DESIGN OF EXPERIMENTS TO IDENTIFY OPTIMAL PARAMETERS OF GENETIC ALGORITHM

SANGHEON HAN *

Received June 5, 2001

Abstract. Validation of an analytical method through a series of experiments demonstrates that the method is suitable for its intended purpose. Due to multi-parameters to be examined and a large number of experiments involved in validation, it is important to design the experiments scientifically so that appropriate validation parameters can be examined simultaneously to provide capabilities of the analytical method. This paper describes identifying the value of Genetic Algorithm (GA)'s parameters to get an optimal result and to reduce the cost. We considered two-GA parameters (crossover probability and mutation probability) by using the design of experiment method. Considerations will lead to the success of the optimization and the efficiency of the approaches to other NP-hard problems. A numerical example is provided to illustrate our method.

1 Introduction Genetic Algorithm was introduced by J.H.Holland(1975) as a method for modelling a complex system and applied by concepts from biological evolution to a mathematical context. The general idea is to start with randomly generated solutions and, implementing a "survival-of-the-fittest" strategy, evolve good solution. See [4],[8]or [14] for details.

GA has been applied to various problems, such as Travelling Sales Man Problem(TSP) [10], Knapsack Problem[13], Scheduling Problem, Multiple-choice Integer Problem, Portfolio Selection Problem and others alike. GA contains several operators called selection, crossover and mutation, being crossover and mutation the ones that specially affect performance of GA. So, it is very important to specify the GA's parameter for getting a good performance. However, it is very troublesome to identify GA-parameters.

2 Genetic algorithm techniques

2.1 Genetic algorithm Since Darwin brought up the theory of natural selection, many academic disciplines have been affected. GA was enlightened and developed by Holland in 1975 and adopted the concept of seeking to survive among living things. A string of words is used to simulate the chromosome of every living thing. According to that chromosome, fitness to environment is computed. Every chromosome in each generations proceeds to crossover and mutation at random so as to give birth to the next generation. Furthermore, general environment, according to the fitness selection of that chromosome, will decide whether it will live or not. This evolution will alternate and last until the final goal is achieved.

GA utilizes three basic operation systems: selection, crossover, and mutation. Through these three evaluation processes, parents give birth to a new generation. The strongest

2000 *Mathematics Subject Classification.* 62K15, 65C50, 68T20.
Key words and phrases. genetic algorithm, design of experiments, NP-hard problem.
*Address correspondence to Sangheon HAN, Graduate School of Economics, Osaka University, Toyonaka 1-16, Osaka 560-0043, Japan. E-mail : sangheon@compsrv.econ.osaka-u.ac.jp

individuals in every generation will have the highest possibilities of survival and will pass part or all to the next generation (Holland 1975, Goldberg 1989). Thus, the basic genetic algorithm may be summarized as follows :

Algorithm GA:

Begin
Initialize population;
Generation := 0;
Repeat
$Generation = Generation + 1$
Selection (*population*);
Crossover (*population*);
Mutation (*population*);
Until Termination Criterion;
End

2.2 Application of Genetic Algorithm GA provides a flow process with a rather simple system structure. However, it can produce a strong research capability for getting solutions, powerful especially as regards the issue of optimum combination (Goldberg, 1989,1994;Srinvas and Patnaik, 1994). In recent years, the applications of GA have been extended not only to biology but also to various areas including financial problems, vehicle scheduling problem and others.

For example, in the investment rule of Bauer (1994), most of the overall economical variables are used to find out the better rule combination with GA and explore the relations with profit in the stock market, basically regarded as a timing strategy of basic analysis. Tayler (1995) used GA to construct an intelligent stock market model.

We are examining the related study in GA included NP-complete class.

3 Design of experiments techniques Design of experiments techniques first gained acceptance as a result of the work of Sir RA Fisher researching into agriculture. In the early 1970s in Japan, Taguchi proposed his own modifications (Taguchi Methods[19]) in order to make these powerful techniques more accessible to non-statisticians. The procedure consists of selecting a design based on the number of set-up parameters (factors) and responses under investigation, perform the experiments and then analyze the result. Each factor under investigation is assigned a number of levels at which to be tested. The design used consists of an array (more correctly called an orthogonal array because of its mathematical properties) which prescribes a series of the variable factors at each of the levels they can take. For example, an array to investigate the effect of two variable factors A and B (like crossover and mutation in our research) on the response of the process where each of the variable factors is tested at three levels would be written as shown below:

Trial	Factor A	Factor B
1	2	2
2	2	2
3	2	2
4	1	1

In this case, all possible combinations of factor A and factor B at each of levels 1 and 2 are tested and referred to as a full factorial experiment. As the number of factors and levels increases, the number of trials and interactions also increases rapidly. Thus the investigation of all possible combinations of m factors each having n levels requires a total of n^m trials. The standard technique, advocated by many authors including Fisher, Taguchi and Logothetis, is an analysis of variance which determines the significance of the effect of factors on the response.

4 Tested Problems and Methodology We adopted as sample problem the small-sized Portfolio Selection Problem(PSP)[14] so as to identify GA's parameters. This sample problem belongs to the class of NP(non-polynomial) complete problems.

PSP Model

This model described in forming an index-like m security portfolio R_m of the n original securities ($n > m$) in the mean-variance framework, satisfying the criterion that the tracking error $E(R_n - R_m)^2$ is minimized, where R_n is the benchmark portfolio with n securities. It should be noted that the tracking error between portfolios R_n and R_m is

$$E\{R_n - R_m\}^2 = Var\{R_n - R_m\} + \{E(R_n) - E(R_m)\}^2.$$

Let $y_i(i = 1, 2, \dots, n)$be a 0-1variable defined as

$$y_i = \begin{cases} 0, & \text{if the } i\text{th security is not included in } R_m \\ 1, & \text{if the } i\text{th security is included in } R_m. \end{cases}$$

Then the problem to minimize the tracking error can be written as

$$\begin{aligned} \underset{x,\ y}{\text{Minimize}} \quad & E(R_n - R_m)^2 \\ & = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(w_i - x_i y_i)(w_j - x_j y_j) \\ \text{subject to} \quad & x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = 1 \\ & y_1 + y_2 + \cdots + y_n = m \\ & y_i = 0 \text{ or } 1. \end{aligned}$$

where,

$R_n = \sum_{i=1}^{n} w_i P_i$ $\left(R_m = \sum_{i=1}^{n} x_i y_i P_i\right)$: random return of $n(m)$ security portfolio,

P_i: random return on security i with mean $E\{P_i\} = \mu_i$, for $i = 1, 2, ..., n$,

a_{ij}=$E(P_i P_j) = Cov(P_i P_j) + E(P_i)E(P_j)$,

x_i: fraction of investor's wealth invested in security i,

w_i: weight of security $i(i = 1, 2, \dots, n)$.

4.1 Preparing for the experimental design In this section, we propose the design of experiments method to test the validity of two-GA parameters. From the result of experimental design, we can know that both parameters are valid in this simulation. In addition, we examine the following crossover and mutation probabilities in the GA-Operators.

$$P_c = 0.0, ..., 1.0, \quad P_m = 0.0, ..., 1.0.$$

In this case, the number of combinations of crossover and mutation probabilities is ∞ . The experimental method can utilize more than two-factors by using *multi-way factorial design* and *table of orthogonal array* and so on, but as the number of factors increase, the number of trials also increase. In this research, we consider above the two main factors that lead to determine the optimal combination of factors and the effectiveness.

We prepare the results calculated by GA-Program coded with $Visual\ C++$ on the Windows operating system. Then, we tested the problem twice at each parameter (Table 1). In this experiment, we consider the Mid-Range(M_R) [1] as the value of parameters to test the problem, $Small = 0.16665$, $Medium = 0.49998$, and $Large = 0.83331$ respectively.

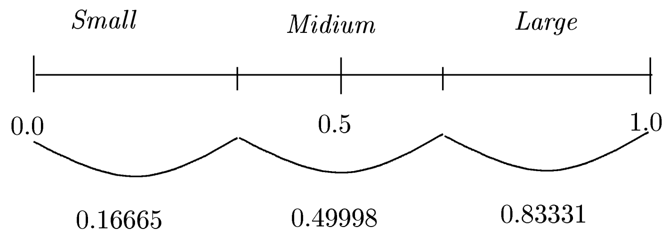

Figure 1: Value of each ranges

Parameters	mutation-Small (B_1)	mutation-Medium (B_2)	mutation-Large (B_3)
crossover-Small (A_1)	0.799171 0.549296	0.76427 0.639657	0.993011 0.973052
crossover-Medium (A_2)	0.972702 0.974271	0.988696 0.979546	0.975576 0.966994
crossover-Large (A_3)	0.99327 0.996157	0.996157 0.995843	0.996157 0.994545

Table 1: Fitness value of PSP

	B_1		B_2		B_3	
	sum	mean	sum	mean	sum	mean
A_1	1.348467	0.674234	1.403927	0.701964	1.966063	0.983032
A_2	1.946973	0.973487	1.968242	0.984121	1.94257	0.971285
A_3	2.397197	1.198599	1.992	0.996	1.990702	0.995351

Table 2: Arranged Data

4.2 Test by experimental design Tables below are experimental design analysis of GA-parameters. We consider two parameters (crossover and mutation) in GA-parameters. So, we test parameters as a *two-way factor design* in experimental design. In these tables, Factor(A) indicates crossover along $0.1, ..., 1.0$ range, and Factor(B) indicates mutation along $0.1, ..., 1.0$ range. Table 3 shows times of experiments, and Table 4 shows the mean of data. Table 5 is the result of analysis of variance. From Table 5, we can say that parameters of crossover and mutation are significant in probability 0.01 and 0.05 respectively. Besides, both parameters have an interaction relation.

The ANOVA table for this data (Table 5) supports the conclusion. In this table, factor

[1] $M_R = \frac{X_{max}+X_{min}}{2}$

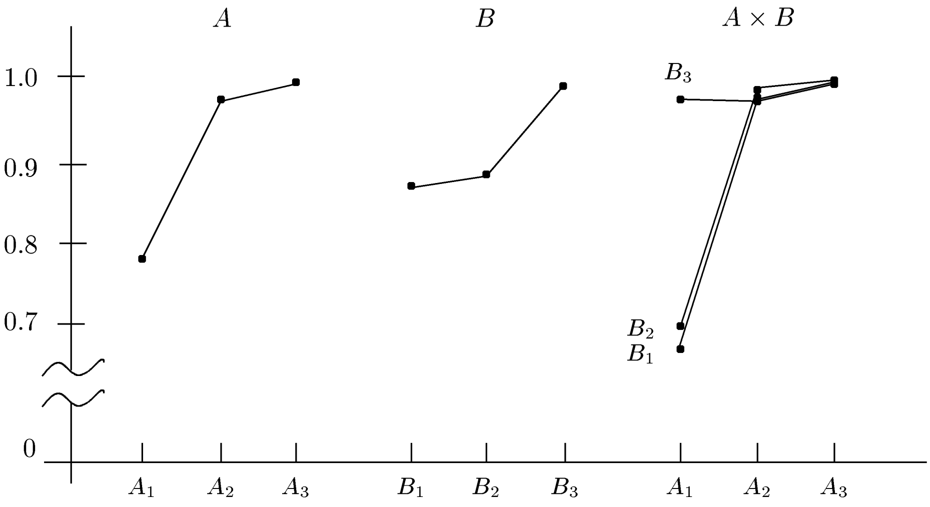

Figure 2: Data graph

parameters	B_1	B_2	B_3	sum
A_1	2	2	2	**6**
A_2	2	2	2	**6**
A_3	2	2	2	**6**
sum	**6**	**6**	**6**	**18**

Table 3: Times of experiment

A shows a Crossover probability and factor B shows a Mutation probability. And, factor A is significant in 0.01, factor B is significant in 0.05 and interaction AB is also effective. Inspection of the factor effects and the ANOVA for this experiment indicate that the optimum combination and level of parameters is A_3B_3 as shown in the Table 4.

4.3 Selection of genetic operators and their probabilities From the result above mentioned, we can observe that GA highly depends on probability specifications. That is, the performance of the GA is sensitive to the parameter specifications. We can adjust the adaptable parameters to this problem by the result of experimental design (table 4 and table 5). In addition, we compared this result with another one that has better conditioned parameters as large *pop_size* (500), and calculated many generations (1,500 times). Figure 3 and Figure 4 show good results out of the performed simulations.

parameters	B_1	B_2	B_3	sum
A_1	0.674234	0.701964	0.983032	**0.78641**
A_2	0.973487	0.984121	0.971285	**0.976298**
A_3	0.994714	0.996	0.995351	**0.995355**
sum	**0.880811**	**0.894028**	**0.983223**	**0.919354**

Table 4: Mean of Data

param	degree of freedom	Sum of squares error	Unbiased variance	Variance ratio	P	Test
Total	17	0.316365				
Factor(A)	2	0.160158	0.080079	18.3538	0.000667	[**]
Factor(B)	2	0.037237	0.018618	4.267277	0.049724	[*]
Factor(AB)	4	0.079703	0.019926	4.566912	0.027374	[*]
Error	9	0.039268	0.004363			

Table 5: ANOVA

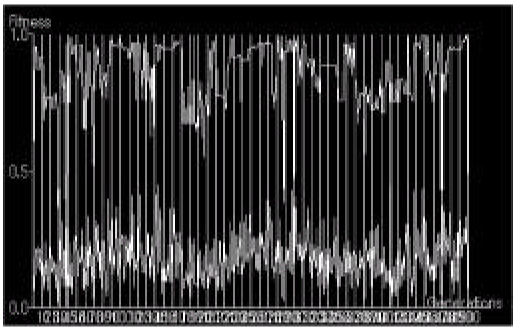

Figure 3: Adapted Experimental Design

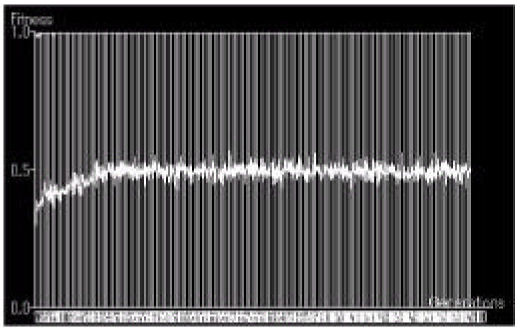

Figure 4: Adapted Good Condition

pop_size	= 21
crossover rate	= 0.83331
mutation rate	= 0.83331
generation	= 500
Best fitness	= 0.996157
Tracking Error	= 48.2706

pop_size	= 500
crossover rate	= 0.5
mutation rate	= 0.5
generation	= 1500
Best fitness	= 0.998197
Tracking Error	= 22.5823

5 Concluding remarks In the described research, we developed an efficient GA to design an index fund with a given number of securities that minimizes the tracking error between the benchmark portfolio and the index fund. The proposed method can render the solutions as optimal or at least sub-optimal. By computer simulations, we pointed out that combination of high performance crossover and mutation operator did not always lead to high performance genetic algorithm. We performed many runs of the genetic algorithm with various specifications of crossover and mutation operators in order to find the best specification. Moreover, we could identify the effectiveness of GA parameters (crossover probability and mutation probability) by the design of experiments. Besides, by means of experimental design, we could know the correlation between crossover and mutation probabilities. This method is very useful not only for the described case but also in another applications. Many problems are left for future research. In particular, it is necessary further theoretical analysis, simplification of operators and setting of parameters for getting a better GA convergency, and it is also necessary to test the accuracy of this algorithm.

References

[1] Balas E.& Martin. 1980. Pivot and complement - a heuristic for 0-1 programming, *Management Science*. 26(1):86–96.

[2] Bauer Jr.,Rechard J. 1994. *Genetic Algorithms and Investment Strategies*. John Wiley & Sons:103–213.

[3] David A Coley. 1999. *An Introduction to Genetic Algorithms for Scientists and Engineers*. World Scientific.

[4] Davis L. 1991. *Handbook of Genetic Algorithm*, Van Nostrand, New York.

[5] Dawsland K.A. 1996. Genetic Algorithm–a tool for OR?. *Journal of the Operational Research Society* 47:550–561.

[6] Fisher RA. 1966. *Design of Experiments*, Oliver & Boyd:Edinburgh (Now Longman, Harlow).

[7] Gen M., Cheng R. 1997. *Genetic Algorithms and Engineering Design.* Wiley, New York.

[8] Goldberg D.E. 1989. *Genetic Algorithm in Search, optimization, and Machine Learning.* Addison-Wesley.

[9] Holland J.H. 1992. *Adaption in Natural and Artificial Systems.* 2nd ed. Cambridge:MIT.

[10] Karp.R and Steel.J 1985. *Travelling Salesman Problem : A Guided Tour of Combination.*Wiley.

[11] Kitano H. 1995. *Genetic Algorithm 1,2,3.* In Japanese. Tokyo:Sangyou Tosho.

[12] Markowitz H. 1952. Portfolio Selection, *Journal of Finance*, 7:77–91.

[13] Martello S. and Toth P. 1990. *Knapsack Problems : Algorithms and Computer Implementation.*Wiley, chichester.

[14] Michalewicz Z. 1994. *Genetic Algorithm + Data Structure = Evaluation program.* 2nd ed. Berlin:Springer.

[15] Pukelsheim F. 1993. *Optimal Design of Experiments.*Wiley, New York.

[16] Schaffer, J.D., (ed.). 1989. Proceedings of the 11^{th} International Joint Conference on Artificial Intelligence, Organ Kaufmann, San Mateo. 750-755.

[17] Sheldon M.Ross. 1997. *Simulation*, 2nd ed. San Diego:Academic Press.

[18] Tabata Y. and Takeda E. 1995. Bicriteria Optimization Problem of Designing an Index Fund, *Journal of Operational Research Society.* 46:1023–1032.

[19] Taguchi G. 1986. *Introduction to Quality Engineering.* ASI Center for Taguchi Methods 1987. Asian productivity Organization, and UNIPUB/Karms International publications.

[20] Tayler P. 1995. *Modelling Artificial Stocks Markets Using Genetic Algorithms.* Edited by Suran Goonatilake and Philip Treleaven, Intelligent Systems for Fiance and Business, John Wiley & Sons, NY.271–288.

Graduate School of Economics, Osaka University, Toyonaka 1-16, Osaka 560-0043, Japan.
E-mail : sangheon@compsrv.econ.osaka-u.ac.jp