Real-coded Genetic Algorithm and Application in the Automatic Composing the Test Paper＊

WANG Yu-Fen，GUO Xiao-Juan,

School of Information Engineering
Henan Institute of Science and Technology
Xinxiang, Henan Province 453003, China
E-mail: wangyufen@hist.edu.cn

Abstract—**The genetic algorithm is one kind of recent development parallel optimization algorithm, it suitable to solve the problem of automatic composing test paper. But binary-coded genetic algorithm has some disadvantages, such as lower operating speed, precocious convergence and so on. For these flaws, based on real-coded genetic algorithm has been put forward, and has been applied in the automatic composing test paper system of the computer basis, its performance is superior to binary-coded genetic algorithm.**

Keywords-genetic algorithm; real-coded; automatic composing test paper

I.   INTRODUCTION

The automatic composing test paper is one core goals of test system automation, according to the requirement of the teacher and the teaching, the computer chooses the test question automatically from test library, and composes the examination paper conformed to the request, this is the important component of computer aided teaching. How to guarantee the inborn examination paper can satisfy the various requirements of users, and it has   randomness, the scientific and the rationality, which is a difficulty to realize.

The genetic algorithm is one kind of parallel optimization algorithm which recent development, it very suitable to solve automatic composing test paper [1]. Using the genetic algorithm to choose the test question, can optimize the search processing and speed up the rapidity of the composing test paper. In order to overcome the disadvantages of the genetic algorithm of the binary code, such as running slow and premature restraining, we proposed the genetic algorithm based on real code according to the characteristic and the request of the automatic composing test paper.

II.   THE GENETIC ALGORITHM BASED ON REAL CODE

A.   Genetic Algorithms

The genetic algorithm, the shortened form is GA, it is one kind of search and optimized method, which developed by simulating the mechanism of life evolution. It was proposed by the Michigan University's Professor John Holland in his bookmaking "Adaption in Natural and Artificial System". At present, GA has been used broad in various fields, such as the PID control, the linearity and the non-linearity, optimization, robust, auto-adapted, the sliding mode control, the fuzzy logic, the neural network, the parameter estimation and the system identification, the model linearization and the controller depreciate, robot arm controlling and track layout and so on.

The genetic algorithm is one search algorithm, which may use in the complicated optimizing system with the robust. Compare with the traditional optimized algorithm, such as parsing law, the numeration, the enumeration, the stochastic search and so on , it mainly has the following characteristic:① The genetic algorithm takes the decision variable's code as it＇s operation object .②The genetic algorithm takes the fitness as it＇s search information, and does not need other auxiliary information like differential coefficient and so on other.③The genetic algorithm uses the search information of many spots, which has the concealed parallelism.④Genetic algorithm use probability search technique, but not the non-definite rule.

1） The basic principle of genetic algorithm

In the process of nature evolutionary, though the heredity (propagates, resemble of descendant and person of father's generation) and the variation (descendant and person of father's generation resemble not completely), the organism adapt the external environment, using the rule of superior win and the inferior wash out by a generation of another generation of place, to evolve. GA simulates the above evolution phenomenon, it takes the search space (solution space of wanting the search space to solve question) to map the heredity space, namely each possibility's solution code is a vector, it is called a chromosome or the individual, it expressed the binary or the decimal, each element of the vector is called the gene, all genomes become the community or the group; According to the fitness, some heredity operations are implemented to various chromosomes, such as choice, exchange, variation and so on. and eliminating the chromosome of lowly sufficiency lowly (performance not good), leaves some chromosome that high sufficiency (performance fine). Thus the new community can be obtained. Because the new community's members are the previous generation community's outstanding candidates, which has inherited previous generation's fine condition, thus surpasses the previous generation obviously; GA repeatedly like this iterates, turns to the direction evolution of more superior solution, until satisfies some kind of predetermined optimized target.

2） The flaw of Binary code genetic algorithm

IEEE
computer society

Using the method of the binary encoding, the genetic algorithm adopts some heredity operation like selection, crossover and the mutation to simulate natural evolution, and complete the optimization. Regarding the complex optimization of multivariable, the SGA code string is longer, and it causes some problem, such as more computation, lower accuracy , longer computing time, unitary heredity operator and so on, the algorithm is bad to the different question's compatibility, the process of optimization easily to fall into partially superior and appears convergence precociously[2][4].

B. The genetic algorithm based on real number code

In order to overcome the shortcoming of binary code, we use the decimal which can carry on the code directly. So we can import the enlightened information relative question field to increase the search ability of the genetic algorithm. Using the real code, some operation like mapping the original question and the decoding of computing fitness, and the most of all, it reduces the length of string code. The experiment proved that for the optimization question of the majority of value, and the heredity operator introduction design specially, the average efficiency of real code compared to the binary code are higher [4]. Moreover, regarding the optimized question of continuous variable, real number expression has more accuracy of computation and will not produce the influence to the encoding method

1） Initial population's design

The size of population as one of controlled parameter of genetic algorithm, it is influential for the efficiency to exert of the genetic algorithm. If population is big, the model processed by the heredity operation is bigger, and it will produce the meaningful gene block and change into the optimal solution by the evolution the opportunity to be higher. In other words, the community scale is bigger, the community's multiplicity is higher, it is smaller for danger that the algorithm falls into the partial solution. But the community scale is too bigger, the algorithm computation load also is increased. Thinking from the speed of computation, the community number of individual scope is generally several dozens to several hundred in the practical application. In addition, if the initial population is produced by the predetermined constraints, which will form subspace of the limited chromosome constitution, the population has feasible solution. Thus may reduce the search space greatly, and enhance the convergence rapidity.

2） Fitness function design

Because the genetic algorithm needs to choose probability based on the individual sufficiency foundation, therefore the fitness function's value must be positive. Thus it is necessary that, maps the objective function to the form of the maximum value or the minimum value and the adaptability function with nonnegative function value.

The model of he automatic composing test paper is the minimum question, generally, the transformation methods as following:

$$f(x) = \begin{cases} C_{max} - g(x) & g(x) < C_{max} \\ 0 & \text{when } x \text{ is invalid chromosome} \end{cases}$$

in the formula: f(x) is the fitness function; g(x) is the objective function; C_{max} may be a suitable input value, and use the maximum value of g(x) of evolution process [4].

3） Selection design

Generally speaking, the selection strategy will affect the algorithm performance and the result. This article has used the selection strategy which combined the fitness proportion method with the best individual preservation method. The fitness proportion method is the most foundation and the most commonly used selection method at present genetic algorithm. In this method, each individual's selection probability is proportional to its fitness value.

Supposes the population size is n, the fitness value of individual i is f_i, the selected probability P_i:　$P_i = \dfrac{f_i}{\sum\limits_{j=1}^{n} f_i}$

Probability P_i reflects the individual fitness proportion which occupies in the entire individual fitness total [4]. According to the size of individual fitness value, the value is bigger, its probability P_i is bigger, the opportunity which is selected is also more. Thus, the possibility that its gene structure is inherited to the next generation is also bigger. But, the fitness proportion method has certain probability in the heredity process to cause the most superior individual to eliminate, therefore this article combines the fitness proportion method with the best individual preservation method, each generation most superior individual directly substitute poorest individual, this method guarantees the genetic algorithm to the most superior direction evolution.

4） Crossover design

The crossover can explore the new gene space, thus enables the new community individual to have the multiplicity, so it plays the most role in the genetic evolution. The genetic algorithm mainly has sole hybrid way like the arithmetic hybrid, the inspiration hybrid and the simple hybrid and so on. This article has used the mix hybrid which is composed of the above three kind of hybrid. Namely according to certain frequency this article simultaneously uses three kind of hybrid, which increases the crossover effect, simultaneously to protect previous generation's best individual, after undergoing the choice in the new population, the design will retain a best individual but not crossover. The binary-coded genetic algorithm only uses the sole crossover way, as the algorithm runs to certain generation, the individual tends consistent, the crossover has weaken the new individual ability.

5） Mutation design

In order to strengthen the mutation partial search ability, individual which fitness is big should mutate in the small scope, but individual which fitness is small should mutate in the big scope. As the concrete operations, the most superior individual will reduce its variation scope, simultaneously to avoid algorithm restraining to being partial most superior, the individual of fitness small will increase its variation scope as

well as the variation probability, to guarantee the individual multiplicity in the community.

III. REAL-CODED GENETIC ALGORITHM APPLICATION IN AUTOMATIC COMPOSING TEST PAPER SYSTEM

According to user's requirement, the automatic composing test paper is that extracts certain amount test question automatically from question library using certain algorithm. Automatic composing test paper algorithm quality influence immediately examination paper quality. It is current research popular topic that how to compose test paper satisfied the user's needs and have the high efficiency and the success ratio from question library.

A. Composing test paper mathematical model

1）Composing test paper constraint condition
Composing test paper involves the primary factor:
- Topic type: the test paper contains topic type.
- Knowledge point: each topic belongs to which knowledge of some curriculum and a various knowledge point occupies score and proportion.
- Question quantity: the paper contains question number.
- Difficulty: the overall difficulty coefficient of paper, namely the all topic difficulty of paper is produced by the weighted averaged.
- Distinguishing degree: the examination paper separating capacity, which distinguish tested groups, the paper all distinguishing degree come from all test questions distinguishing degree that is produced by the weighted averaged.
- Answer time: the users use the longest time which completes all test question of the examination paper. In the ordinary circumstances, the answer time and the topic quantity are being related.

According to the different test goal and the type, the constraints may increase or the reduction.

2）The mathematical model of composing test paper
Based on the above analysis of composing test paper constraints, composing test paper mathematical model is build. Assumed that there are six bound variables in an examination paper, all topic of the examination paper constitute a six dimensions variable space: topic type, knowledge point, question quantity, difficulty, distinguishing degree, answer time.

Every topic of test question library constitutes a point of six dimensions space. To compose test is to choose some points from the space, this points constitute need solutions, which should satisfy given condition, namely meet the made request of every dimension.

Therefore, whenever to decide a topic in the examination paper, six dimensions are considered. If an examination paper which contain n topic is formed, it selects n topics from the test question library, n topics composite indicator must meet certain conditions. Each topic measure is regarded as a six dimensions vector (a1, a2, a3, a4, a5, a6), then an examination paper of n topics can be expressed an n × 6 matrix S.

$$S = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ \vdots & & & & & \\ \vdots & & & & & \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & a_{n5} & a_{n6} \end{bmatrix}$$

In order to reduce the algorithm complexity, simplification can be dealt with. Because composing test paper must classify by topic type, in view of the identical topic need not to consider again the topic type question, this will reduce a dimension variable. Also because the topic quantity restraint may obtain need at the beginning of the composing test paper, when to search other restraint in examination paper, so long as maintaining the topic quantity is invariable, the quantity restraint will not need to take further searches, like this reduces a dimension variable, becomes a four dimensions space variable: the identical topic type test question (knowledge point, difficulty, distinguishing degree, answer time).

Here, test question search of composing test paper predigests a search in four dimensions space.

B. Automatic composing test paper system realization process
The process of automatic composing test paper system based on real-coded genetic algorithm is as shown in Figure 1.
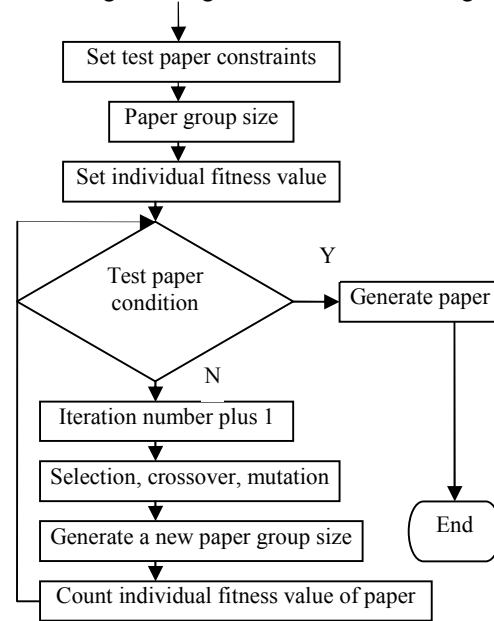

Fig.1 The flow of the automatic composing test paper system based on real -coded genetic algorithm

IV. REAL-CODED GENETIC ALGORITHM WITH BINARY-CODED GENETIC ALGORITHM COMPOSING TEST PAPER PERFORMANCE COMPARISON

Based on the above methods, the author achieves model of two test papers based on real-coded genetic algorithm and binary-coded genetic algorithm for the computer basis examination database. In the experiment, composing test paper uses groups size of 200, the termination of iterations is 200, choose a probability of 0.8, 0.2 crossover probability, mutation probability is 0.05. The experimental result has been shown in table 1.

TABLE I.   THE PERFORMANCE COMPARE OF TWO ALGORITHMS

Number of tests	Iterations	
	Binary-coded genetic algorithm	Real-coded genetic algorithm
1	194	171
2	165	133
3	170	161
4	183	152
5	185	162
Average	179.4	155.8

As can be seen from the table, composing test paper based on real-coded genetic algorithm average iterative number is 155.8, while the use of binary-coded genetic algorithm received to iterations of 179.4, the new algorithm is superior to the old algorithm.

V. CONCLUSION

Research shows that based on real-coded genetic algorithm has made up deficiency of based on binary-coded genetic algorithm, has improved composing test paper performance. Yet the automatic composing test paper system also has the insufficiency, the problem has been further study. Like some dimension have not been considered, the teaching content level which the student grasps, the test question to student knowledge and ability request and so on.

REFERENCES

[1] Mao Bingyi. Research on Database Structure of Composing Test Paper Intelligently System Based on GA[J]. Computer Engineering and Application. 2003 (6): 230-232.

[2] Li Yong, Cao Guangyi, Zhu Xinjian. A Multiple Crossover Real-coded Genetic Algorithm[J]. Computer Simulation, 2006(6): 166-169.

[3] Lynda T, Chrisment C,Mohand B H. Multiple Query Evaluation Based on Enhanced Genetic Algorithm[J].Information Processing and Management(ISSN:0306-4573), 2003, 39(2): 215-231.

[4] Li Peizhi, Fan Ding. Research on Real Coding-based Improved Genetic Algorithm[J]. Journal of Astronautic Metrology and Measurement. 2008 (1): 54-57.

[5] Zhou Ming, Sun Shudong. GA Theory and Application[M]. National Defense Industry Press, 2002.

[6] Li Minqiang. The Basic Theory and Application on GA[M]. Science Press, 2002.