks in TOPIX (Tokyo

Stock Exchange Prices Index) in [2]. The main input variables were historical price average, price deviation, and relative strength index. The 5 years data were used as training samples for making prediction models. Although the result from Mizuno's research was not convincing, two interesting points were raised: 1) the neuron model might not easily predict signals of stock buying/selling, and 2) input data used in the research might not be effective enough.

In 2004, Nygren conducted a research of stock prediction on ANN model with ECNN (Error Correction Neural Networks) with a few major stocks in the Swedish stock index (SXGE) in [3]. In this research, hit rate (HR) and realized potential (RP) were used as benchmark. Nygren compared the result from the ANN model with naive strategy (naive strategy defined as prediction model by using yesterday stock price). The result demonstrated that ECNN showed good results and ECNN outperformed naive strategy.

Moreover, there are more related works such as "Intelligent Technical Analysis Based Equivolume Charting for Stock Trading using Neural Networks" [4], and "Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support" [5].

In recent days, researchers have invented new models and applied to predict stock trend such as GA (Genetic Algorithm), HMM (Hidden Markov Model), and its combined models. This research focuses mainly on an additional preprocessing to support stock forecasting model using the simple Neural Network.

3. ARTIFICIAL NEURAL NETWORK

3.1 Artificial Neural Network Fundamentals

An artificial neural network (ANN) refers to a computational model based on the biological neural networks. ANN is composed of artificial neurons that interconnect among them. ANN is an adaptive system that can learn and be trained to think as one's brain. During learning phase, there is internal and external information flowing through network given by training data. The flow of information makes the network adapt its structure and learn new input/output.

According to the Figure 1, the input *p* is a scalar variable. When input *p* is passed through the network, it is multiplied with weight *w*. It then produces *wp* as an initial output. Next, the *wp* is passed to the transfer function with an additional fixed bias *b* in order to produce the output *a*. The bias could be viewed as additional weights to the network.

The transfer function receives the input n which is the product of $wp + b$. Therefore, $wp + b$ is the only argument to the transfer function *f*. The transfer function could be selected upon the type of output, but it commonly chooses a sigmoid function as the transfer function. When the problem is non-linear, the network can become more complex. Then, hidden neural nodes and hidden layers could be freely added to handle complex problems as demonstrated in Figure 2.

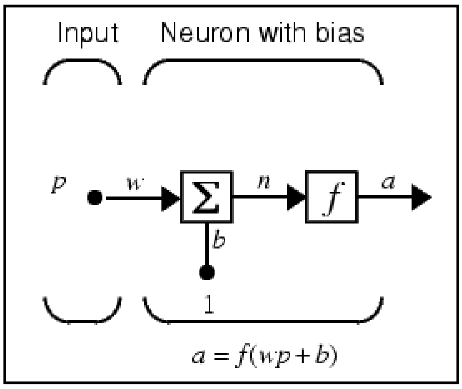

Figure 1. Neural Networks with Bias

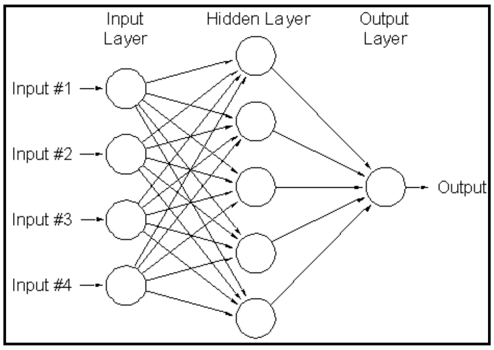

Figure 2. Neural Networks with Hidden Layer

In many prediction models, supervised learning is used to minimize the discrepancy between real data and the output of the network. The backpropagation algorithm belongs to this class and can be described as "an efficient way to calculate the partial derivatives of the network error function with respect to the weights" [3].

The backpropagation is an implementation of Delta rule. It is basically a learning rule for updating the weights of the artificial intelligent network, which is mostly useful for feed-forward networks.

3.2 Performance Measurement

In this study, there are three evaluation metrics including RMSE (Root Mean Squared Error), hit ratio, and total return.

3.2.1. Root Mean Squared Error

RMSE is one of the common evaluation metrics used in ANN predicting system. The equation to calculate RMSE is as in Equation (1).

$$\mathrm{RMSE} = \sqrt{\frac{(t_i - x_i)^2}{N}} \quad (1)$$

The variable t_i is the target value, x_i is the actual network output, and N is the number of samples.

3.2.2. Hit Ratio

Hit ratio is defined as the accuracy of predicting model measured in percentage over test dataset. It is determined by the number of correct signs divided by total number of predicting. The hit ratio formula is as in Equation (2).

$$\text{Hit Ratio} = \frac{\text{The number of correct signs}}{\text{Total number}} \times 100\% \quad (2)$$

3.2.3. Total Return

This metric represents the amount of money return. Total return is easily perceived by investors to see how good the model is, in particular when compared to the bank deposit interest rate. Equation (3) is to calculate the total return.

$$\text{Return Rate} = \frac{\text{Total Profit}}{\text{Initial Investment}} \times 100\% \quad (3)$$

4. STOCK FORECASTING SYSTEM

4.1 Data Preparation

In this research study, both fundamental indicators and technical indicators are chosen as inputs to the predicting system. Inputs for both experiments will use 5-years historical data collected from SET (The Stock Exchange of Thailand) market and financial information providers such as an official broker company in Thailand.

All inputs used in this study are mainly based on indicators used in previous work. Both fundamental and technical indicators are used as inputs in this study. The chosen fundamental indicators for the model include the following.

- Gold Price
- Foreign Exchange (USD/THB)
- Thai Government Bond Yield
- SET 50 Index
- Dow Jones Index

The chosen technical indicators for the model include:

- High Price
- Low Price
- Close Price
- Volume
- MAs (Moving Average)
- RSI (Relative Strength Index)
- Stochastic Oscillator (%K, %D)
- MACD (Moving Average Convergence/Divergence)

Three stocks in banking sector used in this research are selected from SET market. The primary stock is BBL or Bangkok Bank. This study also uses the data of KBANK and SCB stocks. All of them are the most secure banks in Thailand. Only stocks in banking sector are selected because in other sectors there might be other factors affecting the price movement.

4.2 Preprocessing

4.2.1. *Relative Movement*

Relative movement is a part of the data preprocessing to calculate one-day change and two-day change of all inputs. Relative movement introduces the change or differences of today's value and value in one and two days ago.

Table 1. Calculation Example

Day	Absolute Value	1-day Relative Movement	2-day Relative Movement
1	61	N/A	N/A
2	58	-3	N/A
3	57	-1	-4
4	57	0	-1
5	59	2	2

According to Table 1, 1-day relative movement is calculated from the change from the previous day. On the second day, 1-day relative movement is -3 as the subtraction of the value in the second day and the first day. On the third day, 2-day relative movement is -4 as the subtraction of the value in the third day and the first day. This relative movement represents the value change from one and two days ago.

4.2.2. Technical Indicator Calculation

Chosen technical indicators include MAs (MA5, MA25, and MA75), RSI, Stochastic Oscillator (%K and %D), and MACD.

Note that all technical indicators are necessary to compute the relative movement. As a system reads the relative movement as input, the system will not read the exact value except from RSI since RSI is in percentage calculated from a past period.

4.2.3. Principal Component Analysis

Principal Component Analysis or PCA is the orthogonal linear transformation or the eigenvalue decomposition. PCA is commonly used as data preprocessing to remove linear correlation out of the input data before passing processed data to ANN. PCA can also be used in factor analysis or used to remove unimportant factor out of the model; however, in this research PCA is only used to decorrelate the input data.

4.2.4. Z-Scaling Normalization

Z-Scaling normalization is to scale the data to have a zero mean and one standard deviation. The factor in the equation is to freely scale the data in the range approximately between 1 and -1. In this research, the chosen factor value is 4 throughout the research as it is the number that makes all stocks data in a proper range.

$$\text{zscaling} = \frac{x_i - \text{mean (x)}}{\text{stdev (x)} \times \text{factor}} \tag{4}$$

4.3 FANN Library

FANN or Fast Artificial Neural Network library is a free open-source Neural Network library, which implements multilayer artificial neural networks in C with support for both fully connected and sparsely connected networks. FANN is implemented neural network system in this research.

FANN library is commonly used among ANN developers and researchers to implement neural network in many languages such as C, C++, and .NET.

4.4 Training

Training algorithm is used to train Neural Network so that the best generalization is reached. Therefore, the data are separated into training, validating, and testing dataset. There are five steps of training algorithm used in this study.

Step1: Separate data to training, validating and testing set. Training dataset is used in training process where validating dataset is used to measure the generalization of the network. Testing dataset is only used to measure the network performance, and hence it has not been seen while training.

Step2: Create a shortcut network and configure the network. In FANN library, the shortcut network refers to a general backpropagation network which is fully connected and also has shortcuts between nodes. In this step, activation function is chosen as symmetric sigmoid function for both hidden and output layers. Learning rate is set to 0.7 as recommended in [6].

Step3: Initialize weights. The weight initializing algorithm from Nguyen and Widrow [7] is used in this study. The algorithm basically initializes the weight randomly in the range of input data.

Step4: Train network process. The network is trained by the training dataset for several times until the maximum epochs are reached or MSE over validating dataset does not improve for 25 times consecutively. During training, the best network that yields the less MSE over validating dataset is determined and saved.

Step5: Load the best network and measure MSE over testing data set. After training process is completed the best network is loaded to measure a final MSE over testing dataset.

4.5 Evaluation

In this study, the evaluation is conducted with the k-fold cross validation. The k-fold cross validation is to separate dataset into 5 partitions. Three partitions are training dataset, validating dataset, and the testing dataset. Performance is measured for k times by switching validating dataset and testing dataset to all partitions. There are 5 partitions that testing dataset can be while there are 4 partitions that validating dataset can be. Hence, there are 20 folds in total.

4.6 Benchmarks

In this research, the main task is to compare the performance of the proposed model with a traditional model. However, there are also other benchmarks to show whether ANN outperforms naïve methods in general. These benchmarks include 1) buy and hold strategy, and 2) random walk strategy.

The buy and hold strategy is a passive and long-term investment strategy which is not concerned the short-term movement. It starts from buying the stock at the first day and holds it until the last day. The price will be sold at the last day. The profit is gained from the difference of the first and the last day price.

The random walk strategy is to randomly generate buying and selling signals. Then, performance is measured three times to find the hit ratio and total return. The average value is the final result.

5. EXPERIMENTAL RESULTS

5.1 Experimental Environment

The proposed model is implemented using Neural Network model. The inputs of the systems are discussed in section 4.1. Data preprocessing covers relative movement, technical indicator computation, PCA, and Z-Scaling as discussed in section 4.2. The 20-fold cross validation is used in this experiment as described in section 4.5. In the experiment, max epochs are set to 10,000. The learning rate is 0.7. The neural layers start from zero to one layer. The hidden nodes start from 1 to 10 nodes. The best network that yields the best performance is loaded to test its performance.

The data used in experiments are from BBL, KBANK, and SCB stocks from July 2, 2003 to June 30, 2008. In cross validation, each partition is 20% of the whole dataset. Therefore, the test dataset is approximately one year.

In the evaluation, the measurement metrics are test RMSE, hit ratio, and total return.

5.2 Results and Discussion

Table 2 demonstrates the comparison of predicting performance of the proposed model and a traditional model using BBL stock data. The proposed model uses the relative movement as inputs and the model directly predicts future changes; however, a traditional model uses exact value as inputs and the model predicts the future price.

Table 2. BBL Cross Validation Result

Type of Model	Test RMSE	R-square	Hit Ratio (%)	Total Return (%)
Proposed model	0.2439	0.0332	55.94	16.06
Traditional model	0.0465	0.0079	48.59	-1.25

According to the Table 2, Test MSE of the proposed model is higher than a traditional model. Test MSE cannot be fairly compared because errors in both models are in different scale. The MSE in the proposed model refers to the error of the predicted future price change while the MSE in a traditional model refers to the error of the predicted future price.

Both hit ratio (%), and total return (%) of the proposed model show better performance. T-test is also determined to show the significance. For each metric, the hypotheses (paired-test) are:

$$H_0: \mu_{\text{proposed}} = \mu_{\text{traditional}} \quad (5)$$

$$H_A: \mu_{\text{proposed}} \neq \mu_{\text{traditional}} \quad (6)$$

The degree of freedom is 19. The t-test result is shown in Table 3.

Table 3. BBL t-test Result

Metrice	Mean difference	t-value	p-value (2-tailed)
R-square	0.0253	5.91	0.000011
Hit ratio	7.35	4.09	0.000627
Total return	17.30	5.86	0.000012

For both hit ratio and total return, Table 3 shows that the means between two models are significantly different (p-value > alpha, and alpha = 0.05). Therefore, the proposed model outperforms a traditional model using BBL stock data.

Table 4. BBL Benchmarks Comparison

Type of Model	Hit Ratio (%)	Total Return (%)
Proposed model	55.94	16.06
Traditional model	48.59	-1.25
Random walk	49.57	-21.56
Buy & hold	-	10.02

Table 4 shows that the proposed model outperforms other benchmarks including a traditional model, random walk, and buy & hold strategy. For buy & hold strategy, only total return can be computed as it buys and sells once.

This research study also conducts the same test with two more stocks in banking sector using KBANK and SCB data. The result confirms the similar trend as reported with BBL stock data.

Although the result shows a good performance of the proposed model on stocks in banking sector, it might not show a good predicted result on stocks in some sectors having other substantial impacts. For example, the stock of oil and natural gas company very much correlates to oil and gas price in the country.

6. CONCLUSION

6.1 Conclusion

Nowadays, stock analysis plays an important role among investors. Over many years, both fundamental analysis and technical analysis are used as a tool for financial professionals to predict stock trends.

Currently, AI-based stock forecasting system is an alternative tool to predict stock trends. ANN is the most common and accurate model among AI-based models. However, there are several possible ways to choose input and data preprocessing for ANN stock forecasting model. This research selects both fundamental and technical indicators as inputs. The general data preprocessing including PCA and Z-Scale is applied.

This study proposed additional data preprocessing by using relative movement to improve performance of stock forecasting system using Neural Network. The performance of the model was mainly compared with a traditional model which did not calculate relative movement. The evaluation metrics using in this paper were hit ratio and total return. The t-test was also used to show the significance. The evaluation was done with k-fold cross validation. Other benchmarks included random walk and buy & hold strategy.

The experiment shows the promising performance. The proposed model outperforms a traditional model and all benchmarks for both hit ratio and total return. In BBL stock data, the proposed model yields 55.94% hit ratio and 16.06% total return while a traditional model yields 48.59% hit ratio and -1.25% total return. When compared with other benchmarks, the proposed model yields the best performance for both hit ratio and total return. In addition, the result shows the similar trends on stocks of KBANK and SCB.

Although the proposed model does not show a very high hit ratio (less than 60%), the total return shows a good outcome as it gives profit more than earning interest from banks. A traditional model and random walk model yield approximately the same hit ratio or about 50%; however, a traditional model gives a much better result on total return than the random walk model. Therefore, to earn high profit is not always necessary to have a high hit ratio. It is more important to be able to predict a positive change when there is a dramatic change on future price.

The limitation of the proposed model is that the model might not be able to predict stocks that have other substantial direct impact. For example, these stocks are national resource distributor. The stock's price mainly depends on an oil and gas price in the country.

6.2 Future Work

In stock market, automated trading system is not worth money without a good stock forecasting system. This study attempts to improve the Neural Network stock forecasting system by implementing data preprocessing using relative movement. The result shows a good performance of the system; however, there are a few interesting topics as further studies.

Given the stock forecasting system of this study, the yearly profit can be improved if more than one stock are bought and sold in the same evaluation period. When handling more than one stock, investment portfolio needs to be managed to get the optimal profit. Therefore, the portfolio management could be very interesting topic for future work to improve an overall performance.

Furthermore, the user interface improvement can be further implemented. This study implements the system based on C language with DOS-like interface, which is not quite user-friendly. Also, the cross validation process could be further improved. In this research, the full cross validation was conducted. The further work should aim to find the better cross validation method for evaluation of stock forecasting system.

7. ACKNOWLEDGMENT

I would like to thank Nonthawatr Anusornpanich and Supakit Chawchumnum for providing data and stock market knowledge.

8. REFERENCES

[1] B. Egeli, M. Ozturan, and B. Badur, "Stock Market Prediction Using Artificial Neural Networks," March 2008, http://www.hicbusiness.org/biz2003proceedin gs/Birgul%20Egeli.pdf.

[2] H. Mizuno, M. Kosaka, and H. Yajima, "Application Of Neural Network To TechnicalAnalysis Of Stock Market Prediction," March 2008, http://sce.uhcl.edu/ boetticher/ML_DataMining/Kimoto.pdf.

[3] K. Nygren, "Stock Prediction – A Neural Network ApproachStock Prediction – A Neural Network Approach," ," March 2008, http://www.f.kth.se/~f98-kny/thesis.pdf.

[4] T. Chavarnakul and D. Enke, "Intelligent Technical Analysis Based Equivolume Charting for Stock Trading using Neural Networks," *Expert Systems with Applications*, vol. 34, pp. 1004-1017, 2006.

[5] I. Kaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series," *Neurocomputing*, vol. 10, pp. 215–236, 1996.

[6] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights", *Proceedings of the International Joint Conference on Neural Networks*, vol. 3, pp. 21-26, 1990.