A Tentative Approach for Stock Selection Using Neural Network

Hui-Jen Yang[1] Wen-Yu Tsao[1] Yun-Long Lay[2]

***[1] Department of Information Management**
***[2] Department of Electronic Engineering**
National Chin-Yi Institute Technology College

Abstract

Self-organization map neural network technology can be used to differentiate stocks based on investors' requirement to group the similar stocks together. The purpose of this article is to present a tentative approach for investors to select the stocks using neural networks approach. Eight input vectors are selected as group criteria to generate the output stock selection portfolio. We hope to generate suggestions and make options available to investors when faced with the decision to consider stock selection portfolio.

Key Words: Stock Selection; Self-Organization Map(SOM); Neural Network; Stock investment; Multiple Discriminant Analysis (MDA);

類神經網路之試驗性股票選擇系統

楊惠貞[1] 曹文瑜[1] 賴雲龍[2]

*[1]資訊管理科
*[2]電子工程系
國立勤益技術學院

中文摘要

自組織映射神經網路可以依照投資者對股票需求的條件參數選出適合的股票群組。這篇文章的主要目的是提出一個選擇股票的應用系統雛型。以八個參數來規範股票挑選的準則，希望用這些參數經由類神經架構將適合投資的股票篩選出來，做爲投資者選股的決策參考。

1. Introduction

Stock investment in Taiwan is almost " the action of all people". Investors are eager to get profits form stock investment and have been busy counting their profits to worry about potential losses. We all know "a rising tide lifts all boats" when the stock markets are bullish. Investors will not pay much attention to the risk of stock investment. However, one might say that "a sinking ship will gets everybody wet" when the investors are pessimistic and uncertain for the stock market. Investors will encounter more risks on stock investment. In order to reduce the losses of stock investment, how to choose the good performance stock is a critical point for the investors? The Chinese proverb said "don't put all the eggs in one basket". In case the basket falls, all the eggs are all broken. Ideally, all available information pertinent to stock selection decision will be implemented in timely buy, hold or sell decisions and actions. But differences in the criteria used to select stocks often result in conflicts that delay such decisions and actions, hence result in a potential loss of predictive information. This is our primary motivation to present a tentative approach helping the investors to select the good performance stocks using neural network. We also hope to generate suggestions and make options available to investors when faced with the decision to consider stock selection portfolio.

The study of neural networks has grown significantly and is also a very conspicuous research field in the last decade. From the literature reviewed, we know that there are many special issues published for the research of neural networks such as the parallel distributed-processing systems, the study of brain-style computation, artificial neural systems, and so on [Rumelhart etc., 1994]. Around the 1990s, the industrial and commercial applications of neural networks have come into use in volume, especially for the financial service firms, which are becoming more and more

自組織映射神經網路可以依照投資者對股票需求的條件參數選出適合的股票群組。這篇文章的主要目的是提出一個選擇股票的應用系統雛型。以八個參數來規範股票挑選的準則，希望用這些參數經由類神經架構將適合投資的股票篩選出來，做爲投資者選股的決策參考。

1. Introduction

Stock investment in Taiwan is almost " the action of all people". Investors are eager to get profits form stock investment and have been busy counting their profits to worry about potential losses. We all know "a rising tide lifts all boats" when the stock markets are bullish. Investors will not pay much attention to the risk of stock investment. However, one might say that "a sinking ship will gets everybody wet" when the investors are pessimistic and uncertain for the stock market. Investors will encounter more risks on stock investment. In order to reduce the losses of stock investment, how to choose the good performance stock is a critical point for the investors? The Chinese proverb said "don't put all the eggs in one basket". In case the basket falls, all the eggs are all broken. Ideally, all available information pertinent to stock selection decision will be implemented in timely buy, hold or sell decisions and actions. But differences in the criteria used to select stocks often result in conflicts that delay such decisions and actions, hence result in a potential loss of predictive information. This is our primary motivation to present a tentative approach helping the investors to select the good performance stocks using neural network. We also hope to generate suggestions and make options available to investors when faced with the decision to consider stock selection portfolio.

The study of neural networks has grown significantly and is also a very conspicuous research field in the last decade. From the literature reviewed, we know that there are many special issues published for the research of neural networks such as the parallel distributed-processing systems, the study of brain-style computation, artificial neural systems, and so on [Rumelhart etc., 1994]. Around the 1990s, the industrial and commercial applications of neural networks have come into use in volume, especially for the financial service firms, which are becoming more and more relying on advanced computer technologies as a competitive weapon to set up and maintain competitive advantage in a global economic environment [Widrow, etc., 1994]. Neural networks play a key role with the wide scope of financial and investing applications, which cover from routine credit assessment operations to the driving of large-scale portfolio management strategies. In essence, the primary purpose of this article is to provide just a tentative method using neural network to generate the stock selection portfolio.

The remainder of the article is unfolded as follows. First is to do the literature review of stock investment techniques. In the next section we introduce the data and methodology. We explain why the input vectors are selected and how the vectors are normalized. The following section discusses the neural network model. The self-organization map neural network will be applied in the stock selection process. The learning process included the learning process and training samples are also introduced in this section. We sincerely hope that from this learning process to generate good performance stocks as a reference for investors. A testing sample is introduced in result section to generate the stock selection portfolio for investors. Finally we draw a conclusion and describe three advantages of SOMNN model.

2. Literature Review

Most of the previous researches of stock-picking strategy focused on multiple discriminant analysis (MDA), which has been used with quantitative data to help forecast stock price performance [Swales etc., 1992]. Discriminant analysis techniques are used to classify a set of independent variables into two or more mutually exclusive categories [Yoon, 1991]. Although several mathematical models have been developed to predict the performance of stocks, the results have been disappointing. Other method using nonlinear approaches may further enhance forecasting ability, such as neural network. Artificial neural network technology can be used to differentiate between stocks that perform well and those that perform poorly. To predict stock performance, artificial neural network model performed significantly better than linear multiple discriminant analysis approach [Swales etc., 1992].

3. Data and Methodology- input vectors & normalization

In order to simplify the analysis process, we only analyze last five years data for each company. The basic data are obtained from the Taiwan Stock Exchange (TSE) for 1994-1998. In general, a value-oriented screening process might use price/earnings ration, price-to-book-value ratio, and dividend yield to select "good buys" stocks [Wilcox, 1994]. In order to reduce the investment risk of stock selection, we add the stock turnover rate, earning per share (EPS), profit margin, debt ratio, and variance in this article. Therefore, eight input vectors are chosen to analyze the performance of stocks. They are described below.

(1) **Turnover rate denotes as tr.** The turnover rate, which is computed as the average five years' stock turnover rate, is obtained from the Taiwan Stock Exchange. In general, a higher turnover ratio of stock suggests higher risky by investors.

(2) **EPS denotes as eps.** The measure used most widely to appraise a company's operations is earning per share (EPS) of common stock. EPS is equal to earnings available to common stockholders divided by the weighted-average number of shares of common stock outstanding [Hermanson etc., 1995]. The dispersion, or range, of EPS forecasts has been found to be useful in predicting return [Bercel, 1994]. The last five years EPS is computed as an average EPS in this study.

(3) **P/E denotes as pe.** The price/earning (P/E) ratio is simply the market price of the firm's common stock divided by its annual earning per share [Lee, 1997]. In this paper, the P/E ratio is computed as the average market price of last five years divided by average annual earning per share of five years. The P/E ratio shows how much the investors are willing to pay for each dollar of the firm's earnings per share. P/E ratio measures the market's valuation of the firm relative to the income statement value for per-share earning. If the P/E ratio of the company is smaller than that of the industry, which implies the company's earnings have a lower value than its competitors. Comparing a firm's P/E to that of the stock market as a whole indicates the market's perception of the true value of the company.

(4) **Dividend yield denotes as div.** The history of stock returns is measured by the dividends, which are the only factor to determine firm value [Lee, 1997]. According to the Gordon model [Lee, 1997], if the firm increases its cash dividend, the price of its stock will increase. The investors or investment community can use this figure as a signal of good performance with the firm or a result of good management. This implies that an increase in dividends may lead investors to perceive a promising future and share price may increase. A drop in dividends may lead investors to fear a less promising future, resulting in a drop in share price. Dividend yield is usually a screen criteria to select "good buys" [Wilcox],1994. The dividend yield is simply the current dividend of firm's common stock divided by its current market price of stock [Lee, 1997]. The dividend yield in this article is computed as average dividend yields for last five years.

(5) **Debt ratio denotes as dr.** Debt ratio tells us the proportion of the company's assets that it has financed with debt [Horngren etc., 1994]. If the debt ratio is 1, then debt has been used to finance all the assets. A debt ratio of 0.50 means company has used debt to finance half its assets. The owners of the business have financed the other half. The higher the debt ratio, the higher the strain of interest pays each year and the principal amount at maturity. The lower the debt ratio, the less the business's future obligations have. The debt ratio is computed as the total liabilities divided by the total assets. The average of last five years' debt ratio will be computed in this paper as one of seven input vectors.

(6) **Profit margin denotes as pm.** The profit margin represents the proportion of each sales dollar that becomes profit or net income to the firm. Profit margin is computed as last five years' average net income divided by last five years' average sales.

(7) **Price-to-book-value-ratio denotes pbv.** Price-to-book-value ration measures the market's valuation relative to balance sheet equity. The price-to-book-value rat'eio is the market price per share divided by the book value of equity per share [Lee, 1997]. A higher ratio suggests that investors are more optimistic about the market value of stocks. A lower ratio indicates that investors are more pessimistic about the

market value of stocks. The last five years' average price-to-book-value ratio is used in this study.

(8) Variance denotes as vr. The higher the variance, the higher the risk of common stock return [Lee, 1997]. The total risk of common stocks can be measured by its variance of its return. That implies the risk of common stock can be measured by its variability of its net income in relation to this average, called variance. The last five average annual variances are computed in this article as one of eight input vectors. From the ratio of variance, this indicates that an increase in variance may lead investors to perceive a uncertain future. The share price of stocks may be more changeable and more risky in the future.

After aforesaid computation, each average input vector has its own maximum value. The eight average input vector values should be divided by maximum value separately to obtain their own normalization value, which is from 0 to 1.

4. Neural Network Model

4.1 Design of the Network- Self-Organization Map

The SOMNN algorithm as shown in Figure 1, [Kohonen, 1988 & 1990] is to transform an input signal vector of arbitrary dimension into one or two dimension discretion maps which display the important statistical characteristics of the input vector. After an input vector is calculated by SOMNN, a best matching or winning neuron is found in the output map. It shows that similar input vectors activate the selected neuron and its neighbors simultaneously. That indicates the similar characteristics of neuron will group together. The input vector, representing the set of input signals, is denoted by:

$$x = [x_1, x_2, \ldots x_p]^T \qquad (1)$$

Input

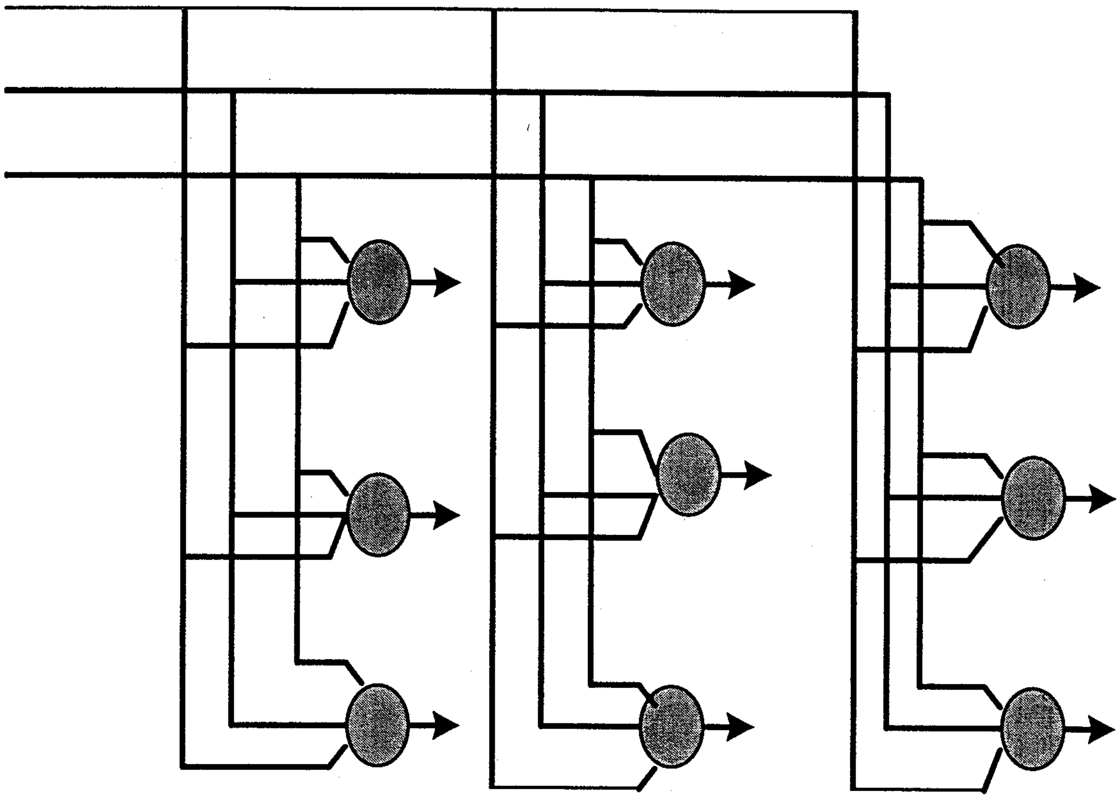

Figure 1: The Structure of SOM Neural Network

We use the following formula to calculate the distance between the vector of the output layers and that of the input layers.

$$||x_p - w_{pij}|| = [x_p - w_{pij}]' . [x_p - w_{pij}] = \sum_{pij} [x_p - w_{pij}]^2 \tag{2}$$

where

$||x_p - w_{pij}||$: Euclidean distance between vectors;

x_p : The input vector of the pth element;

w_{pij} : The weight vector between input and output layers.

The best-matching criterion is equivalent to the minimum Euclidean distance between vectors.

4.2 Learning Process and Training sample

Data Selection. We believe that the stock prices are determined by time-space patterns of economic index. In essence, each time-space pattern can be displayed by a vector, which is shown as Table 1.

Training Data. There are eight input vectors. There are 483 stocks cited in Taiwan Stock Exchange, OTC (Over the Counter) stocks exclusively. Therefore, the dimension of the training data is a 8 x 483 matrix. We develop a self-organization map (SOM) neural network model to group the suitable stocks that match our criteria and worthy to invest. The network architecture has only an input layer, which has eight input lines, and an output layer with 12 x 12 neurons. Assuming the model of training sample is as Xn = [tr,eps,pe,div,dr,pm,pbv,vr], then the training example is shown as Table 2.

Table 1: Input Vectors

1. Stock Turnover Rate
2. Earning Per Share
3. Price/Earning Ratio
4. Dividend Yield
5. Debt Ratio
6. Profit Margin
7. Price-to-book-value ratio
8. Variance

Table 2. Example for Training Data

vector no.	Turnover rate (tr)	EPS (eps)	P/E (pe)	dividend (div)	debt ratio (dr)	profit margin (pm)	price/book (pbv)	variance (vr)
X1	0.71	0.47	0.67	0.93	0.63	0.57	0.66	0.45
X2	0.57	0.46	0.77	0.88	0.48	0.59	0.57	0.60
X3	0.88	0.64	0.60	0.93	0.25	0.84	0.88	0.78
X4	0.45	0.74	0.34	0.78	0.40	0.90	0.93	0.82
X5	0.23	0.89	0.85	0.81	0.22	0.88	0.77	0.92
….	…	…	…	…	…	…		…
X473	0.67	0.22	0.13	0.21	0.78	0.28	0.44	0.38

After training by SOMNN, the output is an x and y coordinates file. Each input vector maps a neuron of the x, y coordinate. From the output data, it is obvious that if the criteria of two stocks are similar, their coordinates are close. If not, their coordinates are far apart. The output file of the training stocks is shown as Table 3.

4. Results

Table 3. Output data of Stock Selection

Vector no.	Ray Neuron no. x	Column Neuron No. y
X1	8	2
X2	10	6
X3	10	10
X4	3	9
X5	9	6
X6	7	6
/	.....	.....
/	.....	.....
/	.....	.....
X470	6	10
X471	4	5
X473	1	7

We assume a criteria testing data as Xt= [0.84,0.76,0.69,0.47, 0.77, 0.74, 0.83] and send the vector to the trained SOMNN. We obtain a coordinate, which is the best match stock. The output neuron of SOMNN has the property of neighbor relationship. We assume a neighbor distance randomly, which depends on investors' requirement, then we can obtain the dispersion, or range, of the stock group. As we see in figure 2, we assume the criteria and its output is the coordinate (9,6), which will map to the fifth stock. If we assume the neighbor distance as 2, then the stock 6 and stock 1 will be included in this group. It indicates that the stock 1, 5, and 6 are the selection portfolio for investors to invest.

6. Conclusion

The SOMNN approach used in this study was able to group the suitable stocks that match our criteria and worthy to invest. Investors who use SOMNN technology should be able to enhance their analysis of stock selection alternative and reduce their investment risk. It indicates an important clue to tend in the stock market for investors. The advantages of SOMNN model are (1) multi-variable comparison. We can feel free to choose variable items, depending on the condition of investor's primary need without changing the architecture of SOMNN; (2) neighbor distance. We can assume the range or dispersion of neighbor distance we need. It depends on the investor's requirement. If the investors are risk-seeker, they can assume a large range of neighbor distance, which will group more stocks under these criteria. If the investors are risk-aversive, they can assume a small range of neighbor distance, which will group less stocks under this criteria; (3) fast speed. The calculation speed of neural model is very fast. Once the learning process is completed, it is easy to calculate a testing vector to obtain the coordinate. It doesn't need any iterative calculation. It indicates that the neural network is only to do a simple computation.

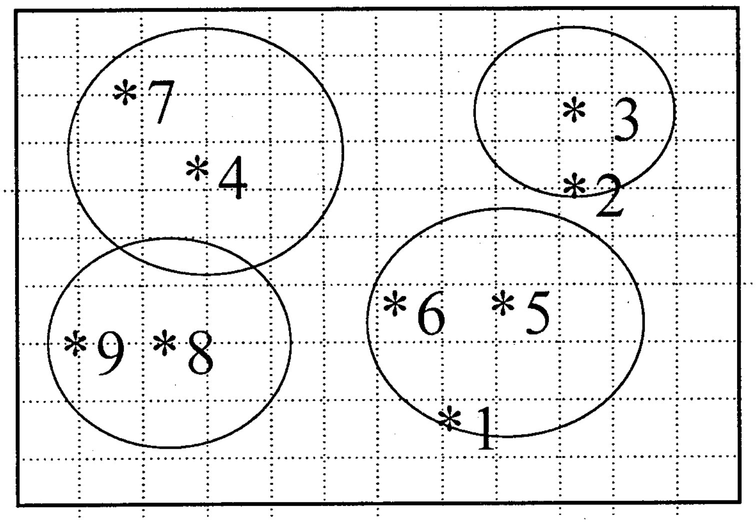

Figure 2. The grouping of stocks

References:

1. Bercel, A., "Consensus Expectations and International Equity Returns," Financial Analysts Journal, 1994 July-August, Vol.50, No.4, pp.76-80.
2. Hermanson, R. H., & Edwards, James Don, Financial Accounting - a Business Perspective, 1995 Sixth Edition, Irwin.
3. Horngren, C., T., Harrison Jr., Walter T., & Robinson, M.A., Principles of financial & Management Accounting, 1994, Prentice Hall.

4. Lee C. F., Finnerty J. E., & Norton E. A., Foundations of Financial Management, 1997, West Publishing Company.
5. Kohonen, T., "Self-Organization and Associative Memory," 3rd Ed. New York:Springer-Verlag, 1988.
6. Kohonen, T., "The Self-Organizing map," Proceeding of the IEEE 78, 1464-1480, 1990.
6. McConnell, D., Haslem, J.A., and Gobson, V.R., "The President's Letter to Stockholders: A New Look ," Financial Analysts Journal, 1986 September-October, pp.66-70.
7. Rumelhart , D.E., Widrow , B., and Lehr, M. A., "The Basic Ideas in Neural Networks," Communication of the ACM, March 1994, Vol.37, No.3, pp.87-92.
8. Swales Jr., G. & Yoon, Y., "Applying Artificial Neural Networks to Investment Analysis," Financial Analysts Journals, 1992 Sept-Oct, pp. 78-80.
9. Widrow, B., Rumelhart, D. E., and Leh, M. A., " Neural Networks: Applications in Industry, Business and Science," Communication of the ACM, March 1994, Vol.37, No.3, pp.93-105.
10. Wilcox, J. W., "Selling Bottlenecks: Causes and Treatments," Financial Analyst Journal, 1994 March-April, Vol. 50, No. 2, pp.49-54.
11. Yoon, Y. & Swales, G., "Predicting Stock Price Performance: A Neural Network Approach," Proceeding of the IEEE 24th Annual International Conference of Systems Sciences, 1991 January, pp. 156-162.