

# A Portfolio Selection Model using Genetic Relation Algorithm and Genetic Network Programming

Yan Chen

Graduate school of Information, Production and Systems  
Waseda University  
Kitakyushu, Fukuoka, Japan  
chenyan15153@akane.waseda.jp

Shingo Mabu

Graduate school of Information, Production and Systems  
Waseda University  
Kitakyushu, Fukuoka, Japan  
mabu@aoni.waseda.jp

Kotaro Hirasawa

Graduate school of Information, Production and Systems  
Waseda University  
Kitakyushu, Fukuoka, Japan  
hirasawa@waseda.jp

**Abstract**—In this paper, a new evolutionary method named genetic relation algorithm (GRA) has been proposed and applied to the portfolio selection problem. The number of brands in the stock market is generally very large, therefore, techniques for selecting the effective portfolio are likely to be of interest in the financial field. In order to pick up a fixed number of the most efficient portfolio, the proposed model considers the correlation coefficient between stocks as strength, which indicates the relationship between nodes in GRA. The algorithm evaluates the relationships between stock brands using a specific measure of strength and generates the optimal portfolio in the final generation. The efficiency of GRA method is confirmed by the stock trading model using genetic network programming (GNP) that has been proposed in the previous study. We present the experimental results obtained by GRA and compare them with those obtained by traditional method, and it is clarified that the proposed model can obtain much higher profits than the traditional one.

**Index Terms**—portfolio selection, genetic relation algorithm, genetic network programming

## I. INTRODUCTION

This paper presents an application of evolutionary computation method named genetic relation algorithm (GRA) to the problem of portfolio selection in the financial field. The conventional portfolio problem in the stock market consists of deciding what brands to include in a portfolio given the investor's objectives and economic conditions, in order to maximize the expected return and minimize risk simultaneously.

Harry Markowitz[1] first proposed a mean-variance optimization model to design an optimum portfolio as the foundation of portfolio selection. In the case of linear constraints, the problem can be solved efficiently by parametric quadratic programming. However, there are many real-world nonlinear constraints which limit the number of different assets in a portfolio. Since the number of brands in the stock market is generally very large, techniques for selecting the effective portfolio are likely to be of interest in the financial field. As a consequence, evolutionary computation was developed to

calculate the optimal portfolio while it makes the search space larger.

Recently, various approaches in the artificial intelligence (AI) field have been applied to several financial problems, especially for stock market activities. Dropsy[2] uses artificial neural networks (ANNs) as a nonlinear prediction tool to forecast international equity risk, in which both linear and nonlinear forecasting results outperform the random work. Oh[3] proposed a new portfolio selection algorithm based on portfolio beta by using genetic algorithm (GA). However, when GA was applied to the portfolio optimization, the problem is that many chromosomes are coded into the same portfolio, or similar chromosomes may be coded into very different portfolios which makes it more difficult for GA to produce better chromosomes from good ones.

Due to such kinds of bottlenecks, we propose the GRA method and apply it to the portfolio selection problem. In order to pick up a fixed number of the most efficient portfolio, the proposed model considers the correlation coefficient between stocks as strength, which indicates the relation between nodes in GRA. The algorithm evaluates the relationships between stock brands using a specific measure of strength and generates the optimal portfolio in the final generation. The efficiency of GRA method is confirmed by the stock trading model using genetic network programming (GNP) that has been proposed in the previous study[4]. Generally speaking, the contributions of our proposed method are as follows: First, the GRA method constructs a model considering the correlation coefficient as the strength between stock brands to optimize the portfolio. Second, the number of stock brands in the best portfolio in the final generation can be flexibly defined by users because the brands correspond to nodes in the GRA individuals.

The outline of this paper is as follows. Section 2 describes the proposed Genetic Relation Algorithm approach in general. In Section 3, we explain the application of Genetic Relation Algorithm to the portfolio selection model. Section 4 presents

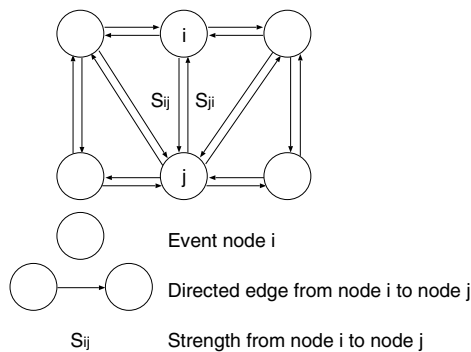
experimental environments, conditions and results using GRA method and conventional stock trading model of GNP. The trading profits are presented and compared with traditional method. Finally, Section 5 concludes this paper.

## II. GENETIC RELATION ALGORITHM

In this section, the outline of GRA is explained briefly. Basically, GRA is an extension of genetic programming (GP)[5] and genetic network programming[6] in terms of gene structures. The original idea is based on the more general representation ability of both directed and undirected graphs. As a new evolutionary computation, GRA is used for determining the best relations between events. There are two kinds of gene structures in GRA, i.e., GRA with directed and undirected edges.

### A. GRA with Directed Edges

Fig. 1 shows the basic structure and genotype expression of GRA with directed edges. GRA is composed of nodes and edges, where nodes represent events and directed edges represent the relations between nodes with their strength. As shown in Fig. 1, node  $i$  has strength  $S_{ij}$  to node  $j$  and node  $j$  has strength  $S_{ji}$  to node  $i$ .



Node $i$	Node gene		Connection gene			
	$ID_i$	$F_i$	$C_{i1}$	$C_{i2}$	.....	$C_{ik}$
			$S_{i1}$	$S_{i2}$	.....	$S_{ik}$

Fig. 1. Basic structure of genetic relation algorithm with directed edges.

Fig. 1 also describes the gene of node  $i$ , then the set of these genes represents the genotype of GRA individuals. Concretely speaking,  $ID_i$  represents an identification number of the node, e.g.,  $ID_i=1$  means node  $i$  has the directed edges with other nodes, while  $ID_i=2$  means node  $i$  has the undirected edges with other nodes.  $F_i$  denotes the function of node  $i$ . In this paper,  $F_i$  represents different stock brands in the portfolio.  $C_{i1}, C_{i2}, \dots, C_{ik}$  show the number of the nodes which are connected from node  $i$  firstly, secondly and so on.  $S_{i1}, S_{i2}, \dots, S_{ik}$  denote the strength of edges from node  $i$  to node  $C_{i1}, C_{i2}, \dots, C_{ik}$ , respectively. All individuals in a population have the same number of nodes.

Like other evolutionary algorithms, selection, crossover and mutation are used as the genetic operators of GRA. The outline of evolution is described as follows:

- Initialize the first population and calculate the fitness of the population;
- Generate new individuals for the next generation by tournament selection and genetic operations of crossover and mutation;
- Calculate the fitness of the new individuals;
- Repeat 2-3 until the terminal condition meets.

The point of GRA is that all the connections between node do not have to be defined, but the connection itself could be evolved.

### B. GRA with Undirected Edges

Fig. 2 shows the basic structure of GRA with undirected edges. Same as directed GRA, the event is also represented by the node, while the relation between nodes is represented by undirected edges with their strength. The relation between node  $i$  and node  $j$  has a strength of  $S_{ij}=S_{ji}$  in GRA with undirected edges, which is different from directed GRA.

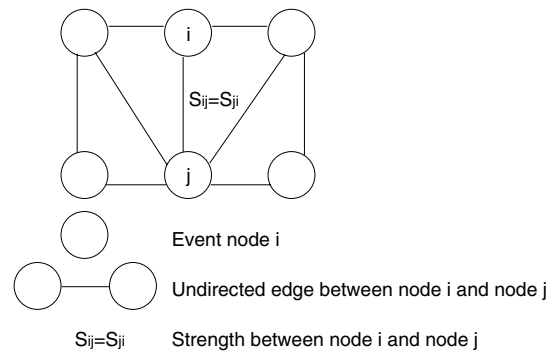


Fig. 2. Basic structure of genetic relation algorithm with undirected edges.

## III. PORTFOLIO SELECTION USING GENETIC RELATION ALGORITHM

In our proposed method, genetic relation algorithm with undirected edges are used to construct the portfolio selection model. As shown in Fig. 3, the basic structure of GRA is described as follows: The nodes in GRA are used to represent different stock brands in a portfolio, and the strength between two nodes are used to indicate the relationship between stock brands, i.e., the value of correlation coefficient. The main point of our proposed model is to select given number of appropriate stocks in a portfolio. In order to maximize the final profit in the buying and selling strategy of GNP[4], we can study what degree of correlation coefficient the stocks should have by GRA method.

### A. Notations and Fitness Function of GRA

- $D$ : set of days
- $S$ : set of stocks

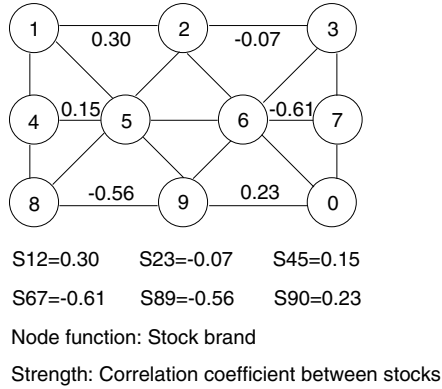


Fig. 3. Genetic relation algorithm for portfolio selection.

- $S(G)$ : set of stocks in GRA
- $S(G_i)$ : set of stocks whose strength is defined between node  $i$  in GRA
- $Price(i, d)$ : price of stock  $i$  on day  $d$
- $\mu_i$ : mean of the price of stock  $i$
- $\sigma_i^2$ : variance of the price of stock  $i$
- $\sigma_{ij}$ : covariance between the prices of stock  $i$  and stock  $j$
- $\rho_{ij}$ : correlation coefficient between the prices of stock  $i$  and stock  $j$
- $\rho$ : target value of the correlation coefficient

The object of GRA is to select appropriate  $|S(G)|$  stocks out of a total number of stocks  $|S|$ , which satisfy a certain value of correlation coefficient, i.e.,  $-1.0 \leq \rho \leq 1.0$ . Therefore, the fitness function of GRA is defined as follow.

$$Fitness = \frac{1}{|S(G)|} \sum_{i \in S(G)} \frac{1}{|S(G_i)|} \sum_{j \in S(G_i)} (\rho_{ij} - \rho)^2, \quad (1)$$

where,

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j},$$

$$\sigma_i^2 = E[(Price(i, d) - \mu_i)^2] = \frac{1}{|D|} \sum_{d \in D} (Price(i, d) - \mu_i)^2,$$

$$\begin{aligned} \sigma_{ij} &= E[(Price(i, d) - \mu_i)(Price(j, d) - \mu_j)] \\ &= \frac{1}{|D|} \sum_{d \in D} (Price(i, d) - \mu_i)(Price(j, d) - \mu_j), \end{aligned}$$

$$\mu_i = E[Price(i, d)] = \frac{1}{|D|} \sum_{d \in D} Price(i, d).$$

In the fitness function of Eq. (1),

- if  $\rho$  is around 1.0, then stock  $i$  and stock  $j$  have positive correlation.
- if  $\rho$  is around -1.0, then stock  $i$  and stock  $j$  have negative correlation.
- if  $\rho$  is around 0.0, then stock  $i$  and stock  $j$  have no correlation.

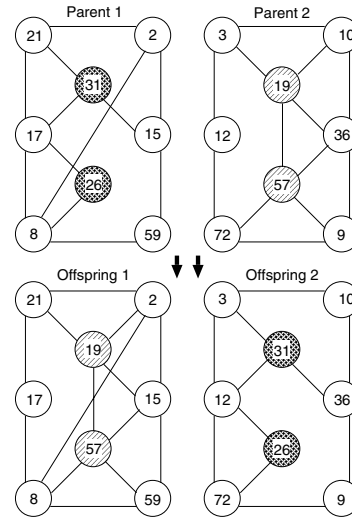


Fig. 4. Crossover.

The fitness function evaluates the GRA individuals so that the strengths between stocks have the target value of correlation coefficient  $\rho$ . Generally, according to the portfolio theory, it is preferable to select  $|S(G)|$  stocks which have small correlations. It is our interest to find out the target value of the correlation coefficient  $\rho$  in the fitness function. By the portfolio selection model of GRA, the stocks having large correlations with each other will be eliminated, as they always cause high risk in a portfolio.

### B. Genetic Operators of GRA

In this sub-section, the genetic operators in the evolution phase are introduced. In order to get the best individual, the function of nodes in GRA should be changed, which can be realized effectively by genetic operations. GRA has three kinds of genetic operators: selection, crossover and mutation. In GRA, mutation operation could be executed not only on the connections between nodes but also on the node functions.

1) *Selection*: At each generation, all of the individuals are ranked by their fitness values and the best individual in the current generation is preserved for the next generation by elite selection. Then, tournament selection of individuals is carried out for reproducing the next generation.

2) *Crossover*: As shown in Fig. 4, crossover is executed between two parents and two offspring are generated. The procedure of crossover is as follows.

- Select two individuals using tournament selection twice and produce them as parents.
- Each node is selected as a crossover node with the probability of  $P_c$ .
- Two parents exchange the genes of the corresponding crossover nodes.
- Generated new individuals become the new ones of the next generation.

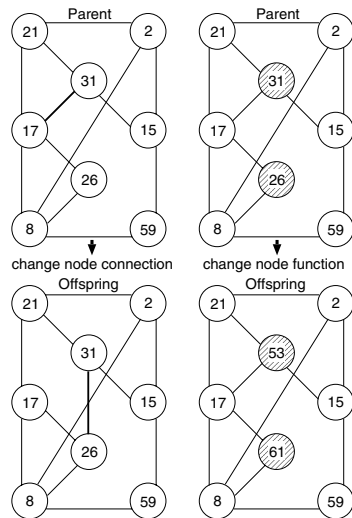


Fig. 5. Mutation.

3) *Mutation*: Fig. 5 shows an example of the mutation operator. Mutation is executed in one individual and a new one is generated. The procedure of mutation is as follows.

- Select one individual as a parent using tournament selection.
- Mutation operation
  - change connection: Each node edge ( $C_{i1}, C_{i2}, \dots, C_{ik}$ ) is selected with the probability of  $P_m$ , and the selected edge is reconnected to another node.
  - change node function: Each node function ( $F_i$ ) is selected with the probability of  $P_m$ , and the selected function is changed to another one.
- Generated new individual becomes the new one of the next generation.

4) *Flowchart of GRA*: Fig. 6 shows the flowchart of GRA. For the first GRA population, each individual is generated assigning a certain stock brand selected randomly to one of the nodes of GRA. It is ensured that all nodes are different within one individual. In the next, evaluation of the individuals is carried out according to their fitness values. At the reproduction phase, selection, crossover and mutation are used as genetic operators to generate the population for the next generation. This process is repeated until the last generation. Finally, after obtaining the best individual in the last generation, it is tested by the stock trading model of GNP[4].

#### IV. EXPERIMENTAL RESULTS

In order to confirm the effectiveness of GRA in the portfolio selection model, we carried out the trading simulations by GNP using the best GRA individual that was obtained in the last generation. The simulation is divided into two stages: one is used for the training of GRA and the other is used for the training and testing of GNP.

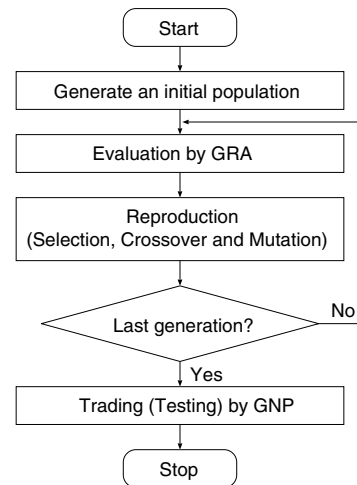


Fig. 6. Flowchart of GRA.

TABLE I  
PARAMETER CONDITIONS FOR EVOLVING GRA

Number of individuals=300 (mutation:179, crossover:120, elite:1)
Number of generations=300
Number of nodes=10
$P_c=0.3, P_m=0.1$

- Training (GRA): January 4, 2001—December 30, 2003 (737 days)
- Training (GNP): January 4, 2001—December 30, 2003 (737 days)
- Testing (GNP): January 5, 2004—December 30, 2004 (246 days)

#### A. Performance of Genetic Relation Algorithm

1) *Experimental Conditions of GRA*: Table I shows the parameters of the evolution of GRA. The total number of nodes in each individual of GRA is 10 which indicate 10 different stocks in a portfolio. Those stocks are selected from the 500 companies listed in the first section of Tokyo stock market in Japan. The content  $F_i$  in each node are determined randomly at the beginning of the first generation, and changed appropriately by evolution.

The initial connections between nodes are also determined randomly at the first generation. At the end of each generation, 179 new individuals are produced by mutation, 120 new individuals are produced by crossover, and the best individual is preserved. The other parameters for crossover and mutation are the ones showing good results in the simulations. The terminal condition is 300 generations.

2) *Simulation Results of GRA*: Fig. 7 shows the average processing time when the number of edges of GRA individuals is changed. It is an example when the target correlation coefficient  $\rho$  is set to 0.1. It is clear from Fig. 7 that when the

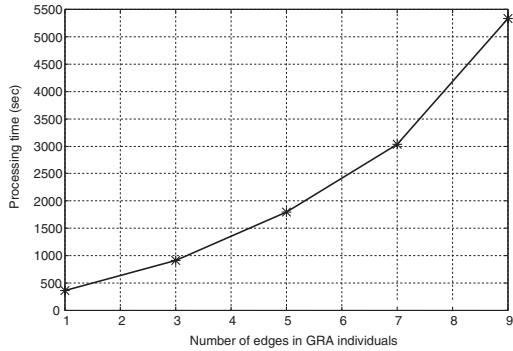


Fig. 7. Processing time when changing the number of edges of nodes.

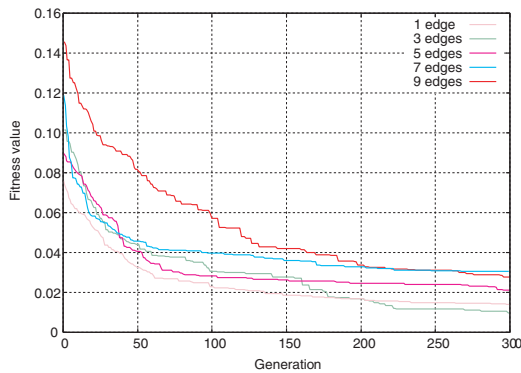


Fig. 8. Average fitness value when changing the number of edges of nodes.

number of edges increases, the average processing time also increases because of the complexity of the network structures.

Fig. 8 shows the average fitness values when the number of edges of GRA individuals is changed using the data from 2001 to 2003, and the curves are the average values over 30 independent simulations. From Fig. 8, we can see that the differences of the fitness values between one edge and the large number of edges become small as the generation goes on. Since a small number of edges can save the processing time as shown in Fig. 7, and comparable fitness value is obtained with one edge in general, it is unnecessary to consider the connections of full edges between nodes. Therefore, only one edge is used for the evolution of GRA.

### B. Validation by the Stock Trading Model of Genetic Network Programming

1) *Experimental Conditions of GNP*: Table II shows the parameters of the evolution of conventional GNP method which was proposed in our previous study[4]. GNP uses the judgment nodes to judge the information from stock markets, and uses the processing node to take buying and selling actions. Five control nodes are assigned to each brand. The total number of nodes in each individual is 80 including 10 processing nodes, 20 judgment nodes and 50 control nodes. The initial connections between nodes are also determined

TABLE II  
PARAMETER CONDITIONS FOR EVOLVING GNP

Number of individuals=300 (mutation:179, crossover:120, elite:1)
Number of nodes=80 (Judgement node=20, Processing node=10, control node=50)
Number of sub-node in each node=2
$P_c=0.1, P_m=0.03, \alpha=0.1, \gamma=0.3, \epsilon=0.1$

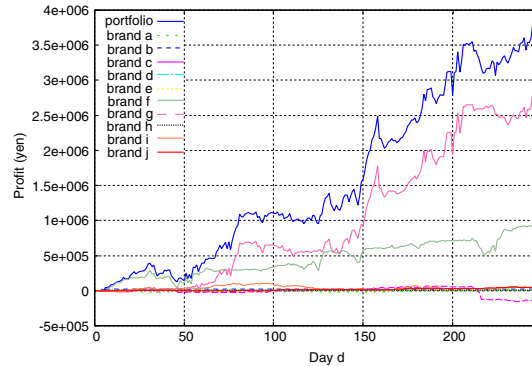


Fig. 9. Profits change of selected 10 brands in the testing period by GNP.

randomly at the first generation. At the end of each generation, new individuals are produced by selection, crossover and mutation.

In the validation phase by the stock trading model of GNP, we suppose that the initial funds are 50,000,000 Japanese yen in both training and testing periods. Especially, when we use GNP to test the best portfolio generated by GRA, one GNP individual has 10 groups of control nodes, each of which deals with one brand, so one GNP can deal with 10 brands in the portfolio.

2) *Simulation Results of GNP*: In order to confirm the efficiency of our proposed method, Table III shows the comparison of the profit between GNP with GRA and conventional GNP. Concretely speaking, in the case of GNP with GRA as shown in the first line, we carry out the simulations by setting the various values of  $\rho$  in the fitness function (1). The value of  $\rho$  indicates positive or negative correlation between different stocks, which is important for the portfolio selection. Table III presents the average profit of the portfolio selected by GRA using stock trading model of GNP over 30 independent simulations when the correlation coefficient  $\rho$  is set to different values. From the results, it is clarified that we can get a good profit in the testing period when  $\rho$  is set to 0.0. Therefore, we set the value of 0.0 for the parameter  $\rho$  in our simulations. In the case of conventional GNP as shown in the fourth line, we randomly select 9 portfolios without the optimization model of GRA from the 500 companies listed in the first section of Tokyo stock market in Japan. From Table III, it is found that the portfolio selected by the GRA optimization model with small  $|\rho|$  can obtain higher profits than conventional GNP, which selected stocks randomly from the stock market.

TABLE III  
COMPARISON OF PROFIT WITH CONVENTIONAL GNP (PROFIT[YEN])

$\rho$	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8
random sequence	1	2	3	4	5	6	7	8	9
GNP with GRA	2,766,084	2,672,986	3,179,176	3,384,061	3,797,128	3,502,800	2,973,384	3,084,725	2,580,135
Conventional GNP	1,738,324	2,631,634	3,337,781	2,987,880	3,148,120	2,602,448	1,992,164	2,733,425	1,376,708

Moreover, in the previous study, it has been confirmed that GNP outperformed the other traditional methods, i.e., GA and Buy&Hold, which are widely used in the financial field[7]. Thus, we didn't carry out the comparisons between GRA and other traditional methods in this paper.

Fig. 9 shows the profits change of selected 10 brands in the testing period by GNP trading model, i.e., the best portfolio obtained with GRA method when the value of parameter  $\rho$  is set to 0.0. We carried out the dealing of these 10 brands by stock trading model of GNP using the data of 2004. From Fig. 9, we can see that the profit keep increasing during the testing period. As a result, by this efficient portfolio optimization system, we can obtain much profits in the trading of those brands.

## V. CONCLUSIONS

In this paper, we proposed the GRA method and applied it to the portfolio selection problem. In order to pick up a fixed number of the most efficient portfolio, the algorithm evaluates the relationships between stock brands using a specific measure of strength and generate the optimal portfolio in the final generation. We carried out the experiments using stock data of selected 10 brands for 4 years. In the experiments, the efficiency of GRA method is confirmed by the stock trading model of GNP that has been proposed in our previous study. Compared to conventional GNP, the advantage of our proposed method is that GRA method considers correlation coefficient as the strength between stock brands to optimize the portfolio, which is different from the conventional method that select stocks randomly for trading. From the results, it is clarified that we can obtain much profits in the trading of those brands.

There remain some further studies in the future. First, the algorithm presented can be further improved by modifying the fitness function. Moreover, we should evaluate the proposed method comparing with other methods in the financial market.

## REFERENCES

- [1] H. Markowitz, *Portfolio selection: Efficient diversification of investments*, New York, John Wiley&Sons, 1959.
- [2] V. Dropsy, "Do macroeconomic factors help in predicting international equity risk premia?" *Journal of Applied Business Research*, vol. 12, pp. 120-132, 1996.
- [3] K. J. Oh, T. Y. Kim, S. H. Min and H. Y. Lee, "Portfolio algorithm based on portfolio beta using genetic algorithm," *Expert Systems with Applications*, vol. 30, pp. 527-534, 2006.
- [4] Y. Chen, E. Ohkawa, S. Mabu, K. Shimada and K. Hirasawa, "A Stock Trading Model for Multi-Brands Optimization Based on Genetic Network Programming with Control Nodes," In Proc. of the SICE Annual Conference 2008, pp. 664-669, Tokyo, August 2008.
- [5] J. R. Koza, *Genetic Programming, on the programming of computers by means of natural selection*. Cambridge, Mass.: MIT Press, 1992.

- [6] S. Mabu, K. Hirasawa and J. Hu, "A graph-based evolutionary algorithm: Genetic network programming and its extension using reinforcement learning," *Evolutionary Computation*, MIT Press, vol. 15, no. 3, pp. 369-398, 2007.
- [7] S. Mabu, Y. Izumi, K. Hirasawa and T. Furuzuki, "Trading Rules on Stock Markets Using Genetic Network Programming with Candle Chart," *SICE Trans.*, vol. 43, no. 4, pp. 317-322, 2007.

## APPENDIX

The list of selected 10 companies in Fig. 9 is as follows.

- (a) Nissin Foods Products Co., Ltd.
- (b) Hitachi Chemical Co., Ltd.
- (c) ToTo Ltd.
- (d) Toshiba Corporation
- (e) Honda Motor Co., Ltd.
- (f) Yamaha Corporation
- (g) Toyota Tsusho Corporation
- (h) All Nippon Airways Co., Ltd.
- (i) Chubu Electric Power Co., Inc.
- (j) Toho Co., Ltd.