Forecasting Stock Price Movement with Semi-Supervised Learning

Kanghee Park and Hyunjung Shin*

Abstract— **Stock price prediction is a field that has been continuously interested. Stock price indexes represent correlation between the company and the others including the influence of oil prices, exchange rates, money interest rates, stock price indexes in other countries, and economic situations, the indexes are sensitively influenced by the fluctuation of these factors. To overcome the complexity, this paper proposes a network based method incorporating the relations between the stock prices and the other factors by using a graph-based semi-supervised learning algorithm. For verifying the significance of the proposed method, it was applied to the prediction problems of company stock prices listed in the KOSPI from January 2007 to August 2008.**

I. INTRODUCTION

Interests on the stock price prediction have been continued according to the public understanding in stock investment. Factors that affect stock prices are oil prices, exchange rates, interest rate, stock price indexes in other countries, and economic situations. Studies on stock price prediction methods based on these factors have been variously conducted [1-4].

Various stock price prediction methods using a time series analysis method have been presented. Jeantheau (2004) predicted stock prices using an ARCH model, and Amilon (2003) and Liu et al. (2009) proposed a prediction method using a GARCH model based on the Skewed-GED Distribution for Chinese stock markets [2-3]. As these methods perform the prediction using a time series analysis method based on the past stock price validity, an assumption in which the future stock price will be varied as similar to that of the past is a basis. The time series data obtained from some natural phenomena, such as numbers of sunspot cycles, rain falls, temperature, and others, nicely follows such an assumption. It is possible to obtain excellent results using the time series analysis method.

Although these various factors mentioned above affect stock prices directly, they have influence on the stock price indirectly through a complex interrelation between these factors. For instance, although interests rates and exchange rates directly affect stock price fluctuations, they have influence on the stock price based on the reciprocal relationship between these two factors.

The authors would like to gratefully acknowledge support from Post Brain Korea 21 and the research grant from National Research Foundation of Korean government (KRF-2010-0007804).

Kanghee Park (can17@ajou.ac.kr) is in the Ph.D course of Industrial & Engineering, Ajou University, Suwon, 443-749 Korea

* Corresponding author: Hyunjung Shin (shin@ajou.ac.kr) is a professor of the department of Industrial & Information Systems Engineering, Ajou University, Suwon, 443-749, Korea

However, there are some methodological limitations that include the relationship between these factors and reciprocal complexity to a time series model specifically and its formalization [5-6]. Also, many studies on the stock price prediction in the machine learning have been conducted. The artificial neural network (ANN) and support vector machine (SVM) methods have been frequently used as a typical model [7-9]. Tay and Cao (2001) proposed a method that introduces financial time series data to the SVM, and Kanas (2003) attempted the prediction of the S&P500 index using the ANN model [10-11]. Also, Yang et al. (2001) proposed an early warning system of commercial bank loan risks using the ANN model, and Bekiros and Georgoutsos (2008) analyzed that how uncertain news, which show a difficulty in identifying bullish and bearish factors, affect the NASDAQ index using the ANN model [12-13].

Although the methods using ANN and SVM include the interrelation and complexity between the stock price and these factors in its modeling specifically, it is still insufficient. It does not formalize the interrelation between factors even though the factors in fluctuating stock prices and their interrelation are expressed in the model [14]. For instance, ANN and SVM can represent how the fluctuations in interests, exchange rates, and oil prices affect the fluctuations in stock prices primarily. However, it is somewhat difficult to express how a drop in interests affects the exchange rates and then how these changes affect the next situation, i.e., how the second, third, and additional higher level interrelations affect the stock prices eventually. In addition, it is not easy to identify how the changes in stock prices caused by such a sequential process affect these factors again. That is, there is a limitation in presenting the complexity between factors. In this study, a stock price prediction method that uses semi-supervised learning (SSL), which has been recently attracted in the field of machine learning methods, is proposed to solve this limitation [15-16].

SSL that is one of recently developed machine learning methods is an analysis method through defining the interrelation between factors to a network [15-16]. SSL can consider the interrelation and complexity between factors through a network. It connects individual networks using the similarities between factors and extracts the influence of the final similarities in the connected input factors and responded factors as its prediction value. In this study, a stock prediction model that considers the interrelation and multi-dimensional causal complexity in various economic indexes by combining time series data to SSL is proposed. The proposed model was applied to the stock price prediction for individual companies listed to KOSPI from January 2007 to August 2008 and its performance was also verified.

This study consists of five sections. Section 2 describes the methodology of SSL. Section 3 proposes a method that combines time series data to SSL. Section 4 represents experiments and results. Finally, Section 5 shows the conclusion of this study.

II. Semi-supervised learning (SSL)

In graph-based SSL algorithm, a data point $x_i \in R^M$ *(i =* 1,…, *n)* is represented as a node *i* in a graph, and the relationship between data points is represented by an edge where the connection strength from each node *j* to each other node *i* is encoded as w_{ij} of a weight matrix *W* [17]. Fig. 1 presents a graphical representation of SSL.

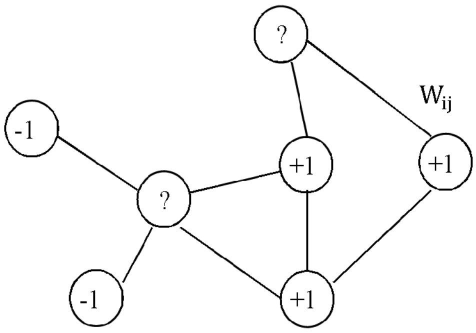

Fig. 1 Graph-based semi-supervised learning (SSL).

A weight w_{ij} can take a binary value (*0* or *1*) in the simplest case. Often, a Gaussian function of Euclidean distance between points with length scale σ is used to specify connection strength:

$$w_{ij} = \begin{cases} \exp\left(-\frac{(x_i-x_j)^T(x_i-x_j)}{\sigma^2}\right) & \text{if } i\sim j \text{ ('k' nearest neighbors)} \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

Usually, an edge $i\sim j$ is established when node *i* is one of *k*-nearest neighbors of node *j* or node *i* is within a certain Euclidean distance *r*, $\|x_i-x_j\| < r$. The labeled nodes have labels $y_l \in \{-1,1\}$ $(l = 1,\dots,L)$, while the unlabeled nodes have zeros $y_u = 0$ *(u=L+1,...,L+U)*. The algorithm will output an n-dimensional real-valued vector $f = [f_l^T\ f_u^T]^T = (f_1,\dots,f_L,f_{L+1},\dots,f_{L+U})^T$ which can be thresholded to make label predictions on $f_{L+1},\dots,f_{L+U}$ after learning. It is assumed that (a) f_i should be close to the given label y_i in labeled nodes and (b) overall, f_i should not be too different from its adjacent nodes f_j. One can obtain *f* by minimizing the following quadratic functional:

$$\mathrm{Min}_f\ (f\text{-}y)^T(f\text{-}y) + \mu f^T L f, \quad (2)$$

where $y = (y_1,\dots,y_l,0,\dots,0)^T$, and the matrix L, called the graph Laplacian, is defined as $L = D - W$, $D = \mathrm{diag}(d_i)$, and $d_i = \sum_j \omega_{ij}$. The first term corresponds to the loss function in terms of condition (a), and the second term represents the smoothness of the predicted outputs in terms of condition (b). The parameter μ represents trades between loss and smoothness. The solution to (2) is obtained as

$$f = (I + \mu L)^{-1} y, \quad (3)$$

where *I* is the identity matrix.

III. Proposed Method

To apply the graph-based SSL to time series prediction, we propose a method of graph representation for time series data, and a procedure for obtaining predicted values from the graph. For instance, assume that multiple time series are given as the input for the prediction problem of the stock price of Hyundai Motors: the stock price of LG chem, the stock price of KIA Motors, WTI intermediate oil price, *etc.* To apply SSL to this problem, the proposed method begins with a re-designed graph as in Fig. 2.

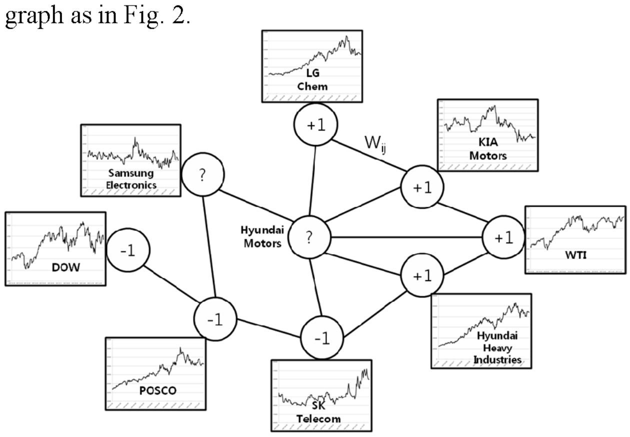

Fig. 2 Graph SSL representation for time series prediction.

The nodes in the graph represent the time series variables that influence the stock price of Hyundai Motors, e.g., the stock price of LG chem, the stock price of KIA Motors, WTI intermediate oil price and other external factors. Then the edge between any two nodes $i\sim j$ stands for the similarity of the two sets of time series, represented as '$w_{ij} \in W$'. The label 'y_t' on each node presents either 'up' (+1) or 'down' (-1) of the time series at time point *t*. In the graph of <Figure 2>, the labels of Hyundai Motors are not known yet at time point *t*, and hence are unlabeled. To estimate the label y_t, the similarity matrix of SSL was calculated at time point t-1, W_{t-1}. Based on this set-up, we explain how to measure the similarity 'w_{ij}' of a weight matrix *W* and how to set the value for label 'y'.

A. Similarity Matrix

The design of the similarity matrix W plays a critical part in the aspect of performance when using SSL[16, 18]. In the matrix W, each element represents how strongly the two nodes are related, with larger elemental value being associated with greater nodal similarity. In the proposed method, the time-series data are transformed by building technical indicators (TIs). The general process of constructing the similarity matrix is described in Fig. 2.

TIs are frequently used in financial forecasting as they offer the advantages of removing the noise (oscillatory noise) inherent in time series and illustrating the underlying structure, i.e., the tendencies and structural factors affecting variation[5-6, 19]. Stock prices and other economic indices exist as time series data by the nature of the variables, and each of them is defined as a sequence as

$$X_t = \{x_1, x_2, \dots, x_i, \dots, x_t\}, \tag{4}$$

where t represents the current time point, and x_t is the corresponding value. The existence of X_t as time series data induces several problems in the direct application of SSL to the data. As shown in Fig. 1, each of the nodes on the graph has its own time series, as shown in (4). For instance, the Hyundai Motors node has $X_t^{\text{Hyundai Motors}}$ and the LG chem node also has $X_t^{\text{LG chem}}$. The problem is that it is difficult to draw the similarity between them directly from the two sets of series data. Therefore, individual time series are transformed into structural characteristics of time point t, i.e., $S_t^{Hyundai\ Motors}$ and $S_t^{LG\ chem}$, representing the tendencies and factors for variation of individual series. Table I summarizes the TIs used in this study. The similarity between the two nodes is measured by using the seven-tuple vector $S_t = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7\}$ composed of MA, BIAS, OSC, ROC, K, D, and RSI.

Using the TIs enables the time series data to be transformed into TIs-type data, while maintaining the time associations of the series, and thus eases their application to SSL.

TABLE I THE DEFINITION OF TECHNICAL INDICATORS (TIS)

	TIs	Meaning
s_1	$\mathrm{MA_p}(\mathrm{X_t}) = \frac{1}{p}(\mathrm{x_t}) + \frac{p-1}{p}\mathrm{MA_p}(\mathrm{X_{t-1}})$	p-moving average (exponential smoothing)
s_2	$\mathrm{BIAS_p}(\mathrm{X_t}) = \frac{\mathrm{x_t} - \mathrm{MA_p}(\mathrm{X_t})}{\mathrm{MA_p}(\mathrm{X_t})}$	The change rate of $\mathrm{x_t}$ relative to $\mathrm{MA_p}(\mathrm{X_t})$
s_3	$\mathrm{OSC_{p,q}}(\mathrm{X_t}) = \frac{\mathrm{MA_p}(\mathrm{X_t}) - \mathrm{MA_q}(\mathrm{X_t})}{\mathrm{MA_p}(\mathrm{X_t})}$	The change rate of $\mathrm{MA_q}(\mathrm{X_t})$ relative to $\mathrm{MA_p}(\mathrm{X_t})$
s_4	$\mathrm{ROC_p}(\mathrm{X_t}) = \frac{\mathrm{x_t} - \mathrm{x_{t-p}}}{\mathrm{x_t}}$	The relative rate of change for $\mathrm{X_t}$ between p consecutive time points
s_5	$\mathrm{K_t^p} = \frac{\mathrm{x_t} - \mathrm{Min}_{i=t-p-1}^{t}(\mathrm{x_i})}{\mathrm{Max}_{i=t-p-1}^{t}(\mathrm{x_i}) - \mathrm{Min}_{i=t-p-t}^{t}(\mathrm{x_i})}$	Standardization of $\mathrm{x_t}$
s_6	$\mathrm{D_t^p} = \mathrm{MA_3}(\mathrm{K_t^p})$	3- Moving Average of $\mathrm{K_t^p}$
s_7	$\mathrm{RSI_t^p} = \frac{\sum_{i=t-p-1}^{t}(\lvert x_i - x_{i-1}\rvert)}{\sum_{i=t-p-t}^{t}(\lvert x_i - x_{i-1}\rvert)},$	The relative strength index.

B. Label

The label on the node in the SSL graph in Fig. 1 is designed to explain whether the predicted value of the corresponding variable is thumbs-up or down. It can be formulated as follows:

$$\mathrm{y_t} = \mathrm{sign}(\mathrm{x_t} - \mathrm{MA_5}(\mathrm{x_t}))\,. \tag{5}$$

For instance, if the total amount of the Hyundai Motors' stock price (t) exceeds the 5-days moving average, (5) will give a '$\mathrm{y_t} = +1$' label. On the contrary, the node is labeled as '$\mathrm{y_t} = -1$' for the opposite case. And '$\mathrm{y_t} = 0$' if there is no information about the movement of the corresponding time series at time point t, the label is to be predicted. In the proposed method, we set the label of the target variable to '0'. Given label $\mathrm{y_t}$, equation (3) provides the predicted value $\mathrm{f_t}$ for every node, which can take on a real number unlike the values of label $\mathrm{y_t}$.

If '$\mathrm{f_t} > 0$', it means the stock price will increase relative to the average of the MA(5), therefore one can take the position of "buy" for the stock. On the other hand, one can take the position of "sell" otherwise. This procedure is described in Fig. 3.

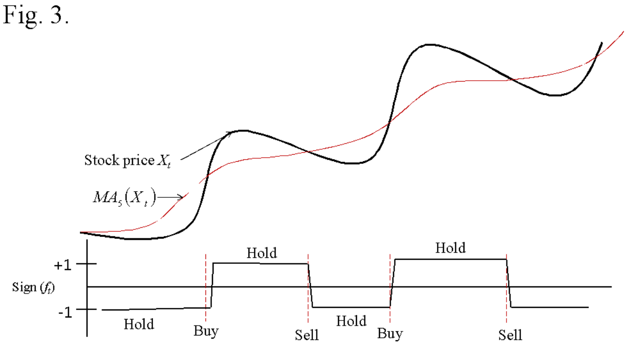

Fig. 3 Interpretation for forecasted values and simple trading strategy

IV. EXPERIMENT

A. Data

The data used in this experiment was presented by a total of 403 daily data points from January 2007 to August 2008. The factors employed as variables were the major global economic indexes, such as Dow-Jones average (DOW), National association of securities dealers automated quotations (NASDAQ), Japanese stock market index (NIKKEI), Hang seng index (HSI), Shanghai composite index (SSE), Taiwan stock exchange corporation (TSEC), Financial times security exchange (FTSE), Deutscher aktien index (DAX), continuous assisted quotation index (CAC), Bombay stock exchange portmanteau of sensitive and index (BSE_SENSEX), Indice bovespa (IBOVESPA), Australia all ordinaries index (AORD), Korea composite stock price index (KOSPI), exchange rate(KRW-USD), the west texas intermediate oil price (WTI), and the certificate of deposit (CD). Also, the stock prices of 200 companies listed to

KOSPI200 were included. Table 1-(Appendix) shows the list of these 200 companies.

B. Experimental Setting

The SSL model proposed in this experiment was compared with the ANN and SVM models. The ANN model used a multilayer perception function. The SVM model used an RBF kernel function that has been known as an excellent performance model relatively. A total of 103 daily data from January 2007 to May 2007 were determined as a training and validation period and the performance for a total of 300 daily data from June 2007 to August 2008 was compared. SSL was predicted using a rolling forecast method [20]. The rolling forecast method predicts a point of t+1 using the data from a point of 1 to a point of t and applies a learning data period from a point of 2 to the point of t+1 for the prediction of a point of t+2. In general, the ANN and SVM models represent higher performance as its learning data periods are highly determined. However, the learning data is very insufficient as the rolling forecast is applied. Thus, as shown in Fig. 4, the learning data periods employed in these models were gradually increased before the prediction point.

The parameters that are to be determined to the SSL model are k and μ and these parameters represent the number of neighbor node presented in Eq. (1) and the loss-smoothness tradeoff presented in Eq. (2), respectively. Also, the parameter values used in this experiment were determined as an optimal combination for the validation set presented in the range of $\{k, \mu\} \in \{2,3,4,5\} \times \{0.01,0.1,0.3,0.5,0.7,1,10,100\}$. In addition, the optimal value of the hidden node in ANN was determined in the range of {3~50} [5] and the parameters of kernel width (gamma) and misclassification tradeoff (C) in SVM were determined as an optimal combination of the values in the range of $\{\text{gamma}, C\} \in \{0.01,0.1,0.3,0.5,0.7,1,10,100\} \times \{0.01,0.1,0.3,0.5,0.7,1,10,100\}$[21].

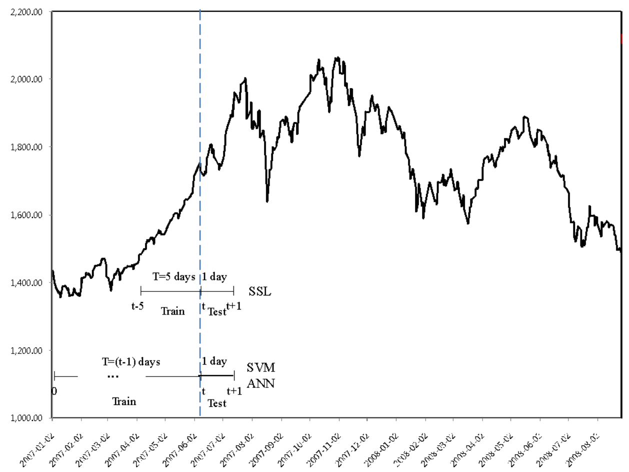

Fig. 4 Experimental setting

V. Results

To measure the prediction performance, the area under the curve (AUC), which is defined as the area under the receiver operating characteristic (ROC) curve [22-23] is used. The ROC curve plots true positive rate as a function of false positive rate for differing classification thresholds as shown in Fig. 5. The AUC measures the overall quality of the model for all possible values of threshold rather than the quality at a single value of threshold. The closer the curve follows the left-hand border and then the top-border of the ROC space, the larger value of AUC the model produces; i.e., the more accurate the model is.

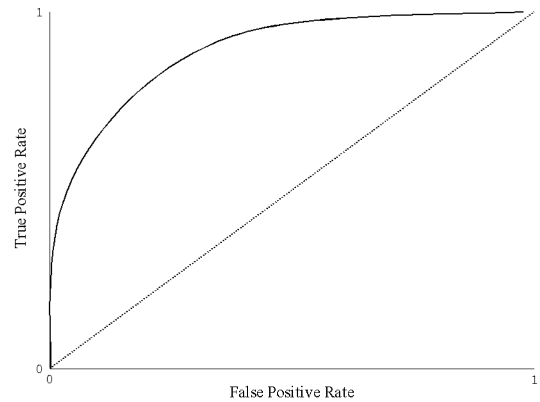

Fig.5 ROC curve

Fig. 6 shows the graph of the values of AUC for the three models used in the test period. Points presented in the graph represent the average section values of AUC in which a section has 10 time points. The average AUC values in SVM and ANN for the total 30 sections were 0.58(±0.08) and 0.51(±0.01), respective, but the value in SSL was 0.72(±0.05). Although ANN represented low volatility based on the standard deviation of 0.01, the AUC value was small compared to other models. In the case of SVM, although it showed partially higher AUC values than SSL, it represented a very high deviation in its accuracy. Whereas, SSL showed stable and high accuracy in most sections compared to that of ANN and SVM. A t-test was applied to verify the significance that SSL represents better performance statistically than that of SVM and ANN. As a result, the difference in the performance between them showed statistical significance as shown in the upper right box in Fig. 6.

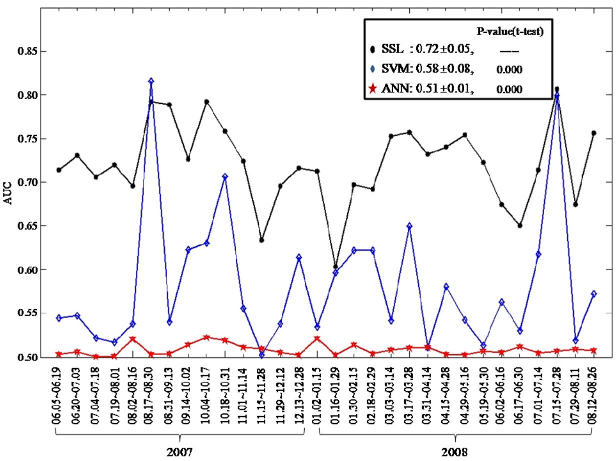

Fig. 6 AUCs comparisons with different methods

VI. Conclusions

In this study, a stock prediction method using time series data to SSL was proposed. The proposed method has the advantage that does not predict stock prices by considering the time series characteristics of the stock price in businesses like the conventional models but makes possible to predict the stock price using a network based on the fluctuation in other companies' stock prices and the economic index that affect the change in stock prices. Regarding the technical issue in the proposed method, the method used SSL and that leads to improve its predictability by including not only the influences on input variables and target variables but also the interrelation between input variables. Based on the combination of these advantages, it was possible to obtain the values of AUC as 0.72. Furthermore, the method proposed in this study can apply for predicting the fluctuation in stock prices for various stock items. Therefore, it is possible to expect profits and stabilities in investments as the results obtained in this study are combined with a portfolio optimization method.

Appendix

Table 1 200 listed stock in KOSPI.

Foods & Beverages	Samyang Corporation, Hite Brew, Doosan Corporation, CJ Corp, Daehan Flour Mills Co, Daesang Corporation, Orion Corporation, Lotte Samkang Co, Namyang Dairy Product Co, Samyang Genex, Nong Shim Co, Lotte Confectionery Co, Bing-grae Co, Lotte Chilsung Beverage Co, Ottogi, Crown Confectionary, Dongwon F&B
Textile & Apparel	Kyungbang Co, FnC Kolon Corp, Nasan, Handsome, The Basic House, LG Fashion Corporation
Paper & Wood	Hankuk paper Mfg, Hansol Paper Co, Seha, Moorim paper.
Chemicals	Woongjin Chemical, Hankook Tire Co, Hanwha Co, Cheil Industries Inc, Kokon, Nexen Tire Co, KCC, Tae Kwang Industrial Co, Samsung Fine Chemicals Co, Hyosung SK Chemicals Co Capro Korea Petro Chemical Aekyung Petrochemical Youl Chon Chemical Hanwha Chemical OCI S-Oil Honam Petrochemical Korea Kumho Pertrochemical SKC UNID KPX Chemical, KPX Chemical Namhae Chemical, LG Household & Health Care, LG Chem, KP Chemical Corporation, Huchems Fine Chemical Corporation, Kumho Tires, Amore Pacific, Foosung
Medical Supplies	Yuhan Corporation, Ildong Pharmaceutical Co, Dong A Pharmaceutical Co, JW Pharmaceutical, Chong Kun Dang Pharmaceutical, Bukwang Pharm, Ilsung Pharmaceuticals Co, Yungjin Pharm, Dong Wha Pharm, Green Cross, Il-Yang Pharm, Hanmi Pharmaceutical, Kwang Dong Pharm, LG Life Sciences Ltd, Daewoong Pharm
Non-metalic Minerals	Chosun Refractories Co, Dongyang Mechatronics, Hanglas, Asia Cement Co, Hanil Cement Co, Ssangyong Cement Industrial Co, Sung Shin Cement Co, Samkwang Glass Ind Co, Hyundai Cement Co, Hankuk Glass Industries
Iron & Metals	Young Poong Co, Dong Kuk Steel Mill Co, Se Ah Be steel Co, Kisco, Kiswire, SeAh Steel, Union Steel, Hyundai-Steel Co, BNG Steel Co, Posco, Poongsan, Korea Zinc, Hyundai Hysco, Dongbu Steel, Daehan Steel
Machinery	Dongbu Hannong Co, KC Cottrell, Shinsung ENG, DKME, Hyundai Elevator, Hankuk carbon, Halla Climate Control co, Doosan Heavy Industries & Construction, Doosaninfracore, Hanmi Semiconductor, STX Engine, Sewon Cellontech, S&TC
Electrical & Electronic Equipment	Hynix Semiconductor Inc, Kumho Electric, Taihan, Daeduck GDS Co, Hansol Lcd Inc, Samyoung Electronics Co, Samsung Electronics, LS Industrial System, Samsung SDI, Daeduck Electronics, Korea Technology Industry, Samsung Engineering, LS, Celrun, Dongwon Systems, Iljin Electric, Korea electric terminal co, Sindoh, LG Display, Hyundai Autonet, LG Electronics
Medicalprecisio	Samsung Techwin, K.C. Tech
Transport Equipment	Hyundai Motors, KIA Motors, Hanjin Heavy Industries, S&T Dynamics, Ssangyong Motor, Hyundai Heavy Industries, Samsung Heavy Industries, Hyundai Mipo Dockyard, Myungsung, Hyundai Mobis, Dongyang Mechatronics, Daewoo Shipbuilding & Marine Engineering, S&T Daewoo STX Offshore & Shipbuilding
Other manufacturing	Fursys, KT&G
Electrical & gas	Kepco, Kogas
Construction	Daelim Industrial Co, Hyundai Engineering & Construction Co, Kumho Industrial Co, GS Engineering & Construction, Hyundai Development, Daewoo E&C
Distribution & Service	Samsung C&T, LG International, SK Networks Co, Amorepacific, LG, SK, Shinsegae Co, STX, S1, Dae Kyo, Coway, Lotte Shopping, Samsung Engineering, Cheil Worldwide Inc, SBS, Kangwon Land, NCsoft, Daewoo International, Hyundai Department Store, GS Holdings
Transport & Storage	Hanjin Shipping Co, Korean Air Lines, Hyundai Merchant Marine
Communication	SK Telecom, KT, KTF
Finance	Samsung Fire & Marine Insurance, Hyundai Securities, Korean Reinsurance, Daegu Bank, Busan Bank, Woori Investment & Securities, Daewoo Securities, Samsung Securities, Industrial Bank of Korea, Mirae Asset Securities, Woori Finance Group, Shinhan Financial Group, Korea Investment Holdings, Hana Financial Group , KB

Acknowledgment

The authors would like to gratefully acknowledge support from Post Brain Korea 21 and the research grant from National Research Foundation of Korean government (KRF-2010-0007804/2012-0000994).

References

[1] C.-J. Huang, D.-X. Yang, and Y.-T. Chuang, "Application of wrapper approach and composite classifier to the stock trend prediction," *Expert Systems with Applications,* vol. 34, pp. 2870-2878, 2008.

[2] H.-C. Liu, Y.-H. Lee, and M.-C. Lee, "Forecasting China Stock Markets Volatility via GARCH Models Under Skewed-GED Distribution," *Journal of Money, Investment and Banking,* pp. 5-14, 2009.

[3] H. Amilon, "GARCH estimation and discrete stock prices: an application to low-priced Australian stocks " *Economics Letters,* vol. 81, pp. 215-222, 2003.

[4] N.-F. Chen, R. Roll, and S. A. Ross, "Economic Forces and the Stock Market," *Journal of Business,* vol. 59, pp. 383-403, 1986.

[5] K.-j. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing,* vol. 55, pp. 307-319, 2003.

[6] K. Park, T. Hou, and H. Shin, "Oil Price Forecasting Based on Machine Learning Techniques," *Journal of the Korean Institute of Industrial Engineers,* vol. 37, pp. 64-73, 2011.

[7] W. Huang, Y. Nakamori, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," *Computers & Operations Research,* vol. 32, pp. 2513-2522, 2005.

[8] Q. Cao, K. B. Leggio, and M. J. Schniederjans, "A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market," *Computers & Operations Research,* vol. 32, pp. 2499-2512, 2005.

[9] A.-S. Chen, M. T. Leung, and H. Daouk, "Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index," *Computers & Operations Research,* vol. 30, pp. 901-923, 2003.

[10] F. E. H. Tay and L. Cao, "Application of support vector machines in financial time series forecasting " *The International Journal of Management Science,* vol. 29, pp. 309-317, 2001.

[11] A. KANAS, "Non-linear Forecasts of Stock Returns," *Journal of Forecasting,* vol. 22, pp. 299-315, 2003.

[12] B. Yang, L. X. Li, and J. Xu, "An early warning system for loan risk assessment using artificial neural networks " *Knowledge-Based Systems,* vol. 14, pp. 303-306, 2001.

[13] S. Bekiros and D. Georgoutsos, "Direction-of-Change Forecasting using a Volatility- Based Recurrent Neural Network," *Journal of Forecasting,* vol. 27, pp. 407-417, 2008.

[14] P. M. Tsang, P. Kwok, S. O. Choy, R. Kwan, S. C. Ng, J. Mak, J. Tsang, K. Koong, and T.-L. Wong, "Design and implementation of NN5 for Hong Kong stock price forecasting," *Engineering Applications of Artificial Intelligence*, vol. 20, pp. 453-461, 2007.

[15] X. Zhu, "Semi-Supervised Learning with Graphs," *Ph.D. dissertation, Carnegie Mellon University,* 2005.

[16] H. Shin, N. J. Hill, A. M. Lisewski, and J.-S. Park, "Graph sharpening," *Expert Systems with Applications,* vol. 37, pp. 7870-7879, 2010.

[17] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Sch¨olkopf, "Learning with local and global consistency " *Advances in Neural Information Processing Systems* vol. 16, pp. 321-328, 2004.

[18] H. Shin, A. M. Lisewski, and O. Lichtarge, "Graph sharpening plus graph integration: a synergy that improves protein functional classification," *Bioinformatics,* vol. 23, pp. 3217-3224, 2007.

[19] K.-j. Kim, "Artificial neural networks with evolutionary instance selection for financial forecasting," *Expert Systems with Applications,* vol. 30, pp. 519-526, 2006.

[20] M. O'Connor, W. Remus, and K. Griggs, "Does updating judgmental forecasts improve forecast accuracy?," *International Journal of Forecasting,* vol. 16, pp. 101-109, 2000.

[21] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery,* vol. 2, pp. 121-167, 1998.

[22] J. A. Hanley and B. J. McNeil, "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve," *Radiology,* vol. 143, pp. 29-36, 1982.

[23] M. Gribskov and N. L. Robinson, "Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching " *Computers & Chemistry,* vol. 20, pp. 25-33, 1996.