International Journal of Fuzzy Logic and Intelligent Systems, vol. 3, no. 1, June 2003 pp. 7-12

A Comparative Study on the Prediction of KOSPI 200 Using Intelligent Approaches

Hyeon Bae, Sungshin Kim, Haegyun Kim*, Kwang Bang Woo**

School of Electrical and Computer Engineering, Pusan National University, Busan, Korea
***Air Conditioning Division, Digital Appliance Company, LG Electronics, Gyeongnam, Korea**
****Automation Technology Research Institute, Yonsei University, Seoul, Korea**

Abstract

In recent years, many attempts have been made to predict the behavior of bonds, currencies, stock or other economic markets. Most previous experiments used the neural network models for the stock market forecasting. The KOSPI 200 (Korea Composite Stock Price Index 200) is modeled by using different neural networks and fuzzy logic. In this paper, the neural network, the dynamic polynomial neural network (DPNN) and the fuzzy logic employed for the prediction of the KOSPI 200. The prediction results are compared by the root mean squared error (RMSE) and scatter plot, respectively. The results show that the performance of the fuzzy system is little bit worse than that of the DPNN but better than that of the neural network. We can develop the desired fuzzy system by optimization methods.

Key words : KOSPI 200, fuzzy, neural network, dynamic polynomial neural network

1. Introduction

Recently, the stock price index is being focused because this represents the economical development and conditions. The prediction of the stock price index is to predict the price trend of the future for preparing the dynamical changes of the markets. Practically the economist or statistics are using the charts or history data to predict the stock price index [1]. In general, the prediction of the stock price is separated into the technical analysis and fundamental analysis. But in the real market, it is very difficult and complicated to predict because there are the complexity and uncertainty in the market. It causes the incorrect predicted results. The technical analysis focuses on market movement, while the fundamental analysis concentrates on economical balance of supply and demand that is the cause of the price pattern as rise, steadiness, and fall. Both analyzing approaches have the same goals to identify the direction of the price movement. But the approach way is difference that fundamental analyzers study the causes of the market movement while technical analyzers research the effects of the market movement.

The artificial techniques are applied for the prediction system of the stock price and introduced as the commercial application for the prediction [2]. The neural networks are broadly employed for the prediction system that uses just history data. But the performance can be worse with respect to the market conditions. On the other hand, the fuzzy logic can use the humans knowledge, so if we can have good market information, the fuzzy model shows the better performance under the changeable market conditions [3], [4]. Also the neural network model and the dynamic polynomial neural network (DPNN) model are introduced in this paper [5], [6]. The inputs of both neural networks are the stock price of the industrial fields for the prediction of the tomorrows stock price that affects the composite stock price index. Also the fuzzy model is applied to do the same purpose. The inputs of the fuzzy model are extracted based upon the moving average that is one of typical prediction approaches of the stock price index. In the experimental results, the prediction performance is presented that is produced by the neural network model, the DPNN model, and the fuzzy model using the KOSPI 200. The major goal of the results is to compare all of the performance of the models. Also the merits of the fuzzy system will be shown in the paper. Simulated annealing called the derivative free optimization method is employed to optimize the fuzzy membership functions [7], [8], [9].

Manuscript received May. 15, 2003; revised May. 1, 2003

2. Background of System

2.1 Neural Network

An important difference between neural networks and standard IT solutions is their ability to learn. The main advantages of ANNs stem from their ability to recognize patterns in data. This can be achieved without a priori knowledge of causal rrelationships, as would be necessary in knowledge based systems. ANNs ability to generalize relationships from input patterns make them less sensitive to noisy data than other approaches. Their ability to represent non-linear relationships makes them well suited for a large variety of applications, such as some industrial control systems or financial forecasting, where linear relationships do not hold.

In this paper, three layers are constructed consisting of input layer (7 neurons), hidden layer (10 neurons), and output layer (1 neuron). The procedure of finding a gradient vector in a network structure is generally referred to as backpropagation because the gradient vector is calculated in the direction opposite to the flow of the output of each node. The input variables consist of seven numbers of values that are the index of the manufacturing, telecommunication, construction, financial service, wholesale, retail & other services, and the KOSPI 200 index of yesterday and today. The tomorrow's KOSPI 200 index is predicted by using these inputs.

2.2 Dynamic Polynomial Neural Network (DPNN)

2.2.1 Basic Structure of DPNN

PNN based on the GMDH algorithm is a useful method to model the system from many observed data and input variables. It is widely employed for the modeling of dynamic systems, prediction, and artificial intelligent control because of the advantages in data handling. The basic structure of DPNN is shown in Fig. 2 that includes the recurrent inputs with one-to-n time delayed output variables. Therefore, this kind of PNN is called as DPNN [10]. In DPNN system, a reference function is Volterra functional series known as Kolmogorov-Gabor polynomial.

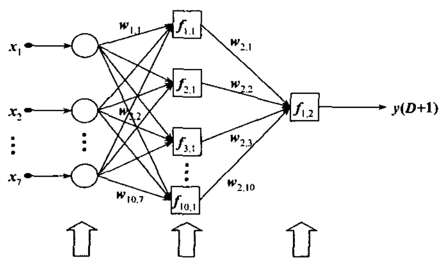

Fig. 1. Structure of neural network.

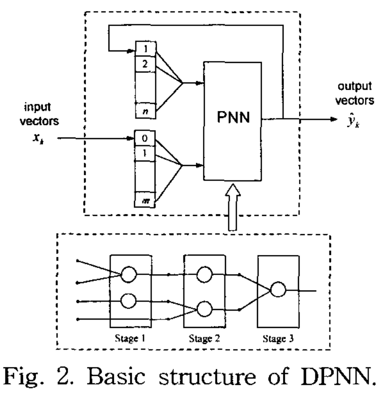

Fig. 2. Basic structure of DPNN.

$$y = a_0 + \sum_{i=1}^{m} a_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij}x_ix_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} a_{ijk}x_ix_jx_k + \ldots, \quad (1)$$

The least square estimator (LSE) is applied for estimating the parameters at an each node to minimize an objective function. If there is m number of data, the output equations at each node are formed as follows:

$$\begin{bmatrix} z_{(1)} \\ z_{(2)} \\ \vdots \\ z_{(m)} \end{bmatrix} = \begin{bmatrix} 1 & x_{1(1)} & x_{2(1)} & x_{1(1)}x_{2(1)} & x_{1(1)}^2 & x_{2(1)}^2 \\ 1 & x_{1(2)} & x_{2(2)} & x_{1(2)}x_{2(2)} & x_{1(2)}^2 & x_{2(2)}^2 \\ \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_{1(m)} & x_{2(m)} & x_{1(m)}x_{2(m)} & x_{1(m)}^2 & x_{2(m)}^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_5 \end{bmatrix} \quad (2)$$

where $x_0 = 1$, $y_{ij} = \Phi\,\omega$.

Statistical learning networks have no loops. Therefore the network is a tree of interconnected functions that implements a single input/output function. In the literature, there are described several composition schemes for network functions and corresponding estimation algorithms [11]. The parameters are estimated by

$$J = \sum_{k=1}^{n}(z(k) - \hat{z}(k))^2 = \| \mathbf{z} - \mathbf{\Phi w} \|^2 \quad (3)$$

$$\mathbf{w} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{z} \quad (4)$$

All of parameters are estimated in this step and then the polynomial equation is completely determined for the node's output in each layer.

2.2.2 Performance Criterion (PC)

One of the most common and difficult problems in the empirical modeling arena is the question of when to stop searching the free parameters or adding terms to a model. The goal of modeling is to obtain a predictive model that generalizes across many such samples to the universe at large, and not merely to the sample at hand.

By using two-separated data set, DPNN estimates parameters of each node and composes the network structure of the dynamic system. Training data set is used to adjust the parameter of function of each node and testing data set is employed to evaluate performance of models. DPNN selects the input to the next node by performance indicator and according to training error and testing error by performance criterion, network structure is determined by itself as shown in Fig. 3.

The proposed performance criterion for a model selection is based on the GMDH to minimize the error and at the same time to prevent *overfitting* of the empirical data set. As a proper criterion for the verification of the model, the observed data are divided into two sets, N_A for training and N_B for testing purposes [12]. The performance criterion is calculated as

$$\begin{aligned} e_1^2 &= \sum_{i=1}^{n_A}(y_i^A - f_A(x_i^A))^2/n_A, \\ e_2^2 &= \sum_{i=1}^{n_B}(y_i^B - f_A(x_i^B))^2/n_B, \\ PC &= e_1^2 + e_2^2 + \eta(e_1^2 - e_2^2)^2 \end{aligned} \quad (5)$$

where n_A is the number of data points in the data set N_A, and y_A^i is the real output of data set N_A. $f_A(x_B^i)$ is the estimated output for the data set N_B from the model identified using the data set N_A is a weight factor on the difference between e_1 and e_2. We want to minimize the PC

to find the best model for N_A and N_B.

2.3 Fuzzy Logic

2.3.1 Basic Concept of Fuzzy Logic

Formally, fuzzy logic is a structured, model-free estimator that approximates a function through linguistic input/output associations. Fuzzy rule-based systems apply these methods to solve many types of real-world problems, especially where a system is difficult to model, is controlled by a human operator or expert, or where ambiguity or vagueness is common. A typical fuzzy system consists of a rule base, membership functions, and an inference procedure (see Fig. 4). Some fuzzy logic applications include control, information systems, pattern recognition, and decision support.

3. Prediction Models

3.1 Knowledge Information

3.1.1 KOSPI 200 Index

KOSPI 200 is to deal with two stock price markets. One is the futures market started from Jun 3 1996; the other is the options market opened from July 7 1997 KOSPI 200 is a capitalization-weighted index composed of 200 stocks from a broad range of industries. The component stocks are weighted according to the total market value of their outstanding shares. The impact of a component's price change is proportional to the issue's total market value, which is the share price time the number of shares outstanding. The base value for KOSPI 200 is adjusted to reflect changes in capitalization resulting from mergers, acquisition, stock rights, substitution, etc. The feature and methodology of the KOSPI 200 calculation is virtually identical to the KOSPI calculation except the fact that the KOSPI 200 is a 200-sampled constituent index. The base index is 100 and the base date is January 3, 1990. The KOSPI 200 is calculated by the following formula:

$$KOSPI\ 200 = \frac{\text{Current aggregated market value of component stocks}}{\text{Base aggregated market value of component stocks}} \times 100$$

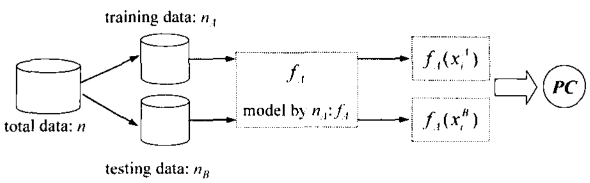

Fig. 3. Data split for training and testing data.

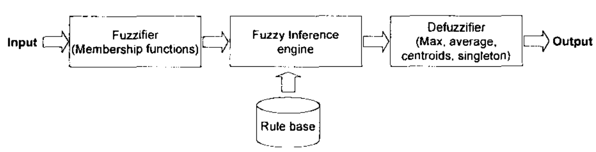

Fig. 4. Basic procedure of fuzzy inference system.

3.1.2 Moving Average

One of the indexes for the stock price is the moving average that presents the trend of the stock price. The main concept of the moving average is to extract the specific pattern from the history data for the future prediction. For each point in time, t, the Moving Average, $A(t)$, is defined to be equal to

$$A(t) = (P(t) + P(t-1) + \cdots + P(t-n-1))/n \tag{6}$$

3.1.3 Trifle Cross Method

Trifle cross method is one of the popular analysis method of the index that combines the moving average of the 4 days-9 days-18 days generally. In this paper, the moving average of 5 days-10 days-20 days is applied for the trifle cross method. Fig. 5 is a simple example of the application. The general knowledge is as follows.

A: the moving average of 5 days is going through the moving average of 10 days and 20 days
$\Rightarrow$ golden cross (rising trend)
B: the moving average of 5days is steady in the moving average of 120 days, 10 days is rising, and 20 days is edge along.
$\Rightarrow$ steadiness
C: the moving average of 5 days is under 10 days and 10 days is under 20 days.
$\Rightarrow$ dead cross (falling trend)

From the cross pattern of the moving averages we can decide the trend of the stock price as rising, steadiness, and falling.

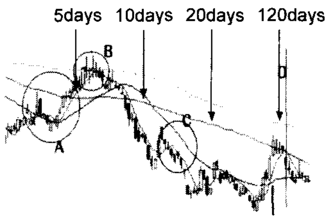

Fig. 5. Moving average crosses.

3.2 Extraction of Knowledge Information

The inputs of the fuzzy system is defined by the general knowledge of the KOSPI 200 that is the pattern of the moving average such as the trifle cross method. This knowledge is called knowledge information. In the fuzzy system, the fuzzy rules have to be extracted by the expert's knowledge or data because the fuzzy logic is one of rule-base model. But in the neural network models, inputs are just data variables. It is easier model to operate than the fuzzy model.

3.2.1 Knowledge Extraction for Fuzzy Rules

As mentioned before, the outputs of the trifle cross method is employed for the inputs of the fuzzy system. Table 1 and 2 indicate the way to extract inputs from the moving average lines. As shown in table 1, the trend of tomorrow is determined by the moving average of today. This basic concept of the stock price index variation can be expressed by the following flow.

Table 1. Basic knowledge for generating rules.

20days 5days 10days Rising 1	10days 5days 20days Falling 1
5days 20days 10days Rising 2	10days 20days 5days Falling 2
5days 10days 20days Rising	20days 10days 5days Falling

Table 2. Quantization of each defined case.

5days-10days (+) Rising 1	5days-10days (-) Falling 1
5days-10days (+) Rising 2	5days-10days (-) Falling 2
5days-20days (+) Rising	5days-20days (-) Falling

Table 3. Fuzzy look-up table of defined rules.

T-1 \ T	Rising 1	Rising 2	Rising	Falling 1	Falling 2	Falling
Rising 1		PB	PB			NS
Rising 2	NS		PB	ZE		
Rising		NS	PB	NS	NB	
Falling 1			PS		NB	NB
Falling 2	ZE			PS		NB
Falling	PS	PB			PS	NB

$$Rising1 \Rightarrow Rising2 \Rightarrow Rising \Rightarrow Falling1 \Rightarrow Falling2 \Rightarrow Falling$$

All change is moving with 6 steps. Table 2 shows the calculation technique that is based upon Table 1. If we want to predict tomorrow's prices index, the first input of the fuzzy system is the difference moving average between the 5 days and 10 days before from today (*T*). The other input is the difference moving average between the 5 days and 10 days before from yesterday (*T*-1). From this procedure, the fuzzy rules are extracted based on the experienced knowledge kind of the trifle cross method.

3.2.2 Fuzzy Rule Generation for the Prediction

Through preprocessing two fuzzy inputs are generated as *T* and *T*-1. And the output of the fuzzy model is the real stock price index of tomorrow. In the fuzzy model Gaussian functions are employed for the fuzzy membership functions and the Mamdani fuzzy inference method is applied for the rule inference. Total rules based on the moving average are 20 numbers as shown in Table 3.

4. Experimental Results

4.1 Simulation Conditions

In this paper, the results of the neural network model, DPNN model, and fuzzy model optimized by simulated annealing are compared together. Both neural networks and DPNNs are the model of the network type and the fuzzy model is a knowledge-based type. The two inputs of the networks are daily KOSPI 200 and the industrial index announced by Korea Stock Exchange. The all inputs consist of today's index and yesterday's index.

The training data of the neural networks are KOSPI 200that is chosen by the data time from Jan. 4 1999 to Dec. 26 2000and the testing data are selected by the data time from Jan. 2 2001 to Aug. 28 2001.

4.2 Prediction Results

Fig. 6, 7, and 8 show the training results of three models using the data from Jan. 4 1999 to Dec. 26 2000. The training results show that three models are trained well by using the training data. It means that the procedure of the training is working properly.

Fig. 9, 10, and 11 are to present the testing results of three models that are taken the data from Jan. 2 2001 to Aug. 8 2001. In the experimental results, the performance of the DPNN model is the best for the prediction of the stock price index. The fuzzy model shows the better performance than the performance of the neural network model.

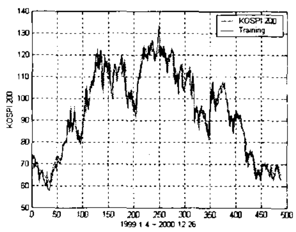

Fig. 6. Training result of the NN model.

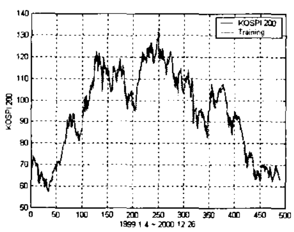

Fig. 7. Training result of the DPNN model.

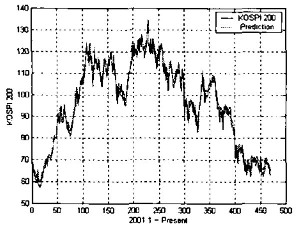

Fig. 8. Training result of the fuzzy model.

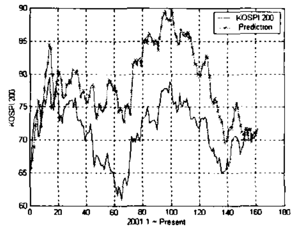

Fig. 9. Testing result of the NN model.

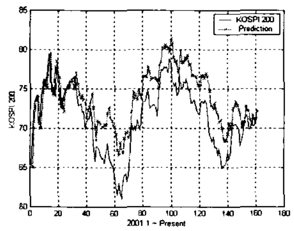

Fig. 10. Testing result of the DPNN model.

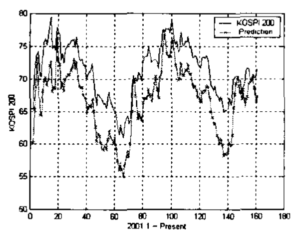

Fig. 11. Testing result of the fuzzy model.

As shown in Fig. 12, 13, and 14, the scatter plot represents the prediction performance comparing the real values and prediction value. The points are grouping around the diagonal line as it has prediction performance. So from the graphs the DPNN model shows the best performance and the fuzzy model has a good prediction performance secondly. The neural network model shows the worst performance.

Table 5 presents RMSE (root-mean squared error) that is one of performance criteria. The performance is better with lower values. In the table, the DPNN model has the lowest RMSE value so this model is the best for the performance. This result is the same with the result of the scatter plot. Next good model is the fuzzy model that has the 4.9131 RMSE value. But when we consider the correlation value fuzzy model is the best model because this model can follow the direction of the real value trend well.

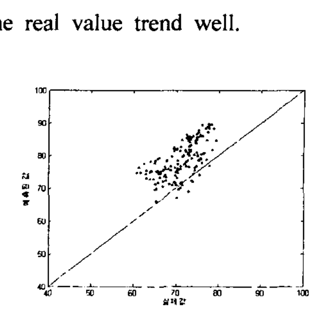

Fig. 12. Scatter plot of the NN model.

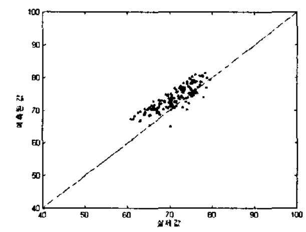

Fig. 13. Scatter plot of the DPNN model.

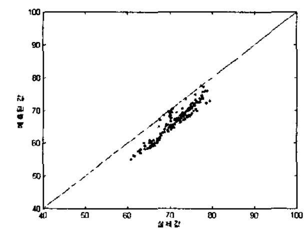

Fig. 14. Scatter plot of the fuzzy model.

Table 5. The performance index using RMSE.

Error \ Model	NN model	DPNN model	Fuzzy model
RMSE	8.3852	3.3703	4.9131
R2	0.6717	0.7213	0.8912

5. Conclusion

The main goal of this research is to design the intelligent models using the history data and knowledge information. For the knowledge-based model, the fuzzy logic is employed and the optimization method is applied for the optimization of the membership functions. In this paper, the prediction target is KOSPI 200 that is the typical non-linear data. The neural network model uses the just history data for the model inputs but the fuzzy model takes the knowledge-based rules that are extracted from data, user's knowledge, and so on. If the knowledge information is more correct to the real world index, the result of the prediction should be getting better. In this paper, the moving average is used for generating rules but it could be change with respect to the knowledge what the user wants to apply for the prediction models.

The scatter plot and RMSE are applied for the performance comparison. Both methods show the same result of the prediction models. Through the results of three models, the fuzzy model is not so bed to predict the stock price index. The DPNN model is the best but it has no capability to infer the various environmental conditions. Therefore the fuzzy model can be applied in the stock prediction because of the merits of the model.

Reference

[1] John J Murphy, *Technical Analysis of the Financial Markets*. Prentice Hall Direct, 1999.

[2] B. Wuthrich, V. Cho, S. Leung, K. Sankaran, and J. Zhang, "Daily stock market forecast from textual web data," In 1998 *IEEE International Conference on Systems, Man, and Cybernetics*, pages 2720--2725, 1998.

[3] C. C. Lee, "Fuzzy Logic in Control Systems: Fuzzy Logic Controller, Part I & II," *IEEE Transactions on Systems, Man and Cybernetics*, 20(2):404--435, 1990.

[4] L. A. Zadeh, "Outline of a New Approach to the Analysis of Complex System and Decision Process," *IEEE Transactions on Systems, Man and Cybernetics*, 3: 28--44, 1973.

[5] Duc Trung Pham and Liu Xing, *Neural Network for Identification, Prediction and Control*. Springer-Verlag Inc, 1995.

[6] A. G. Ivakhnenko, "The Group Method of Data Handling in Prediction Problem," *Soviet Automatic Control*, 9(6):21--30, 1976.

[7] J.-S. R. Jang, C-T. Sun, and E. Mizutani, *Neuro-Fuzzy AND Soft Computing*. Prentice Hall International Edition, 1996.

[8] Y. Hiemstra, "A stock market forecasting support system based on fuzzy logic," In *the Proceedings of the 27th Annual Hawaii International Conference on System Science*, pages 281--287, 1994.

[9] G. Liu and W. Yang, "Learning and Tuning of Fuzzy Membership Functions by Simulated Annealing Algorithm," In *The 2000 IEEE Asia-Pacific Conference on Circuits and Systems*, pages 367--370. 2000.

[10] Sungshin Kim and George J. Vachtsevanos, "An intelligent approach to integration and control of textile processes," *Information Sciences*, 123: 181--199, 2000.

[11] Franklin A. Graybill, *Theory and Application of the Linear Model*. Duxbury Press, CA, 1976.

[12] S. Kim and G. Vachtsevanos, "Polynomial fuzzy neural network for identification and control," *NAFIPS'96, North American Fuzzy Information Proc. Society, Berkeley*, CA, pages 5-9, June 19-22, 1996.

Sungshin Kim

He received the B.S. and M.S. degrees from Yonsei University, and the Ph.D. degree in electrical engineering from Georgia Institute of Technology, Atlanta, in 1984, 1986, and 1996, respectively.

He is currently an Assistant Professor in the School of Electrical & Computer Engineering, Pusan National University. His research interests include intelligent control, fuzzy logic control, manufacturing systems, and data mining.

Phone : +82-51-510-2374

Fax : +82-51-513-0212

E-mail : sskim@pusan.ac.kr

Hyeon Bae

He received the M.S. degree in electrical engineering from Pusan National University in 2001. He is currently pursuing the Ph.D. degree at Pusan National University.

Haegyun Kim

He received the M.S. degree in electrical engineering from Pusan National University in 2002. He is working for LG-Electronics, Korea.

Kwang Bang Woo

He received the B.S. and M.S. degrees from Yonsei University, and the Ph.D. degree in electrical engineering from Oregon State University, in 1957, 1959, and 1964, respectively.

He is currently an Research Professor in Automation Technology Research Institute, Yonsei University. His research interests include intelligent control, system and factory automation.