Stock Prediction Using Multiple Time Series of Stock Prices and News Articles

Daigo Kato and Tomoharu Nagao
Graduate School of Environment
and Information Sciences
Yokohama National University
Yokohama, Japan
Email: daigo@nlab.sogo1.ynu.ac.jp / nagao@ynu.ac.jp

***Abstract*—In the stock market, stock prices of multiple companies interact with each other. For instance, a stock price movement of a company triggers that of another one. Therefore, investors are interested in inter-relationship of multiple companies whose stock prices interact with each other. In recent years, a number of studies are conducted to predict stock price movements in the area of artificial intelligence. Most of them focus on stock prediction but not on explaining the reason why they succeed in stock prediction. In this study, we propose a method to find out a rule that predicts the stock price movement of a target company. We use rate of change of multiple companies' stock prices and a newspaper article about a company in the rule. We explain the reason why the rule succeeds in its prediction by analyzing inter-relationship of these companies and the target company by the use of newspaper articles and stock prices related to Companies Co-occurrence Map. This method is applied to the first section of the Tokyo Stock Exchange and encouraging results are obtained.**

I. Introduction

In the stock market, stock prices of multiple companies interact with each other. For instance, a stock price movement of a company triggers that of another one. Therefore, investors are interested in inter-relationship of multiple companies whose stock prices interact with each other.

In recent years, a number of studies are conducted to predict stock price movements in the area of artificial intelligence [1], [2]. Most of them focus on stock prediction but not on explaining the reason why they succeeded in stock prediction. In the real market, it is difficult for investors to use these prediction methods based on artificial intelligence because they do not know the reason why they succeeded in stock prediction by the use of them.

News articles are used to predict stock price movements in some studies [3]–[9]. Investors predict stock price movements with a lot of informations about these stocks. News articles are one of important informations. The response to the articles varies according to stock market conditions at the times. Therefore, there are some difficult cases to predict stock price movements with only news articles.

In other studies, news and multiple time series of stock prices are used to predict stock price movements. However, inter-relationship of each stock is not taken into account.

In our study, we have two purpose. One is to predict the stock price movement of a target company by the use of multiple companies whose stock prices or newspaper articles influence the target stock price. Another is to explain the reason why the rule succeeds in its prediction. Stock price movements are a consequence of investors' actions that are triggered by informations about the stocks such as news articles. In other words, such informations influence the stock price movements. Then, these informations (the news articles and the stock price movements) influence another stock price. In this study, we propose a method to find out a rule that predicts the stock price movement of a target company. We use rate of change of multiple companies' stock prices and a newspaper article about a company in the rule. We explain the reason why the rule succeeds in its prediction by analyzing inter-relationship of these companies and the target company. Newspaper articles and stock prices related to *Companies Co-occurrence Map* is used for the analysis.

II. The Proposed Method

In the study of finance, Fama proposed an efficient market hypothesis [10]. In the efficient market hypothesis, informations are quickly reflected in an asset price and the asset price becomes equal to the primary price constantly. Actually, however, it is confirmed that the asset price is often different from the primary price. Moreover, it is thought that the asset price moves toward the primary price steadily because it takes some the asset price to factor in informations.

In light of this fact, Fig.1 shows the outline of our concept. We have an assumption that a news article about the company B at the date n influences rate of change of other stock prices (the company C and D) immediately. Then we also have an assumption that these informations (the news article and the stock price movements) influence the stock price of the target company A from the next day $n+1$ onward. Therefore, we predict the stock price movement of the company A from the next day $n+1$ onward.

We propose a method to find out a rule that predicts the stock price movement of a target company based on this concept. We use rate of change of multiple stock prices and articles about one company in the rule. The article influences the multiple stock prices immediately. Then, these informations influence the target company's stock price. This idea is the same as this concept.

978-1-4673-1686-6/12/$26.00 ©2012 IEEE

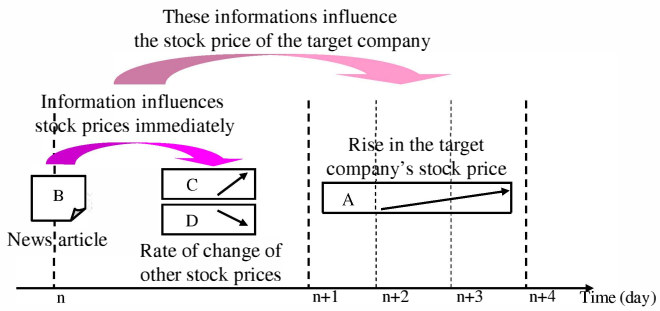


Fig. 1. Outline of our concept.

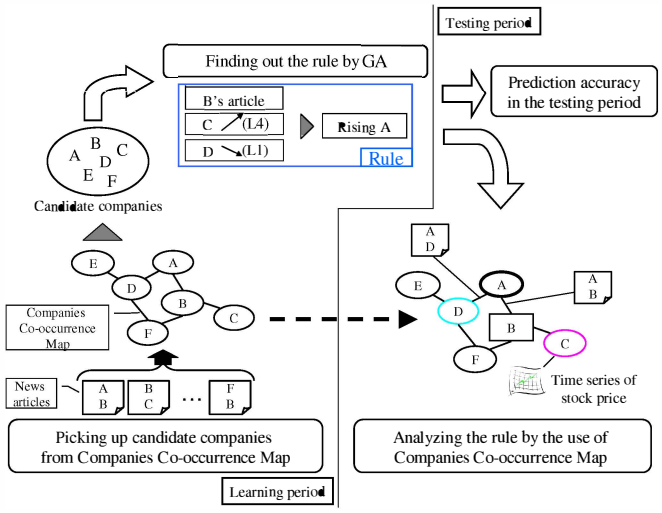


Fig. 2. Flow of our proposed method.

Fig.2 shows the flow of our proposed method. In the learning period, first, we pick up candidate companies. The candidate companies consist of all companies that are included in Companies Co-occurrence Map. Next, we use Genetic Algorithm (GA) [11] to find out a rule. We select the companies whose rate of change of stock prices are used in the rule and the company whose article is used in the rule from the candidate companies. After finding out the rule, prediction accuracy of the rule is measured in the testing period. In addition, we explain the reason why the rule succeeds in its prediction by the use of Companies Co-occurrence Map.

At first, we explain Companies Co-occurrence Map, then explain the way of finding out a rule by GA. Finally, we explain the starting point and the ending point to predict by the found rule.

A. Companies Co-occurrence Map

A node and an edge in Companies Co-occurrence Map indicate a company name and a connection between company names, respectively.

Fig.3 shows the overview of the generating method of Companies Co-occurrence Map. First, we pick out pairs of two company names that are published in the same articles

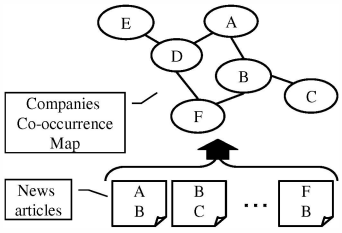


Fig. 3. Overview of the generating method of Companies Co-occurrence Map.

more than a certain times (e.g. more than 5 business days) in the learning period. Then, each company name is represented by a node and company names in each pair are connected by an edge. In doing so, Companies Co-occurrence Map is generated.

We use Companies Co-occurrence Map when we find out a rule. A rule includes company names. The company names are selected from nodes that are included in Companies Co-occurrence Map. We also use Companies Co-occurrence Map when we analyze the reason why the rule succeeds in its prediction. We visually confirm the relationship with companies whose name is included in the rule by the use of nodes and edges included in Companies Co-occurrence Map. Moreover, we confirm the relationship with these companies closely by the use of news articles about these companies and these companies' stock prices. In doing so, we explain the reason why the rule succeeds in its prediction.

B. Way of finding out a rule by Genetic Algorithm

We explain the way of finding out a rule that predicts the stock price movement of a target company.

Fig.4 shows the framework of a rule. A rule consists of *Condition part* and *Prediction part*. Condition part consists of *Article part* and *Change rate part*. Article part has a company name whose article the rule takes notice. Change rate part has names of companies whose rate of change of stock prices the rule takes notice and the levels (level 0 to level 4, $L0$ to $L4$, as hereafter defined) for the rate of change of these stock prices. Prediction part has a target company name and the direction of its stock price movement.

Fig.5 shows the overview of the way of finding out a rule by GA and Fig.6 shows GA's chromosome. In the learning period, first, a target company (e.g. the company A) is chosen from the candidate companies. The candidate companies consist of all companies that are included in Companies Co-occurrence Map. Next, the rule to predict the rise in the target company's stock price is found out. We finds out Condition part of the rule by GA. Concretely speaking, GA selects a company name in Article part and company names in Change rate part and the levels ($L0$ to $L4$) for the rate of change of these stock prices. So does the rule to predict the drop in the target company's stock price.

In a similar way, we find out other rules of all companies in the candidate companies. Finally, we use rules that meet a

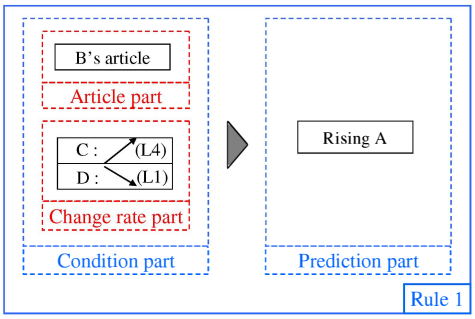


Fig. 4. Framework of a rule (Rule 1).

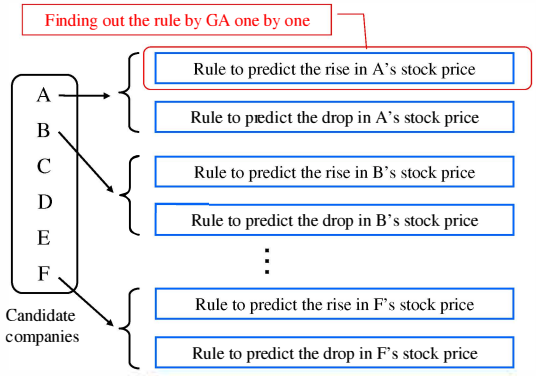


Fig. 5. Way to find out a rule by GA.

certain condition among these found rules in the testing period.

C. Starting point and ending point to predict by the found rule

We explain the starting and the ending point to predict by the found rule. We take *Rule 1* in Fig.4 as an example. Fig.7 shows the overview. Our proposed method has an assumption that the information about the rule's Condition part influences the stock price of the target company.

Therefore, we define the closing price of the day when the present conditions corresponds with Condition part as the starting point to predict. Condition part in Fig.7 indicates that the news article about the company B is published at date n, the company C's stock price movement and the company D's stock price movement are rising ($L4$) and slightly dropping ($L1$) at the same day n, respectively.

Then, we define the point of the day when the information about Condition part finishes influencing as the ending point to predict. Concretely speaking, we define the following three points as the ending points to predict. First, we define the closing price of the day when the present conditions corresponds with Condition part again as the ending point 1. Second, we define the opening price of the day when the target company's article is published as the ending point 2. Third, we define the previous closing of the day when straight-line approximation's error exceeds the threshold as the ending point 3.

The reason why these three points are ending points is that a condition of the target company might change from these

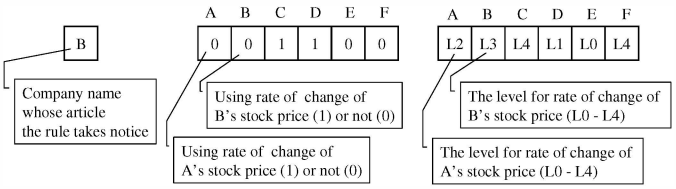


Fig. 6. GA's chromosome.

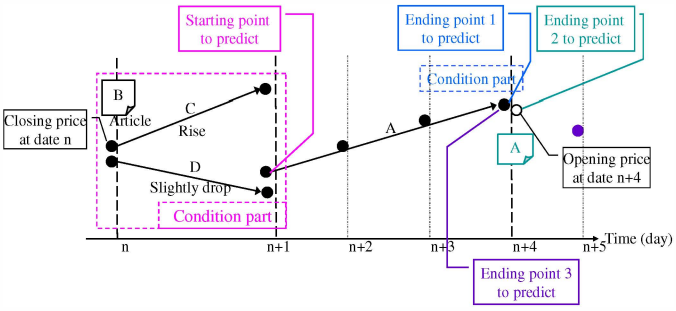


Fig. 7. Starting point and the ending point to predict by Rule 1.

points because the events at these points might influence the stock price of the target company.

In Fig.7, the company A's stock price movement is rising from the starting point to the ending point. Therefore, Rule 1 succeeds in its prediction in this case.

III. Evaluation

We applied our proposed method to the market simulation to evaluate the accuracy of the predictions and analyzed the found rule to explain the reason why the rule succeeds in its prediction.

A. Experiment setup

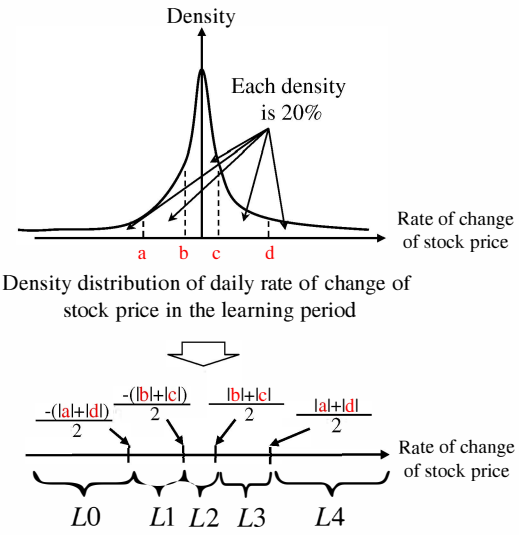
The stock prices and news articles are archived from the first section of the Tokyo Stock Exchange and Keizai Shimbun during 1/1/2005 and 12/31/2006, respectively. The learning period is 80 business days and the testing period is 40 business days. After the testing period, each period is moved forward to 40 business days. Moreover, the learning and the testing are conducted again. Therefore, we have 10 learning and testing periods independently.

When generating Companies Co-occurrence Map, the minimum co-occurrence business days more than that each pair of two company names co-occurrences in the same article is 5 business days in this study. Table I shows the number of companies included in Companies Co-occurrence Map at each period. The error threshold for the straight-line approximation is set to maintain the duration of each approximated line has the mean around 3.

Rate of change of stock price has 5 levels, $L0$ to $L4$. Each level is decided as Fig.8. First, we calculate density distribution of daily rate of change of stock price in the learning period. Next, we identify absolute values of rate of change at the top 20 percent $|d|$, at the top 40 percent $|c|$, at

TABLE I
NUMBER OF COMPANIES SELECTED FROM COMPANIES CO-OCCURRENCE MAP.

Period 1 59 companies	Period 2 52 companies	Period 3 27 companies	Period 4 37 companies	Period 5 53 companies
Period 6 57 companies	Period 7 52 companies	Period 8 26 companies	Period 9 26 companies	Period 10 49 companies



Density distribution of daily rate of change of stock price in the learning period

Fig. 8. Levels for rate of change of stock price.

the top 60 percent $|b|$ and at the top 80 percent $|a|$. Then, level 0 $L0$ to level 4 $L4$ are calculated as:

$$L0 \leq -(|a| + |d|)/2, \tag{1}$$
$$-(|a| + |d|)/2 < L1 \leq -(|b| + |c|)/2, \tag{2}$$
$$-(|b| + |c|)/2 < L2 < (|b| + |c|)/2, \tag{3}$$
$$(|b| + |c|)/2 \leq L3 < (|a| + |d|)/2, \tag{4}$$
$$(|a| + |d|)/2 \leq L4. \tag{5}$$

The parameters of GA are shown in Table II.

B. Prediction accuracy

The prediction accuracy is obtained by checking whether the direction the rule predicts is the same as the actual stock direction of the target company. For instance, if the rule that predicts the direction will rise and the actual stock price is really rising, the prediction is correct. Otherwise, if the rule predicts the direction will rise and the actual stock price is dropping, the prediction is wrong.

In the testing period, we use rules that are found out in the learning period and meet the following two conditions. First, the rule has to predict the stock price movement of the target company at least 12 times in the learning period. Second, we formulate a sign test:

$$H_0 : p = p_0 \tag{6}$$
$$H_1 : p = p_1 \tag{7}$$

TABLE II
PARAMETERS OF GA.

Generation change model	MGG [12]
Crossover model	Uniform crossover
Number of generations	50000
Number of individuals	100
Number of children	40
Number of tournaments	2
Crossover rate	1.0
Mutation rate	0.03

TABLE III
PREDICTION ACCURACY AND THE PROFIT RATE.

Number of rules using in the testing period	20 rules
Number of transaction by all rules	98 times
Prediction accuracy by all transaction	59.05%
Profit rate by all transaction	23.15%
Profit rate by buy and hold	74.23%

where p is the binomial parameter and p_0 is the target value.

p represents the success rate of prediction. p_0 is set to 0.01. If the null hypothesis is rejected, we use the rule in the testing period.

In the real transaction, we can not buy or sell a stock for the previous closing price in case of the ending point 3 described at the previous section. Therefore, we assume that we buy or sell the stock for the present closing price in case of it in the real transaction. In this study, we focus on analyzing the relationship implied in the found rule. Moreover, the prediction accuracy changes according to a transaction strategy. Therefore, we do not care about the real transaction.

In addition to the prediction accuracy, we calculate the profit rate. The profit rate r per prediction is calculate as below:

$$r = \frac{y_j - y_i}{y_i} \tag{8}$$

where y_i is the target company's stock price at the starting point and y_j is the target company's stock price at the ending point. We also do not care about the profit rate because the profit rate changes according to a transaction strategy.

We executed the market simulation 4 times, and the average of the 4 times is shown in Table III and Table IV. The number of rules and the number of transaction in Table III are the summation of the 10 testing periods. The prediction accuracy and the profit rate in Table III and Table IV are calculated by the all transaction of the all rules used in the 10 testing periods. As you can see in Table III, the found rules gained high prediction accuracy.

Fig.9 shows the frequent distribution of the number of transaction and the prediction accuracy in the testing period achieved by the 4 executions. As you can see, the more the number of transaction are, the better the prediction accuracy are. With these factor, we think that our proposed method gained efficient rules to predict.

TABLE IV
PREDICTION ACCURACY AND THE PROFIT RATE IN THE REAL TRANSACTION.

Prediction accuracy by all transaction	56.79%
Profit rate by all transaction	39.53%

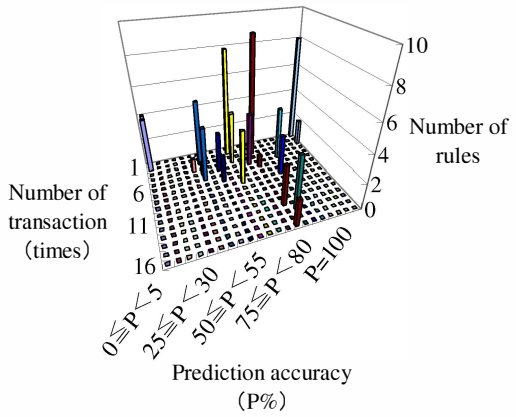


Fig. 9. Frequent distribution of the number of transaction and the prediction accuracy in the testing period (the same condition as the learning period).

C. Analyzing the found rule

We attempt to analyze the found rule to explain the reason why the rule succeeds in its prediction. Fig.10 shows the rule named *Rule A* found out at the period 1. The result of Rule A is shown in Table V and Table VI. Fig.11 shows the part of Companies Co-occurrence Map related to Rule A. The values on the edges in Fig.11 are the number of articles that two company names at the ends of edges are published in. Fig.12 shows the summaries of the part of news articles related to companies in the map of Fig.11 in the learning period. In addition, Fig.13 shows the time series of the target company's stock price, rate of change of other companies' stock prices in the map of Fig.11 and the present conditions' correspondence with Rule A's Condition part or not in the part of learning period.

As you can see the map in Fig.11 and the articles in Fig.12, Mitsubishi Group companies have a strong relationship with each other because of new rebuild plan of Mitsubishi Motors and so on. Hitachi Group companies also have a relationship with each other because of working together for beefing up the sale of IP phone. Hitachi and Mitsubishi Electric Corporation have a relationship with each other because of establishing co-parent company. Hitachi and Mitsubishi Heavy Industries also have a relationship with each other because of aiming at establishing co-parent company. Mitsubishi Corporation and Isuzu Motors have a relationship with each other because of Automobile business abroad and so on.

As you can see the blue frames (solid line) in Fig.13, many companies in the map of Fig.11 drop when the present conditions corresponds with Rule A's Condition part.

From here onward, Rule A is explained that "when the news

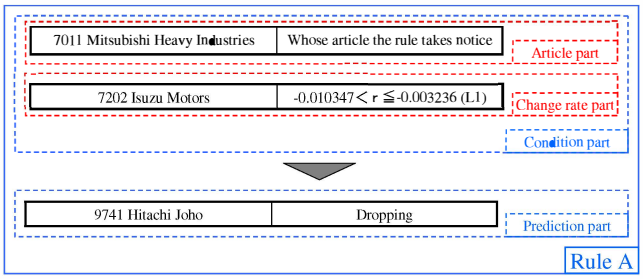


Fig. 10. Rule named Rule A found out at the period 1.

TABLE V
PREDICTION ACCURACY AND THE PROFIT RATE OF RULE A.

	Learning period	Testing period
Prediction accuracy	13/15	3/5
Profit rate	17.06%	2.87%
Profit rate by buy and hold	22.97%	3.12%

about Mitsubishi Heavy Industries that triggers the dropping of Mitsubishi Group companies' stock prices is published, this event influences Hitachi Group companies, and then the stock price of Hitachi Joho drops after making sure of the dropping of Isuzu Motors's stock price that has a relationship with Mitsubishi Corporation".

IV. CONCLUSION

In this study, we proposed the method to find out a rule that predicts the stock price movement of a target company. We used rate of change of multiple companies' stock prices and a newspaper article about a company in the rule. The found rule gained high prediction accuracy. Moreover, we explained the reason why the rule succeeded in its prediction by analyzing inter-relationship of these companies and the target company with newspaper articles and stock prices related to Companies Co-occurrence Map.

In this study, the learning period and the testing period are 80 business days and 40 business days, respectively. If changing the span of these periods, we might obtain a different rule that contains a companies's relationship different from the result of this study. In the previous section, we explained the reason why Rule A succeeded in its prediction. In that regard, many companies in the map of Fig.11 drop when the present conditions corresponds with Rule A's Condition part. Therefore in the future, we change Condition part so as to contain such companies in the rule. In doing so, the explanation becomes easier than that of the result in this study.

REFERENCES

[1] Takashi Kimoto and Kazuo Asakawa, Stock Market Prediction System with modular Neural Networks, *IEEE International Joint Conference on Neural Networks*, vol. 1, pp. 1-6 1990.

[2] Jung-Hua Wang and Jia-Yann Leu, Stock Market Trend Prediction Using ARIMA-based Neural Networks, *IEEE International Conference on Neural Networks*, vol. 4. pp. 2160-2165, 1996.

[3] Di WU, Gabriel Pui Cheong FUNG, Jeffrey Xu YU and Qi PAN, Stock prediction: an event-driven approach based on bursty keywords, *Frontiers of Computer Science in China*, 2009.

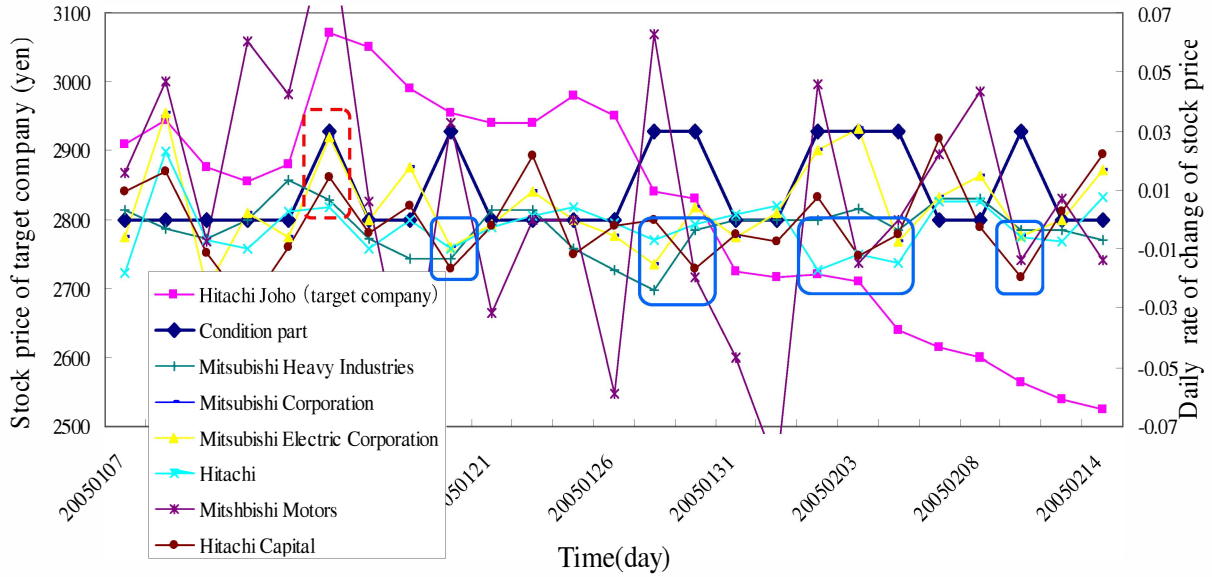


Fig. 13. Time series of the target company's stock price, rate of change of other companies' stock prices in the map of Fig.11 and the present conditions' correspondence with Rule A's Condition part or not in the part of learning period.

TABLE VI
PREDICTION ACCURACY AND THE PROFIT RATE OF RULE A IN THE REAL TRANSACTION.

Prediction accuracy	1/5
Profit rate	-8.50%

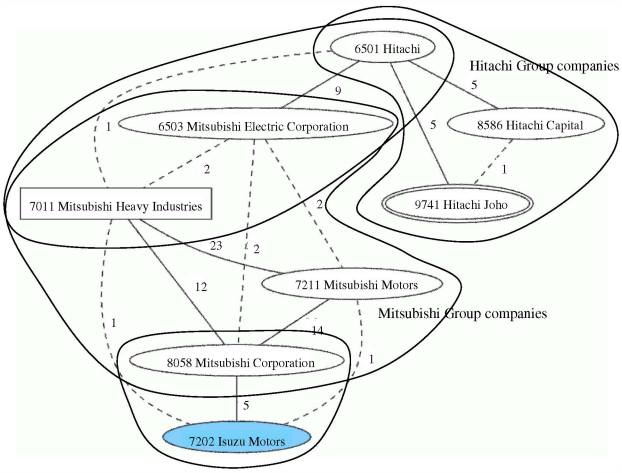


Fig. 11. Part of Companies Co-occurrence Map related to Rule A.

1/5/2005
Renesas Technology jointly established by Hitachi and Mitsubishi Electric Corporation beefed up new line.
3/30/2005
Hitachi and Mitsubishi Heavy Industries aimed at establishing co-parent company builds a business air-conditioner project.

4/7/2005
Hitachi Group companies worked together for beefing up the sale of IP phone.

Mitsubishi Group companies increased Mitsubishi Motors' stake in for new rebuild plan.
1/27/2005
Mitsubishi heavy Industries affiliated Mitsubishi Motors.
2/15/2005
Mitsubishi Electric Corporation beefed up auto engine component. One of deliveries is Mitsubishi Motors.

1/12/2005
Mitsubishi Corporation obtained preferential stock of Isuzu Motors.
2/4/2005
Automobile business abroad of Mitsubishi Corporation and Isuzu Motors.

Fig. 12. Summaries of the part of news articles related to companies in the map of Fig.11 in the learning period.

[4] Gabriel Pui Cheong Fung, Jeffrey Xu Yu and Hongjun Lu, The Predicting Power of Textual Information on Financial Markets, *IEEE Intelligent Informatics Bulletin*, 2005.

[5] Gabriel Pui Cheong Fung, Jeffrey Xu Yu and Wai Lam, News Sensitive Stock Trend Prediction, *Springer-Verlag Berlin Heidelberg*, 2002.

[6] Takatora Honda, Kiyoshi Izumi, Tohgoroh Matsui, Minoru Yoshida, Hiroshi Nakagawa, Tomonari Ishida, Akihiro Nakashima and Toshiharu Sugawara, Stock Fluctuation Forecasting Based on Text mining of Daily Sector News, *SIG-FIN-005-06*, 2010.

[7] Kiyoshi Izumi, Takashi Goto and Tohgoroh Matsui, Analysis of Financial markets' Fluctuation by Textual Information, *The 23rd Annual Conference of the Japanese Society for Artificial Intelligence*, 2009.

[8] Tomoya Ogawa and Isamu Watanabe, Mining of Stock Prices and News Articles, *Information processing Society of Japan*, 2001.

[9] Di WU, Gabriel Pui Cheong FUNG, Jeffrey Xu YU and Wai Lam, Stock Prediction: Integrating Text Mining Approach using Real-Time News, , *IEEE International Conference on Computational Intelligence for Financial Engineering*, pp. 395-402, 2003.

[10] Eugene Fama, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance*, vol. 25, pp. 383-417 1970.

[11] J. P. Holland, Adaptation in natural and artificial systems, *The Univ. Michigan Press*, 1975.

[12] I. O. H. Sato nad S. Kobayashi, A new generation alternation model of genetic algorithms and its assessment, *J. of Japanese Society for Artificial Intelligence*, vol. 12, no. 5, pp. 734-744, 1997(in Japanese).

[13] Eamonn J. Keogh, Selina Chu, David Hart, and Michael J. Pazzani, An online algorithm for segmenting time series, *Proceedings of 2001 IEEE International Conference on Data Mining*, 1996.