An improved genetic algorithm for Intelligent test paper generation

Nie Jun
Department of Computer Science，GuangDong College of science and Technology Dongguan 523083, China
13739149@qq.com

***Abstract*—Considering the problem on generating test papers is multi-objective parameter optimization under multiple constraints.I proposed a new improved genetic algorithm based on the researches of the mathematical model of generating test paper after encoding segmented chromosome, confirming adaptability function, segmented group initialization, altering adaptive crossover probability and mutation probability and conserve optimization individuals. This method implemented generating test paper well, and the experimental results show that this improved genetic algorithm is more practical and effective compared to the common algorithm in the same conditions.**

Keywords: Improved Genetic; Intelligent Test Paper Generation; Adaptive

I. INTRODUCTION

Examination play an important role in teaching, paperless automation test becoming increasingly prevalent in all walks of life, rich in questions, and Intelligent test paper generation is one of the key steps to implement automated test.Intelligent test paper auto-generating is of a high quality paper which satisfy multiple constraints set by author after selecting questions from the bank of questions. Some algorithms have been considered, such as select-random algorithm, backtrack algorithm, and genetic algorithm. But select-random and backtrack algorithm are not of intelligent, they always fail to compose test paper because of they can't conform to local constraints. However, they consuming lots of time and space in the condition of much questions. Genetic algorithm is a natural evolutionary process simulation model which have the characteristics of concurrency, intelligent search, robust, simplicity.

It can well solve the shortage of the above two algorithms. However, the standard genetic algorithm also has poor stability, the defects of conflicts between multiple constraints that cannot achieve the expectation. The method I proposed implemented the algorithm of intelligent test paper generation based on the researches of the mathematical model of generating test paper after encoding segmented chromosome, confirming adaptability function, segmented group initialization, altering adaptive crossover probability and mutation probability and conserve optimization individual. The experimental results show that this improved genetic algorithm is more practical and effective compared to the common algorithm.

II. THE MATHMETICAL MODEL OF GENERATING TEST PAPER

The problem of generating test paper contains the following attributes as constraints: item types, scores, difficulty, their respective points of knowledge, discrimination, estimated answer time, which satisfies these constraints at most according to the algorithm. In the test paper composition, designed an examination paper that is the decision of a matrix(n×m).In this matrix, n is the question number,m show that each question including m kinds of attribute.m represent item types,scores,difficulty,their respective points of knowledge,discrimination,estimated answer time respectively.Therefore,test paper can represented by n×m matrix:

$$X=\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{1m} \\ a_{21} & a_{22} & a_{23} & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & a_{n3} & a_{nm} \end{bmatrix}$$

The distribution of elements in the matrix columns need to meet the general requirements specified by the user, namely item types,scores,difficulty,difficulty,their respective points of knowledge,discrimination,estimated answer time. And the symbol a_{i1} is the question number of every question of 1st column, a_{i2} is the question scores of every question of second column.a_{i3} is the question difficulty of every question of 3th column.a_{i4} is the question decriminalization of every question of fourth column.a_{i5} is the question's points knowledge of fifth column.a_{i6} is the question estimated answer time of every question of sixth column. Then the target matrix should satisfy the following constraints:

(1)Total scores：$S=\sum_{i=1}^{n} a_{i2}$，a_{i2} denotes the scores of the i-th question;

(2)The difficulty coefficient :$W=\sum_{i=1}^{n} a_{i2}a_{i3}/S$，$a_{i2}$ denotes the scores of the i-th question.a_{i3} denotes the difficulty of the i-th question.

(3)Discrimination：$Z=\sum_{i=1}^{n} a_{i2}a_{i4}/S$，$a_{i2}$ denotes the scores of the i-th question.a_{i4} denotes the discrimination of the i-th question.

(4)The coverage of points of knowledge ：$Q=\sum_{i=1}^{n} a_{i5}/K_t$，$K_t$ denotes the total points of knowledge which setted by user;

(5)Total time：$T=\sum_{i=1}^{n} a_{i6}$，a_{i6} denotes the time of each question.

III. TEST PAPER COMPOSITION STRATEGY

The performance of generated initial population at random is poor, the quality individuals is easily damaged

978-1-4799-6636-3/14 $31.00 © 2014 IEEE
DOI 10.1109/ICICTA.2014.25

after selection, the crossover and mutation operations, and the disadvantage of the standard genetic algorithm is unstable and slow convergence. The intelligent test paper composition proposed an improved genetic algorithm to solve these defects, strategies are as follows:

(1)The question types, the number of every question of each types, total scores, difficulty coefficiency, and total points of knowledge, discrimination, and total time are setted by the user as the constraints of the target papers, and then set the termination conditions of the algorithm and the size of initialization population of the papers.

(2)Real-Encoding in segment for the solution of the problem.

(3)Determination of individual fitness function.

(4)Generation of the initial population in some constraints.

(5)Operation of select individual.

(6)Adjustment-auto the crossover probability for next crossover operation.

(7)Adjustment-auto the mutation probability for next mutation operation.

(8)Execution the strategies of optimal conservation.

(9)If the condition of termination is valid, go to step (10), otherwise go to step (5).

(10) Output the optimal test paper after end of algorithm.

IV. The Implementation of Intelligent test paper composition

A. Chromosomes coding

The traditional binary encoding that has large calculation can't show problems directly, the length of chromosome increased quickly and the speed of the algorithm runs slow with the increasing question in papers. This algorithm improved the performance of search and the efficiency of solving problems based on real-encoding in segment. A paper mapped to a chromosome, each question of papers mapped to the gene of chromosome and the value of gene is the number of question. Since the item types of test paper independent ,the question number of every kind of question types remain unchanged.Therefore,each segmented coding execute the operation of crossover and mutation in the same kinds of question types based on segmented coding of question types. Assuming the number of question in a paper is n, question types is k.Hence the papers needs n bits to encode, and the encode divided into k sections according to question types. suppose s_i is the number of question of i-th question types, then the chromosome encoding is C(1)C(2)C(3) …C(s1)C(s1+1)…C(n).Where, $\sum_{i=1}^{n} s_i = n$,C(i) denotes the selected question number of question database.

B. confirmation of individual fitness function

Fitness function which has an affect on the efficiency of test paper composition is used to measure the performance of individuals in papers population. In general, the better the performance of the individual the larger the fitness value, and vice versa. The genetic algorithm main work is to maximize matching three indicators which are difficulty,decriminalization,coverage points of knowledge, because the number and scores of question types in papers matrix settled which conform to requirement proposed by user.So,we use the fitness function: $f(x) = \sum_{i=1}^{4} W_i Q_i(x)$,where f(x) is the fitness function,$Q_i(x)$ is the i-th constraint function of test paper composition,W_i is the weight of Q_i,the more important $Q_i(x)$,the larger the value of W_i and i=1.Where $Q_1(x)$, $Q_2(x)$,$Q_3(x)$ is difficulty function, discrimination function, and coverage points of knowledge function respectively.

$$Q_1(X) = \frac{1}{|D_s - \sum_{i=1}^{n} a_{i2} a_{i3} / \sum_{i=1}^{n} a_{i2} + 0.1|}$$

$$Q_2(X) = \sum_{i=1}^{n} a_{i2} a_{i4} / \sum_{i=1}^{n} a_{i2}$$

$$Q_3(X) = \sum_{i=1}^{n} a_{i5} / K_t$$

Where D_s is the difficulty coefficiency set by users,K_t is total points of knowledge set by users.

C. Population initialization of test paper

Since the test paper is divided into different blocks based on question types and the bit string of chromosome is independent, this algorithm adopted a segmented initial population to avoid the singularity of population. We initial 5 attributes of the population of test paper according to question types, the number of question types of every question, difficulty efficiency, total scores and total time.

The operation of population initialization of test paper as follow:

Step 1: generate a question number in question types which in question database compared to the question number generated in segment. If same generate it again, otherwise, keep it.Repeat this process to produce the total number of test paper.

Step 2: Comparing the answer time settled by user to the answer time generated by paper. If the difference out of the range set by user, then regenerate the test paper, otherwise go to step 3.

Step 4: Comparing the total scores of test paper composition to the total scores set by user, if they are not equal, then regenerate test paper, otherwise go to step 4.

Step 5: Comparing the difficulty efficiency of test paper composition to difficulty efficiency set by user, if the difference in the scope set by user, then generates the test paper, otherwise regenerate it.

Considering the efficiency and performance of test paper composition, we set the initial population value from 70~200,if the value is over larger, it can led to long calculation time, on the contrary, the value is too small, it can appear premature.

D.Genetic operation

1)selection operation

Selection operation is a process of generating new population by selecting strong individuals in population for best individuals in population. This algorithm adopted a random selection of competition through generating two individuals by selective mechanism to compete, and then the higher fitness value will be inherited to the next generation.

2)The operation adaptive crossover and mutation

The probability of crossover P_c and mutation Pm has an important affect on genetic algorithm and determines the convergence of genetic algorithm, then it influence the performance directly.P_c is too large, excellent individuals are easy to destroyed,P_c is too small, the diversity of the population decreased, the slow search of GA is easy to fall into local optimum.P_m is too large, the GA became the problem of search randomly and affect the convergence rate,P_m is too small, the population evolutionary slow and is not easy to produce a new individual.Generally,we set P_c=0.4~0.99,P_m=0.001~0.1.

We use the probability of adaptive crossover and mutation, P_c and P_m can adjust their value in a reasonable range according to the actual situation of population. Set the probability of adaptive crossover as:

$$P_c = \begin{cases} P_{c1} - \dfrac{(P_{c1} - P_{c2}) \times (f_h - \overline{f})}{f_{\max} - \overline{f}} & f_h \geq \overline{f} \\ P_{c1} & f_h < \overline{f} \end{cases}$$

Where, $\overline{f}$ is the average fitness value of current population,f_h is the larger of the two individuals fitness value,f_{max} is the maximum fitness value of the current population, here P_{c1}=0.95,P_{c2}=0.65.Adaptive mutation probability is set as follows:

$$P_m = \begin{cases} P_{m1} - \dfrac{(P_{m1} - P_{m2}) \times (f_{\max} - f)}{f_{\max} - \overline{f}} & f \geq \overline{f} \\ P_1 & f < \overline{f} \end{cases}$$

Where f_{max} is the largest fitness value of the current population, f is the fitness value of individual participate in mutation and is the average fitness value of the current population, here P_{m1}=0.05,P_{m2}=0.001. If paper generates a random probability P less than that of generating adaptive mutation P_m, execute the mutation, otherwise do nothing.

3)The memory strategies of excellent individual

When the probability of the crossover and mutation becomes more and more, it could damage the excellent individuals. This paper adopted the memory strategies of excellent individual based on GA to ensure the diversity of population and accelerate the converges to the optimal solution. The specific process as follows: first, find out the highest and lowest fitness value of offspring in the population and the highest fitness value of parent in the population. Second, Comparing their value, if the highest fitness value of parent population is larger than its son's value, then parent's value will substitute for the lowest fitness value of offspring in the population.

4)The condition of termination

If the fitness value of the individual in papers is greater than(equal to)the fitness value set by user in the number of iterations, the expected test paper is the best one and the algorithm executed successfully; If the fitness value fail to meet the value set by user out of iterations, executed failed and then restart the algorithm.

V. Experimental Analysis

In the experiment,we set the size of initial population, the maximum number of iterations, estimated time, total scores of test paper as 90,200,100,100 respectively. We create a question bank of 1000,where have 400 multiple-choice questions, 200 fill in the blank questions,200 judgment questions and 200 comprehensive questions. The standard GA and improved GA were executed in the experiment for test paper composition that come out the comparison of highest fitness value of the population in different iterations, see figure 1,and the comparison of running time of different iterations in population, see figure 2.

The proposed improved genetic algorithm in this paper, to some extent, improved the efficiency of test paper composition, we can see this improvement from the data in Figures 1 and 2.The convergence rate of improved GA is better than the standard GA,However,the algorithm is not very perfect, where the conformation of parameters need to test over and again by our experience.

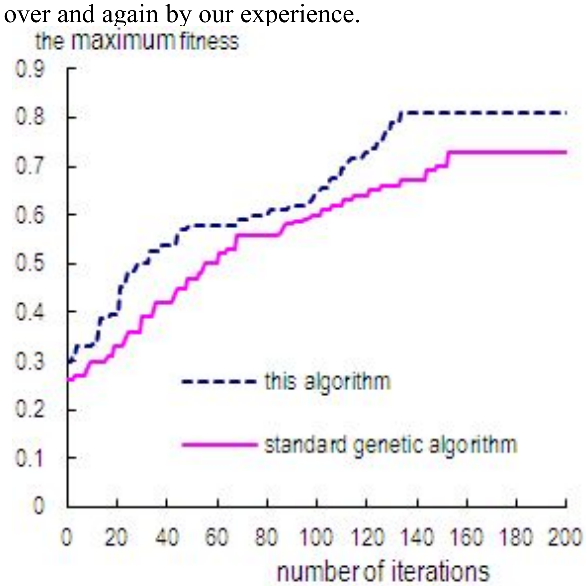

Figure 1 the comparison of the maximum in different number of iterations

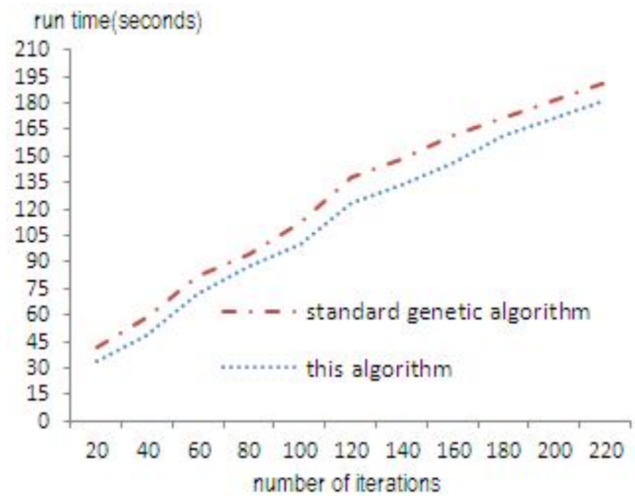

Figure 2 The comparison of the running time in different number of iterations

VI. CONCLUSION

The intelligent test paper composition is a typical multi-constraint optimization problem. This algorithm of intelligent test paper composition generated by improving generation of initial population, individual coding, crossover and mutation in the standard genetic algorithm. It has proved that this improved algorithm raised the performance of optimization search and efficiency of test paper generation in the case of provided appropriate question bank, and meet the requirements better.

REFERENCES

[1] ZhU Jian-bing,LI Zhan-huai,ZhAO Na. Research on the Problem of Auto- composing Test Paper Based on Hybrid Genetic Algorithm, ,2009,26(5):328~35

[2] Xu Yong-da.An intelligent generating examination paper method based on improved genetic algorithm: Computer & Digital Engineering, 2013（2）:176-178,207

[3] Wang Yi-ping,Qu Wei-jian,Pan Hai-zhu. Research on adaptive item bank system based on genetic algorithm. Journal of Qiqihar University(Natural Science Edition), 2008,24(5):1-4

[4] Geng Xia .Research of genetic algorithm automatic ttest paper composition based on knowledge point constraint.Microcomputer & Its Applications, 2010: 2010(7):27-33

[5] DU Hua,WANG Suo-zhu.Web sites structure optimization model and algorithm analysis.Computer Engineering and Design, 2008,29(21):5591-5594

[6] WU Xu,HONG Lian-Xi,WEI De-zhi,GUO Jian-ping.Improved AFSA for solving intelligent test problem.Joumal of Jimei University(atural Science),2012,17(5):394-400