Parameter Settings of Genetic Algorithm Based on Multi-Factor Analysis of Variance

Zhengfeng Cao
The school of Economics and Management
Beijing Institute of Petrochemical Technology, BIPT
Beijing, China
e-mail:caozhengfeng@bipt.edu.cn

ZhiHui Zhang
The school of Economics and Management
Beijing Institute of Petrochemical Technology, BIPT
Beijing, China
e-mail: zhangzhihui@bipt.edu.cn

***Abstract*—Six primary factors have been analyzed which influence the performance of the Genetic Algorithm. The concrete method has been designed of setting algorithm parameters by using multi-factor analysis of variance. And it illustrates the implementation process of the methods by specific examples of applications. In the end with the data gained from the experiment it has proved that the performance of the algorithm can be markedly improve after setting algorithm parameters by using the analysis of variance.**

Keywords-Genetic Algorithm;Multi-factor Analysis of variance;Orthogonal Design;Parameter Setting

I. INTRODUCTION

Genetic algorithm is an auto-adapted global optimization search algorithm which simulates the heredity and evolution process of creatures. Since Dr. John Holland brought up the algorithm officially in 1975， Genetic Algorithm has gradually developed into a computational model which was used to resolve optimization questions by simulating natural evolution process. But in practical applications, setting operation parameter may affect the performance of the algorithm, and illogical setting may contribute to poor stability of the algorithm, and prevent the algorithm from achieving its optimal solution.

This paper analyses many factors affecting the performance of algorithm, and confirms importance of these factors by using multi-factor analysis of variance in statistics, and thus get specific value for each parameter which is beneficial to improve the stability of algorithm and avoid setting parameter randomly so as to achieve the goal of quantifying parameter setting of Genetic Algorithm, and advance the performance of the algorithm significantly.

II. THE ANALYSIS OF THE RELATION OF THE ALGORITHM PARAMETER AND THE PERFORMANCE OF ALGORITHM

Having designed the Genetic Algorithm, the operating parameters which need to be set in application are mainly population size, crossover operator, mutation operator etc.. The parameters affecting performance of Algorithm are as following:

A. Crossover operator

Crossover operator affects the probability of genomic hybridization in two chromosomes of an individual. The Crossover operator is marked as Pc. The value of Pc is generally from 0.2 to 1. If the value of Pc is too small, it will be difficult for the algorithm to search forward, and reduce the performance, or else destroy the high fitness individual.

B. Mutation operator

The probability of gene mutation in chromosome of one individual is affected by the mutation operator. The mutation operator is marked as Pm. The value of Pm is generally from 0.0001 to 0.1. Likewise, Mutation operator must be adaptive. If the value of Pm is too small, it cannot make new individuals, or else will make the algorithm become a simple random search and produce too many new individual.

C. Population size

It is generally given by experience according to demands. Population size is marked as S, and the value of S is generally from 50 to 160. If the value of S is too small, it will be difficult to get the optimal solution. If too large, it will augment the convergence time.

D. Maximum number of generations

The algorithm may iterate endlessly, but for the sake of operation in practice, it is normally to set maximum number of generations, which is marked as E. The value of E is from 50 to 160.If the value of E is too small it will be difficult to get the optimum solution and make algorithm terminate early, or else will augment the complexity of operation.

E. Chromosome length

The chromosome length of each individual varies when faced with practical questions. Chromosome length is marked as L, and the value of L is generally from 20 to 100. If the value of L is too small, it will be difficult to get the optimum solution. If too large, it will augment the convergence time.

F. Maximum fitness

It implies that the algorithm will be terminated when the best individual reaches the Maximum fitness. Maximum fitness is marked as F. The value of F varies according to the actual application. Generally, it will be determined by subtracting an appropriate numbers according to the target value of the algorithm. The further the value of F is away from the target value, the more difficult to obtain the

978-0-7695-4281-2/10 $26.00 © 2010 IEEE
DOI 10.1109/ICGEC.2010.82

optimum solution. The nearer the value of F is to the target value, the more complex the algorithm is.

The analysis above shows that the six main factors of the genetic algorithm affect the performance of algorithm extremely. But in the practical work, these parameters are usually determined by individual experience, which makes the performance vary from person to person and extend the randomness. How to set parameter in practical applications becomes one of the important factors of applying algorithm successfully.

III. METHOD OF SETTING ALGORITHM PARAMETERS BY USING MULTI-FACTOR ANALYSIS OF VARIANCE.

Analysis of variance is a method which analyzes the relationship between variables by dividing the error sources. The basic idea of analysis of variance is to divide the variance of all observed numbers into several parts. In other words, it is that the sum of squares of deviation from mean which is attained from variety of all described observed-numbers are departed into parts and random sampling errors, and then each corresponding mean square is calculated, and a test of F statistics has been constructed to process the test of statistics. Multi-factor analysis of variance is a kind of analysis of variance which analyzes the effects about an independent variable which is subject to one or more variables.

Method of Setting algorithm parameters by using multi-factor analysis of variance is: Firstly, it is to confirm the operation scope of the six factors above and build multiple orthogonal design chart to obtain level number in operating scope; Secondly, it is to carry on the experimental statistics based on the orthogonal design chart, and obtain the results of the experiments. And then they are input into the statistical software to do the analysis of variance. Finally, the main influential factor can be obtained from the result and then the average value of K is calculated by the character of orthogonal design chart. And the settings of the parameters of the algorithm have been completed by setting the maximum average value of K the best level of the factors.

IV. EXAMPLES OF APPLICATIONS

In order to make a further explanation of the process of setting algorithm parameter by using the analysis of variance, an example of Genetic Algorithm optimizing function is to introduce the concrete measures. The function is adopted from the example of DR. De Jong's paper. The mathematical model are as follows: [3]

$$\begin{cases} \text{Min}(f(x_1,x_2)) = 100(x_1^2 - x_2)^2+(1 - x_1)^2 \\ -2.048 \leqslant x_i \leqslant 2.048 \quad i = 1, 2 \end{cases} \qquad (1)$$

This model is a two-dimensional single-peak function. It is morbid and hard to maximize, and it always sticks at local optima. The optimum is 0.

The Genetic Algorithm in the example adopts binary coding, the principal of gamblers selecting, fixed crossover operator, fixed mutation operator. The operation parameters needed to be set are population size (S), crossover operator (Pc), mutation operator (Pm), maximum number of generations (E), max fitness (F) and chromosome length (L). At the same time, there are five levels for each factor, so the parameter is set into "six factors- five levels" of optimum design. Five levels of each factor are as following:

TABLE I. THE LEVEL OF EACH FACTOR

Factor		Level				
Mark	*Name*	*1*	*2*	*3*	*4*	5
Pm	crossover operator	0.05	0.01	0.005	0.002	0.001
Pc	mutation operator	0.6	0.65	0.7	0.75	0.8
S	population size	60	64	68	72	76
E	maximum number of generations	320	340	360	380	400
L	chromosome length	20	22	24	26	28
F	max fitness	0.98	0.985	0.99	0.995	1

According to the table above, six factors-5 levels orthogonal array chart $L25(6^5)$ is built. By the chart, two tests to each scheme are processed, that is to say there are 50 tests totally. The result of test is the fitness of optimal individual. The target value is 1 in the example. That is, the closer the individual fitness is to 1, the better. The orthogonal array chart and the results are as following:

TABLE II. THE ORTHOGONAL ARRAY CHART AND THE RESULTS

Scheme	*Pm*	*Pc*	*S*	*E*	*L*	*F*	*First test*	*Second test*
1	1	1	1	1	1	1	0.845704	0.80738
2	1	2	2	2	2	2	0.927249	0.928955
3	1	3	3	3	3	3	0.854666	0.874238
4	1	4	4	4	4	4	0.799569	0.911474
5	1	5	5	5	5	5	0.852717	0.667609
6	2	1	2	3	4	5	0.846791	0.807485
7	2	2	3	4	5	1	0.995275	0.975804
8	2	3	4	5	1	2	0.848118	0.98566
9	2	4	5	1	2	3	0.899518	0.990071
10	2	5	1	2	3	4	0.436369	0.591061
11	3	1	3	5	2	4	0.840299	0.992838
12	3	2	4	1	3	5	0.978018	0.852121
13	3	3	5	2	4	1	0.918663	0.98886
14	3	4	1	3	5	2	0.988301	0.862865
15	3	5	2	4	1	3	0.828048	0.785862
16	4	1	4	2	5	3	0.971705	0.977454
17	4	2	5	3	1	4	0.961688	0.80933
18	4	3	1	4	2	5	0.656494	0.859372
19	4	4	2	5	3	1	0.545238	0.570177
20	4	5	3	1	4	2	0.595676	0.86243
21	5	1	5	4	3	2	0.999099	0.978921
22	5	2	1	5	4	3	0.972693	0.929261
23	5	3	2	1	5	4	0.985816	0.997557
24	5	4	3	2	1	5	0.836473	0.995119
25	5	5	4	3	2	1	0.916593	0.976243

According to the value of the table, we put the values into the software to analysis. Markway statistical analysis software is used in the example. Check list of effects among groups for Analysis of variance is as following:

TABLE III. THE CHECK LIST OF GROUP EFFECTS OFTHE ANALYSIS OF VARIANCE

source	*sum of square*	*df*	*Mean square*	*F*	*Sig*
crossover operator	0.1842	4	0.0461	6.9665	0.0007
mutation operator	0.2096	4	0.0524	7.9271	0.0003
population size	0.1197	4	0.0299	4.5288	0.0069
maximum number of generations	0.0312	4	0.0078	1.1818	0.3429
chromosome length	0.1449	4	0.0362	5.4793	0.0026
max fitness	0.0501	4	0.0125	1.8935	0.1429
crossover operator	0.1653	25	0.0066		
Totality	0.9050	49			

In the table above, we can see that mutation operator, crossover operator, population size and chromosome length which are the main factors affecting the performance of algorithm as their significances are below 0.05.

The other factors have little influence on the algorithm. Obviously more attention should be put on setting the value of main four factors.

According to the table II and the character of orthogonal test, the mean of K of each factor is as following：

TABLE IV. THE MEAN OF K OF EACH FACTOR

	Pm	Pc	S	E	L	F
K1	0.846956	0.906768	0.79495	0.881429	0.870338	0.853994
K2	0.837615	0.933039	0.822318	0.857191	0.898763	0.897727
K3	0.903587	0.896944	0.882282	0.88982	0.767991	0.908352
K4	0.780956	0.83988	0.921695	0.878992	0.86329	0.8326
K5	0.958777	0.751261	0.906648	0.820461	0.92751	0.83522

We inform that the level 5 is best for Pm, the level 2 is best for Pc, the level 4 is best for S, and the level 5 is best for L. So the optimum condition of keeping the algorithm stability is that: population size is 72, and mutation operator is 0.001, and crossover operator is 0.65, and the number of chromosome is 28.

V. THE VERIFICATION OF THE OPTIMIZATION DEGREE OF THE ALGORITHM

Adopting the method of comparing experimental data may intuitively reflect the above method. The mean is to set the parameter randomly and then get the test results used to compare with the results by using the method above.

Random settings: Pm0.05, Pc0.8, Population size 60, maximum number of generations 380, chromosome length 24, max fitness 0.995.

Analysis of variance settings： Pm0.001， Pc0.65， Population size 72， maximum number of generations 360， chromosome length 28， max fitness 0.995。

The algorithm uses the same hardware and software. The difference is only the algorithm parameter. The results are as following:

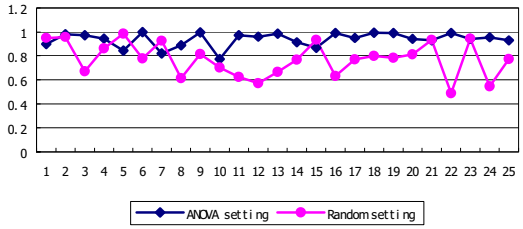

Figure 1. The sketch map of experiment results

We can learn from the chart that the adaptive value of optimal individual by randomly setting is from 0.4 to 1 and fluctuates severely; but after optimization setting, the adaptive value of optimal individual is from 0.8 to 1 and fluctuates rarely. So settings the parameters by the analysis of variance can make algorithm stable and easy to get the optimum solution.

VI. CONCLUSION

The article provides an method of Genetic Algorithm parameters settings based on Analysis of variance and it utilizes the well-rounded Statistics Tools, and the complete hypothesis means has been obtained. The solutions make the algorithm get perfect performance and make the settings operate easily without the subjective experience, and it has very good practical application value.

REFERENCES

[1] HE Da-kuo,WANG Fu-li,ZHANG Chun-mei.Establishment of Parameters of Genetic Algorithm Based on Uniform[J].Journal of Northeastern University (Natural Science),2003,24(5):409-411.

[2] Shitomatu. Fuzzy satisfactory method for electric power plant coal put chase using genetic algorithms [J].European Journal of Operational Research, 2000,126(1):218-230.

[3] YuYuLu,JianNengChen, Xiang Zhang,An improved genetic algorithm and its convergence.[J]Journal of Anhui Agricultural University,2007,34(4), 608-612

[4] Li M J,Fan S S,Tong T S. The application of partheno-genetic algorithm in pattern clustering problem [J].Engineering Applications of Artificial Intelligence, 1999,12(2):175-184.

[5] XiaoDongCHEN,HongYuWANG,A exam-PaPer generating algorithm based on imProved genetic aigorithm.[J]Journal of Harbin Institute of Technology,2005,37(9),1174-1176.

[6] Hui Wang, Joyce Liang and C.-C. Jay Kuo, A Secure Steganography Method based on Genetic Algorithm, Journal of Information Hiding and Multimedia Signal Processing, Vol. 1, No. 1, pp. 28-35, Jan. 2010.