Extracting Trading Rules from the Multiple Classifiers and Technical Indicators in Stock Market

Kyoung-jae Kim and Ingoo Han

Graduate School of Management, Korea Advanced Institute of Science and Technology

John S. Chandler

Department of Accountancy, University of Illinois at Urbana-Champaign

Abstract

This study intends to mine reasonable trading rules by classifying the up/down fluctuant direction of the price for Korea Stock Price Index 200 (KOSPI 200) futures. This research consists of two stages. The first stage is classifying the fluctuant direction of the price for KOSPI 200 futures with several technical indicators using artificial intelligence techniques. And the second stage is mining the trading rules to resolute conflict among the outputs of the first stage using the inductive learning. To verify the effectiveness of proposed approach, this study composes four comparable models and performs statistical test. Experimental results show that the classification performance of the proposed model outperforms that of other comparable models. In addition, the proposed model yields higher profit than other comparable models and buy-and-hold strategy.

Key words: Rule extraction, Conflict resolution, Inductive learning, Artificial neural network, Case-based reasoning.

I. Introduction

It is known that stock market shows a nonlinear pattern. By this reason, there may be some limitations to analyze the stock market with linear model. Therefore, artificial intelligence (AI) techniques are often used for the nonlinear pattern analysis such as stock market analysis. Several studies show that artificial neural network and case-based reasoning can be applied to stock market predictions, but most of them have focused mainly on prediction of spot market and time series data (Ahmadi, 1990, Kamijo and Tanikawa, 1990, Kimoto et al., 1990, Yoon and Swales, 1991). They also did not bare outstanding prediction accuracy because tremendous noise and non-stationarity of data.

Korea launched trading in index futures market (KOSPI 200) on May 1996, then more people became attracted to this market, because the investor can avoid or reduce the risk of stock investment via index futures. The lack of the research on the index futures market in Korea has not met the people's interest. By this reason, this research intends to classify the daily up/down fluctuant direction of the futures price for Korea stock index (KOSPI 200) to meet this recent surge of interest.

The classification methodologies employed in this research are the artificial neural network (ANN) and case-based reasoning (CBR). Through the experiment, we can get two outputs from each classifier.

Many cases of the two models indicate a consistent signal, but some cases produce an inconsistent signal between two models. Prior studies presented several kinds of ensemble methods between the conflict of multiple classifiers (Maclin and Shavlik, 1995, Hansen and Salamon, 1990, Lincoln and Skrzypek, 1990, Perrone and Cooper, 1994, Rost and Sander, 1993, Zhang et al., 1992, Wolpert, 1992). But most of them, the basic idea is to combine multiple neural network models, not the heterogeneous classifiers. In addition, most of them focused on prediction or classification accuracy by combining results of each classifier. Moreover, they did not explain how these outputs were produced. In this study, we try to resolute conflict between two outputs by trading rules, not to combine it. Trading rules can resolute these conflicts. ANN and CBR, however, is not an outstanding tool for rule generation. Therefore, we employ inductive learning to generate the trading rules. In this way, we can take following benefits. We can resolute conflict between the outputs of multiple models. In addition, investors in the stock market can get reasonable trading rules. Fund managers of trading company or consultants of consulting firm also can modify this rule by adding their preferred factors in the analysis of investment.

To verify the effectiveness of proposed approach, this study composes four comparable models and performs statistical test. As to statistical test, we perform the McNemar test to examine whether the classification performance of proposed approach is significantly higher than that of competitive approaches. In addition, we perform a simulation of buying and selling of stocks to verify the profitability of proposed approach.

The rest of the paper is organized into four sections. The next section reviews related prior researches. In the third section, we propose the research model and execute experiments. In the fourth section, the results are verified and discussed. In the following section, the conclusions and future research issues are presented

II. Prior Research

2.1 AI Applications in Stock Market

Kimoto et al. (1990) and Kamijo and Tanikawa (1990) used several learning algorithm and prediction method for the Tokyo stock exchange prices index (TOPIX) prediction system. Ahmadi (1990) was tried to test the 'Arbitrage pricing theory (APT)' by ANN. Yoon and Swales (1991) performed prediction using mixed qualitative and quantitative data. The architecture of neural network model was a four-layered network. Lee et al. (1989) developed the intelligent stock portfolio management system (ISPMS). They are attempted to take advantage of optimization models and expert systems, integrated two models. Trippi and DeSieno (1992) executed daily prediction of up and down direction of S&P 500 Index Futures using ANN. They performed composite rule generation procedure to generate rules for combining outputs of networks. Duke and Long (1993) also executed daily prediction of German Government Bond Futures using feedforward backpropagation neural network. Choi et al. (1995) also performed daily prediction of up/down direction of S&P 500 Index Futures.

In summary, above three studies showed the availability of the artificial intelligence to predict future price for equity index. However, they did not obtain outstanding prediction accuracy. In addition, they did not take variables specific to futures market like basis or open interest (OI) into consideration in selecting input variables.

2.2 Ensemble Methods among Multiple Classifiers

Prior studies presented several kinds of ensemble methods between classifiers. But most of them, the basic idea is combine multiple neural network models, not the different classifiers. The basic idea in combining neural networks is to train a number of networks, and then somehow use the collection to increase generalization (Maclin and Shavlik, 1995). Hansen and Slamon (1990) used voting schemes, Lincoln and Skrzypek (1990) and Perrone and Cooper (1994) used simple average scheme and weighted average scheme respectively. Others used scheme for training combiner (Rost and Sander, 1993, Zhang et al., 1992, Wolpert, 1992).

III. Research Model and Experiments

3.1 Domain Knowledge based Categorical Preprocessing

The Problem of analyzing data using statistical method or AI techniques is separable to trend prediction and pattern classification problem. Trend prediction problem usually treated continuous single or multiple time series data as input variable. It mainly aims to capture temporal patterns between the data on the time lag. The examples of trend prediction are stock price prediction, interest rate prediction or economic forecasting by regression or time series analysis (ARIMA etc.). However, pattern classification problem such as bond rating or credit evaluation is usually uses multiple discrete or continuous data as input variable. It chiefly aims to grasp the correlation between the data on the same point of time.

When analyzing the time series data using ANN, considering temporal patterns between the data on the time lag is very important. A temporal pattern, however, can be difficult to train because the multi-layer perceptron has the risk of learning the unnecessary random correlation and noise, because it has an outstanding ability of fitting. Weigned et al. (1991) used weight-elimination, and Jhee and Lee (1993) used recurrent neural network to prevent the overfitting problem. Moreover, time series prediction requires a long computational time because it uses a large number of complex relationships.

Because of above reasons, this study incorporates the categorical approach based on domain knowledge for data preprocessing. Traditional data preprocessing method generally includes linear scaling to [0,1] or [-1,1]. While the categorical preprocessing in this study mean categorizing continuous input variables to some discrete categories. The categorical classification criteria are expert's knowledge. For example, market technicians usually regard below 25 of stochastic %K level as the signal of a bear market, and above 75 as the signal of a bull market and between 25 and 75 as the signal of a neutral market.

The categorical approach can convert continues time series data into discrete and symbolic one for experiments. By this procedure, we can perform pattern classification in stock market analysis rather than time series prediction. With pattern classification problem, we can overcome the limitations of prediction problem. The categorical preprocessing brings several advantages: This approach effectively filters the data and training the classifier, and can extract the rules from the classifier easily. The superiority of this approach to traditional preprocessing was justified by prior studies (Kim and Han, 1997, Kim and Han, 1998). Table 1 shows the examples of categorical classification criteria.

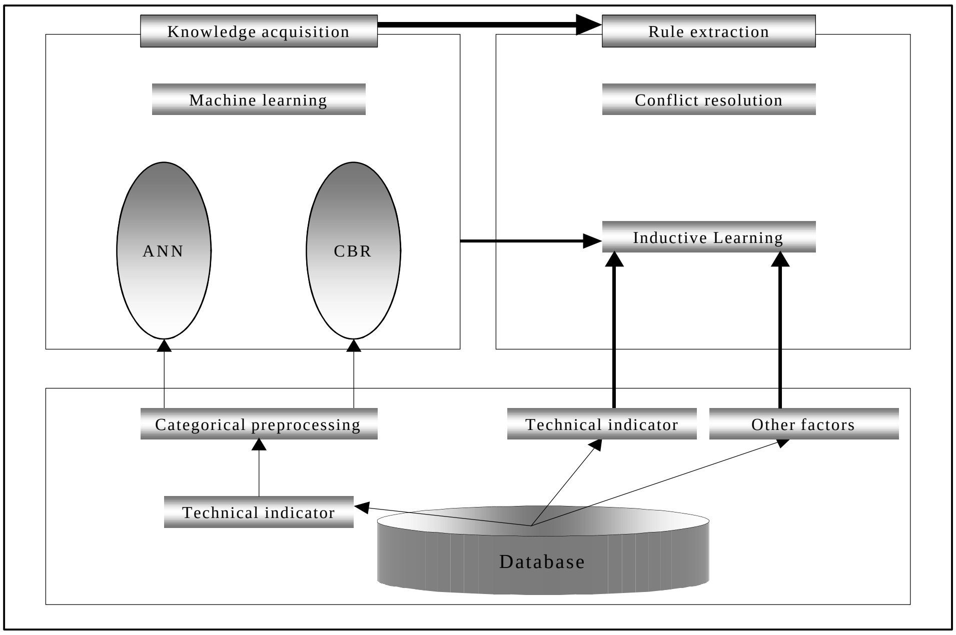
Indicator \ Category	Category 1	Category 2	Category 3
Stochastic %K	below 25	25-75	above 75
Momentum		(-)	0 or (+)
CCI (commodity channel index)		(-)	0 or (+)
OSCP (price oscillator)		(-)	0 or (+)
PVI (positive volume index)		below MA5 of PVI	above MA5 of PVI
Stochastic Slow %D	below 25	25-75	above 75
RSI (relative strength index)	below 30	30-70	above 70
ROC (rate of change)		below 100	0 or above 100
A/D Oscillator		below 0.5	0 or above 0.5

<Table 1> The examples of categorical classification criteria

3.2 Research Framework

This study is composed of two phases. The Objective of the first phase is knowledge acquisition using machine learning. ANN and CBR are two components of machine learning tool. Besides, categorical data preprocessing is presented for new preprocessing method. The second phase is to resolve conflict between the outputs of ANN and CBR when both give inconsistent signals. By this procedure, this study can generate reasonable trading rules from the multiple classifiers and technical indicators in stock market. In this phase, the outputs of ANN and CBR are used as one of the technical indicators like CCI, RSI etc.

The research framework of this study is shown in Figure 1.



<Figure 1> The research framework

3.3 Research Data and Experiments

3.3.1 Research data

The research data used in this research is futures price for KOSPI 200 from May 1996 through November 1996. Futures are the standard forms that decide the quantity and price in the certified market (trading place) at certain future point of time (delivery date). General functions of futures market are supplying information about future price of commodities, function of speculation and hedging (Kolb and Hamada, 1988). Being different from the spot market, futures market does not have continuity of price data. That is because futures market has price data by contract. So, in futures market analysis, nearest contract data method is mainly used and incorporated in this research.

Many previous stock market analyses have used technical or fundamental indicator. In general, fundamental indicators are mostly used for long-term trend analysis while technical indicators are for short-term pattern analysis. In this research, we use the technical indicators as input variables.

Initial available variables are technical indicators such as PVI (positive volume index), Stochastic %K, Stochastic %D, Stochastic Slow %D, Basis, Open interest, Momentum, ROC (rate of change), LW%R (Larry William's %R), A/D Oscillator (accumulation / distribution oscillator), ADL (Accumulation / distribution Line), Disparity 5days, CCI (commodity channel index), OSCP (price oscillator) and RSI (relative strength index). The description and formula of technical indicators are presented in Appendix.

Previous researches usually used the statistical method such as correlation test, factor analysis, stepwise regression analysis etc. This study uses stepwise regression analysis and genetic algorithms (GA) for input variable selection. In the first place, this study screens the candidate for input variables by stepwise regression analysis and selects appropriate variables. Then GA is executed for selected variables from regression analysis. GA is performed by NeuralWorks Predict (NeuralWare, Inc., 1995). The crossover probabilities and probability of mutation assumed to be 0.7 and 0.05 respectively. In addition, we use two kinds of AI techniques; one is ANN and the other is CBR.

Consequently, final input variables for each AI techniques are summarized at Table 2. GA selects these variable sets, because each set reveals the best evaluation values respectively.

ANN	Stochastic %K, Momentum, CCI, OSCP, PVI
CBR	Stochastic Slow %D, RSI, ROC, A/D Oscillator, PVI

<Table 2> Input variables for ANN and CBR

3.3.2 Experiments

3.3.2.1 Phase I: Knowledge Acquisition

In Phase I, experiments are implemented by NeuralWorks Predict (NeuralWare, Inc., 1995) for ANN modeling and KATE 5.02 (AcknoSoft, 1996) for CBR modeling. The backpropagation algorithm and sigmoid function are used in ANN. Learning rate and momentum is 0.1 respectively and initial weight is 0.3. The 10% of data for testing, 20% for holdout, and 70% for training mutually and exclusively used in order to avoid overfitting. And 50,000 learning events since minimum average error of test set are permitted. In experiments for CBR, this study uses nearest-neighbor method as case retrieval algorithm and employs the Euclidean distance as a measure of similarity.

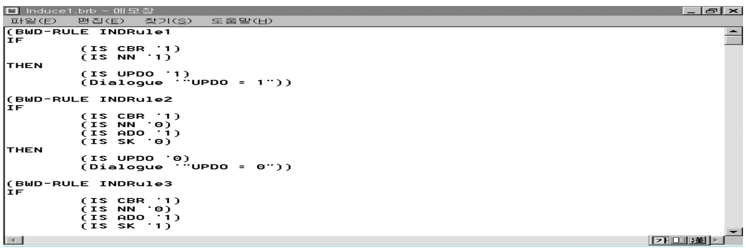
We use cross-validation method to solve insufficiency problem in the number of data and to generalize the experimental results. The cross-validation error rate estimator is an almost unbiased estimator of the true error rate of a classifier (Weiss and Kulikowski, 1991). In this study, first we classifying the mutually exclusive five sets which is composed of 30 of all 150 data respectively. Then it uses four sets for training and testing and the other one set for holdout. We repeated this procedure for five times. Finally, we compose data set of 600 data for training and testing and 150 data for validating.

3.3.2.2 Phase II: Conflict Resolution through Rule Extraction

As the results of Phase I, we get the two experimental outputs by ANN and CBR for each case. If above two outcomes give the same signals, then investors follow that. If above two, however, do not give same signals, investors may be come into a conflict. Moreover, some investors may wish the model consider the individual preferred factors of decision making about investment, for example rainfall, specific technical indicator, climate etc. Because the model selects its input variables through the structured way, such as statistical test or heuristic search, these factors cannot be reflected in the model entirely. In this study, we use the inductive learning to consider the user-preferred factors and to resolute conflict among the outputs of model and above factors. In this way, we can take following benefits. We can resolute conflict among the output of each model and user-preferred factors. In addition, fund managers in the Securities company or consultants at consulting firm can modify this rule by adding their preferred factors in the analysis for investment. Moreover, investors in the stock market can get reasonable trading rules in investment. The conflict resolution rule is as follows.

> "**If** the output of CBR and the output of ANN give the same signal
>
> **Then** follow this signal
>
> **Else** adapt the solution by the rules from the inductive learning"

As to the induction method, ID3 is used. ID3 is a simple decision tree learning algorithm developed by Quinlan (1986) and the most frequently used method among the inductive algorithms. We use KATE 5.02 (AcknoSoft, 1996) to implement the ID3 model. After the experiments, several rules generated from each validation set. We validate these rules and adapt the solutions using conflict resolution algorithm. By this procedure, We can get reasonable trading rules. Figure 2 is the example of generated trading rules by ID3.



<Figure 2> The example of the generated rules by ID3

IV. Results

To verify the effectiveness of proposed approach, this study composes four comparable models and performs statistical test. Model_A uses nine technical indicators as input variables for inductive learning. These indicators are selected as input variables for ANN and CBR modeling through GA. Model_B includes three technical indicators, which are not selected ANN and CBR modeling. All indicators used for Model_A and Model_B are included in Model_C. Among the indicators, nine indicators are used as input variables of ANN and CBR modeling. Model_D employs output values of ANN and CBR, and three technical indicators, which are not used for ANN and CBR modeling. Input variables for each model are summarized in Table 3.

	Input variables
Model_A	Stochastic %K, Stochastic Slow %D, Momentum, ROC, RSI, PVI, AD Oscillator, CCI, OSCP
Model_B	Disparity5, Stochastic %D, Larry William's %R
Model_C	Stochastic %K, Stochastic Slow %D, Momentum, ROC, RSI, PVI, AD Oscillator, CCI, OSCP, Disparity5, Stochastic %D, Larry William's %R
Model_D	Output values of ANN and CBR, Disparity5, Stochastic %D, Larry William's %R

<Table 3> Input variables for each model

Table 4 describes the average classification accuracy for each competitive model. The classification accuracy of Model_D is 77.33% and Model_A and Model_C is 69.33% and 68.00%, respectively. In addition, Model_B is 58.00%. On the average, Model_D outperforms the other models by about 9~19% of classification accuracy.

	Set 1	Set 2	Set 3	Set 4	Set 5	Average
Model_A	80.00	76.67	56.67	70.00	63.33	69.33
Model_B	63.33	53.33	60.00	66.67	46.67	58.00
Model_C	80.00	70.00	56.67	70.00	63.33	68.00
Model_D	76.67	83.33	76.67	73.33	76.67	77.33

<Table 4> Average classification accuracy (hit ratio, %)

As to statistical test, we employ the McNemar tests to examine whether the classification performance of proposed approach is significantly higher than that of competitive approaches. The results show that Model_D performs significantly better than Model_B and Model_C at the 1% significance level and Model_A at the 5% significance level. Therefore, we can conclude that our proposed approach outperforms other competitive approaches with statistical significance. Table 5 is the results of statistical test.

	Model_A	Model_B	Model_C	Model_D
Model_A		6.568 (0.010)***	-[a] (0.250)	-[a] (0.027)**
Model_B			4.780 (0.029)**	20.103 (0.000)***
Model_C				-[a] (0.004)***

<Table 5> Chi-square value (P value) from McNemar test

a. Binomial distribution used.

* significant at the 10% level

** significant at the 5% level

***significant at the 1% level

We also perform a simulation of buying and selling of stocks to verify the profitability of proposed approach. Buying and selling is simulated under following assumptions. If generated rule indicates bullish market, all available money is used to buy stocks at a time. And we hold it until the rule produces the signal of bearish market. If generated rule indicates bearish market, all stocks are sold at a time. Total amounts of the trading simulation and return on investment (ROI) are summarized in Table 6 and Figure 3.

Trading Strategy	Total amount (ROI) (assume initial investment of 1,000,000 won)
Following Buy-and-hold strategy	760,510 won (-23.9%)
Following the rules of Model_A	1,278,039 won (27.8%)
Following the rules of Model_B	986,678 won (-1.3%)
Following the rules of Model_C	1,246,296 won (24.6%)
Following the rules of Model_D	1,409,400 won (40.9%)

<Table 6> Total amount and ROI of buying and selling simulations

<Figure 3>Performance of simulations

While the underlying index decreased about 24% during the period, we can yield the high level of profit using the proposed approach. In addition, our proposed approach yields higher profit than other competitive approaches, such as Model_A, Model_B, Model_C, and buy-and-hold strategy. These results demonstrate that proposed approach is promising methods for extracting profitable trading rules.

V. Conclusion and Future Research Issues

This study suggests that the outputs of ANN and CBR can be proper substitutes for technical indicators. This study also proposes inductive learning can be a tool of resolving conflict between multiple classifiers. Inductive learning can effectively resolve the conflict between the inconsistent output values of ANN and CBR model and can provide reasonable and profitable trading rules to investors in stock market.

Following issues are further research needed.

Additional classifiers such as genetic algorithm and fuzzy neural network can be the candidates for base classifier. In this study, only two classifiers are used as base classifier, but additional classifier can provide synergistic effect to the integrated system.

Appendix: Technical indicator

Name	Description	Formula
PVI	The Positive Volume Index (PVI) focuses on days when the volume increased from the previous day.	$PVI_{t-1} + (\frac{C_t - C_{t-1}}{C_{t-1}} \times PVI_{t-1})$
Stochastic %K	The Stochastic Oscillator compares where a security's price closed relative to its price range over a given time period.	$\frac{C_t - L_n}{H_n - L_n} \times 100$
Stochastic %D	The Stochastic Oscillator compares where a security's price closed relative to its price range over a given time period. It is a moving average of %K.	$\frac{\sum_{i=0}^{n-1} \%K_{t-i}}{n}$
Stochastic Slow %D	The Stochastic Oscillator compares where a security's price closed relative to its price range over a given time period. It is a moving average of %D.	$\frac{\sum_{i=0}^{n-1} \%D_{t-i}}{n}$
Basis	The basis is the current spot price of a particular commodity minus the price of a particular futures contract for the same commodity. The basis can be used to predict future spot prices of the commodities that underlie the futures contract.	Current spot price – Futures price
Open Interest	Open Interest is the number of open contracts of a given futures contract. An open contract can be a long or short contract that has not been exercised, closed out, or allowed to expire.	*N/A*
Momentum	The Momentum indicator measures the amount that a security's price has changed over a given time span.	$C_t - C_{t-4}$
ROC	The Price Rate-of-Change (ROC) indicator displays the difference between the current price and the price *x* periods ago.	$\frac{C_t}{C_{t-n}} \times 100$
Larry William's %R	Larry William's %R is a momentum indicator that measures overbought/oversold levels.	$\frac{H_n - C_t}{H_n - L_n} \times 100$
A/D Oscillator	The A/D Oscillator measures the accumulation and distribution of market power. It means relative strength of bullish and bearish market.	$\frac{H_t - C_{t-1}}{H_t - L_t}$
Disparity 5days	The Disparity means the distance of current spot price and moving average.	$\frac{C_t}{MA_n} \times 100$

CCI	The Commodity Channel Index (CCI) measures the variation of a security's price from its statistical mean. High value of CCI indicates prices are unusually high compared to average prices and low value of CCI indicates that prices are unusually low.	$\frac{-}{\times}$
		$\frac{MA_5 - MA_{10}}{MA_5} \times 100$
RSI	The RSI is price following oscillator that ranges from 0 to 100. The RSI usually tops above 70 and bottoms below 30.	$100 - \frac{100}{1 + \frac{\sum_{i=0}^{n-1} Up_{t-i} / n}{\sum_{i=0}^{n-1} Dw_{t-i} / n}}$

<Table A> Technical indicators (Kolb and Hamada, 1988, Achelis, 1995)

Note) C: Closing price, L: Low price, H: High price, Volume: Trading volumes

MA: Moving average of price, M_t : $\frac{(H_t + L_t + C_t)}{3}$, SM_t : $\frac{\sum_{i=1}^{n} M_{t-i+1}}{n}$

D_t : $\frac{\sum_{i=1}^{n} |M_{t-i+1} - SM_t|}{n}$, Up: Upward price change, Dw: Downward price change

Reference

Achelis, S. B., *Technical Analysis from A to Z*, Probus Publishing, 1995.

AcknoSoft, *CaseCraft™: The KATE™ Toolbox for reasoning from Cases*, 1996.

Ahmadi, H., 'Testability of the arbitrage pricing theory by neural networks.', *Proceedings of the International Conference on Neural Networks*, 1990, pp. 385-393.

Choi, J. H., Lee, M. K., and Rhee, M. W., 'Trading S&P 500 stock index futures using a neural network.', *The Third Annual International Conference on Artificial Intelligence Applications* *on Wall Street*, 1995, pp. 63-72.

Duke, L. S., and Long, J. A., 'Neural network futures trading - A feasibility study.', *Adaptive Intelligent Systems*, Elsevier Science Publishers, 1993, pp. 121-132.

Hansen, L., and Salamon, P., 'Neural network ensembles.', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, 1990, pp. 993-1001.

Jhee, W. C., and Lee, J. K., 'Performance of neural networks in managerial forecasting.', *Intelligent Systems in Accounting, Finance and Management*, Vol. 2, 1993, pp. 55 ~71.

Kamijo, K., and Tanigawa, T., 'Stock price pattern recognition: A recurrent neural network approach.', *Proceedings of the International Joint Conference on Neural Networks*, 1990, pp. 215-221.

Kim, K. J., and Han, I. G., 'Prediction of the price for stock index futures using integrated artificial intelligence techniques with categorical preprocessing.', *Proceedings of the Korean Operations Research and Management Science Society Conference*, November, 1997, pp. 105-108.

Kim, K. J., and Han, I. G., 'AI application to futures price for Korea stock index with categorical preprocessing.', *Working Paper*, KAIST, 1998.

Kimoto, T., Asakawa, K., Yoda, M., and Takeoka, M., 'Stock market prediction system with modular neural network.', *Proceedings of the International Joint Conference on Neural Networks*, 1990, pp. 1-6.

Kolb, R.W., and Hamada, R. S., *Understanding Futures Markets*, Scott, Foresman and Company, 1988.

Lee, J. K., Kim, H. S., and Chu, S. C., 'Intelligent stock portfolio management system.', *Expert Systems*, Vol. 6, No. 2, 1989, pp. 74-87.

Lincoln, W., and Skrzypek, J., 'Synergy of clustering multiple back propagation networks.', *Advances in Neural Information Processing Systems*, Vol. 2, Morgan Kaufmann, 1990, pp. 650-659.

Maclin, R., and Shavlik, J. W., 'Combining the predictions of multiple classifiers: Using competitive learning to initialize neural networks.', *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1995.

NeuralWare, Inc., *NeuralWorks Predict: Complete Solution for Neural Data Modeling*, 1995.

Perrone, M., and Cooper, L., 'When networks disagree: Ensemble method for neural networks.', *Artificial Neural Networks for Speech and Vision*, Chapman and Hall, 1994.

Quinlan, J. R., 'Induction of decision trees', *Machine Learning*, Vol. 1, 1986, pp. 81-106.

Rost, B., and Sander, C., 'Prediction of protein secondary structure at better than 70% accuracy.', *Journal of Molecular Biology*, 232, 1993, pp. 584-599.

Trippi, R. R., and DeSieno, D., 'Trading equity index futures with a neural network.', *The Journal of Portfolio Management*, 1992, pp. 27-33.

Weigned, A. S., Rumelhart, D. E., and Huberman, B. A., 'Generalization by Weight Elimination applied to Currency Exchange Rate Prediction.', *Proceedings of the International Joint Conference on Neural Networks*, 1991, pp. 837 - 841.

Weiss, S., and Kulikowski, C., *Computer Systems That Learn*, Morgan Kaufmann Publishers, Inc., 1991.

Wolpert, D., 'Stacked generalization.', *Neural Networks*, Vol. 5, 1992, pp. 241-259.

Yoon, Y., and Swales, G., 'Predicting stock price performance: A neural network approach.', *Proceedings of the 24th Annual Hawaii International Conference on Systems Sciences*, 1991, pp. 156-162.

Zhang, X., Mesirov, J., and Waltz, D., 'Hybrid system for protein secondary structure prediction.', *Journal of Molecular Biology*, 225, 1992, pp. 1049-1063.