

2012 AASRI Conference on Computational Intelligence and Bioinformatics

Test Paper Generating Method Based on Genetic Algorithm

Li Yan^{a,*}, Li Shuhong^b, Li Xiurong^c

^a*Continuation Education College, Henan Institute of Science & Technology, Xinxiang, 453003, China*

^b*Training Department, Xinxiang Vocational and Technical College, Xinxiang, 453006, China*

^c*Guangshan Experimental Elementary School, Guangshan, 465450, China*

Abstract

The continuation education is very important for people who have left school to work to increase their competence and skills. To avoid the disadvantages of the common test paper generating methods, genetic algorithm is used to generate the test paper automatically. The concrete design process of test paper generating based on genetic algorithm is discussed in this paper, and some corresponding parameters setting have been compared and defined. The application results demonstrated that the genetic algorithm was an effective tool in the test paper generating.

2012 Published by Elsevier B.V. Selection and/or peer review under responsibility of American Applied Science Research Institute Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Continuation education; Test paper generating; Genetic algorithm; Genetic operator

1. Introduction

In the continuation education, the exam, which is the key stage in the course of the whole teaching process, plays an important role in estimating the competence and learning effects of the students, and is also an effective method of measuring the teaching effects of the particular teacher. However, there are some disadvantages in the common method of test paper generating, and thus leading to obvious deviation from the reasonable distribution of the different parts of the specific subject.

* Corresponding author. Tel.: +86-373-3040135.
E-mail address: lxr6060@163.com.

With the rapid development and application of the modern computer technology, a new kind of cross subject—computer aided teaching (CAT)—is coming into use in the continuation education field. By means of this kind of technology, the test paper can be generated automatically with the help of the computer, thus the disadvantages of the traditional test paper generating method can be overcome, and the teachers' time and efforts can be saved.

The test paper generating system is an expert system to generate the test paper automatically according to such specified parameters as the number and the type of the exam questions, the degree of difficulty, the scope of knowledge, etc. Most of the traditional test paper generating algorithms are easy to lead to local optimal solution and slow convergence. In this paper, we present an efficient intelligent paper generating method based on the genetic algorithm according to the mathematical model and better effects have been obtained.

2. Principle of Genetic Algorithm

Genetic Algorithm (GA) is a kind of stochastic search process that is capable of adaptive and robust searching over a wide range of searching space, which first appeared in the Ph.D. thesis of Bagay in 1967. In 1975, genetic algorithm was first introduced by Prof. Holland for use and admitted formally. A typical genetic algorithm in its simple form works with a set of N data structures called individuals, each representing a separate solution to the specific problem [1-2].

In genetic algorithm, the calculations are performed with a population of individuals in following steps

- (1) Generate randomly an initial population of possible solutions (candidates chromosomes);
- (2) Decode the chromosomes to obtain a set of parameters for each individual;
- (3) Select a pair of individuals according to the fitness function, which is a measure of the quality of this solution;
- (4) Parent individuals are selected for reproduction proportional to a fitness value and mutate each bit of each parents with mutation probability;
- (5) Each pair of parents creates a pair of new individuals called offsprings, partially by recombining (crossover), partially by mutating chromosomes of the best-fitted individuals;
- (6) Repeat steps 4 and 5 until the stopping criterion is satisfied.

3. Design method

3.1. Description of Algorithm

Suppose y_i , h_i , j_i ($j=1, 2, 3 \dots k$, k is the levels of the teaching difficulties, here $k=5$) represent the ratio of scores with difficulty i required by the customer, the ratio of scores with difficulty i in the practical generated test paper, and the permission error of the ratio of scores with difficulty i , which is evaluated according to formulas (1) and (2).

$$f_1 = \sum_{i=1}^k t_i \quad (1)$$

$$t_i = \begin{cases} 0, & |y_i - b_i| \leq h_i \\ |y_i - b_i - h_i|, & |y_i - b_i| > h_i \end{cases} \quad (2)$$

It can be concluded from formulas (1) and (2) that the smaller f is, the smaller the error between the generated test paper and what the customer requires is. The test questions in the test question base may be indexed according to the types of the test questions, it's easy to meet the customers' demands if there are

enough amounts of test questions in the test question base [3]. Therefore, the object function of solving the problem of test paper generating can be obtained to be

$$\text{Min}(f)=f_1 \quad (3)$$

3.2. Design Process

(1) Chromosomes coding and population initialization

To solve a problem with genetic algorithm, the solutions space of the problem should first be mapped to be a set of codes. The binary code is used in the classical genetic algorithm, which “1” means a question item is selected, and “0” means a question item is not selected. This is an easy coding method, but the question number can not be controlled exactly and the coding is very long when the test base is very large. If the real number coding is utilized, the optimization efficiency will be increased effectively.

(2) Adaption function

In genetic algorithm, the bigger the adaption values, the better the individual. The test paper generating model is a minimal problem in this paper. The object function can be transformed to be the adaption function F' according to formula (4):

$$F' = \begin{cases} 100 - f, & f < 100 \\ 0, & f \geq 100 \end{cases} \quad (4)$$

The exponent ratio will allow the individuals to have more duplicating chances, and it also restrict the duplicating numbers [4-6]. So the above formula can be transformed to be the following formula:

$$F = e^{-\beta F'} \quad (5)$$

where β is selected to be 0.03.

(3) Design of genetic operator

Compute the number of times that the individuals are expected to be selected according to formula (6)

$$N_i = \frac{MF_i}{\sum_{i=1}^M F_i} \quad (i = 1, 2, \dots, M) \quad (6)$$

where M is the population scale, F_i is the adaption value of the i^{th} individual.

Match the individuals that are selected according to the above method in pairs randomly, every pair is crossed evenly and exchanged at a certain cross probability P_c , therefore two new individuals are generated [7, 8]. First randomly generate a real number r_1 which is between 0 and 1, if $r_1 < P_c$ and the exchange condition is satisfied, the two individuals are exchanged and two new individuals are generated, otherwise they are not exchanged.

The effective evaluating index of a particular test paper is difficulty D , differentiation degree, reliability, and effectiveness. For the difficulty of a particular test paper, it can be described as the mean score rate P which is shown in formula (7), or the difficulty weighted mean of the test problem which is shown in formula (8), or the mean score of the high scores H and the low scores L which is shown in formula (9).

$$D = P = \frac{\bar{X}}{C} \quad (7)$$

$$D = \frac{\sum_{j=1}^m D_j X_j}{C} \quad (8)$$

$$D = \frac{\bar{H} + \bar{L}}{2C} \quad (9)$$

where \bar{X} is the mean score, m is the number of the test problem, D_j is the difficulty of the j^{th} test problem, X_j is the given score of the j^{th} test problem, C is the full mark of the test paper, H and L are the mean scores of the high scores and the low scores, respectively.

Differentiation degree Q of the test paper is usually expressed according to formula (10) or (11) [9-11]

$$Q = \frac{\sum_{j=1}^m D_j X_j}{C} \quad (10)$$

$$Q = \frac{\bar{H} + \bar{L}}{C} \quad (11)$$

Reliability R is one of the key factors of evaluating a specific test paper, and R is usually denoted to be formula (12)

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (12)$$

where \bar{X} is the mean score, X_i is the test score of the i^{th} student, Y_i is the test score of the i^{th} student in the equivalent test, \bar{Y} is the mean score in the equivalent test.

(4) Standardization of test paper generating

The index of the test problems is the important factor in the automatic test paper generating, including the full mark FM , the test range, the test time TT , and the types of the test problems, etc.

The distribution curve, which is also called the index curve, is the ratio of the corresponding index in the test paper. The corresponding distribution curves are shown in table 1, 2, and 3, respectively.

Table 1. Type-score distribution LT

Type of test problem	T1	T2	T3	...	Tn
Number of test problem	N1	N2	N3	...	Nn
Score	S1	S2	S3	...	Sn

where $S1+S2+S3+\dots+Sn=FM$

Table 2. Difficulty distribution LD

Difficulty level	D1	D2	D3	...	Dn
Score	S1	S2	S3	...	Sn
Error	E1	E2	E3	...	En

where $S1+S2+S3+\dots+Sn=FM$

Table 3. Teaching requirement-score distribution LR

Teaching requirement	1	2	3	...	n
Score	S1	S2	S3	...	Sn
Error	E1	E2	E3	...	En

where $S1+S2+S3+\dots+Sn=FM$

Applying the above genetic algorithm and the corresponding parameters setting to the automatic test paper generating system, a better effect has been obtained in several subjects of the test paper. In addition, the overall efficiency has been increased to a certain extent.

4. Conclusion

In the design of the test paper generating system based on genetic algorithm, an independent coding strategy is used, which is fitted to the practical condition of generating test paper. By means of selecting such parameters as the variation operator, the adaption crossing, and the weighted adaption function properly, there are no needs to control the problems' type, the problems' number, or the problems' mark, and the efficiency has been increased to a great extent. On the basis of the effects obtained from practical application, it was demonstrated that the genetic algorithm is a powerful tool for parameter optimization, which can be successfully applied in many fields.

References

- [1] J.D. Bagley, The behavior of adaptive systems which employ genetic and correlation algorithms, Dissertation Abstracts International 1968; 28 (12).
- [2] D.E. Goldberg, Genetic Algorithm—In Search, Optimization and Machine Learning, Addison Wesley, New York, 1989.
- [3] Murthy, C.A., Chowdhury, N., 1996. In search of optimal clusters using genetic algorithm. Pattern Recognition Lett. 17, 825-832.
- [4] Yu Xiaoqiang, Liu Bai, Xie Yiwu. Strategy and realization of auto-generating paper based on genetic algorithm. Microelectronics and Computer, 2006, 23(12):189-191.
- [5] J.R. Koza, Genetic Programming II: Automatic Discovery of Reusable Programs, MIT Press, Cambridge, 1994.
- [6] W. Banzhaf, P. Nordin, R. Keller, F. Francone, Genetic Programming—An Introduction, Morgan Kaufmann Publishers Inc, San Francisco and dpunkt Verlag, Heidelberg, 1998, 330-334.
- [7] D. Abramson, J. Abela, A parallel genetic algorithm for solving the school timetabling problem, in: Proceedings of the Fifteenth Australian Computer Science Conference (ACSC-15) 14, 1992, 1-11.
- [8] Sven E. Eklund. A massively parallel architecture for distributed genetic algorithms. Parallel Computing. 2004; 30: 647-676.
- [9] D. Abramson, G. Mills, S. Perkins, Parallelization of a genetic algorithm for the computation of efficient train schedules, Proceedings of the 1993 Parallel Computing and Transputers Conference, 1993, 139-149.
- [10] Dybowski R, Weller P, Chang R, Gant V. Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. Lancet 1997;347:1146-50.
- [11] Dorsey, R.E., Mayer, W.J. Genetic algorithms for estimation problems with multiple optima, nondifferentiability and other irregular features. Journal of Business and Economic Statistics, 1995; 13: 53-66.