INTELLIGENT SYSTEMS IN ACCOUNTING, FINANCE AND MANAGEMENT
Intell. Sys. Acc. Fin. Mgmt. **12**, 167–176 (2004)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/isaf.252

ARTIFICIAL NEURAL NETWORKS WITH FEATURE TRANSFORMATION BASED ON DOMAIN KNOWLEDGE FOR THE PREDICTION OF STOCK INDEX FUTURES

KYOUNG-JAE KIM*
Department of Information Systems, Dongguk University, Seoul, Korea

SUMMARY

A feature transformation method based on domain knowledge for artificial neural networks (ANNs) is proposed. The method of feature transformation based on domain knowledge converts continuous values into discrete values in accordance with the knowledge of experts in specific application domains. This approach effectively filters data, trains the classifier, and extracts the rules from the classifier. In addition, it reduces the dimensionality of the feature space, which not only decreases the cost and time in the operation but also enhances the generalizability of the classifier. The experimental results of the proposed approach will be compared and tested statistically with the results of the linear transformation method. The results show that the method of feature transformation based on domain knowledge outperforms the linear transformation in modelling of ANNs. Copyright © 2004 John Wiley & Sons, Ltd.

1. INTRODUCTION

It has been widely accepted that most financial variables are nonlinear. Recently, artificial neural networks (ANNs) have been popularly applied to the problems of finance, such as bankruptcy prediction, corporate bond rating, etc. The reason is that they can model nonlinear relationships among variables.

Several studies on stock market prediction using artificial intelligence (AI) techniques have been carried out during the past decades. Stock market prediction was a typical problem of financial time-series prediction. Prior studies used various types of ANN model to predict accurately the stock index and the direction of its change. In one of the earliest studies, Kimoto *et al.* (1990) used several learning algorithms and prediction methods for the Tokyo stock exchange prices index (TOPIX) prediction system. Their system used modular neural networks to learn the relationships among various factors. Kamijo and Tanigawa (1990) used recurrent neural networks and Ahmadi (1990) used backpropagation neural networks with the generalized delta rule to predict the stock market. Yoon and Swales (1991) also performed predictions using qualitative and quantitative data. Some researchers investigated the issue of predicting the stock index futures market. Trippi and DeSieno (1992) and Choi *et al.* (1995) predicted the daily direction of change in the S&P 500 index futures using ANNs. Duke and Long (1993) executed the daily predictions of the German government bond futures using feedforward backpropagation neural networks.

* Correspondence to: Kyoung-jae Kim, Department of Information Systems, Dongguk University, 3-26, Pil-dong, Chung-gu, Seoul 100-715, Korea. E-mail: kjkim@dongguk.edu
Contract/grant sponsor: Dongguk University Research Fund.

Recent research tends to include novel factors and to hybridize several AI techniques. Hiemstra (1995) proposed fuzzy expert systems to predict stock market returns. He suggested that ANNs and fuzzy logic could capture the complexities of functional mapping because they do not require the specification of the function to approximate. A more recent study of Kohara *et al.* (1997) incorporated prior knowledge to improve the performance of stock market prediction. Tsaih *et al.* (1998) integrated the rule-based technique and ANNs to predict the direction of the S&P 500 stock index futures on a daily basis.

However, they did not bear outstanding prediction accuracy, partly because of the tremendous noise and nonstationary characteristics in stock market data. Training ANNs tends to be difficult with highly noisy data: the networks then fall into a naive solution, such as always predicting the most common output (Lawrence *et al.*, 1996).

This study proposes a feature transformation approach based on domain knowledge. Because data preprocessing is an essential step for knowledge discovery and eliminates some irrelevant and redundant features, many researchers in the society of data mining have a broad interest in feature transformation (Liu and Motoda, 1998). In many applications, the amount of data is so large that learning of patterns may not work well. Reducing and transforming irrelevant and redundant features can also reduce the learning time and yield more general results (Dash and Liu, 1997). Feature transformation, in this study, means transforming continuous values into discrete ones in accordance with the knowledge of experts in the application domain. This approach effectively filters data, trains the classifier, and extracts the rules from the classifier. In addition, it reduces the dimensionality of the feature space, which not only decreases the cost and time in the operation but also enhances the generalizability of the classifier. The experimental results of the proposed approach will be compared and tested statistically with the results of the linear transformation method.

The rest of the paper is organized as follows. Section 2 reviews feature transformation methodologies. In Section 3 we propose the method of feature transformation based on domain knowledge and describe the benefits of this approach. In Section 4 we describe the design of this research and execute experiments using ANNs. In Section 5 the results are summarized and discussed. In Section 6, conclusions are presented with the assessment of our approach.

2. FEATURE TRANSFORMATION METHODS IN DATA MINING

Feature transformation is the process of creating a new set of features (Liu and Motoda, 1998). Feature transformation differs from feature subset selection in that the latter does not generate new features and selects a subset of original features (Blum and Langley, 1997; Dash and Liu, 1997). Feature transformation methods are classified as endogenous (unsupervised) versus exogenous (supervised) and local versus global (Dougherty *et al.*, 1995; Scott *et al.*, 1997; Susmaga, 1997).

Endogenous (unsupervised) methods do not take into consideration the value of the decision attribute, whereas exogenous (supervised) methods do. Local methods discretize one attribute at once, whereas the global ones discretize all attributes simultaneously. The methods of endogenous feature transformation include discretization using a self-organizing map (Lawrence *et al.*, 1996), the percentile method (Scott *et al.*, 1997; Buhlmann, 1998), and the clustering method (Scott *et al.*, 1997; Kontkanen *et al.*, 1997). Basak *et al.* (1998) proposed the neuro-fuzzy approach using a feature evaluation index and Piramuthu *et al.* (1998) suggested a decision-tree-based approach as an endogenous method. Although these methods have the advantage of simplicity in the transformation process, they do not give consideration to the correlation among each independent and dependent feature. Prediction performance, however, is enhanced by the ability of discrimination from not only a single feature but also the association among features. For this limitation, endogenous methods do not provide an effective way of forming categories (Scott *et al.*, 1997).

On the other hand, the methods of exogenous feature transformation include maximizing the statistical significance of Cramer's *V* between other dichotomized variables (Scott *et al.*, 1997), the entropy minimization heuristic in inductive learning, and the *k*-nearest neighbor methods (Fayyad and Irani, 1993; Ting, 1997; Martens *et al.*, 1998). Exogenous methods also include functional links found by the genetic algorithm (GA) for C4.5 (Vafaie and De Jong, 1998). These methods transform an independent variable to maximize its association with the values of dependent and other independent variables.

3. FEATURE TRANSFORMATION BASED ON DOMAIN KNOWLEDGE

Data analysis using statistical methods or AI techniques includes trend (or index) prediction and pattern classification. Trend prediction usually treats continuous time-series data as input variables. It aims to capture temporal patterns between the time lag of historical data. The examples of trend prediction are the prediction of stock prices, interest rates, and economic indices. They were traditionally analyzed by linear regression or time-series analysis, such as the autoregressive integrated moving average process (ARIMA). Pattern classification, such as bond rating and credit evaluation, however, usually uses discrete or continuous cross-sectional data as input variables. It aims to grasp the causality in the data simultaneously.

As mentioned above, it is very important to consider the temporal patterns in the data based on time lag when analyzing the time-series data using ANNs. A temporal pattern, however, is difficult to train because the multi-layer perceptron has the risk of learning the unnecessary random correlation and noise because it has the outstanding ability of fitting. Weigned *et al.* (1991) used weight-elimination and Jhee and Lee (1993) used recurrent neural networks to prevent the overfitting problem. In addition, time-series prediction requires much computational time because it uses a large number of complex relationships.

In this study, we propose a domain-knowledge-based discretization of continuous time-series data as a method of feature transformation. This approach effectively filters data, trains the classifier, and extracts the rules from the classifier. In addition, it reduces the dimensionality of the feature space, which not only decreases the cost and time in the operation but also enhances the generalizability of the classifier. Feature transformation based on domain knowledge is classified as an endogenous, local, parameterized, and hard method. This method discretizes input variables into some discrete categories. The discretizing criterion is domain knowledge in the stock market. Fund managers and investors in the stock market generally accept and use the criteria in Table I as the signal of future market trends. Even if a feature represents a continuous measure, the experts usually interpret the values in qualitative terms as bullish/bearish or low/medium/high (Slowinski and Zopounidis, 1995). For 'stochastic %*K*', a value of 75 is basically accepted by stock market analysts as a strong signal; when the value exceeds 75, the market is regarded as an overbought situation or a bullish market. On the other hand, if it drops below 25 it is considered as an oversold situation or the signal of a bearish market. When the value of 'stochastic %*K*' is placed between 25 and 75, it is regarded as the signal of a neutral market (Edwards and Magee, 1997). Table I reviews interpretation thresholds

Table I. Interpretation threshold

	Murphy (1986)	Achelis (1995)	Choi (1995)	Chang *et al.* (1996)	Edwards and Magee (1997)
PVI		Indicator relative to moving average	Indicator relative to moving average		
Stochastic %*K*	30/70	20/80	20/80	25/75	20–25/75–80
Stochastic %*D*	30/70	20/80	20/80	25/75	20–25/75–80
Stochastic slow %*D*	30/70	20/80	20/80	25/75	20–25/75–80
Momentum	0			0	
ROC	100		100		
LW %*R*	20/80	20/80	20/80	10/90	
A/D oscillator				0.5 or 0.2/0.8	
Disparity 5 days			100		
CCI	0 or −100/+100	−100/+100	0 or −100/+100	0 or −100/+100	
Price oscillator	0	0			
RSI	30/70	30/70	30/70	30/70	20–30/70–80

Table II. Discretizing criteria

Indicator	Category A	Category B	Category C
PVI	(−∞, MA5[a] of PVI]		(MA5[a] of PVI, ∞)
Stochastic %*K*	[0, 25]	(25, 75]	(75, 100]
Stochastic %*D*	[0, 25]	(25, 75]	(75, 100]
Stochastic Slow %*D*	[0, 25]	(25, 75]	(75, 100]
Momentum	(−∞, 0]		(0, ∞)
ROC	(−∞, 100]		(100, ∞)
LW %*R*	[0, 20]	(20, 80]	(80, 100]
A/D oscillator	[0, 0.5]		(0.5, 1]
Disparity 5 days	(−∞, 100]		(100, ∞)
CCI	(−∞, 0]		(0, ∞)
Price oscillator	(−∞, 0]		(0, ∞)
RSI	[0, 30]	(30, 70]	(70, 100]

[a] Moving average for 5 days.

for some technical indicators and Table II shows discretizing criteria in this study. These criteria are produced based on Table I.

4. RESEARCH DESIGN

The research data used in this study are the futures price for the Korea stock index (KOSPI 200) from May to November 1996. The plot of the original time series is presented in Figure 1.

Futures are the standard forms that decide the quantity and price in the certified market at a certain future point in time. The general functions of the futures market are supplying information about the future price of commodities, speculation, and hedging (Kolb and Hamada, 1988).

Being different from the spot market, the futures market does not have continuity of price data. This is because the futures market has price data by contract. The nearest contract data method is used in this research because this method is popular in futures market analysis.

Initial data are technical indicators such as the positive volume index (PVI), stochastic %*K*, stochastic %*D*, stochastic slow %*D*, momentum, rate of change (ROC), Larry William's %*R* (LW

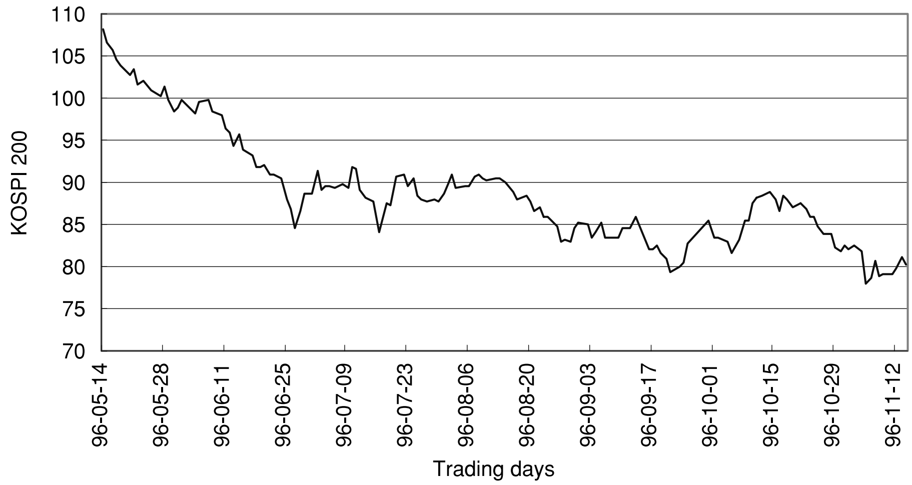

Figure 1. The plot of original time series

%*R*), accumulation/distribution (A/D) oscillator, disparity 5 days, commodity channel index (CCI), price oscillator and relative strength index (RSI). These indicators are generally used in the stock market. The formulas of the indicators are presented in Table III (Murphy, 1986; Achelis, 1995; Gifford, 1995; Chang *et al.*, 1996; Edwards and Magee, 1997).

In this study, a GA is used for the step of feature subset selection and optimizing the network structure of ANNs. This study compares the result of the feature transformation based on domain knowledge with that of linear transformation to test the significance of the difference. Linear transformation means the linear scaling of the data to the range 0.0 to 1.0. Linear transformation is usually used to enhance the performance of ANNs because most ANN models accept numeric data only in the range of 0.0 to 1.0 or −1.0 to +1.0 (Bigus, 1996). In feature subset selection, the GA selects different subsets, because each subset reveals the best evaluation values on differently transformed data. The final input variables are stochastic %*K*, PVI, momentum, CCI, and price oscillator for domain-knowledge-based transformed data and stochastic %*D*, LW %*R*, disparity 5 days, CCI, and Price oscillator for linear transformed data.

The difference in hit ratio between feature transformation based on domain knowledge and linear transformation is compared. The backpropagation algorithm and sigmoid function are used in the modeling of ANNs. The learning rate and the momentum are both 0.1 and the initial weight for the link between layers is a random value within the range from −0.3 to 0.3. Among the data, 10% of the data are used for testing, 20% for the hold-out, and 70% for training in order to avoid overfitting. In addition, only 50,000 learning events are permitted after the minimum average error of the test set is measured. The predicted value of the ANN is the direction of daily change of the index and it is categorized as '0' or '1'. A '0' means that the next day's index is lower than today's index, and a '1' means that the next day's index is lower than today's index.

We use the fivefold cross-validation method to settle the insufficiency problem in the amount of data and to generalize the experimental results. The cross-validation error-rate estimator is an almost completely unbiased estimator of the true error rate of a classifier (Weiss and Kulikowski, 1991). Finally, the data set is composed of 600 pieces of data for modeling and 150 pieces of data for validation. Table IV presents the summary statistics for each variable.

Table III. Technical indicators[a]

Name	Formula	Name	Formula
PVI	$\mathrm{PVI}_{t-1} + \left(\frac{C_t - C_{t-1}}{C_{t-1}} \times \mathrm{PVI}_{t-1}\right)$	Stochastic %K	$\frac{C_t - L_n}{H_n - L_n} \times 100$
Stochastic %D	$\left(\sum_{i=0}^{n-1} \%K_{t-i}\right) \Big/ n$	Stochastic slow %D	$\left(\sum_{i=0}^{n-1} \%D_{t-i}\right) \Big/ n$
Momentum	$C_t - C_{t-4}$	ROC	$\frac{C_t}{C_{t-n}} \times 100$
LW %R	$\frac{H_n - C_t}{H_n - L_n} \times 100$	A/D oscillator	$\frac{H_t - C_{t-1}}{H_t - L_t}$
Disparity 5 days	$\frac{C_t}{\mathrm{MA}_5} \times 100$	CCI	$\frac{M_t - \mathrm{SM}_t}{0.015 \times D_t}$
Price oscillator	$\frac{\mathrm{MA}_5 - \mathrm{MA}_{10}}{\mathrm{MA}_5} \times 100$	RSI	$100 - \cfrac{100}{1 + \cfrac{\left(\sum_{i=0}^{n-1} \mathrm{Up}_{t-i}\right) \Big/ n}{\left(\sum_{i=0}^{n-1} \mathrm{Dw}_{t-i}\right) \Big/ n}}$

[a] C: closing price; L: low price; H: high price; MA: moving average of price; $M_t = (H_t + L_t + C_t)/3$; $\mathrm{SM}_t = \left(\sum_{i=1}^{n} M_{t-i+1}\right) \Big/ n$; $D_t = \left(\sum_{i=1}^{n} |M_{t-i+1} - \mathrm{SM}_t|\right) \Big/ n$; Up: upward price change; Dw: downward price change.

5. EXPERIMENTAL RESULTS

Two models are compared according to the feature transformation method. As mentioned earlier, we have two ways to transform the features: one way is transformation by linear scaling to the range of 0.0 to 1.0 and the other is to discretize using domain knowledge. The first model linearly transforms data (model 'LT' in Tables V–VII) and the second model transforms the data based on domain knowledge (model 'FT' in Tables V–VII).

Tables V and VI describe the average hit ratio of each transformation method for the in-sample and the hold-out sample data. The average hit ratio is represented by the following equation:

$$\text{Hit ratio} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{CR}_i \quad (i = 1, 2, \ldots, n) \tag{1}$$

$$\begin{cases} \text{if } \mathrm{PO}_i = \mathrm{AO}_i & \mathrm{CR}_i = 1 \\ \text{otherwise} & \mathrm{CR}_i = 0 \end{cases}$$

where CR_i is the prediction result for the ith trading day, which is denoted by 0 or 1, PO_i is the predicted output from the model for the ith trading day, and AO_i is the actual output for the ith trading day.

Table IV. Summary statistics

Name	Statistic	Value	Name	Statistic	Value
PVI	Max	99.4539	Stochastic %K	Max	100.0000
	Min	80.2881		Min	0.0000
	Mean	92.0117		Mean	38.5567
Stochastic %D	Max	93.5324	Stochastic slow %D	Max	92.6448
	Min	0.0000		Min	3.4058
	Mean	38.572		Mean	38.4541
Momentum	Max	6.2500	ROC	Max	107.6220
	Min	−6.7500		Min	92.6431
	Mean	−0.9210		Mean	99.0566
LW %R	Max	100.0000	A/D oscillator	Max	1.0000
	Min	0.0000		Min	0.0000
	Mean	61.4433		Mean	0.43119
Disparity 5 days	Max	103.7817	CCI	Max	0.4557
	Min	96.0235		Min	−0.4780
	Mean	99.6169		Mean	−0.0556
Price oscillator	Max	2.5910	RSI	Max	100.0000
	Min	−2.3165		Min	0.0000
	Mean	−0.5315		Mean	37.6026

Table V. Average hit rate (%) for the in-sample data

Model	Set 1	Set 2	Set 3	Set 4	Set 5	Average
Benchmark	52.50	48.33	53.33	55.00	50.83	52.00
LT	64.17	60.00	68.33	62.50	57.50	62.50
FT	75.00	76.67	77.50	78.33	74.17	76.33

Table VI. Average hit rate (%) for the hold-out data

Model	Set 1	Set 2	Set 3	Set 4	Set 5	Average
Benchmark	53.33	66.67	46.67	40.00	56.67	52.67
LT	50.00	56.67	60.00	53.33	66.67	57.33
FT	73.33	73.33	70.00	66.67	83.33	73.33

The hit ratio in Tables V and VI is measured for the five different sets because this study uses the fivefold cross-validation. The benchmark model in Tables V and VI assumes that the pattern of the next day takes the same pattern of the current day. This shows that feature transformation based on domain knowledge outperforms a linear transformation.

We examine the statistical significance of whether the FT model outperforms the LT model. The two-sample test for proportions is executed. This test is designed to distinguish between two proportions (Harnett and Soni, 1991). Table VII shows the standardized normalized test statistic, Z values, when the prediction accuracy of the left-vertical methods is compared with those of the right-horizontal methods.

With respect to the feature transformation method, the FT model performs significantly better than the LT model and the FT model also performed better then benchmark at the 1% significance level.

Table VII. Z values for the pairwise comparison of performance between models (for the hold-out data)

	LT	FT
Benchmark	0.812	3.707*
LT		2.912*

* Significant at the 1% level.

Therefore, we can conclude that feature transformation based on domain knowledge outperforms the linear transformation method at a statistically significant difference.

6. CONCLUSIONS

We presented the method of feature transformation based on domain knowledge in AI applications to predict the pattern of stock market trends. In this approach, domain knowledge in the stock market is used to discretize the original data. We show that this approach effectively filters data, trains the classifier, and extracts the rules from the classifier. In addition, we conclude from the empirical results that this approach reduces the dimensionality of the feature space, enhances the generalizability of classifier.

The results of the experiment show that domain-knowledge-based feature transformation significantly outperforms linear transformation in classifying the fluctuation of the index futures. It appears that feature transformation based on domain knowledge supports the learning of noisy patterns better than linear transformation does. It can also effectively produce reasonable trading rules using an inductive learning method, such as a decision tree, because it has discrete feature values rather continuous ones. The implications of this study suggest that ANNs applications of pattern classification are valid with the method of feature transformation based on domain knowledge. This implies a high potential for the appropriate feature transformation method for data.

Though the method of feature transformation based on domain knowledge produces valid results, the approach has some limitations. First, feature transformation based on domain knowledge always needs domain-specific knowledge. It is difficult, however, to extract domain-specific knowledge from an unstructured domain. In addition, extracted domain knowledge sometimes may be subjective and arbitrary. The second limitation is that the method of endogenous feature transformation, as mentioned earlier, does not give consideration to the correlation between each independent variable and dependent variable. Prediction performance, however, is enhanced by the ability to discrimination only from not a single variable but also the association among variables. For these limitations, endogenous methods do not provide an effective way of forming categories. Future studies are expected to focus on the method of exogenous feature transformation with objective and systematic characteristics.

ACKNOWLEDGEMENTS

This work is supported by the Dongguk University Research Fund.

REFERENCES

Achelis SB. 1995. *Technical Analysis from A to Z*. Probus Publishing: Chicago.

Ahmadi H. 1990. Testability of the arbitrage pricing theory by neural networks. In *Proceedings of the International Conference on Neural Networks*; 385–393.

Basak J, De RK, Pal SK. 1998. Unsupervised feature selection using a neuro-fuzzy approach. *Pattern Recognition Letters* **19**(11): 997–1006.

Bigus JP. 1996. *Data Mining with Neural Networks.* McGraw-Hill: New York.

Blum AL, Langley P. 1997. Selection of relevant features and examples in machine learning. *Artificial Intelligence* **97**(1–2): 245–271.

Buhlmann P. 1998. Extreme events from the return-volume process: a discretization approach for complexity reduction. *Applied Financial Economics* **8**(3): 267–278.

Chang J, Jung Y, Yeon K, Jun J, Shin D, Kim H. 1996. *Technical Indicators and Analysis Methods*. Jinritamgu Publishing: Seoul.

Choi J. 1995. *Technical Indicator*. Jinritamgu Publishing: Seoul.

Choi JH, Lee MK, Rhee MW. 1995. Trading S&P 500 stock index futures using a neural network. In *Proceedings of the Third Annual International Conference on Artificial Intelligence Applications on Wall Street*, New York, Freedman RS (ed.); 63–72.

Dash M, Liu H. 1997. Feature selection for classification. *Intelligent Data Analysis: An International Journal* **1**(3): 131–156.

Dougherty J, Kohavi R, Sahami M. 1995. Supervised and unsupervised discretization of continuous features. In *Proceedings of the Twelfth International Conference on Machine Learning*, Prieditis A, Russell S (eds). Morgan Kaufmann Publishers: San Francisco; 194–202.

Duke LS, Long JA. 1993. Neural network futures trading—a feasibility study. In *Adaptive Intelligent Systems*, Society for Worldwide Interbank Financial Telecommunications (ed.). Elsevier Science Publishers: Amsterdam; 121–132.

Edwards RD, Magee J. 1997. *Technical Analysis of Stock Trends.* John Magee Inc.: Chicago.

Fayyad UM, Irani KB. 1993. Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*; 1022–1027.

Gifford E. 1995. *Investor's Guide to Technical Analysis: Predicting Price Action in the Markets*. Pitman Publishing: London.

Harnett DL, Soni AK. 1991. *Statistical Methods for Business and Economics*. Addison-Wesley Publishing Company, Inc.: Reading, MA.

Hiemstra Y. 1995. Modeling structured nonlinear knowledge to predict stock market returns. In *Chaos & Nonlinear Dynamics in the Financial Markets: Theory, Evidence and Applications*, Trippi RR (ed.). Irwin: Chicago; 163–175.

Jhee WC, Lee JK. 1993. Performance of neural networks in managerial forecasting. *International Journal of Intelligent Systems in Accounting, Finance and Management* **2**(1): 55–71.

Kamijo K, Tanigawa T. 1990. Stock price pattern recognition: a recurrent neural network approach. In *Proceedings of the International Joint Conference on Neural Networks*; 215–221.

Kimoto T, Asakawa K, Yoda M, Takeoka M. 1990. Stock market prediction system with modular neural network. In *Proceedings of the International Joint Conference on Neural Networks*; 1–6.

Kohara K, Ishikawa T, Fukuhara Y, Nakamura Y. 1997. Stock price prediction using prior knowledge and neural networks. *International Journal of Intelligent Systems in Accounting, Finance and Management* **6**(1): 11–22.

Kolb RW, Hamada RS. 1988. *Understanding Futures Markets*. Scott, Foresman and Company.

Kontkanen P, Myllymaki P, Silander T, Tirri H. 1997. A Bayesian approach to discretization. In *Proceedings of the European Symposium on Intelligent Techniques*; 265–268.

Lawrence S, Tsoi AC, Giles CL. 1996. Noisy time series prediction using symbolic representation and recurrent neural network grammatical inference. Technical Report UMIACS-TR-96-27 and CS-TR-3625. Institute for Advanced Computer Studies, University of Maryland.

Liu H, Motoda H. 1998. Feature transformation and subset selection. *IEEE Intelligent Systems and Their Applications* **13**(2): 26–28.

Martens J, Wets G, Vanthienen J, Mues C. 1998. An initial comparison of a fuzzy neural classifier and a decision tree based classifier. *Expert Systems with Applications* **15**(3–4): 375–381.

Murphy JJ. 1986. *Technical Analysis of the Futures Markets: A Comprehensive Guide to Trading Methods and Applications*. New York Institute of Finance/Prentice-Hall: New York.

Piramuthu S, Ragavan H, Shaw MJ. 1998. Using feature construction to improve the performance of neural networks. *Management Science* **44**(3): 416–430.

Scott PD, Williams KM, Ho KM. 1997. Forming categories in exploratory data analysis and data mining. In *Advances in Intelligent Data Analysis*, Liu X, Cohen P, Berthold M (eds). Springer-Verlag: Berlin; 235–246.

Slowinski R, Zopounidis C. 1995. Application of the rough set approach to evaluation of bankruptcy risk. *International Journal of Intelligent Systems in Accounting, Finance and Management* **4**(1): 27–41.

Susmaga R. 1997. Analyzing discretizations of continuous attributes given a monotonic discrimination function. *Intelligent Data Analysis—An International Journal* **1**(3): 157–179.

Ting KA. 1997. Discretization in lazy learning algorithms. *Artificial Intelligence Review* **11**(1–5): 157–174.

Trippi RR, DeSieno D. 1992. Trading equity index futures with a neural network. *Journal of Portfolio Management* **19**: 27–33.

Tsaih R, Hsu Y, Lai CC. 1998. Forecasting S&P 500 stock index futures with a hybrid AI system. *Decision Support Systems* **23**(2): 161–174.

Vafaie H, De Jong K. 1998. Feature space transformation using genetic algorithms. *IEEE Intelligent Systems and Their Applications* **13**(2): 57–65.

Weigned AS, Rumelhart DE, Huberman BA. 1991. Generalization by weight elimination applied to currency exchange rate prediction. In *Proceedings of the International Joint Conference on Neural Networks*; 837–841.

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

Yoon Y, Swales G. 1991. Predicting stock price performance: a neural network approach. In *Proceedings of the 24th Annual Hawaii International Conference on System Sciences*; 156–162.

