Stock Market Prediction System with Modular Neural Networks

Takashi Kimoto and Kazuo Asakawa
Computer-based Systems Laboratory
FUJITSU LABORATORIES LTD., KAWASAKI
1015 Kamikodanaka, Nakahara-Ku, Kawasaki 211, JAPAN

Morio Yoda and Masakazu Takeoka
INVESTMENT TECHNOLOGY & RESEARCH DIVISION
The Nikko Securities Co., Ltd.
3-1 Marunouchi 3-Chome, Chiyoda-Ku Tokyo 100, JAPAN

Abstract: This paper discusses a buying and selling timing prediction system for stocks on the Tokyo Stock Exchange and analysis of internal representation. It is based on modular neural networks[1][2]. We developed a number of learning algorithms and prediction methods for the TOPIX(Tokyo Stock Exchange Prices Indexes) prediction system. The prediction system achieved accurate predictions and the simulation on stocks trading showed an excellent profit. The prediction system was developed by Fujitsu and Nikko Securities.

1. Introduction

Modeling functions of neural networks are being applied to a widely expanding range of applications in addition to the traditional areas such as pattern recognition and control. Its non-linear learning and smooth interpolation capabilities give the neural network an edge over standard computers and expert systems for solving certain problems.

Accurate stock market prediction is one such problem. Several mathematical models have been developed, but the results have been dissatisfying. We chose this application as a means to check whether neural networks could produce a successful model in which their generalization capabilities could be used for stock market prediction.

Fujitsu and Nikko Securities are working together to develop TOPIX's a buying and selling prediction system.

The input consists of several technical and economic indexes. In our system, several modular neural networks learned the relationships between the past technical and economic indexes and the timing for when to buy and sell. A prediction system that was made up of modular neural networks was found to be accurate. Simulation of buying and selling stocks using the prediction system showed an excellent profit. Stock price fluctuation factors could be extracted by analyzing the networks.

Section 2 explains the basic architecture centering on the learning algorithms and prediction methods. Section 3 gives the outcome of the prediction simulation and results of the buying and selling simulation. Section 4 compares statistical methods and proves that stock prices can forecast by internal analysis of the networks and analysis of learning data.

2. Architecture
2.1 System Overview

The prediction system is made up of several neural networks that learned the relationships between various technical and economical indexes and the timing for when to buy and sell stocks. The goal is to predict the best time to buy and sell for one month in the future.

TOPIX is an weighted average of market prices of all stocks listed on the First Section of the Tokyo Stock Exchange. It is weighted by the number of stocks issued for each company. It is used similar to the Dow-Jones average.

Figure 1 shows the basic architecture of the prediction system. It converts the technical indexes and economic indexes into a space pattern to input to the neural networks. The timing for when to buy and sell is a weighted sum of the weekly returns. The input indexes and teaching data are discussed in detail later.

2.2 Network Architecture
2.2.1 Network Model

Figure 2 shows the basic network architecture used for the prediction system. It consists of three layers: the input layer, the hidden layer, and the output layer. The three layers are completely connected to form a hierarchical network.

Each unit in the network receives input from low-level units and performs weighted addition to determine the output. A standard sigmoid function is used as the output function. The output is analog in the [0,1] section.
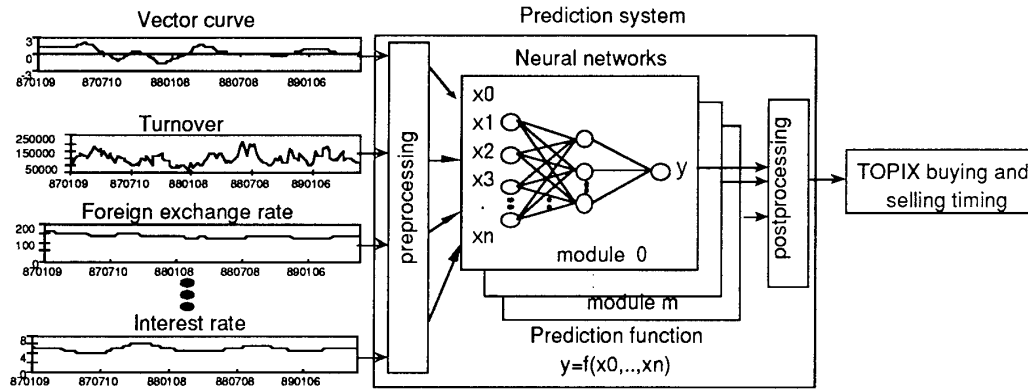
Figure 1 Basic architecture of prediction system

2.2.2 High-speed Learning Algorithm

The error back propagation method proposed by Rumelhart[3] is a representative learning rule for hierarchical networks. For high-speed learning with a large volume of data, we developed a new high-speed learning method called supplementary learning[4].

Supplementary learning, based on the error back propagation, automatically schedules pattern presentation and changes learning constants.

In supplementary learning, the weights are updated according to the sum of the error signals after presentation of all learning data. Before learning, tolerances are defined for all output units. During learning, errors are back-propagated only for the learning data for which the errors of output units exceed the tolerance. Pattern presentation is automatically scheduled. This can reduce the amount calculation for error back propagation.

Figure 2 Neural network model

As learning progresses, learning data for which tolerances are exceeded are reduced. This also reduces the calculation load because of the decreased amount of data that needs error back propagation. High-speed learning is thus available even with a large amount of data.

Supplementary learning allows the automatic change of learning constants depending on the amount of learning data. As the amount of learning data changes and learning progresses, the learning constants are automatically updated. This eliminates the need for changing learning parameters depending on the amount of learning data.

With supplementary learning, the weight factor is updated as follows:

$$\Delta w(t) = - (\varepsilon / \text{learning\_patterns}) \, \partial E / \partial W + \alpha \Delta w(t-1)$$

Where
  ε : learning rate
  α : momentum
  learning_patterns : number of learning data items that require error back propagation

The value of ε is divided by the number of learning data items that actually require error back propagation. The required learning rate is automatically reduced when the amount of learning data increases. This allows use of the constants ε regardless of the amount of data.

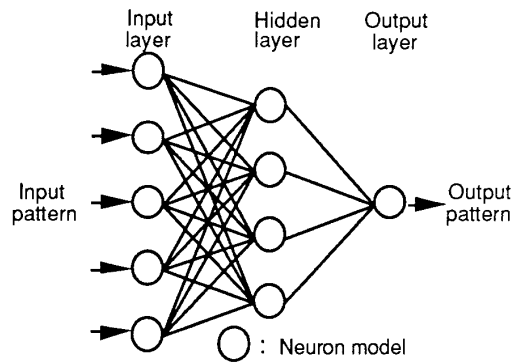As learning progresses, the amount of remaining learning data decreases. This automatically increases the

I - 2

learning rate. Using this automatic control function of the learning constants means there is no need to change the constants ($\varepsilon = 4.0$, $\alpha = 0.8$) throughout simulation and that high-speed learning can be achieved by supplementary learning.

2.3 Learning Data
2.3.1 Data Selection
We believe stock prices are determined by time-space patterns of economic indexes such as foreign exchange rates and interest rates and of technical indexes such as vector curves and turnover.

The prediction system uses a moving average of weekly average data of each index for minimizing influence due to random walk.

Table 1 lists some of the technical and economic indexes used. The time-space pattern of the indexes were converted into space patterns. The converted indexes are analog values in the [0,1] section.

Table 1 Input indexes

1. Vector curve	2. Turnover
3. Interest rate	4. Foreign exchange rate
5. New York Dow-Jones average	6- Others

2.3.2 Teaching Data
The timing for when to buy and sell is indicated as an analog value in the [0,1] section in one output unit. The timing for when to buy and sell used as teaching data is weighted sum of weekly returns. When the TOPIX weekly return is r_i, teaching data $r_N(t)$ is defined as:

$$r_t = \ln(\text{TOPIX}(t) / \text{TOPIX}(t-1)) \qquad \text{TOPIX}(t) : \text{TOPIX average at week } t$$

$$r_N(t) = \Sigma_i \phi^i r_{t+i} \qquad \phi : \text{Weight}$$

2.4 Preprocessing
Input indexes converted into space patterns and teaching data are often remarkably irregular. Such data is preprocessed by log or error functions to make them as regular as possible. It is then processed by a normalization function which normalizes the [0,1] section, correcting for the irregular data distribution.

2.5 Learning Control
In the TOPIX prediction system, we developed new learning control. It automatically controls learning iterations by referring to test data errors, thereby preventing overlearning. The learning control allows two-thirds of data in the learning period to be learned and uses the rest as test data in the prediction system. The test data is evaluation data for which only forward processing is done during learning, to calculate an error but not to back propagate it.

Our learning control is done in two steps. In the first step, learning is done for 5,000 iterations and errors against test data are recorded. In the second step, the number of learning iterations where learning in the first step suggests a minimum error against the test data is determined, and relearning is done that number of iterations. This prevents overlearning and acquires a prediction model involving learning a moderate number of times. In the second step, learning is done for at least 1,000 iterations.

2.6 Moving Simulation
For prediction of an economic system, such as stock prices, in which the prediction rules are changing continuously, learning and prediction must follow the changes.

We developed a prediction method called moving simulation. In this system, prediction is done by simulation while moving the objective learning and prediction periods. The moving simulation predicts as follows.
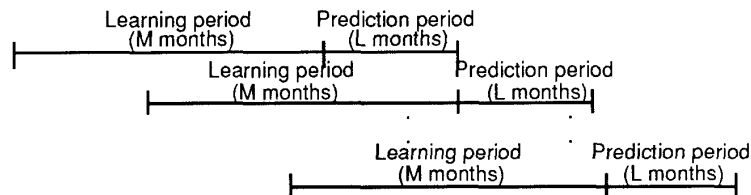

Figure 3 Moving simulation

As shown in Figure 3, the system learns data for the past M months, then predicts for the next L months. The system advances while

repeating this.

3. Result of Simulations
3.1 Prediction Simulation
We verified the accuracy of the prediction system by simulating prediction of the timing for when to buy and sell. We used historical data of stock prices, technical indexes and economical indexes.

The TOPIX prediction system improves its prediction accuracy by averaging prediction results of modular networks that learn for different learning data items. Four independent modular networks learn for four types of different learning data. Moving simulation is used with L as one month. The average of prediction outputs from these networks became the prediction output from the system. Prediction was thus repeated by moving simulation for each month to verify accuracy. Prediction was done for 33 months from January 1987 to September 1989.

Figure 5 shows the correlation coefficient between the predictions and teaching data and those of individual networks and prediction system. The prediction system uses the average of the predictions of each network. Thus the prediction system could obtain a greater correlation coefficient for teaching data than could be obtained with neural network prediction.

Table 2 Correlation coefficient

	Correlation Coefficient
Network1	0.435
Network2	0.458
Network3	0.414
Network4	0.457
System	0.527

3.2 Simulation for Buying and Selling Simulation
To verify the effectiveness of the prediction system, a simulation of buying and selling of stock was done. Buying and selling according to the prediction system made a greater profit than the buying and holding.

Buying and selling was simulated by the one-point buying and selling strategy, so performance could be clearly evaluated. One-point buying and selling means all available money is used to buy stocks and means all stocks held are sold at a time. In the prediction system, an output of 0.5 or more indicates buy, and an output less than an 0.5 indicates sell. Signals are intensified as they get close to 0 or 1.

The buying and selling simulation considered "buy" to be an output above some threshold and "sell" to be below some threshold. Figure 4 shows an example of the simulation results. In the upper diagram, the buy-and-hold performance (that is, the actual TOPIX) is shown as dotted lines, while the prediction system's performance is shown as solid lines. The TOPIX index of January 1987 was considered as 1.00, it was 1.67 by buy-and-hold at the end of September 1989. It was 1.98 by the buying and selling operation according to the prediction system. Use of the system showed an excellent profit.
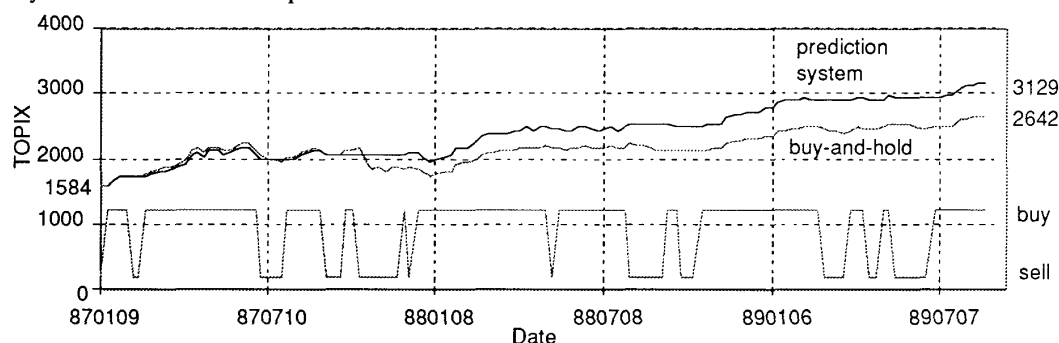

Figure 4 Performance of the prediction system

4. Analysis
4.1 Comparison with Multiple Regression Analysis
The timing for when to buy and sell stocks is not linear, so statistical methods are not effective for creating a model. We compared modeling with the neural network and with multiple regression analysis. Weekly learning data from January 1985 to September 1989 was used for modeling. Since the objectives of this test were comparison of learning capabilities and internal analysis of the network after learning, the network learned 100,000 iterations.

The hierarchical network that had five units of hidden layers learned the relationships between various

I - 4

economic and technical indexes and the timing for when to buy and sell. The neural network learned the data well enough to show a very high correlation coefficient, the multiple regression analysis showed a lower correlation co-efficient. This shows our method is more effective in this case. Table 3 shows the correlation coefficient produced by each method. The neural network produced a much higher correlation coefficient than multiple regression.

Table 3 Comparison of multiple regression analysis and neural network

	Correlation coefficient with teaching data
Multiple regression analysis	0.543
Neural network	0.991

4.2 Extraction of Rule

The neural network that learned from January 1985 to September 1989 (Section 4.1) was analyzed to extract information on stock prices stored during that period.

Cluster analysis is often used to analyze internal representation of a hierarchical neural network [5][6]. In 1987 stock prices fluctuated greatly. The hidden layer outputs were analyzed to cluster learning data. The cluster analysis was applied to the output values in the [0,1] sections of the five units of hidden layers. Clustering was done with an inter-cluster distance determined as Euclidian, and the clusters were integrated hierarchically by the complete linkage method. Figure 5 shows the cluster analysis results of the hidden layers in 1987. It indicates that bull, bear, and stable markets each generate different clusters.

From cluster analysis, characteristics common to data that belong to individual clusters were extracted by analyzing the learning data. Figure 6 shows the relationships between TOPIX weekly data and six types of clusters in 1987.

This paper analyzes the factors for the representative bull (2)(3) and bear (6) markets in 1987 as follows.

Data (2) and (3) belong to different clusters but have similar characteristics. Figure 7 shows learning data corresponding to clusters (2) and (3). The horizontal axis shows some of the index-
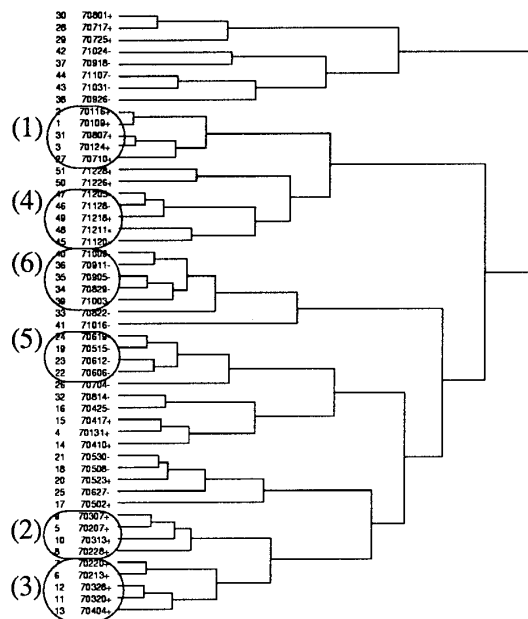

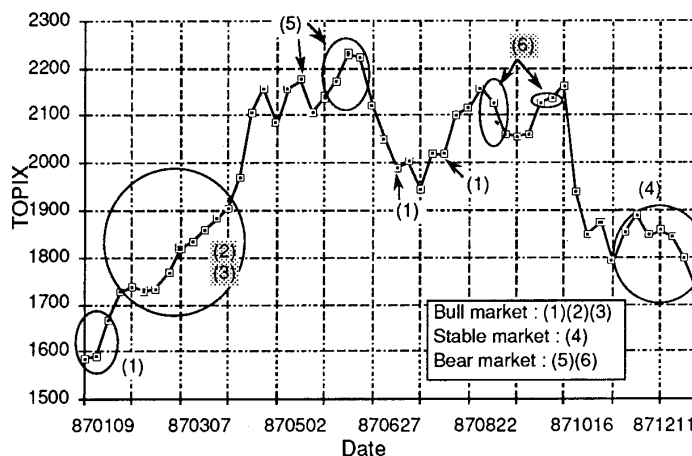Figure 5 Cluster analysis results from 1987

Figure 6 TOPIX in 1987

es of the neural network. The vertical axis show the value of each index. For example, New York Dow-Jones average is low when it is close to 0 and is high when close to 1.

This diagram suggests that the vector curve in the bull market during February to the beginning of April in 1987 were high enough to indicate a high-price zone. At the same time, however, the high turnover kept the market going up. Also, the low interest rates and high New York Dow-Jones average helped pushed up stock prices.

Figure 8 shows the learning data corresponding to (6) in Figure 6. It is obvious that the high interest rate pulled prices down.
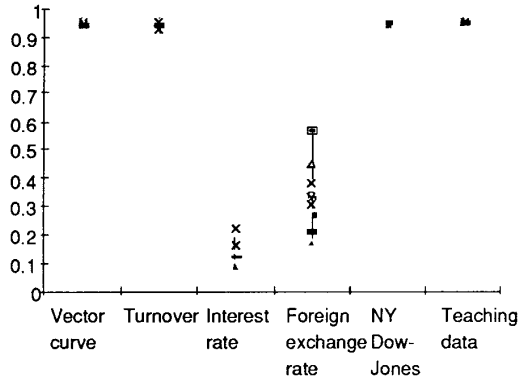
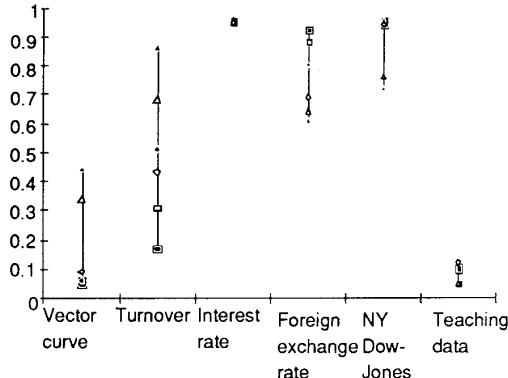Figure 7 Input indexes for bull market(2)(3)

Figure 8 Learning data in bear market (6)

Part of the learning data in 1987 was analyzed. It was proved that the causes of stock price fluctuation could be analyzed by extracting the characteristics common to the learning data in the clusters obtained by cluster analysis of the neural network.

5.0 Further Research

The following subjects will be studied.
- Using for actual stock trading

The current prediction system uses future returns to generate teaching data. A system in which teaching data is generated in combination with a statistical method must be developed.
- Adaptation of network model that has regressive connection and self-looping

The current prediction system requires much simulation to determine moving average. Automatic learning of individual sections requires building up a prediction system consisting of network models fit to time-space processing.

6.0 Summary

This paper has discussed a prediction system that advises the timing for when to buy and sell stocks. The prediction system made an excellent profit in a simulation exercise. The internal representation also was discussed and the rules of stock price fluctuation were extracted by cluster analysis.

For developing the prediction system, Nikko Securities offered investment technology and know-how of the stock market and Fujitsu offered its neural network technology. Fujitsu and Nikko Securities are studying further to build up more accurate economic prediction systems.

References:
[1] S. Nagata, T. Kimoto and K. Asakawa, Control of Mobile Robots with Neural Networks, INNS, 1988, 349
[2] H. Sawai, A.Waibel, et al., Parallelism, Hierarchy, Scaling in Time-Delay Neural Networks for Spotting Japanese Phonemes/CV-Syllables, IJCNN vol II, 1989, 81-88
[3] D. E. Rumelhart, et al., Parallel Distributed Processing vol. 1, The MIT Press, 1986
[4] R.Masuoka, et al., A study on supplementary learning algorithm in back propagation, JSAI, 1989, 213-217.
[5] T. J. Sejnowski,C. R. Rosenberg, Parallel Networks that Learn to Pronounce English Text, Complex Systems, 1, 1987
[6] R. Paul Gorman, T. J. Sejinowski, Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets, Neural Networks, Vol 1, No 1, 1988, 75-90