1998 International Computer Symposium
Workshop on Artificial Intelligence
December 17-19, 1998, N.C.K.U., Tainan, Taiwan, R.O.C.

A GENETIC AND NEURAL NETWORK APPROACH TO SIGNIFICANT INDICATORS SELECTION

Anthony J.T. Lee, Y.S. Chen, and J.Z. Chen

Department of Information Management
National Taiwan University, Taipei, Taiwan R.O.C

ABSTRACT

The determination of significant indicators to forecast the value of a stock or mutual fund has been a tough and confusing problem. In this paper, we propose a fund net asset value (NAV) forecasting model, which combines a neural network and a genetic algorithm, providing useful investment reference to investors. First, we use our proposed model to find out the significant indicators for forecasting the NAVs of mutual funds. Next, we use those significant indicators to forecast the NAVs of mutual funds. According to the experiments on Taiwan mutual funds, we found that there did exist common significant indicators for forecasting the NAVs of mutual funds.

1 INTRODUCTION

A mutual fund by definition is the money gathered from lots of investors, and it is manipulated by professional fund managers to invest in money markets, bond markets, and stock markets. In comparison to stocks, the advantage of mutual funds is its lower risk and more stable profitability. Investors can make use of small amount of money to earn more. So, mutual funds have been booming substantially for the past few years. In Taiwan, the scale of domestic mutual funds has approached 70 billion NT dollars in late July 1997, expanding 150 times from 1985 to 1997, according to Investment and Trust Report[14].

Mutual funds are designed to invest various kind of financial products. Take a stock mutual fund for example, its target is the stocks of public corporations. According to Chou[5] and Lin[9], they found that the macro-economic indicators (such as money supply, exchange rate, proceeding indicators) did have close relationship with the movements of stock markets. Therefore those indicators influence the performance of stock mutual funds to some degree, but there may exist some other variables (such as financial reports of funds, stock technical indicators, etc.) that may also have close relationship with mutual funds. If we can pick those variables out, it will be a great help in forecasting the performance (or NAVs) of mutual funds.

In recent years, neural network applications shine in the field of finance. Various neural network-based financial applications have been proposed to forecast the direction of stock movement[3][6][8][10], to build a security transaction-decision system [2][7][11], or to forecast corporation bankruptcy[1][13]. Chiang et. al.[4] proposed a back-propagation neural network(BPN) model to forecast the NAVs of U.S. mutual funds. They took 15 macro-economic indicators as inputs to the BPN. According to their experiment, it showed that their model outperformed traditional regression models.

The determination of significant indicators to forecast the NAV of mutual funds has been a tough and confusing problem. Most of neural network-based forecasting systems select randomly or heuristically macro-economic or stock technical indicators as inputs to a neural network. However, very little is known about selecting systematically significant indicators to a neural network-based financial forecasting system. In this paper, we propose a fund net asset value (NAV) forecasting model, which combines a neural network and a genetic algorithm, providing useful investment reference to investors. First, we use a genetic algorithm and a neural network to find out the significant indicators for forecasting the NAVs of mutual funds. Next, we use those significant indicators to forecast the NAVs of mutual funds.

The outline of this paper is as follows. In section 2 we present the input and output variables of our system model. In section 3 we propose our system model which shows the fundamental structure of a neural network, the methods to find out the significant indicators and to use those significant indicators to forecast the NAVs of mutual funds. In section 4 we give some experimental results on Taiwan mutual funds. In section 5 we list the conclusions and future work.

2 INPUT AND OUTPUT VARIABLES

2.1 Data Sample

In our mutual fund NAV and performance prediction model, twelve mutual funds are selected as analytic samples. Five of them are of aggressive-growth type, five are of growth type, and the rest of them are of growth-income type. Their types are identified in the fund prospectus, which shows their investment objectives. Table 1 is the list of these funds.

In order to have the same experiment intervals, the founded date of the latest fund is chosen as our starting time. The period of our research interval covers four years, from April 1, 1994 to December 31, 1997. We partition this interval into two sub-periods: April 1994~September 1996 and October 1996~December 1997. Data in the former period are used as the training set to provide the network enough "knowledge". Data in the latter period are served as the testing set to prove the performance of our model.

2.2 Input Variables and Output

Our model focuses on predicting the NAVs of mutual funds. It is apparent that the forecasting outcomes are the future values of the NAVs of those sample funds. The NAV predicted by our model is the average NAV of one month, computed as follows:

$$\mu_{nav} = \frac{\sum_{i=1}^{N} NAV_i}{N}$$

where μ_{nav} denotes monthly average of NAV and NAV_i denotes NAV of the i-th day of a month, N is the number of operation days during that month.

All data used here are monthly-based, so the output forecasting value is the monthly average NAV of next month.

Our model attempts to discover the indicators that influence significantly the NAVs of mutual funds. If we can find these indicators and construct a function from them, then as long as the values of these indicators are known, the monthly average NAV can be easily computed. This is illustrated by the following equations:

$$\overline{\mu_{nav_{i,t}}} = \mu(x_{1,t}, x_{2,t}, \ldots, x_{n,t})$$

where $\overline{\mu_{nav_{i,t}}}$ denotes the predicted average NAV of the t-th month of the i-th fund.

Here we select 46 indicators that might have important impact on the NAVs of mutual funds. These indicators consist of macro-economic indicators, stock technical indicators, and the quarterly fund reports. These indicators are listed in Table 2.

All the data of the variables listed in Table 2 are collected monthly. Before those data can be processed by a neural network, the data in [$\mu-3\sigma$, $\mu+3\sigma$] are normalized into the range of [0,1], where μ is the mean of 45-month data of a variable and σ is the standard deviation of 45-month data of that variable

Fund Company	Fund Name	Objectives	Founded Time	Initial Fund Size (unit: NT$ 100 million)
China Securities Investment & Trust Co.	China Phonix	Aggressive-growth	1993/12/21	10
Kwang-Hua Securities Investment & Trust Company	Kwang-Hua Aggressive-growth	Aggressive-growth	1994/01/12	50
First Global Securities Investment &Trust Company	First Global Duo-Fu	Aggressive-growth	1994/03/16	2.3348
Core Pacific Securities Investment & Trust Company	Core Pacific Aggressive-growth	Aggressive-growth	1994/03/22	7.6431
Entrust Securities Investment & Trust Company	Entrust Chong-Long	Aggressive-growth	1994/03/23	30
International Securities Investment & Trust Company	International First	Growth	1986/01/04	20
Kwang-Hua Securities Investment & Trust Company	Kwang-Hua Kwang-Hua	Growth	1987/04/29	20
National Securities Investment & Trust Company	NITC	Growth	1987/07/11	20
China Securities Investment & Trust Co.	China China	Growth	1987/09/05	20
Fortune Securities Investment & Trust Company	Fortune United Dragon	Growth	1994/03/09	50
Polaris International Securities Investment & Trust Company	Polaris Chronicle 2001	Growth-income	1993/02/18	50
Fubon Securities Investment & Trust Company	Fubon Fortune	Growth-income	1994/03/15	20.2850

Table 1: Sample mutual funds

3 SYSTEM MODEL

3.1 Selecting Indicators

Now, we have 46 candidate indicators served as input variables. However, we do not want all of them to be sent into a neural network at the same time. We want to select a combination (a subset) of those indicators so that the selected combination has a better outcome than other combinations, including the combination of all 46 indicators. So, we use a genetic algorithm to filter and choose the optimal combination. The process of filtering indicators by the genetic algorithm is shown in Figure 1.

3.2 Network Training and Testing

Our model is based on the recurrent neural network proposed by Kamijo and Tanigawa [8], including a input layer, a hidden layer, a output layer, and a context layer which receives replicated data from the hidden layer. The recurrent neural network is shown in Figure 2.

Fundamental data of mutual funds					
ID	variable	ID	variable	ID	variable
M1	Net asset of fund	M2	Stock share holding ratio	M3	β value
M4	P/E ratio of fund	M5	Average NAV for last month	M6	Sharpe index
M7	Trenyor index	M8	Jensen index	M9	Performance of last month

Economic Indicators					
ID	variable	ID	variable	ID	variable
F1	GNP	F2	CPI	F3	WPI
F4	Unemployment rate	F5	Money supply:M1	F6	Money supply:M2
F7	Inflation rate	F8	One-month rate	F9	Three-year rate
F10	NT/US exchange rate	F11	Taiwan Stock Index	F12	Don Jones Index

Stock Technical Indicators					
ID	variable	ID	variable	ID	variable
T1	$DIF^{12,26}$	T2	$MACD^{12,26}$	T3	K^3
T4	D^3	T5	K^3-D^3	T6	$BIAS^{10}$
T7	$BIAS^{30}$	T8	$BIAS^{72}$	T9	Average $Volume^6$
T10	Average $Volume^{12}$	T11	Average $Volume^6$ − Average $Volume^{12}$	T12	AR^{26}
T13	BR^{26}	T14	MTM^{10}	T15	OSC^{10}
T16	RSI^6	T17	RSI^{12}	T18	RSI^6-RSI^{12}
T19	PSY^{13}	T20	VR^{14}	T21	$WMS\%R^{12}$
T22	OBV^{12}	T23	$TAPI^{10}$	T24	$(+DI^{14})$-$(-DI^{14})$
T25	ADX^{14}				

Table 2: Input variables

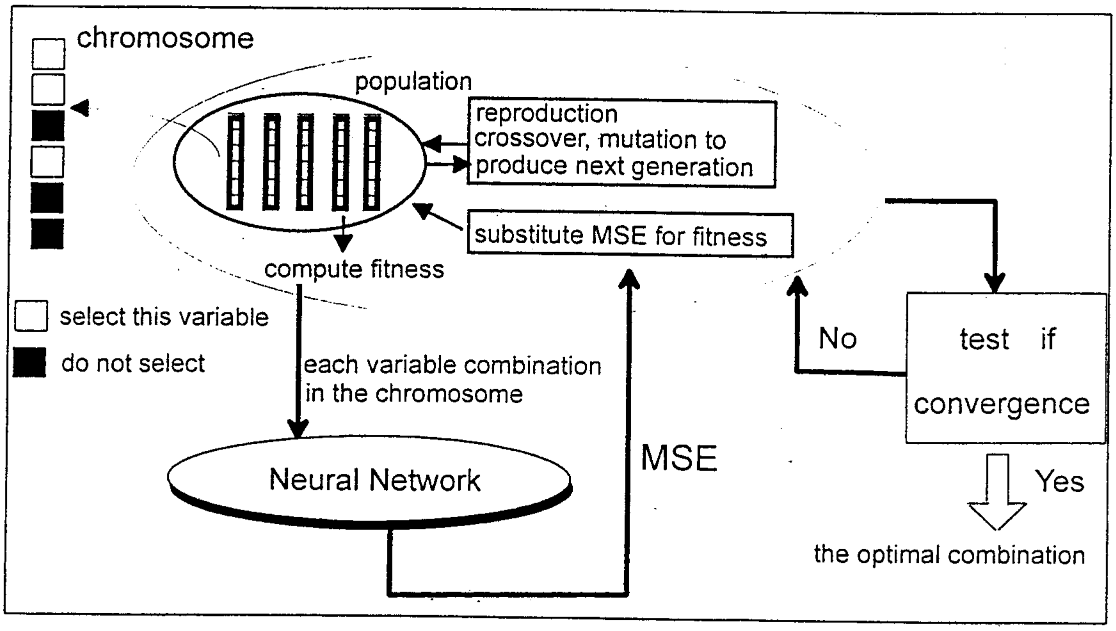

Figure 1: Genetic algorithm

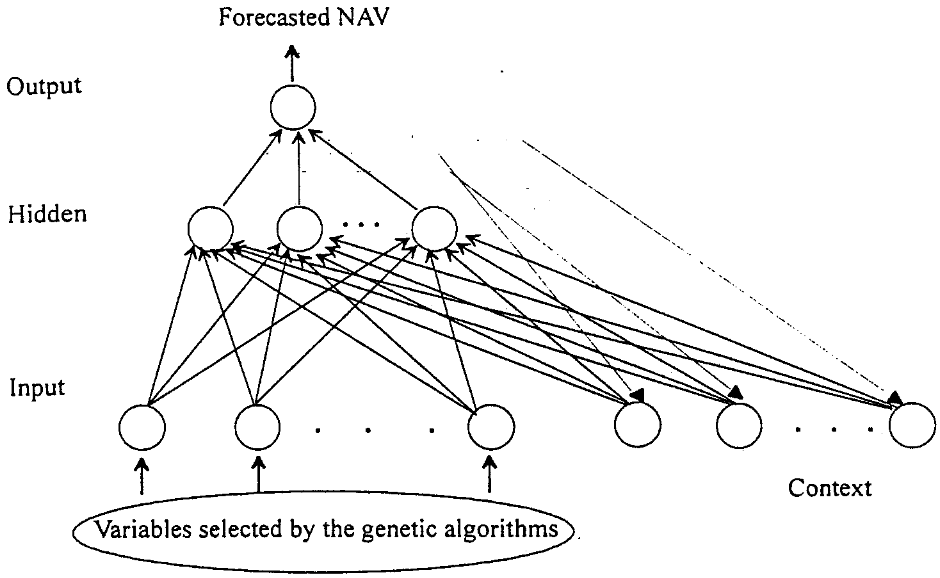

Figure 2: Recurrent neural network

3.2.1 Network training

Our process of training the neural network is listed as follows: each chromosome represented by some combination of selected variables will be used as a network input. Here the number of input neurons equals to the number of variables (denoted as a while square box in Figure 1) contained in each chromosome. In addition, two files containing training and testing data for the network will be created for each chromosome. The training file contains thirty data of the corresponding variables and a target output value. The network then learns from the data and computes the difference between the target value and forecasted value for each datum and then adjusts the weights in the network. In our experiment, each chromosome is learned 1000 epoches to ensure that the relationship between input variables and a target output value is fully captured.

3.2.2 Network testing

After the training process, the testing file containing fifteen monthly data of corresponding variables in each chromosome is input to the trained network to test the ability of prediction. In addition to the output of forecasting value, a value summing the difference between target and forecasting value of each datum in the file is also produced to serve as the fitness of each chromosome. The value will in turn be feedbacked into the genetic algorithm to select better combinations of variables. We use a mean square error (MSE) to calculate the forecasting error for each testing datum. MSE is defined as follow.

$$MSE = \sqrt{\frac{\sum_{i=1}^{15}(O_i - T_i)^2}{15}}$$

where O_i is the forecasting value for the NAV of the ith data in the testing file, T_i is the actual value for the NAV of the ith data.

4 RESULTS

In our experiment, we used the neural network software of PDP++ V1.2 built by Carnegie Mellon University. The program of the genetic algorithm was constructed by Linux C[12].

The selected variables for each mutual fund are shown in Table 3. The relative error for each mutual fund are displayed in Table 4. Figure 3 is the list of the actual and predicted NAVs of the best two and the worst two mutual funds. The common significant indicators selected by each mutual fund are: the net asset value of last month, performance of last month, the Beta value of funds, GNP, M1, Inflation rate, three-year rate, and BR.

If we take a closer look at each type of fund, we will find that there exists some commonly selected indicators for each type. We can further filter indicators selected by the types of funds. All we will do next is to collect the "significant indicators" and then add other "less important" indicators to find the fitter combinations. By "significant indicators" we mean that those indicators are selected by more than 10 funds. Then other "less important" indicators, which are selected by more than 6 (half of 12) funds but less than 10 funds, are added one by one to the significant indicators to form a new combination. Each new combination is trained and tested by the neural network so that the fittest combination that influences each type of funds most is then picked up. Table 5 shows the indicators that influence each type of funds most, respectively.

Var. ID	China Phonix	Kwang-Hua Agg.	First Global Duo-Fu	Core Pacific Agg.	Entrust Chong-Long	International First	Kwang-Hua Kwang-Hua	NITC	China China	Fortune United Dragon	Polaris Chronicle 2001	Fubon Fortune
M1												
M2												
M3	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
M4												
M5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
M6	✓	✓	✓		✓		✓	✓			✓	
M7			✓			✓			✓	✓		✓
M8						✓			✓	✓		✓
M9	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	
F1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
F2												
F3												
F4												
F5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
F6	✓	✓		✓	✓		✓	✓				
F7	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
F8			✓			✓			✓	✓		
F9	✓	✓		✓	✓		✓	✓	✓	✓	✓	✓
F10						✓						
F11												
F12												
T1												
T2			✓									✓
T3					✓							
T4			✓	✓								✓
T5												
T6						✓			✓	✓		✓
T7	✓	✓	✓		✓		✓	✓			✓	
T8												
T9												
T10				✓				✓				
T11			✓	✓		✓				✓		✓
T12						✓				✓		✓
T13	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
T14												
T15						✓		✓				
T16	✓	✓					✓				✓	
T17												
T18				✓								
T19				✓				✓	✓		✓	
T20												
T21												
T22			✓	✓		✓				✓		✓
T23												
T24												
T25										✓		

Table 3: Selected indicators

Fund Name	China Phonix	Kwang-Hua Aggres-sive-growth	First Global Duo-Fu	Core Pacific Aggres-sive-growth	Entrust Chong-Long	Interna-tional First	Kwang-Hua Kwang-Hua	NITC	China China	Fortune United Dragon	PolarisChroni-cle 2001	Fubon Fortune	Average relative error
Rela-tive error	5.96%	5.71%	5.48%	5.54%	7.17%	6.77%	6.15%	6.13%	3.80%	4.92%	5.80%	7.37%	5.90%

Table 4: Relative error for each mutual fund

Fund type	Aggressive-growth	Growth	Growth-income	All funds
Optimal combination	M3, M5, M6, M9, F1, F5, F6, F7, F9, T7, T13	M3, M5, M9, F1, F5, F7, F9, T7, T13	M3, M5, F1, F5, F7, F9, T13	M3, M5, M6, M9, F1, F5, F7, F9, T7, T13
Relative error for the above combination	8.32%	7.10%	12.93%	8.99%

Table 5: Significant indicators for each type of mutual funds

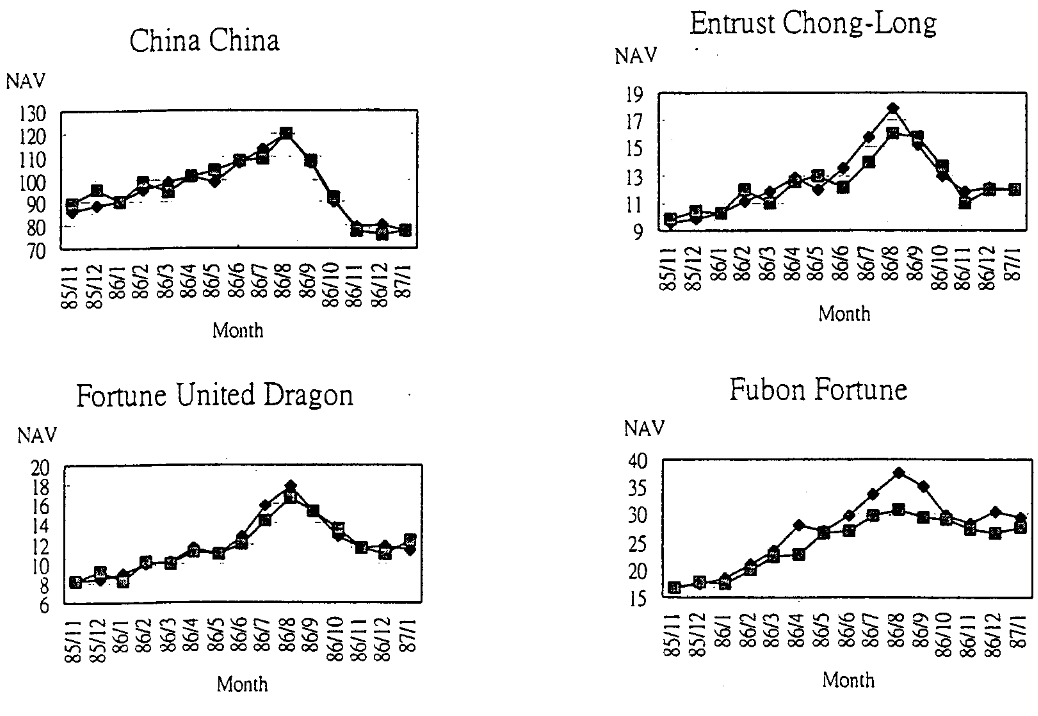

Figure 3: Actual and predicted NAV of the best two and the worst two mutual funds, actual value —▮— , predicted value —▮—

5 CONCLUSIONS

This paper aims at seeking the relationship among macro-economic indicators, monthly and yearly reports of fund companies, and the NAVs of mutual funds. Then we introduce a model making use of a neural network and a genetic algorithm to provide a forecasting system of mutual funds. We find that the common significant indicators for forecasting the NAVs of mutual funds are the net asset value of last month, performance of last month, the Beta value of funds, GNP, M1, Inflation rate, three-year rate, and BR. They are eight in total. The significant indicators for forecasting each type of mutual funds are also listed; the results appear that there is only little distinction among them.

Here are the suggestions of future work: (1) We have made the hypothesis that the behavior and moral of fund managers is constant and won't affect the NAV of funds. However, it is usually not the case. If these human factors can be quantified and gave reasonable values, it is helpful for making more accurate forecasts. (2) In Taiwan, the scale and number of mutual funds are still in progress and the samples for research are too scarce to come to a general and complete conclusion. If data samples are enlarged, a more precise and more general prediction is foreseeable. (3) Some methods, like wavelet, are useful to filter out noises. Since our data may contain noises that influence our results, it is recommended to "purify" the original time series data.

REFERENCES

[1] Back, B., Laitinen, T., and Sere, K., "Neural Networks an Genetic Algorithms for Bankruptcy Predictions." *Expert Systems with Applications*, Vol. 11, No. 4, 1996.

[2] Bandy, H. B., "Neural Network-based Trading System Design: Prediction and Measurement Tasks." *NeuroVe$T*, Vol. 2, No. 5, 1994, pp.26-32.

[3] Caldwell, "Design of Neural Network-based Financial Forecasting Systems: Data Selection and Data Processing." *NeuroVe*$t, Vol. 2, No. 3, 1994.

[4] Chiang, W.C., Urban, T.L., and Baldridge, G.W., "A Neural Network Approach to Mutual Fund Net Asset Value Forecasting." *International Journal Management Sicence*, Vol. 24, No. 2, 1995.

[5] Chou, Y.W., "Constructing Taiwan Stock Forecasting Model by Using a Neural Network." *Master Thesis*, National Taiwan University, 1996.

[6] Jurik, "A Primer on Market Forecasting with Neural Networks." NeuroVe$t, Vol. 1, No. 3, 1993.

[7] Katz and McCormick, "Neurogenetics and Its Use in Trading System Development." *NeuroVe$t*, Vol. 2, No. 2, 1994.

[8] Kamijo, K. and Tanigawa, T., "Stock Price Pattern Recog-nition -- A Recurrent Neural Network Approach." *IEEE International Joint Conference on Neural Networks*, 1990, pp.215-221.

[9] Lin, C.F., "Using a Neural Network to Support Stock Investment Decisions." *Master Thesis*, National Taiwan University, 1995.

[10] Malliaris, "A Neural Network Supports the Chaotic Paradigm for the S&P 500 Index." *NeuroVe$t*, Vol. 3, No. 2, 1995.

[11] March, "Supervised Evolution of the Neural Trader Component of a Stock Portfolio Trading System." *NeuroVe$t*, Vol. 3, No. 2, 1995.

[12] Welstead, S. T., *Neural Network and Fuzzy Logic Applications in C/C++*, John Wiley & Sons Inc., 1994.

[13] Wilson, R.L., and Sharda, R., "Bankruptcy Prediction Using Neural Networks." *Decision Support Systems*, Vol. 11, No. 5, 1994.

[14] *Summary of Taiwan Mutual Funds*, Investment and Trust Report, May 1997.