Computational Neural Network for Global Stock Indexes Prediction

Dr. Wilton.W.T. Fok, IAENG, Vincent.W.L. Tam, Hon Ng

Abstract - In this paper, computational data mining methodology was used to predict four major stock market indexes. Two learning algorithms including Linear Regression and Neural Network Standard Back Propagation (SBP) were tested and compared. The models were trained from two years of historical data from January 2006 to December 2007 in order to predict the major stock prices indexes in the United States, Europe, China and Hong Kong. The performance of these prediction models was evaluated using two widely used statistical metrics. The comparison showed that using Neural Network Standard Back Propagation algorithm resulted in better prediction accuracy then using linear regression algorithm. In addition, traditional knowledge shows that a longer training period with more training data could help to build a more accurate prediction model. However, as the stock market in China has been highly fluctuating in the past two years, this paper shows that data collected from a closer and shorter period could help to reduce the prediction error for such highly speculated fast changing environment.

Index Terms - Neural network, Computational Data Mining, Stock Market Forecast

I. INTRODUCTION

In the past, stock market indexes were forecasted by simple technical analysis such as moving average and linear regression models. The stock market index trends were assumed to be linear, or were moving along a smooth rolling average. However, the stock market indexes in the real world are not linear. They are often influenced by numerous complex factors and many interrelated variables, such as closing price, highs, lows, and volume, influence stock prices, interest rate, bonds rate and so on. Traditional forecasting methods are limited in their effectiveness as they make assumptions about the distribution of the underlying data, and often fail to recognize the interrelatedness of variables [1].

Stock market index prediction is one of the challenging applications of modern time series forecasting. Financial parameters such as currency rates and stock indexes are inherently noisy, non-stationary and chaotic [5, 13]. These characteristics suggest that there is no complete information that could be obtained from the past behavior of such indexes to fully capture the dependency between the future stock indexes and that of the past. One general assumption in such cases is that the historical data incorporate all those behaviors. As a result, the historical data is the major player in the prediction process. Although the well-known conventional forecasting techniques provide predictions, for many stable forecasting systems, of acceptable quality, these techniques seem inappropriate for non-stationary and chaotic system such as stock market indexes and exchange rates. The purpose of this paper is to investigate the use of data mining techniques for prediction of stock market indexes. We planned to experiment with two algorithms under different architectures for several stock market indexes.

Recently, researchers have focused on using artificial intelligence and data mining techniques to analyze historical data and to recognize subtle relationships between variables in the financial market. Neural networks methodology is good at pattern recognition, generalization, and predicting trends. It could tolerate imperfect data, and does not require formulas or rules. It could also determine the relationship between variables and detect relevant patterns in the data. In this paper, Neural Network technique will be used for building a neutral network model for the prediction of the major stock market indexes in the United State, Europe, China and Hong Kong.

Neural Network is commonly used as an estimation function in data mining for data prediction and system modeling. Recently it has been widely applied in time-series analysis and forecasting [11-14]. It enables multivariate analysis so that not only the lagged time series is being forecasted but also other indicators (such as technical, fundamental, inter-marker for financial market) can be combined to act as predictors. In addition, Neural Networks is more effective in describing the dynamics of non-stationary time series due to its unique non-parametric, non-assumable, noise-tolerant and adaptive properties. It can map any nonlinear function without a *priori* assumptions about the data [3].

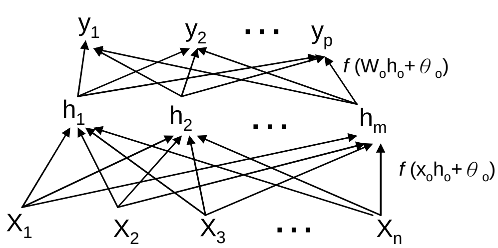
It had been shown that Neural Network performs better than other linear models, specifically, for more irregular series and for multiple-period-ahead forecasting. Tang and Fishwich [9], Kamruzzaman and Sarker [6] and Wang and Leu [10] provided a general introduction of how a neural network model should be developed to model financial and economic time series. A similar analysis but on currency exchange rate had also been conducted by Zhang and Hu [14], Kamruzzaman and Sarker using back propagation neural network and artificial neural network [1].

In this paper, Neural Network data mining techniques were applied to predict the four major stock indexes, namely United States, Europe, China and Hong Kong, using their historical data. The standard back propagation data mining model and the common regression model was used and compared. A total of 510 historical indexes data (the closing price on each trade day), for each of the four stock market indexes were collected and used as inputs to build the prediction models in this research. The prediction results were compared based on two common evaluation indicators namely the Mean Absolute Error (MAE) and Normalized Mean Square Error (NMSE).

II. DATA MINING MODEL AND ALGORITHM

One of the most commonly used data mining forecasting methodology is multilayer feed-forward network. There are input layer, output layer and one or more hidden layers between the inputs and outputs in such model. All the nodes at each layer are connected to each node at the upper layer by interconnected weights. A training algorithm is used to attain a set of weights that minimizes the difference between the predicted value and the actual output through the network. In this research, the standard Back Propagation (BP) algorithm was used to predict the indexes. The performance and accuracy of the predicted results in different markets were evaluated.

Standard Back Propagation (SBP) [8] could update the weights iteratively to map a set of input vectors $(x_1,x_2,\ldots,x_p)$ to a set of corresponding output vectors $(y_1,y_2,\ldots,y_p)$. The input x_p is presented to the network and then multiplied by the weights. All the weighted inputs to each unit of upper layer are summed up to produce the output which is governed by the following equations.



$y_p = f(W_o h_p + \theta_o)$ (1)
$h_p = f(W_h x_p + \theta_h)$ (2)

where W_o and W_h are the output and hidden layer weight matrices, hp is the vector denoting the response of hidden layer for pattern *'p'*, θ_o and θ_h are the output and hidden layer bias vectors, respectively and $f(.)$ is the sigmoid activation function. The cost function to be minimized in Standard Back Propagation is the sum of squared error defined as:

$$E = \tfrac{1}{2} \Sigma (t_p - y_p)^T (t_p - y_p) \quad (3)$$

where t_p is the target output vector for pattern '*p*'. The algorithm uses gradient descent technique to adjust the connection weights between neurons. Denoting the fan-in weights to a single neuron by a weight vector w, its update in the t^{th} iteration is governed by the following equations.

$$\Delta w_t = -\eta \nabla E(w)\big|_{w = w(t)} + \alpha\, \Delta w_{t-1} \quad (4)$$

The parameters η and α are the learning rate and the momentum factor, respectively. The learning rate parameter controls the step size in each iteration. In general, a high learning rate values could decrease the chance of getting stuck in a sub optimal solution, or the local minimum, but a low learning rate values could increase the change of finding the bottom of any minimum. In this analysis, the learning rate is set to 0.8.

III. THE ANALYSIS PROCESS

The analysis process consists of five stages: data collection, data benchmarking, neural network model building, generation of prediction result and performance metrics evaluation.

A. Data Collection

The data used in this study is the major stock indexes in different markets including Europe, the United States, China and Hong Kong from January 4, 2006 to December 20, 2007 made available from Yahoo! Finance. These markets include: the United State Dow Jones Industrial Average Index (DJI), the London Financial Times Stock Exchange (FTSE), the Chain Shanghai B-Share Index (SHB) and the Hong Kong Hang Seng Index (HSI). 510 trade days data was considered. For a fair comparison on their growth in the past two years, their indexes on 3 Jan 2006 were taken as the base number and were normalized to 100. The plots of these four normalized historical stock indexes are shown in Fig. 1.

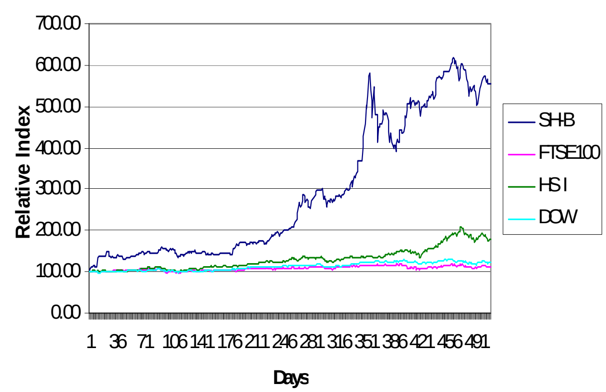


Fig 1. Historical stock indexes for Dow Jones Industrial Average, FTSE100, Hang Seng Index, and Shanghai B Index.

B. Data Benchmarking by using 10-day Moving Average

Similar to some other finance time series models, the stock market exhibits its own trend, cycle, season and irregularity. Many financial analysts used moving

ISBN:978-988-17012-3-7

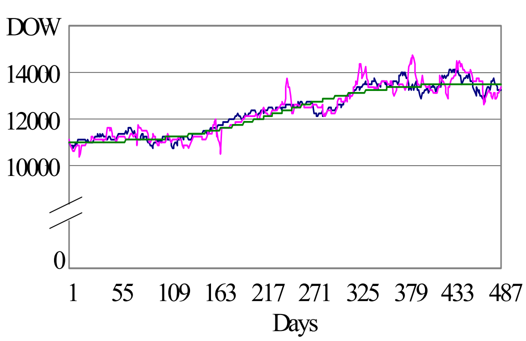
average as technical data to analyze and predict the trend. The advantage of using moving average is its tendency to smooth out some of the irregularity that exits between market days [12]. For a comparison with the performance of Standard Back Propagation, the 10-day moving average values of the four indexes from January 4, 2006 to December 20, 2007 were computed to build a linear regression model for predicting the indexes value after the next 10 days. This prediction is used to benchmark the performance of the prediction model using neural network in the following step.

C. Neural Network Model Building

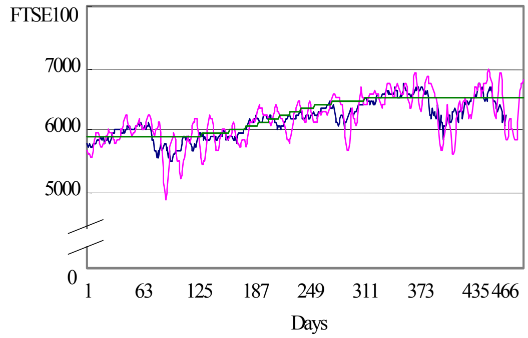
There were 510 historical data being used to train the model using the Tiberius Data Mining Software [2]. The final set of weights to which a network settles down (and hence its performance) usually depends on a number of factors such as the initial weights chosen, different learning parameters used during training and the number of hidden units. In this case, the number of hidden neuron was selected to 3 in order to obtain a better optimal performance. Introducing too many hidden units may create additional parameters, introduces redundancy and deteriorates the performance. The number of iterations, that is the number of times when all of the training patterns have been presented for learning, was set to 100. Once the model is trained, the trained model is then used to generate the prediction results.

D. Generation of Prediction Results

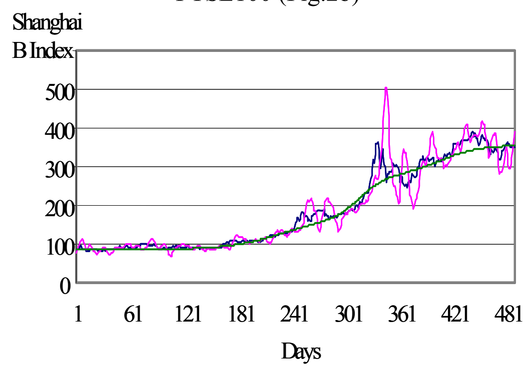
The models were used to generate the prediction result by the tool. The results are shown on Fig. 2(a) to Fig. 2(d) representing the actual and forecasted data of the Dow Jones Industrial Average, FTSE100, Hang Seng Index, and Shanghai B Index respectively. In the figures, day 1 and day 510 represented January 4, 2006 and December 20, 2007 respectively and all non-trade days were omitted. The results show that SBP can perform better than linear regression model in both performance metrics.



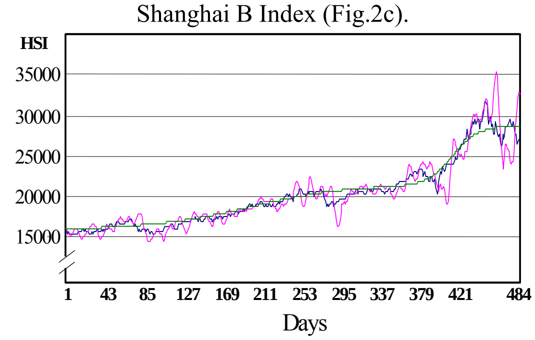
Dow Jones Industrial Average (Fig.2a)



FTSE100 (Fig.2b)



Shanghai B Index (Fig.2c).



Hang Seng Index (Fig.2d)

Actual Forecast Linear Forecast NN

Fig.2. Comparison of the actual data (blue line), forecasted data by linear regression data (pink line) and neural network forecasted data (green line) among the Stock market indexes.

E. Performance Metrics Evaluation

The prediction results were compared against the actual data to evaluate the predictive power of the models. The forecasting performance of the above regression and neural network models are evaluated against two widely used statistical metric, namely, Normalized Mean Square Error (NMSE) and Mean Absolute Error (MAE). These metrics are defined as follow:

Root Mean Square Error (RMSE) = $[\,\Sigma\,(y_k - \hat{y}_k)^2\,]^{½}\ / N$

Mean Absolute Error (MAE) = $|y_k - \hat{y}_k| / N$

Where y_k and $\hat{y}_k$ are the actual and predicted values, respectively. NMSE and MAE measure the deviation between actual and forecast values. Smaller values of these metrics indicate higher accuracy in forecasting. As the orders of magnitude of these stock indexes are different, their RMSE and MAE are divided by the mean of the actual data to normalize the error so that they can be compared in the same order of magnitudes. The original and normalized performance metrics are computed in the Table 1 and Table 2 respectively.

Performance Metrics		Market Indexes			
		SH-B	FTSE 100	HS I	DOW
MAE	Linear	20.2	183.6	923.3	266.7
	NN	12.0	130.1	595.2	240.8
RMSE	Linear	34.1	240.6	1406.5	351.4
	NN	19.2	165.0	790.9	294.0

Table 1: Original performance metrics MAE and RMSE of the four market indexes

Metrics		Market Indexes			
MAE	Linear	0.11	0.03	0.05	0.02
	NN	0.06	0.02	0.03	0.02
RMSE	Linear	0.181	0.039	0.070	0.028
	NN	0.102	0.027	0.039	0.024

Table 2: Normalized performance metrics MAE and RMSE of the four market indexes

In this case of predicting the stock indexes over 2 years (510 trade days), both values of MAE and RMSE achieved by neural network are quite low and result in only 60% to 80% of values achieved by regression model. This means that neural network algorithm is more capable of predicting the stock market than regression or moving average.

The metrics showed that prediction errors for the Shanghai Stock B Index are much higher than all other markets. One possible explanation is that the stock market in China has been growing rapidly and was being speculated in an unpredictable way. Many news policies introduced during the last two years have set up new rules and investment environment for the stock market in China. For example, factors such as the appreciation of the Reminbi currency, establishment of Qualified Domestic Institutional Investor (QDII) scheme and Qualified Foreign Institutional Investor (QFII) scheme all changed the stock market environment. Therefore, predictions using the old model which is trained by historical data and based on the past policies, were no longer valid. Thus, in order to provide a better prediction model for the China market, shorter timeframe and more recent training data should be used.

	2 year data	100 days data
Original MAE	12.0	9.72
Normalized MAE	0.06	0.028

Table 3: Comparison of model accuracy using 100-day trained model and 2-year trained model

Another neural network model using SBP algorithm with data extracted from the recent 100 trade days from July 11, 2007 to November 27, 2007 was built. The normalized MAE predication error could be reduced to half showing that using too many outdated historical data could indeed deteriorate the model accuracy, especially when the environment has changed significantly.

IV. CONCLUSION

This paper has presented and compared, with different data mining algorithms, the results of using regression and neural network models to perform stock market indexes prediction. The results show that a neural network model with improved learning technique is a promising methodology for stock market indexes prediction. Results in this study show that neural network model achieves close prediction in terms of RMSE and MAE metrics.

The rapid changing environment in China requires the training model to be frequently updated in order to maintain the level of accuracy. Chan *et al.* [4] has applied Probabilistic Neural Network (PNN) for forecasting stock index with a view that training PNN is faster and therefore, it enables the user to develop a frequently updated training scheme. In actual applications, retraining of a forecasting model with the most recent data may be necessary to increase the chance of achieving better accuracy. Fast learning may be an advantage but it does not always guarantee an improved performance [1]. Therefore, finding a better way of fast training for stock index prediction needs to be investigated and remains the focus of future research.

REFERENCES

[1] Joarder Kamruzzaman, Ruhul A. Sarker, "ANN-Based Forecasting of Foreign Currency Exchange Rates", Neural Information Processing - Letters and Reviews Vol. 3, No. 2, May 2004, pp/49

[2] Phil Brierley, "Knowledge Discovery and Data Mining", Web-site http://www.philbrierley.com/

[3] L. Cao and F. Tay, "Financial forecasting using support vector machines," *Neural Comput. & Applic*, vol. 10, pp.184-192, 2001.

[4] A. Chen, M. Leung and H. Daouk, "Application of neural netwroks to an emerging financial market: forecasting and trading the Taiwan Stock Index," *Computer & Operations Research*, vol. 30, pp. 901-903, 2003.

[5] G. Deboeck, *Trading on the Edge: Neural, Genetic and Fuzzy Systems for Chaotic Financial Markets*, New York Wiley, 1994.

[6] J. Kamruzzaman and R. Sarker, "Forecasting of currency exchange rates using ANN: a case study," *Proc. of IEEE Int. Conf. on Neural Networks & Signal Processing (ICNNSP03)*, pp. 793-797, 2003.

[7] B. D. Klein and D. F. Rossin, "Data quality in neural network models: effect of error rate and magnitude of error on predictive accuracy," *OMEGA: Int. Journal of Management Science*, vol. 27, pp 569-582, 1999.

[8] D.E Rumelhart, J.L. McClelland and the PDP research group, *Parallel Distributed Processing*, vol. 1, MIT Press, 1986.

[9] Z. Tang and P. A. Fishwich, “Backpropagation neural nets as models for time series forecasting,” *ORSA Journal on Computing*, vol. 5, no. 4, pp 374-385, 1993.

[10] J. H.Wang and J. Y. Leu, “Stock market trend prediction using ARIMA-based neural networks,” *Proc. of IEEE Int. Conf. on Neural Networks*, vol. 4, pp. 2160-2165, 1996.

[11] J. Yao, Y. Li and C. L. Tan, “Option price forecasting using neural networks,” *OMEGA: Int. Journal of Management Science*, vol. 28, pp 455-466, 2000.

[12] J. Yao and C.L. Tan, “A case study on using neural networks to perform technical forecasting of forex,” *Neurocomputing*, vol. 34, pp. 79-98, 2000.

[13] S. Yaser and A. Atiya, “Introduction to Financial Forecasting,” *Applied Intelligence*, vol. 6, pp 205-213, 1996.

[14] G. Zhang and M. Y. Hu, “Neural network forecasting of the British Pound/US dollar exchange rate,” *OMEGA: Int Journal of Management Science*, vol. 26, pp. 495-506, 1998.

ABOUT THE AUTHORS

Dr. Wilton W.T. Fok, Ph.D, MBA, M.Sc. (Engg), M.Sc. (Env.Mgnt.), MHKIE, CEng

Dr. Fok is a Teaching Consultant in the Department of Electrical and Electronic Engineering in the University of Hong Kong. He is the lecturer of the subject on Engineering Economics and Finance with expertise of using quantitative analysis for solving financial problems. Dr. Fok received his B.Eng degree in Electrical and Electronic Engineering from the University of Hong Kong. Dr. Fok also received two other M.Sc. degrees in Industrial Engineering and Environmental Management from HKU in the 90’s. Dr. Fok also holds a MBA degree and a PhD degree from the University of Cambridge and the Renmin University of China respectively. Dr. Fok is a Chartered Engineer with broad industry experience in telecom, RFID, digital TV and data networking technologies. He published papers in different areas including RFID, mobile communication and data mining.

Dr. Vincent W.L. Tam BA (CompStud), M.Phil (CS), Ph.D (CS)

Vincent Tam completed his Ph.D. in Computer Science from the University of Melbourne for his work on integrating stochastic solvers into the framework of Constraint Logic Programming. Since 1998, he joined the School of Computing, the National University of Singapore (NUS), as an academic fellow to pursuit the investigation on local search methods including artificial neural networks and evolutionary algorithms for resource allocation and scheduling. In addition, Vincent was selected the winner of the Innovative Teaching Award 2000 in NUS. From 2001 until now, Vincent is teaching in the Department of Electrical and Electronic Engineering, the University of Hong Kong.

Mr. Hon Ng, MBA, M.Sc(Info Engg), B.Engg

Mr. Hon Ng is the Business Development Manager of ETI Consulting Limited (a wholly-owned subsidiary of the University of Hong Kong). He is responsible for the commercialization of innovative technologies (e.g., mobile messaging, 3G and RFID). During his career, Mr. Ng has spent significant periods establishing start-up businesses, developing start-up products and implementing emerging technologies. Mr. Ng graduated with First Class Honours in Electrical and Electronic Engineering from the University of Hong Kong. He also holds a Master degree in Information Engineering from the Chinese University of Hong Kong and a Master degree in Business Administration from the University of Oxford. He is currently reading a doctoral degree at Sun Yat-sen University and is affiliated with the Business Administration Research Institute of Sun Yat-sen University.