TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.4, April 2014, pp. 3240 ~ 3246
DOI: http://dx.doi.org/10.11591/telkomnika.v12i4.4888

Application of Genetic Algorithm in Intelligent Examination Paper Composition

Yan Yan*[1], Dali Yin[2]
Department of Computer Science and Technology, ChangChun University of Science and Technology
ChangChun, China
*Corresponding author, e-mail: yanyanxueer@126.com[1], yindali2008@126.com[2]

Abstract

With the development of computer technology and artificial intelligence, the research into examination paper composition receives more and more concern, and the artificial composition of examination paper based on the genetic algorithm gradually demonstrates its superiority. However, due to the complexity of the restraints of paper composition, it has many shortcomings. This paper focuses on the application of the genetic algorithm in paper composition, and proposes an improved tactic of arithmetic crossing, which enlarges the searching space of algorithm to a considerable extent, avoiding the defects of the slow convergence of the simple genetic algorithm and regional convergence so that the efficiency and quality of paper composition are raised.

Keywords: *genetic algorithm, intelligent examination paper composition, mathematical model, arithmetic crossing*

1. Research Background

With the development of computer technology in every fields, as a new subject, the intelligent computer aided instruction (ICAI), which unites computer science, education, mathematics and other discipline in itself, receives more and more concern, while intelligent examination paper composition is an important part of the ICAI, therefore it has a significant research value. The intelligent examination paper composition is a question of optimization of multi-object parameters under certain restraints, whose efficiency and quality are wholly determined by the design of the test paper base and its question-drawing algorithm. The genetic algorithm has the advantages of simplicity, strong robustness and overall optimum-seeking, and applies to the question of multi-object solution. All these qualities determine that the genetic algorithm is very adaptable to solving problems in paper composition. However, the present genetic algorithm cannot deal with the problem of restrictive contradiction between the speed of convergence and the capacity of searching. The present research of the genetic algorithm aims at improving the genetic mechanism of algorithm, for example, revision of the genetic operators of the algorithm.

2. Intelligent Paper Composition

2.1. Examination Question Attributes

The composition of a paper involves many factors, and there are 6 common attributes that are related to paper composition shown in the following:

1) Question number
The unique code of question in the question base.

2) Type of question
According to content, there are usually multi-choice questions, blank filling-out questions, true-or-false questions, calculation questions and application questions, etc.

3) Score
Score of each question.

4) Coefficient of difficulty
Represent the index of difficulty of examination question.

5) Differentiation

Received September 3, 2013; Revised November 5, 2013; Accepted December 11, 2013

The paper with high differentiation can discriminate the tested of different levels effectively. There are many calculations of differentiation, of which the common one is the analysis of grouping high scores and low scores.

6) Knowledge spots belonging to

In order to be adapted to different textbooks, the content of a certain course can be divided into different knowledge testing spots according to the curriculum.

2.2. Model of Intelligent Paper Composition

The composition of paper usually takes many restraints into account, but too many restraints will add to the difficulty of composition. For that reason, this paper selects some core attributes as the restraints for paper composition (score, difficulty, differentiation, and knowledge spot). In this case, if we want to compose a paper containing m questions, an object-status matrix of m×4 should be set up:

$$S = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & a_{m3} & a_{m4} \end{bmatrix}$$

This is a solution of question of object-status matrix, and the object-status is not unique. Each column represents an attribute, needing to satisfy four restraints:

1) Total score of paper = $\sum_{i=1}^{m} a_{i1}$

2) Coefficient of difficulty of paper = $\sum_{i=1}^{m} a_{i1} a_{i2} \div totalscore$

3) Differentiation of paper = $\sum_{i=1}^{m} a_{i1} a_{i3} \div totalscore$

4) Coverage of knowledge spots of paper= $\sum_{i=1}^{m} a_{i4} \div$ total knowledge spots

In which the coverage of knowledge spots of paper refers to the proportion of the sum of knowledge spots in a paper to the total knowledge spots, while the total knowledge spots is given by the subscriber in advance.

3. Application of the Improved Genetic Algorithmin Intelligent Paper Composition

3.1. Coding of Chromosome and Genus Group Initialization

This paper adopts the tactic of sectional real number coding [1]. Sectional real number coding is to project a paper as a chromosome, with its gene value represented by its question number. The variables represented in this method are simple in operation, clear in meaning, unnecessary to be decoded, which can effectively improve the complexity of the genetic algorithm. In order to keep the question amount of each type of question in the genetic operation unchanged, we can divide the whole questions in the question base into different sub-blocks according to types of question, each sub-block corresponding to a type of question, and every type of question being independent from each other. Likewise, each chromosome is divided into different sub-blocks according to types of question.

Complete adoption of random can reduce the speed of constringency, so this paper will produce initial genus groups according to certain restraints, that is: the initial genus groups are produced according to the five restraints of type of question, amount of question, whether containing repeated question, the differential between the requirement of difficulty from the subscriber and the difficulty of the produced paper, the size of initial genus groups.

3.2. Function of Adaptation

By the analysis of the paper composition model, this paper sets the adaptation function as shown in formula (1):

$$F(x) = \sum_{i=1}^{3} W_i \ F_i(x) \tag{1}$$

In formula (1), F(x) stands for the function of adaptation, W(i) is the weight put on the i restraint in paper composition. The more important the restraint is, the bigger the value of weight will be. $F_{i(}x)$ is the i restraint, satisfying $\sum_{i=1}^{3} W_i$ =1. Specifically, $F_1(x)$ stands for the difficulty function, $F_2(x)$ stands for differentiation function, and $F_3(x)$ stands for the function of coverage of knowledge spots. In which:

$$F_1(x) = \frac{1}{|difficulty \operatorname{Re} quiredByPaper - AverageDifficulty + 0.1|} \tag{2}$$

In formula (2), the average difficulty = $1 - \frac{PeopleRightAtTheQuestion}{TotalTestedSample}$.

$$F_2(x) = \sum_{i=1}^{m} a_{i1} a_{i3} \div totalscore \tag{3}$$

In formula (3), m stands for the total number of questions in a paper, a_{i1} score of each question, while a_{i3} differentiation of each question.

$$F_3(x) = \frac{SumOfKnowledgeSpotsInAPaper}{TotalKnowledgeSpots} \tag{4}$$

In this paper, the increase of the difficulty function, differentiation function, knowledge spot coverage function of examination papers is favourable to the increase of F(x), the bigger F(x) is, the property of chromosome will be better, and the more suitable it will be to paper composition.

3.3. Genetic operator

3.3.1. Selection of Operators

This paper adopts the method of roulette selection[2], the probability of each individule's being chosen is directly proportional to its adaptation, whose operation is as follows:

a. Calculate the adaptation value of each individual in the genus group *F(i)* , $i = 1$,2,…,M, M is the size of the initial genus group;

b. Calculate the probability of each individual's being copied to the next generation, and express it as the formula:

$$P(i) = \frac{F(i)}{\sum_{i=1}^{M} F(i)}; \tag{5}$$

c. Calculate the accumulated probability of each roulette's interval, and express it as the formula: $[P'(j), P'(j+1)]$, in which, $P'(1) = P(1)$,when $j \geq 2$, $P'(j) = P(j) + P'(j-1)$, $j \in [2, M]$;

d. in the interval of [0,1] there will be a random number r;

e. If r< $P'(1)$, select the individual 1; otherwise, if $P'(j)$< r < $P'(j+1)$ holds, select the individual j+1;

f. Repeat the steps d, e, for the required times.

3.3.2. Cross Operator

This paper advances a kind of improved cross operator after summarizing the strong points and weak points of every kind of cross operators [5, 8]. Suppose any parental individuals are:

$$x(t) = \{x_1(t), x_2(t), \ldots, x_i(t), \ldots, x_m(t)\}$$
$$y(t) = \{y_1(t), y_2(t), \ldots, y_i(t), \ldots, y_m(t)\}$$

The filial individuals after crossing are:

$$x(t+1) = \{x_1(t+1), x_2(t+1), \ldots, x_i(t+1), \ldots, x_m(t+1)\} \quad (6)$$

$$y(t+1) = \{y_1(t+1), y_2(t+1), \ldots, x_i(t+1), \ldots, y_m(t+1)\} \quad (7)$$

The equations for the i dimension of the two filial operators in the improved cross operators:

$$x_i(t+1) = ax_i(t) + (1-a)y_i(t) \quad (8)$$

$$y_i(t+1) = ax_i(t) - (1-a)y_i(t) \quad (9)$$

In which, $a = \frac{F(x)}{F(x)+F(y)}$, and the adaptation value *F(x)* of the individual *x(t)* is bigger that the adaptation value F(y) of y(t), that is to say, the individual x(t) is better.

In the improved cross operators, the individuals produced in the formula (8) can not exceed the interval of parental individuals. When the individuals in the formula (9) exceeds the fixed range, a examination question can be generated randomly in its corresponding type of question, which can not only satisfy the reasonable requirement of the individual, but also ensure the searching ranges between the superior individuals, raising greatly the speed of the algorithm solution and avoiding the “prematurity” [12] of most individual’s tending to be uniformity.

3.3.3. Variant Operators

The variant operation in this paper is to randomly change certain question number in a paper with a very little probability, and adopts the method of variation in the same section of type of question, whose solution is as follows:

1) Suppose the probability of self-adaptation variation is pm;
2) Results in random probability p;
3) If p<p_m, then execute the variant operation, otherwise not execute the operation.In the execution of the operation, a random number will be generated, representing the position of individual’s variation, and then a question number will be generated randomly in each type of question, to replace the question number of the former individual’s variant position with it;
4) if the generated question number is the repetition of the existing question number of the individual, then it will generate a random number again, until there is no repetition of question number in each section of type of question.

3.4. Optimal Saving Tactics

When the adaptation of individual [6, 7] is too big, it will spoil the optimization of solution. With regards to this question, this paper adopts the optimal saving tactics [3] to save the optimum individual. The method is as follows: first calculate the maximum and minimum adaptation values of the former and new generation of genus group [15]. If the adaptation of the new generation is smaller than the former generation, then we will use the individuals of the former generation with the maximum adaptation to substitute for the individuals of the new generation with minimum adaptation. The adoption of the optimal saving tactics enables the present optimal individuals not to be damaged due to the execution of crossing and variant operation, which is an important safeguard for the improvement of convergence of self-adaptation heredity algorithm [9, 10].

3.5. Setting of Controlling Parameters

3.5.1. Scale of Genus Groups

According to the specific research findings, this paper set the value M at 100, that is, at the initialization of genus groups, 100 papers will be selected from the question base according to the question selection function.

3.5.2. Self-adaptation Crossing and Variant Probability

The fixed crossing probability has its defects [11]. At the early stage of evolution, cross operation can increase the variety of genus groups, and strengthen the searching capacity, but at the later period of evolution, it will unfortunately spoil the individual mode of the approaching optimum [13, 14] . As for this, this paper adopts self-adaptation cross probability. According to the characteristics of the heredity algorithm and observation of many time's experiment data, this paper design the following self-adaptation cross probability: if the adaptation of individuals is bigger or equal to the average adaptation of the parental individuals that involves in cross operation, then let the self-adaptation cross probability P_c=0.7*[$F(x)_{max}$-F(i)], otherwise P_c=0.9. In which, $F(x)_{max}$ represents the maximum adaptation of individuals of parental generation that involves in cross operation, F(i) stands for the adaptation of the i individual that involves in cross operation.

Likewise, the fixed variant probability also has its shortcomings. In the early stage of evolution, the adaptation of individual is not high, and variation can make it more likely for genus groups to generate superior individuals. But at the later period of evolution, it will damage some superior individual solutions. Therefore, the variant probability adopted in this paper is also self-adaptation probability.

The self-adaptation variant probability is designed as follows: when the adaptation of the individual that executes variant operation is bigger or equal to the average adaptation of all variant individuals, let P_m=0.2*[$F_2(x)_{max}$-$F_2(i)$], otherwise P_m =0.2. in which, $F_2(x)_{max}$ stands for the maximum adaptation of the individual in the former variant filial generation, $F_2(i)$ stands for the adaptation of the individual of the i filial generation before variation. The adaptation of individual will be determined by the adaptation of individual, which can well cure the defects of the fixed probability value.

3.5.3. Iterative Termination Conditions

This paper adopts cyclical fixed times as the condition of termination of algorithm, setting the maximum iterative times for 200, and attaches an extra condition, requiring the optimum adaptation of individual to reach a certain fixed value that is determined by the adaptation function according the question attribute, which is a functional value of larger adaptation. And what's more, this value can be adjusted according to the practical requirements in paper composition.

3.6. Process of Algorithm Realization

Process of algorithm is shown as Figure 1:

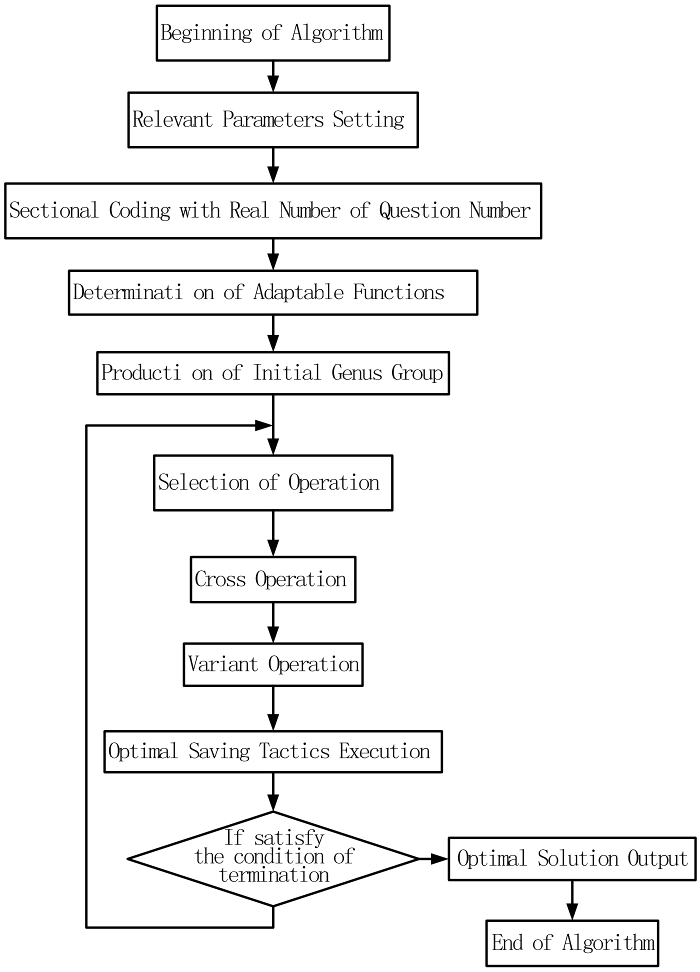

Figure 1. Flow Chart of the Improved Heredity Algorithm

3.7. Experiment Results and Analysis

The improved genetic algorithm has been realized in the simulated paper composition system. The paper composition system is developed in the .NET environment, with the background language of C#, and database SQL Serve 2000.

Some people in Anhui University proposes a paper composition algorithm of adopting grouped real number coding, roulette selection, sectional single-cross operators, sectional single variant operators, self-adaptation probability and optimal saving tactics [4]. In order to verify the superiority of the improved genetic algorithm in this paper, we have the two methods compared with each other. Table 1 is the comparison after ten time's operations of the two algorithms.

Table 1. Comparison of Paper Composition Algorithms

Algorithm	Average iterative times	Time of paper composition（s）	Average optimal adaptation
Compared algorithm	37.5	10.40	2.231
Improved algorithm in the paper	12.5	9.6747	2.246

From Table 1 we can see that the application of the improved genetic algorithm in paper composition improves the average iterative times, as well as the average time of paper composition and the average optimal adaptation. This is because the improved arithmetic crossing adopts the omnidirectional cross tactics, which enables the generated filial individuals to surround the superior ones and increase the searching space, which makes the searching adjust itself to the direction of the most likely optimum solutions, to the advantage of the evolution of genus groups, so that the best solution can easily be found, curing the defects of the narrow range of single-point cross and disadvantage of production of superior individuals.

4. Conclusion

This paper discusses the application of the genetic algorithm in the intelligent paper composition, improving the cross tactics of the algorithm. The improved cross tactics of algorithm greatly increases the searching space and ensures the variety of genus groups, making it easy for the algorithm to find solutions and raising the quality of the papers. It proves to be an effective method of paper composition.

References

[1] Wan Wenjun, Peng Huiuyun. An Improved Genetic Algorithm with Real Number and its Application. *Journal Vocational College of Electric Technology of Jiangxi.* 2007; 20(2): 71-73.
[2] Xia Guimei, Zeng Jianchao. A Random Particle Group Algorithm of Genetic Algorithms based on Roulette Selection. *Computer Engineering and Science*. 2007; 29(6): 51-54.
[3] Hop, Nguyen Van, Tabucanon, Mario T. Adaptive genetic algorithm for losing problem with self-adjustment operation rate. *Production Economics*. 2005; 98(2): 129-135.
[4] Wu Xiaoqin. Study and Application of Intelligent Paper Composition based on Genetic Algorithm. Thesis for the master's degree. Anhui: Anhui University. 2007.
[5] D Beasley, DR Bull, RR Martin. An over view of Genetic Algorithm: Part2, Research Topic. University Computing. 1993; 15(4): 170181.
[6] Michalewicz Z. *A* Modified Genetic Algorithm for Optimal Control Problems. *Computers Math Application*. 1992; 23(12): 83-94.
[7] Lynda, C Chrisment, Boughanem Mohand. Multiple Query Evaluation Based on Enhanced Genetic Algorithm. *Information Processing and Managent.* 2003; 39(2): 215-231.
[8] Felix TS, Chan, SH Chung, PLY Chan*.* An adaptive genetic algorithm with dominated genes for distributed scheduling problem. *Expert Systems with Applications*, 2005; 29(2): 364-371.
[9] Yuan Xiao-hui, Cao Ling, Xia Liang-zheng. Adaptive Genetic Algorithm with the Criterion of Premature Convergence. *Southeast University*. 2003; 19(1): 40-44.
[10] LIMei-yi, CAIzi-xing, SUNGuo-rong. An Adaptive Genetic Algorithm with Diversity-guided Mutation and its Global Convergence Property*. Central South University of Technology*. 2004; 11(3): 323-327.
[11] Mashhadi H, Shane chi HM, Lueas CA. New genetic algorithm with Lamarckian individual learning for generation scheduling. *IEEE Transaction Son Power Systems.* 2003; 18(3): 1181-1186.
[12] Xuesong Ya, Qinghua Wu. An Improved Genetic Algorithm and Its Application*. TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2012; 10(5): 1081-1086.
[13] VENKATRAMANS, YEN GG. A generic framework for constrained optimization using genetic algorithms. *IEEE Transactions on Evolutionary Computation.* 2005; 9(4): 424-435.
[14] CHIU CC, HSU PL. A constraint-based genetic algorithm for mining classification rules. *IEEE Transactions on SMC.* 2005; 35(2): 205-220.
[15] Jian Ni, Ning-Ning Yang. Genetic Algorithm and Its Application in Scheduling System. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2012; 11(4): 1934-1939.