http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (22): 6807-6811, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.6807.6811

Analysis of Variance for A Influence of Genetic Probability on the Convergence times of Genetic Algorithm

Nie XingXin, Lu CaiWu, Liu ShuXiang and Shen Yang Yang
College of management Xi'an University of Architecture and Technology, China

Abstract: The three-factor about replication, Exchange, mutation probabilities and their interaction is important on the convergence of the genetic algorithm. In order to analysis the effect. This study designed an three-factor cross-group experiment of three probability about reproduction, crossover and mutation, we use analysis of variance on empirical datum to obtain the effect of reproduction, crossover and mutation on convergence of Genetic algorithm. Conclusion of study show that, the influence of reproduction is the most significant, the influence of crossover is the second, the influence of mutation is not significant. But, as the main way of producing new individuality, Reproduction probability is important for getting overall situation optimal solution. In addition, the instrumental error is great which show that the calculate result of Genetic algorithm is randomness.

Key words: Genetic algorithm, variance analysis, Genetic probability, convergence times, crisscross inheritance

INTRODUCTION

Three basic operations of genetic algorithm are reproduction, crossover and mutation which guaranteed searching global optimal solution directionally (Yun *et al*., 1997). The three Operations dominated by certain probabilities which conclude reproduction probability (P_y), crossover probability (P_c), mutation probability (P_m), whether the three big probabilities choose reasonable or not directly affect the stand or fall of genetic algorithm convergence(Chao, 2007) The long period of study found that different genetic probability have a big influence on genetic algorithm, the result mainly displays in: first, converging quickly, but converging to local solution, not the global optimal; Second, converging too slowly, maybe a few hundred generations, even thousands of generations and sometimes simply cannot converge. Therefore, genetic algorithm convergence can not be too short or too long, it must converge to the global optimal solution in the appropriate times (as short as possible).

From the group, in order to measure the good or bad of genetic probability three factors must be considered: convergence times (g), the convergence rate (α), convergence of maximum fitness value (f_{max}). That is, in the proper convergence times converge to the global optimal solution according to the requirements in the convergence rate. This is also the standard that chooses genetic probability.

The convergence times of genetic algorithm follow lognormal distribution (Su, 2012), but it is effected by many factors, it's hard to see how much influence genetic probability on it from a variety of measurement results. We can see various factors to the influence degree of the genetic algorithm only by the analysis of variance for genetic probability.

Considering the many factors, the differences of determination value are bound to exist in different factors and different level, this calls variation. The variation is usually caused by the change of various factors and system error. We usually express it by the sum of deviation squares between the determination value x_i and their average (Yu and Zuo, 2010):

$$Q = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad (1)$$

Meanwhile, the test results are influenced by many factors, so each factors could make corresponding contribution to the total deviation(Yu and Zuo, 2010). Sum of deviation squares is used to represent the errors of test results, then the total sum of deviation squares equal to the sum of sum of deviation squares caused by a single factor, called the additivity of the sum of deviation squares which is the basis of analysis of variance. According to the principle of the additivity of sum of deviation squares, on the basis of sum of deviation squares decomposition and with the aid of F tests, we analyze the influences of the each genetic probability and their interaction on the total sum of deviation squares, this calls variance analysis. The variance analysis is one of the basic method of dealing with many factors experimental data(Chen *et al*., 2006).

Corresponding Author: Nie XingXin, College of management Xi'an University of Architecture and Technology, China

COMPREHENSIVE EXPERIMENTAL DESIGN BASED ON THREE GENETIC PROBABILITIES CROSS GROUP

Genetic probability includes reproduction probability, crossover probability and mutation probability, so the variance analysis based on comprehensive experimental data of three genetic probabilities cross classification could be used to determine the influences of the three genetic probabilities and their interaction on convergence times of genetic algorithm (Zhang *et al*., 2005) The experimental arrangement is shown in Table 1.

Each of the sum of deviation squares mainly includes (Yun *et al*., 1997):

- Total sum of deviation squares Q_T
- Sum of deviation squares Q_r caused by the factors P_y change in test level
- Sum of deviation squares Q_c caused by the factors P_c change in test level
- Sum of deviation squares Q_m caused by the factors P_m change in test level
- Sum of deviation squares Q_{rc}, Q_{rm}, Q_{cm} caused by factors P_r and P_c, P_y and P_m, P_c and P_m mutual influence
- Sum of deviation squares Q_{rc} caused by factors P_r, P_c and P_m mutual influence
- Test error effect Q_E.

The calculation method of each of sum of deviation squares is as follows:

$$Q_T = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{m}\sum_{q=1}^{n}(x_{ijkq}-\bar{x})^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{m}\sum_{q=1}^{n}x_{ijkq}^2 - \frac{T^2}{N}$$

$$Q_r = cmn\sum_{i=1}^{r}(\bar{x}_i-\bar{x})^2 = \frac{1}{cmn}\sum_{i=1}^{r}T_i^2 - \frac{T^2}{N}$$

$$Q_c = rmn\sum_{j=1}^{c}(\bar{x}_j-\bar{x})^2 = \frac{1}{rmn}\sum_{j=1}^{c}T_j^2 - \frac{T^2}{N}$$

$$Q_m = rcn\sum_{k=1}^{m}(\bar{x}_k-\bar{x})^2 = \frac{1}{rcn}\sum_{k=1}^{m}T_k^2 - \frac{T^2}{N}$$

$$Q_{rc} = mn\sum_{i=1}^{r}\sum_{j=1}^{c}(\bar{x}_{ij}-\bar{x}_i-\bar{x}_j+\bar{x})^2 = \frac{1}{mn}\sum_{i=1}^{r}\sum_{j=1}^{c}T_{ij}^2 - \frac{1}{cmn}\sum_{i=1}^{r}T_i^2 - \frac{1}{rmn}\sum_{j=1}^{c}T_j^2 + \frac{T^2}{N}$$

$$Q_{rm} = cn\sum_{i=1}^{r}\sum_{k-1}^{m}(\bar{x}_{ik}-\bar{x}_i-\bar{x}_k+\bar{x})^2 = \frac{1}{cn}\sum_{i=1}^{r}\sum_{k=1}^{m}T_{ik}^2 - \frac{1}{cmn}\sum_{i=1}^{r}T_i^2 - \frac{1}{rcn}\sum_{k=1}^{m}T_k^2 + \frac{T^2}{N}$$

$$Q_{cm} = rn\sum_{j=1}^{c}\sum_{k=1}^{m}(\bar{x}_{jk}-\bar{x}_j-\bar{x}_k+\bar{x})^2 = \frac{1}{rn}\sum_{j=1}^{c}\sum_{k=1}^{m}T_{jk}^2 - \frac{1}{rmn}\sum_{j=1}^{c}T_j^2 - \frac{1}{rcn}\sum_{k=1}^{m}T_k^2 + \frac{T^2}{N}$$

Table 1: Comprehensive test schedule based on the three genetic probabilities cross group and same testing times

		P_{m1}	P_{m2}	…	P_{mm}
P_{r1}	P_{c1}	$X_{1111}X_{1112}\cdots X_{111n}$	$X_{1121}X_{1122}\cdots X_{112n}$	…	$X_{11m1}X_{11m2}\cdots X_{11mn}$
	P_{c2}	$X_{1211}X_{1212}\cdots X_{121n}$	$X_{1221}X_{1222}\cdots X_{111n}$		$X_{12m1}X_{12m2}\cdots X_{12mn}$
	…	…	…		…
	P_{cc}	$X_{1c11}X_{1c12}\cdots X_{1c1n}$	$X_{1c21}X_{1c22}\cdots X_{1c2n}$		$X_{1cm1}X_{1cm2}\cdots X_{1cmn}$
P_{r2}	P_{c1}	$X_{2111}X_{2112}\cdots X_{211n}$	$X_{2121}X_{2122}\cdots X_{212n}$	…	$X_{21m1}X_{21m2}\cdots X_{21mn}$
	P_{c2}	$X_{2211}X_{2212}\cdots X_{221n}$	$X_{2221}X_{2222}\cdots X_{211n}$		$X_{22m1}X_{22m2}\cdots X_{22mn}$
	…	…	…		…
	P_{cc}	$X_{2c11}X_{2c12}\cdots X_{2c1n}$	$X_{2c21}X_{2c22}\cdots X_{2c2n}$		$X_{2cm1}X_{2cm2}\cdots X_{2cmn}$
…	…	…	…	…	…
P_{rr}	P_{c1}	$X_{r111}X_{1112}\cdots X_{r11n}$	$X_{r121}X_{r122}\cdots X_{r12n}$	…	$X_{r1m1}X_{r1m2}\cdots X_{r1mn}$
	P_{c2}	$X_{r211}X_{1212}\cdots X_{121n}$	$X_{r221}X_{r222}\cdots X_{r11n}$		$X_{r2m1}X_{r2m2}\cdots X_{r2mn}$
	…	…	…		…
	P_{cc}	$X_{rc11}X_{rc12}\cdots X_{rc1n}$	$X_{rc21}X_{rc22}\cdots X_{rc2n}$		$X_{rcm1}X_{rcm2}\cdots X_{rcmn}$

$$Q_{rcm} = cn\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{m}(x_{ijk}-\bar{x}_{ij}-\bar{x}_{ik}-\bar{x}_{jk}+\bar{x}_i+\bar{x}_j+\bar{x}_k-\bar{x})^2$$
$$= \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{m}T_{ijk}^2 - \frac{1}{mn}\sum_{i=1}^{r}\sum_{j=1}^{c}T_{ij}^2 - \frac{1}{cn}\sum_{i=1}^{r}\sum_{k=1}^{m}T_{ik}^2 - \frac{1}{rn}\sum_{j=1}^{c}\sum_{k=1}^{m}T_{jk}^2$$
$$+ \frac{1}{cmn}\sum_{i=1}^{r}T_i^2 + \frac{1}{rmn}\sum_{j=1}^{c}T_j^2 + \frac{1}{rcn}\sum_{k=1}^{m}T_k^2 - \frac{T^2}{N}$$

$$Q_E = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{m}\sum_{q=1}^{n}(x_{ijkq}-\bar{x}_{ijk})^2 = \sum_{i=1}^{r}\sum_{j=}^{c}\sum_{k=1}^{m}\sum_{q=1}^{n}X_{ijkq}^2 - \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{m}T_{ijk}^2$$

According to the additivity of the sum of deviation squares (Yun *et al*., 1997):

$$Q_T = Q_r+Q_c+Q_m+Q_{rc}+Q_m+Q_{cm}+Q_{rcm}+Q_E$$

The above r, c, m represent the level for factor p_r, factor p_c, factor p_m. n is the number of repeat test frequency.

Among them:

$$T = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{m}\sum_{q=1}^{n}x_{ijkq} \qquad T_{ij} = \sum_{k=1}^{m}\sum_{q=1}^{n}x_{kq}$$
$$T_i = \sum_{j=1}^{c}\sum_{k=1}^{m}\sum_{q=1}^{n}x_{jkq} \qquad T_{ik} = \sum_{j=1}^{c}\sum_{q=1}^{n}x_{jq}$$
$$T_j = \sum_{i=1}^{r}\sum_{k=1}^{m}\sum_{q=1}^{n}x_{ikq} \qquad T_{jk} = \sum_{i=1}^{r}\sum_{q=1}^{n}x_{rq}$$
$$T_k = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{q=1}^{n}x_{ijq} \qquad T_{ijk} = \sum_{q=1}^{n}x_q$$
$$N = rcmn$$

The degrees of freedom of each of the sum of deviation squares is: the degrees of freedom of the factor P_y, P_c, P_m (Jian *et al*., 2013) are $f_y = r-1$, $f_y = r-1$, $f_m = m-1$; The degrees of freedom of the interaction between factor P_y and P_c (Fallah-Jamshidi *et al*., 2010), factor P_y and P_m, factor P_c and P_m are: $f_{rc} = (r-1)(c-1)$, $f_m = (r-1)(m-1)$, $f_{cm} = (c-1)(m-1)$; the degree of freedom of the interaction between all of them is $f_{rmc} = (r-1)(c-1)(m-1)$ (Tan *et al*., 2013); the degrees of freedom of the sum of deviation squares

which has deviated effects is f_{rmc} = (r-1)(c-1)(m-1); the degrees of freedom of the total sum of deviation squares is f_T = rcnm-1.

VARIANCE ANALYSIS OF SAMPLE DATA

Steps of variance analysis: The steps of variance analysis based on the data of three genetic probabilities and same testing times(Bai *et al*., 2009):

- Calculate the sum of deviation square Q according to the sample data
- Calculate the degree of freedom f for each the sum of deviation square Q
- Calculate each the estimated variance S^2
- Calculate each F value according to the estimated variance and the degree of freedom;

Choose the significance level α, then look up the critical value F_α from F distribution table according to the corresponding degree of freedom;

Compare F with F_α, $F<F_\alpha$ means it has no obvious effect on the convergence of genetic algorithm, $F\geq F_\alpha$ means it has a obvious effect on the convergence of genetic algorithm. meanwhile, $F\geq F_{0.05}$ means the effect is of obviousness, $F\geq F_{0.01}$ means it has a significant effect on the convergence of genetic algorithm(Jin and Shan, 2013).

Variance analysis: Variance analysis is showed in Table 2.

Determination of the test conditions: The parameters are selected within reason, in order to ensure that the genetic algorithm can converge in normal times. Select 0.04,0.05 and 0.06 for Reproduction probability p_r,0.5,0.55 and 0.6 for crossover probability p_c; 0.005,0.006 for mutation probability p_m, then r = 3, c = 3, m = 2. 10 for repeated test times in each level, n = 10.

Result analysis: Convergent times that in different index and level are showed in Table 3 base on perform simulation.

Because of the convergence time of the genetic algorithm follow lognormal distribution, the sample data should be translated according to the Eq. 2 before discuss them with analysis of variance:

Table 2: Variance analysis schedule based on the three genetic probabilities and same testing times

Source of variation	Sum of deviation square	Degree of freedom	Estimated variance	F value
P_r	Q_r	f_r = m-1	$S^2_r = Q_r/f_r$	S^2_r/S^2_E
P_c	Q_c	F_c = c-1	$S^2_c = Q_c/f_c$	S^2_c/S^2_E
P_m	Q_m	F_m = m-1	$S^2_m = Q_m/f_m$	S^2_m/S^2_E
$P_r\times P_c$	Q_{rc}	f_{rc} = (r-1)(c-1)	$S^2_{rc} = Q_{rc}/f_{rc}$	S^2_{rc}/S^2_E
$P_r\times P_m$	Q_{rm}	f_{rm} = (r-1)(m-1)	$S^2_{rm} = Q_{rm}/f_{rm}$	S^2_{rm}/S^2_E
$P_c\times P_m$	Q_{cm}	F_{cm} = (c-1)(m-1)	$S^2_{cm} = Q_{cm}/f_{cm}$	S^2_{cm}/S^2_E
$P_r\times P_c\times P_m$	Q_{rcm}	f_{rcm} = (r-1)(c-1)(m-1)	$S^2_{rcm} = Q_{rcm}/f_{rcm}$	S^2_{rcm}/S^2_E
Trial error	Q_E	F_E = rcm(n-1)	$S^2_E = Q_E/f_E$	
Sum	Q_T	F_T = rcmn-1		

Table 3: Convergent time in different genetic probability

	Pr1(0.04)			Pr2(0.05)			Pr3(0.06)		
	P_{c1}(0.5)	P_{c2}(0.55)	P_{c3}(0.6)	P_{c1}(0.5)	P_{c2}(0.55)	P_{c3}(0.6)	P_{c1}(0.5)	P_{c2}(0.55)	P_{c3}(0.6)
Pm (0.005)	43	147	51	32	50	49	84	34	34
	53	52	63	49	32	29	49	35	59
	40	44	81	25	28	64	44	52	49
	34	36	85	48	32	40	52	27	61
	61	110	70	57	74	49	27	35	32
	47	110	121	46	46	45	43	71	65
	58	22	56	40	63	32	76	39	27
	53	23	74	22	64	26	75	33	49
	56	62	40	34	37	48	24	50	50
	84	45	38	45	38	61	33	30	41
Pm (0.006)	59	47	88	75	76	28	28	67	134
	107	126	61	48	72	52	46	110	81
	108	61	85	44	28	64	75	78	105
	75	54	77	107	103	106	68	68	71
	45	71	45	120	49	65	94	46	66
	95	120	44	66	58	28	110	110	59
	161	37	164	70	31	67	87	76	37
	80	113	86	38	67	28	76	51	121
	81	75	98	61	41	59	38	30	60
	155	55	50	91	68	54	37	90	46

Table 4: Computation sheet

		Sum of squares
T	183.74	33759.50
T_i	71.41 53.25 59.08	11425.04
T_j	62.39 59.12 62.22	11259.95
T_k	75.80 107.94	17396.29
T_{ij}	24.41 22.37 24.63 18.56 17.96 16.73 19.42 18.79 20.87	3815.45
T_{ik}	30.88 40.53 21.80 31.44 23.12 35.97	5888.12
T_{jk}	24.26 38.13 24.35 34.77 37.18 35.04	5811.25
T_{ijk}	9.39 15.01 9.87 12.50 11.61 13.02 6.43 12.13 7.86 10.10 7.51 9.22 8.44 10.99 6.22 12.17 8.06 12.81	1978.07

i = 1, 2, 3; j = 1, 2, 3; k = 1, 2

Table 5: Result of three-factor variance analysis

Source of variation	Sum of deviation square	Degree of freedom	Estimated variance	F-value	$F_\alpha(n1, n2)$
P_r	2.845134930	2	1.432256746	9.076885569	3.06
P_c	0.113045911	2	0.056522956	0.358212591	3.06
P_m	5.739348533	1	5.739348533	36.37295478	3.84
$P_r \times P_c$	0.242187722	4	0.060546930	0.383714415	2.37
$P_r \times P_m$	0.113937278	2	0.056968639	0.361037096	3.06
$P_c \times P_m$	0.303084855	2	0.151542428	0.960395738	3.06
$P_r \times P_c \times P_m$	0.878376841	4	0.219594210	1.391671935	2.37
Trial error	25.56224722	162	0.157791650		
Sum	35.81674186	179			

$$x'_i = \ln(x_i) \quad (2)$$

Analysis of variance are carried out for the translated data, the process and the result are showed in Table 4 or 5.

May see from the result of Table 5:

- The F value of P_m is $36.373 > F_{0.01}(1,162)$ that means it has a significant effect on the convergence of genetic algorithm
- The F value of P_r is $9.0769 > F_{0.01}(2, 162)$ that means it has a significant effect on the convergence of genetic algorithm
- The F value of P_c is $0.3582 < F_{0.05}(2, 162)$ that means it has no obvious effect on the convergence of genetic algorithm

According to estimated variance, trial errors take up a great proportion in the genetic algorithm that means the genetic algorithm has much randomness. There are many uncontrolled factors in the genetic algorithm is the main reason that causing the random error. For example, choosing mutations, creating the first generation, choosing crossover points, etc.

CONCLUSION

In the genetic algorithm, reproduction, crossover or mutation is of randomness, so that the result has the obvious random. In order to measure the convergence of the result, convergence rate, convergence time and convergence of maximum fitness value could be treated as the main basis for researching the genetic algorithm.

Using variance analysis proved that mutation probability have the most significant effect on the convergence time of the genetic algorithm. The second is reproduction probability. And the effort of the crossover probability is not noticeable. Although crossover probability has no obvious effect on the convergence time, but it has an important contribution for searching the global optimal solution comes from crossover probability as the primary way to creat a new generation. It mainly displays in the p_c value is two or three amount levels higher than the p_r or p_m value. The rapid creation of new generation is attributed to the higher crossover probability. Studies show that if p_c is small, then the global optimal solution of genetic algorithm is hard to find.

In addition, errors of the estimated variance take up a great proportion in the whole calculation process. The result fully proves the randomness of the calculation in genetic algorithm. So the preferred method to obtain the conclusions is statistical analysis.

ACKNOWLEDGMENT

This study was supported in part by (Shaanxi Funding Project sponsor: No.E08001,E08003,2012JM9004).

REFERENCES

Bai, X.J., G.H. Peng and X. Chen, 2009. Bezier curve fitting based on genetic algorithm and steepest descent algorithm. Comput. Eng. Design.

Chao, B.Z., 2007. Statistical Theory and Methods for Surveying Data Processing. Press Ganey Associates Inc., South Bend, IN., pp: 15-18.

Chen, G.L., X.F. Wang and Z.Q. Zhuang, 2006. Theory and Applications of Genetic Algorithms. People's Posts and Telecommunications Press, Beijing, pp: 32-36.

Fallah-Jamshidi, S., M. Amiri and N. Karimi, 2010. Nonlinear continuous multi-response problems: A novel two-phase hybrid genetic based metaheuristic. Applied Soft Comput., 10: 1274-1283.

Jian, J., X. Wang and X.L. Ge, 2013. The optimization of capacity constraints base on adaptive genetic algorithm for vehicle dispatching. J. Chongqing Univ., 3: 112-114.

Jin, F. and R. Shan, 2013. Hybrid genetic algorithm based on an effective local search technique. J. Liaoning Techn. Univ. (Natl. Sci.), 32: 43-44.

Su, K., 2012. Genetic algorithm based research on constrained optimization problems with discretely distributed decision spaces. North China Electr. Power Univ., 2: 33-47.

Tan, Y.H., S.H. Chen, G.M. Zhang and Z.T. Xiong, 2013. Adaptive impedance matching using quantum genetic algorithm. J. Central South Univ., 20: 977-981.

Yu, S.Y. and S.Q. Zuo, 2010. Convergence and convergence rate analysis of elitist genetic algorithm based on martingale approach. Control Theory Appl., 7: 843-848.

Yun, Q.X., G.Q. Huang and Z.Q. Wang, 1997. Genetic Algorithm and Genetic Programming. Metallurgical Industry Press, Beijing, pp:24-32.

Zhang, X.W., Z.D. Xing and J.M. Dong, 2005. A hybrid genetic algorithm to seeking the optimum solution of a class of unconstrained optimization. J. Northwest Univ. (Natl. Sci. Edn.), 35: 130-132.