PERGAMON

Expert Systems with Applications 22 (2002) 33–39

Expert Systems with Applications

www.elsevier.com/locate/eswa

Predicting stock price using fuzzy grey prediction system

Y.-F. Wang

Department of Information Management, Chang Gung Institute of Nursing, Kwei-Shan, Tao-Yuan, Taiwan

Abstract

The purpose of this paper is to predict the stock price instantly at any given time. One problem with predicting stock prices is that there may be a large or small difference in two continuous sets of data. The other problem is that the volume of stock data is so large that it affects our ability to use it. To solve these problems, we constructed a *data mart* to reduce the size of stock data and combined fuzzification techniques with the grey theory to develop a *fuzzy grey prediction* as one of predicting functions in our system to predict the possible answer immediately. To demonstrate that our system is working correctly, we used our prediction system to analyse stock data and to predict the stock price promptly at a specific time. The system can effectively help stock dealers deal with day trading. © 2001 Published by Elsevier Science Ltd.

Keywords: Stock price; Fuzzy grey prediction

1. Introduction

Investors have been trying to find a way to predict stock prices accurately, but have had less than successful results. Kuo, Chen and Hwang (2001) showed that there are many studies which address prediction of stock price that have generally employed the time series analysis techniques (Kendall & Ord, 1990) and multiple regression models. Most of them only consider quantitative factors like technical indexes. Recently, artificial intelligence techniques like artificial neural networks (ANNs) and genetic algorithms (GAs) have been applied to this area. However, the above-mentioned concern still exists (Baba & Kozaki, 1992; Mahfoud & Mani, 1996). Kim and Han (2000) showed that ANNs had some limitations in learning the patterns because stock price data has tremendous noise and complex dimensionality. Moreover, the sheer quantity of stock data sometimes interferes with the learning of patterns.

Kuo et al. (2001) pointed out that numerous factors, such as macro-economical and political events, can have a major influence on stock prices. Even the psychology of investors can generate index oscillation. Therefore, it is not sufficient to restrict stock price prediction to just some technical indexes in such a complicated environment. This explains why experienced stock experts and brokers can make more accurate decisions than average investors, since they do not only consider the technical indexes but also qualitative factors based on their experience and knowledge.

The timing of buying/selling stock is based on determining the best time to buy and sell stocks given the constant fluctuation of stock prices. However, humans have a difficult time doing this because of the complexity of the stock market, as showed in Lee and Jo (1999).

The problem with predicting stock prices is that the volume of data is too huge to influence the ability of using information (Fayyad, Shapiro & Smyth, 1996; Widom, 1995). Analyzing stock price data over several years may involve only a few thousand records, but these must be selected from millions. A stockbroker who serves tens of thousands of customers each year may generate up to 170 GB of stock data at any given time. Multiyear trend analysis of the stock price thus still presents a problem due to the vast amount of data involved. It is, therefore, important to devise efficient methods to analyse and predict stock prices. For this reason, we constructed a data mart (Demarest, 1994), a relational database, to clean and reduce the size of the stock data so only the useful data is downloaded and reformatted into the data mart (Liu & Setiono, 1996). Consequently, the size of these stock data is reduced from approximately 600 MB per day to approximately 1 MB per day. The advantage of using reformatted data is that reformatted data is more easily understood and used by users, as shown in Table 1. Since our database stores the reformatted data supported, the minimum time interval is 5 min.

The grey theory that was first proposed by (Deng, 1982, 1989a,b) avoids the inherent defects of conventional statistic methods and only requires a limited amount of

E-mail address: yfwang@cc.cgin.edu.tw (Y.-F. Wang).

0957-4174/01/$ - see front matter © 2001 Published by Elsevier Science Ltd.
PII: S0957-4174(01)00047-1

Table 1
Reformatted relational database

StockNo	Date	Time	Open	High	Low	Close	Vol	Invol	OutVol	Amt
2311	2000/07/24	09:10:00	8650	8750	8550	8750	192	118	74	16627
2311	2000/07/24	09:15:00	8750	8750	8650	8650	123	104	19	10694
2311	2000/07/24	09:20:00	8650	8700	8600	8650	88	48	40	7621
2311	2000/07/24	09:25:00	8650	8700	8650	8700	35	22	13	3032
2311	2000/07/24	09:30:00	8650	8650	8650	8650	64	64	0	5534
2311	2000/07/24	09:35:00	8650	8650	8650	8650	161	161	0	13923
2311	2000/07/24	09:40:00	8650	8650	8600	8650	221	220	1	19011
2311	2000/07/24	09:45:00	8600	8650	8600	8650	139	136	3	11955
2311	2000/07/24	09:50:00	8600	8650	8600	8600	73	58	15	6299
2311	2000/07/24	09:55:00	8650	8650	8600	8600	59	25	34	5089
2311	2000/07/24	10:00:00	8600	8600	8550	8550	100	41	59	8588
2311	2000/07/24	10:05:00	8550	8600	8550	8600	77	69	8	6586
2311	2000/07/24	10:10:00	8600	8600	8550	8600	39	21	18	3342
:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:

data to estimate the behaviour of unknown systems (Deng, 1982, 1989a,b). Therefore, we use the grey theory to develop an efficient method to solve the problems of stock price prediction. In this paper, we adopt the most commonly used GM(1,1) for our fuzzy grey prediction. The forecasting step is then generated dynamically from the most recently system behaviour by the proposed method to reduce predication errors.

The other problem with predicting stock prices is that price difference can vary greatly over 2 h. For instance, it is often difficult to classify patients as fully sick (Kacprzyk & Iwanski, 1992). Therefore, crisp data mining approaches may not be appropriate for these situations. To solve this problem, we employ a fuzzification technique into our fuzzy grey prediction (FGP) as one of predicting functions in our system to predict the stock price in the next hour. Using our system not only enables the user to know the stock price in any given hour, but also to follow stock price trends. In order to demonstrate that our method works correctly, we considered the stock price analysis problem over 300 weekdays of trading.

2. Preliminaries

Dr Deng initialed the grey theory in 1982. He used a grey predictor to predict the system behaviour and fed the predicting information back to the decision-making mechanism to indicate an appropriate control action. The most commonly used GM(1,1) model for the grey prediction is described as follows (Deng, 1989b):

$$x^{(1)} = \mathrm{IAGO} \cdot \mathrm{GM}(1,1) \cdot \mathrm{AGO} \cdot x^{(0)} \tag{1}$$

where $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(k))$ is a non-negative original data sequence and $k \geq 4$; moreover, $x^{(1)}$ is a predicting value of $x^{(0)}$, AGO takes the accumulated generating operation on $x^{(0)}$, and IAGO takes the inverse accumulated generating operation on $x^{(1)}$. Hence, the predicating value of $x^{(1)}(k+p)$ can be calculated by

$$x^{(1)}(k+p) = \left(x^{(0)}(1) - \frac{b}{a}\right)(1 - e^{a})e^{-a(k+p-1)} \tag{2}$$

where p is the forecasting step, a and b are the development coefficient and the grey input respectively (Wong, Liang, Feng & Chiang, 1998). In this paper, we will use formula (2) to predict the stock price.

As pointed out by Deng (1982), the constraint of grey prediction model GM(1,1) is that the ratio, $x^{(0)}(k-1)/x^{(0)}(k)$, must be in the period of [0.1345, 7.389]; otherwise, the grey prediction model GM(1,1) is not suitable to use for predicting the behaviour of the system. Therefore, we propose a fuzzy grey prediction method to overcome this problem.

Essentially, the traditional grey predictions use a fixed forecasting step to predict the system behaviour. It can reduce or prevent overshoot. However, the appropriate forecasting step is hard to find. To solve this problem, Chiang, Wang, Lee and Shi (1999) developed a new method to discover a suitable forecasting step dynamically from the most recent system behaviour. Expanding on formula (2), this formula can be rewritten as:

$$p' = \frac{1}{a}\left(-\log_e x^{(1)}(k+p-1) + \log_e\left(x^{(0)}(1) - \frac{b}{a}\right) - \log_e\left(1 - e^{-a}\right)\right) - k \tag{3}$$

where p' is the new forecasting step.

To predict the value of $x^{(1)}(6)$, only use a series of five historical data, $x^{(0)}(1), \ldots, x^{(0)}(5)$, to discover the appropriate forecasting step for $x^{(1)}(6)$. In formula (3), let that $x^{(0)}(5) = x^{(1)}(k+p-1)$, and the value b/a is computed by $x^{(0)}$, where $x^{(0)} = (x^{(0)}(1), \ldots, x^{(0)}(4))$. Consequently, we can find out a new forecasting step p' for $x^{(1)}(6)$ by formula

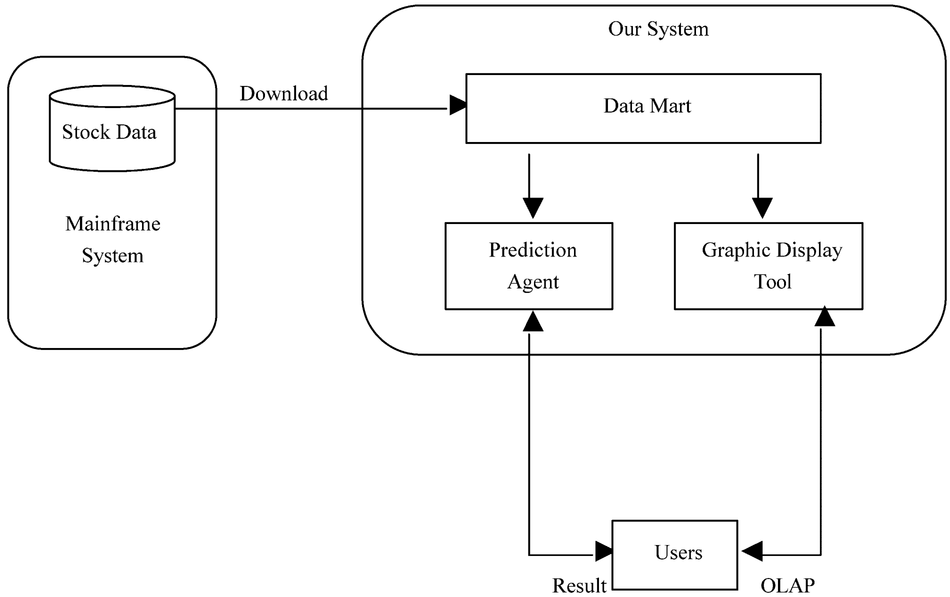

Fig. 1. The simplified architecture of the system.

(3). Then, let $x^{(0)} = (x^{(0)}(2),\ldots,x^{(0)}(5))$ and apply formula (2) to predict the value of $x^{(1)}(6)$.

3. Fuzzy grey prediction system

Our system was written with Visual BASIC on an IBM PC and includes two major modules: graphic display tool, prediction agent. The simplified architecture of the system is given in Fig. 1. In the proposed system, users can check the stock price through the graphic display tool. Since the predicting process is an application oriented process, different applications may need different predicting approaches. In this paper, we introduce only fuzzy grey prediction in the prediction agent.

3.1. Graphic display tool

This subsection provides a brief introduction of the graphic display tool. The graphic display tool was developed to display the specific stock price. Though the graphic display tool is not much more friendly than commercial products, it is a useful tool that can be used to analyse stock price trends effectively. It includes window-based interactive dialog boxes, mouse-driven menus and scroll bars. As shown in Fig. 2, the input screen of the system can be separated into two regions for input selection:

- StockNo/Name region;
- Function groups region.

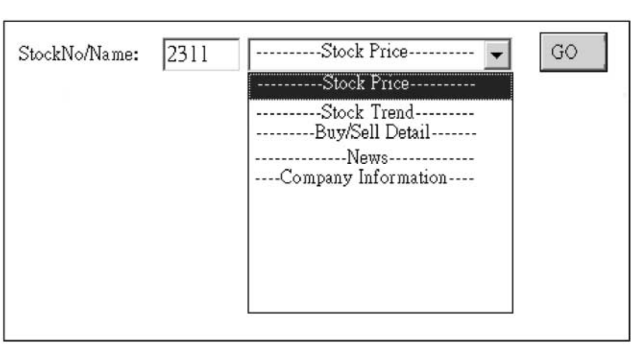

Fig. 2. The input screen of the system.

3.1.1. StockNo/Name region

This includes one text box. Users can input the specific stock number or stock name they are interested in.

3.1.2. Function groups region

This contains one combo box, the drop down list appears after the users press this box, which includes the functional keywords: Stock Price, Stock Trend, Buy/Sell Detail, News, Company Information. Some of partial functions are briefly introduced as follows:

- Stock Price Function displays the selected stock current market dealing price, the price of buying and selling as well as the highest and lowest price and volume and so forth.
- Stock Trend Function displays the selected stock price over the past 6 years diagram. The stock price data can be grouped into daily, weekly, or monthly.
- Buy/Sell Detail Function displays the selected stock price buying or selling detail report, which includes the time of buying or selling, volume, price, etc.

Table 2
Stock price on 5 February 2001

Date	Time	Stock price	$\mu_{\text{price}}(t[\text{price}])$
2001/02/05	09:00	35.81	1
2001/02/05	10:00	35.20	0.9662
2001/02/05	11:00	34.29	0.9169
2001/02/05	12:00	33.48	0.8741
2001/02/05	13:00	35.20	0.9662
2001/02/05	13:30	35.01	0.9777

3.2. Prediction agent using fuzzy grey prediction

The stock price is distinct each hour and can be influenced by numerous factors at any given time. For example, as shown in Tables 2 and 3, many factors contributed to the differences seen in the stock price between 5 February 2001 and 15 August 2000 because of many reasons. To solve this problem, we employ fuzzification techniques on the FGP so that the proposed methods can be used to predict the stock price promptly and effectively. In Tables 2 and 3, the value $\mu_{\text{price}}(t[\text{price}])$ is the degree of the stock price at a specific trading hour on that day, and the membership function of μ_{price} is given as:

$$\mu_{\text{price}}(x) = (x/y)^2,$$

where x is the stock price at a specific trading hour on one day, and y is the highest stock price during trading hours on the same day.

Particularly, the change rate at an hour interval of the stock price may vary slightly every day. Since different times present different specific behaviours, we can pre-classify stock price data into different time groups if we consider only the behaviour of stock price of trading hours on weekdays. We can pre-classify tuples in the database with the 'Time' attribute value so that the tuples are divided into six groups. As shown in Table 4, each group has the same values for the attribute 'Time'. Consequently, stock price has the same behaviour in the same group (Chiang, Chow & Wang, 2000).

The advantages of the results of fuzzification and pre-classification are that this process can highlight the stock price characteristics. Moreover, we can use formula (2) to predict the stock price because the stock price data is pre-classified and fuzzed.

Table 3
Stock price on 15 August 2000

Date	Time	Stock Price	$\mu_{\text{price}}(t[\text{price}])$
2000/08/15	09:00	86.91	1
2000/08/15	10:00	85.23	0.9617
2000/08/15	11:00	83.02	0.9125
2000/08/15	12:00	80.92	0.8669
2000/08/15	13:00	82.47	0.9004
2000/08/15	13:30	82.15	0.9452

Table 4
Table of $\mu_{\text{price}}(t[\text{price}])$

	Time	$\mu_{\text{price}}(t[\text{price}])$
↑	09:00	0.9864
Group 1	:	:
↓	09:00	1
↑	10:00	0.9327
Group 2	:	:
↓	10:00	0.9665
:	:	:
:	:	:
:	:	:
↑	14:00	0.9887
Group 6	:	:
↓	14:00	0.9678

The following is an example of predicting the stock price by using fuzzy grey prediction.

3.2.1. Example

Consider the stock price at 09:00am in Table 5. We want to predict the stock price at 09:00am on 10 February.

Let $x^{(1)}(5) = \mu_{\text{price}}(t[\text{price}])$ at 09:00am on 9 February = 0.774 and $x^{(0)} = (1,\ 0.853,\ 0.652,\ 0.386)$. According to formula (3), we find out that the forecasting step with respect to $\mu_{\text{price}}(t[\text{price}])$ at 09:00am on 9 February is −2.590. Now, let $x^{(0)} = (0.853,\ 0.652,\ 0.386,\ 0.774)$. By formula (2), we know that $\mu_{\text{price}}(t[\text{price}])$ at 09:00am on 10 February is 0.560. Compared with the actual data on 9 February, as shown in Table 6, we get the following error ratio: $(\sqrt{0.560} - \sqrt{0.544}) \div \sqrt{0.544} = 0.0146$. Accordingly, the predicting error ratio is 1.46%.

4. Experimental results

In this paper, all the data was collected from a stockbroker's mainframe in Taiwan stock market recorded every 5 min. The trading hours in Taiwan stock market is from 9:00am to 13:30pm without lunch break on weekdays. To simplify data processing, the data was grouped into hours assuming the trading time is from 9:00am to 14:00pm on weekdays.

To demonstrate how effective our system is, we used data from September 2000 to February 2001 as training examples, and data from March 2001 to April 2001 as testing examples.

Table 5
The $\mu_{\text{price}}(t[\text{price}])$ of last week

Date	Time	$\mu_{\text{price}}(t[\text{price}])$
2001/02/05	09:00	1
2001/02/06	09:00	0.853
2001/02/07	09:00	0.652
2001/02/08	09:00	0.386
2001/02/09	09:00	0.774

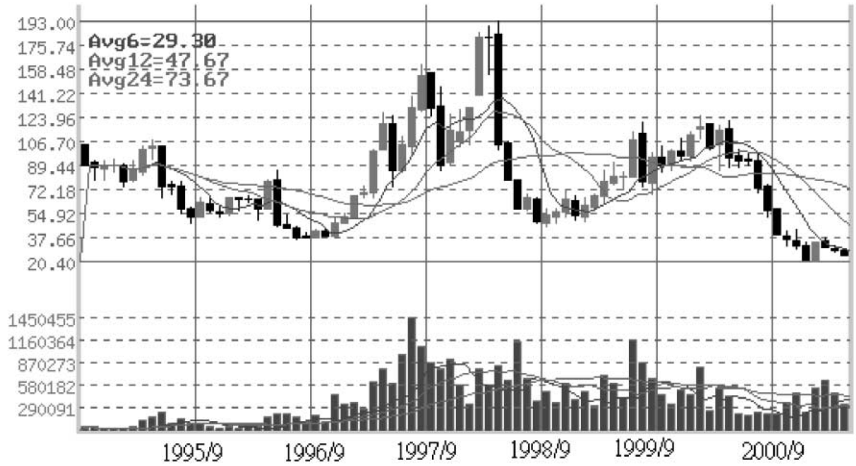

Fig. 3. Stock price from September 1994 to May 2001 (by month).

Table 6
The actual data on 9 February

Date	Time	Stock price	$\mu_{\text{price}}(t[\text{price}])$
2001/02/10	09:00	30.70	0.544
2001/02/10	10:00	32.46	0.608
2001/02/10	11:00	41.20	0.980
2001/02/10	12:00	35.63	0.733
2001/02/10	13:00	35.02	0.841
2001/02/10	13:30	41.63	1

4.1. Checking stock price using the graphic display tool

The graphic display tool can also display both historical and current stock price data for comparison purposes. First of all, after entering StockNo and choosing the ‘Stock Price’ function on the input screen as shown in Fig. 2, press the ‘GO’ button to generate the stock price diagram shown in Fig. 3. In Fig. 3, the bar chart and line chart represent the high/low value and average value of the stock price for each month, respectively.

Double-clicking the diagram area displays the chart shown in Fig. 4, which shows weekly stock price data up to a 1-year period. At this moment, we can double-click again, as shown in Fig. 5, to see the daily stock price over a 4-month period. Double-click again as shown in Fig. 6 to display a line chart for 9–13 h stock prices on 3 May 2001. After seeing these diagrams, we concluded that they might show a relationship between the stock price and time. In other words, the knowledge to be discovered is the relationship between the stock price ranking and time in this case.

4.2. Predicting stock price using prediction agent

The advantage of using fuzzy grey prediction in our prediction agent is that only a few data inputs are needed to predict stock behaviour. To assess the accuracy of our method, we conducted several experiments, similar to Section 3.2.1. Consequently, the value of $\mu_{\text{price}}(t[\text{price}])$ at highest is always equal to 1. Therefore, when we know the $\mu_{\text{price}}(t[\text{price}])$ at 09:00, we can infer the highest stock price. In the Section 3.2.1, the highest stock price is:

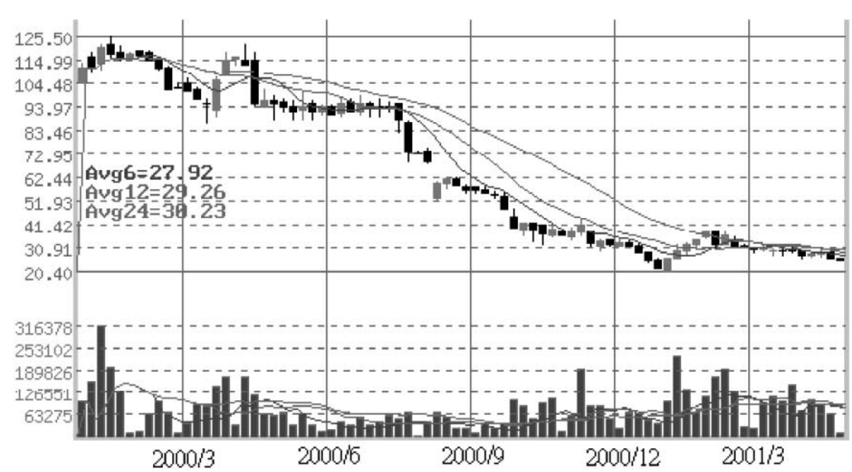

Fig. 4. Stock price from January 2000 to May 2001 (by week).

Fig. 5. Stock price from January 2001 to May, 2001 (by day).

$30.7 \div \sqrt{0.56} = 41.026$. After running 180 trials, the average deviation from the predicted values is less than 9%. The result shows that our fuzzy grey prediction is feasible.

5. Conclusion

In the proposed system, the graphic display tool can assist users in checking the historical stock data or the current situation of the stock market. The prediction agent can be used to predict the stock price at any given moment when after trading begins. In addition, the proposed system can still be used to effectively predict stock prices even without using the graphic display tool.

In our system, most parameters are defaulted and cannot be changed by the user. Therefore, we plan to develop a more flexible interface for users in the near future. At present, we just simply ignore the data for that day when missing data appears. In the future, we will consider the problem of missing data and develop more useful agents for our system.

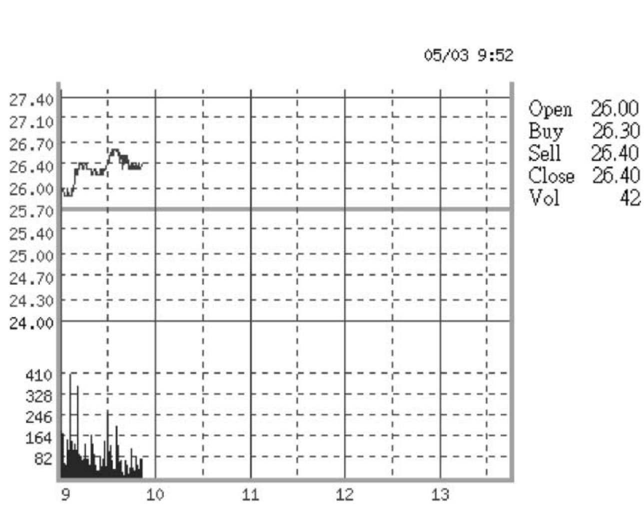

Fig. 6. Stock price on 3 May 2001 (every 5 min).

References

Baba, N., & Kozaki, M. (1992). An intelligent forecasting system of stock price using neural networks, *Proc. IJCNN*.

Chiang, D. A., Wang, Y. F., Lee, S. L., & Shi, C. H. (1999). Analyzing mainframe resource usage by dynamic grey prediction. *The Journal of Grey System*, *11* (2), 97–106.

Chiang, D. A., Chow, L. R., & Wang, Y. F. (2000). Mining time series data by a fuzzy linguistic summary system. *Fuzzy Sets and Systems*, *112* (3), 419–432.

Demarest, M. (1994). Building the data mart. *DBMA Magazine*, *7* (8), 44–45.

Deng, J. (1982). Control problems of grey system. *Systems & Control Letters*, *1* (5), 288–294.

Deng, J. (1989a). Introduction to grey system theory. *The Journal of Grey System*, *1* (1), 1–24.

Deng, J. (1989b). Properties of multivariable grey model GM(1,1). *The Journal of Grey System*, *1* (1), 25–42.

Fayyad, U., Piatesky-Shapiro, G., Smyth, P., & The, K. D. D. (1996). process for extracting useful knowledge from volumes of data. *ACM Comm.*, *39* (11), 27–34.

Kacprzyk, J., & Iwanski, C. (1992). Fuzzy logic with linguistic quantifiers in inductive learning. *Fuzzy logic for the management of uncertainty* (pp. 465–478). New York: John Wiley & Sons Ltd.

Kendall, S. M., & Ord, K. (1990). *Time series*, (3rd ed). New York: Oxford University Press.

Kim, K. J., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications*, *19*, 125–132.

Kuo, R. J., Chen, C. H., & Hwang, Y. C. (2001). An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy Sets and Systems*, *118*, 21–45.

Lee, K. H., & Jo, G. S. (1999). Expert system for predicting stock market timing using a candlestick chart. *Expert Systems with Applications*, *16*, 357–364.

Liu, H., & Setiono, R. (1996). Dimensionality reduction via discretization. *Knowledge-Based Systems*, *9* (1), 67–72.

Mahfoud, S., & Mani, G. (1996). Financial forecasting using genetic algorithms. *Appl. Artificial Intell.*, *10*, 543–565.

Widom, J. (1995). Research problems in data warehousing. *Proceedings of the Fourth International Conference on Information and Knowledge Management, Baltimore, 25–30 November*.

Wong, C. C., Liang, W. C., Feng, H. M., & Chiang, D. A. (1998). Dynamic grey prediction control design. *The Journal of Grey System*, *10* (2), 123–131.