Research on Computer Test Paper Generation Algorithm Based on Gene Expression Programming

Yang Wenyan
Dalian Vocational&Technical College, 116035
Achievements of scientific research projects of Dalian Vocational&Technical College (Dalian Radio and Television University) in 2020
yan171289@dlvtc.edu.cn

Abstract: This article analyzes the common classifications of computer test paper algorithms. This article includes random method, backtracking method, genetic algorithm, random location search method, ant colony algorithm, gene expression algorithm, etc. The author obtains the expected calculated values by studying the technologies of setting control parameters, creating initial populations, calculating individual fitness, optimizing preservation strategies, performing genetic operations, and generating new populations. The author's purpose is to improve the application effect of the computer test paper algorithm and the practicability of the information in the question bank.

1. Introduction
The rapid development of contemporary science and educational technology makes traditional teaching and examination methods gradually unable to adapt to modern teaching requirements. Therefore, the networked teaching test has gradually entered people's field of vision. The key to the networked teaching test is whether it can quickly and intelligently compose test papers. That is to set up the test question bank on the computer, and select the test questions through the appropriate algorithm and form the test paper. This type of algorithm is more efficient than the traditional manual test. The application of procedural gene expression programming to the study of computer paper generation algorithms is conducive to improving the reliability of calculation results.

2. Common Classifications of Computer Paper Generating Algorithms

2.1 Random Method
The random method is to randomly select questions from the question bank based on some random functions provided by the computer. Then, with the help of algorithms, the applicability of the selected questions is judged, and the questions that meet the user's requirements are selected and included in the test paper. In practical application selection, randomness is difficult to rule out repetitiveness, so it is easy to repeat questions. Therefore, when applying this method to compose papers, the questions will be distinguished by setting the question ID to avoid duplication of questions. The specific operation process is shown in Figure 1. After the computer starts to execute the command, it will randomly select a number r in the system, which is the ID of the corresponding question. After the last process, the system substitutes the question into the formula for evaluation. If the test question meets the application requirements, it can be put into the test paper. If the requirements are not met, the random number r will be regenerated, and the process of re-evaluation will be entered. This process continues until the test paper is completely generated.
2.2 Backtracking

The backtracking method is to take the required objective function as the exploration target, and then conduct a tentative search in the question bank. This is a high-depth priority search method. This algorithm can first be explored along a certain application direction in the application. Then we need to explore the desired target object in this direction until it can no longer move forward. From the perspective of practical application, the retrospective method needs to set an established goal in the process of use. After that, with the help of exploration, the whole space is continuously explored, and the optimal solution is obtained and put into the test paper. Finally, we are supposed to generate the required test paper content, stop continuing to explore at this time, and obtain a stable calculation result. The specific operation process is shown in Figure 2. In the existing question bank, select a question type in a question bank to record, and the result obtained can be used as a basic reference and included in the linked list TMP. At the same time, the system also needs to establish a linked list DXF to extract project content. Firstly, the system will extract all the questions in a certain order, and all the extracted questions will be temporarily stored in DXF. Secondly, we should make judgments on the constraint conditions of the questions. If the requirements are met, you can continue to extract the questions and continuously improve the existing questions. Otherwise, it needs to go back, clean up the questions in the DXF linked list, and re-extract the questions. Finally, this method should be implemented until the complete test paper position is output.

2.3 Genetic Algorithm

From the current application situation, the application effect of genetic algorithm is relatively good. Because it adopts the relevant principles of genetics, transforms the entire set of test papers that need to be applied, and continuously optimizes the existing content of the test paper, so as to get the highest
application effect. From the perspective of practical application, the principle of genetic algorithm is to imitate the principle of biological evolution. It follows the evolutionary law of "survival of the fittest" and inherits better and higher-quality genes to the next generation. If we regard the question bank as a population, then each question is an individual. Therefore, after judging its adaptability, the genetic algorithm can screen out the topics that meet the requirements and perform crossover, replacement and other processing. In this way, genetic algorithm improves the reliability of system application results.

2.4 Random Location Search
From the perspective of practical application, the random location search method is a binary search method. That is, a certain starting point is randomly selected in the search question bank as the starting point of the test paper search, and the test paper search is carried out based on this. If during the evaluation process it is found that the question does not meet the test paper requirements, then start a radiation search along the starting point, and stop continuing the search after obtaining the test paper that meets the established requirements. In practical applications, a random number a can be randomly generated as the ID number of the title, and then an optional judgment can be made. If it meets the requirements, we can continue to extract and judge questions along the random number. Otherwise, a deviation value between 0 and 1 will be generated. If the deviation exceeds 0.5, the search range will be searched backward along a. Otherwise, it will search forward to form a complete test paper [1].

2.5 Ant Colony Algorithm
Ant colony algorithm is a major algorithm in the research field of swarm intelligence theory. It searches for the optimal solution by simulating the behavior of the natural ant search path, and successfully solves the classic combination optimization problems such as traveling salesman and graph coloring. Firstly, we need to initialize the pheromone concentration of the test question, and then set an appropriate number of ants to be randomly distributed in the test question bank. Secondly, we use the transition probability function determined according to the goal of the test paper to make the ant state transition and update the pheromone concentration. Finally, we repeat the above process again with ants until the termination condition is met or the number of iterations is reached. Ant colony algorithm itself has advantages in solving complex optimization problems, especially discrete optimization problems. Its advantage is that it can be combined with a variety of heuristic algorithms to improve algorithm performance. Its disadvantage is that the parameters are set improperly, it is prone to stagnation, and it is trapped in defects such as local optimization [2].

2.6 Gene Expression Algorithm
In addition to the algorithm content mentioned above, gene expression algorithm is also an algorithm with relatively large application popularity. In the application, this algorithm combines the application characteristics of genetic algorithm and ant colony algorithm, and it has more application advantages. From the perspective of practical application, the algorithm makes full use of the characteristics of GA fixed-length linear coding, and expresses the evaluation results in expressions. The genes of GEP are composed of multiple sub-contents. If its head is denoted as a and its tail is denoted as b, then the total number of genes can be denoted as n, then the tail b=a×(n-1)+1. Subsequently, the gene expression programming algorithm will re-evaluate the degree of satisfaction to obtain a more accurate evaluation result, thereby improving the reliability of the test paper result [3].

3. Analysis of Computer Test Paper Generation Algorithm Based on Gene Expression Programming

3.1 Set Control Parameters
Relying on the gene expression programming algorithm for computer test papers, the selection of control parameters belongs to the basic application content. It mainly contains the following content in specific applications. First, the collection of functions. The entire question bank of test papers needs to
have a clear specific selection function when selecting. That is, to clarify the operation process and deal with it in a targeted manner, so as to obtain the more reliable test questions [4]. Second, set the terminator. After the test paper is assembled, the operation will be terminated at this time. Moreover, the test paper will be output at this time. The terminator can generally choose numbers, question types, etc. to reduce the error tolerance rate of the screening process (control within 0.1%) and improve the screening results of the test paper.

3.2 Create Initial Population
The initial population refers to the topics that need to be selected. In order to increase the richness of the initial population, in actual selection, we should pay attention to the following points. (1) Diversity of solution schemes. The selected individual should have multiple types of solutions, which are convenient for subsequent comprehensive evaluation. (2) Increase in individual richness. At the beginning of the establishment of the population, the number of each type of population should be ≥50, which can give more room for selection and increase the probability of being selected [5].

3.3 Calculate Individual Fitness
When calculating individual fitness, we need to complete the analysis and decoding of the GEP gene first. Then, we determine the benchmark value for evaluating individual fitness based on the decoding result, so as to obtain a reasonable and effective population evaluation result. Moreover, the evaluation results of individual fitness also need to be backed up in time. Finally, in order to improve the reliability of individual calculation results, and enhance the validity and practical value of the final calculation results [6].

3.4 Optimize Preservation Strategy
As mentioned in the previous chapters, the application of genetic expression programming algorithm combines the application characteristics of genetic algorithm. Therefore, this algorithm will imitate the principle of biological evolution and follow the evolutionary law of "survival of the fittest" to select better and higher quality genes. The question bank can be regarded as a population, and each question is an individual. After judging its adaptability, we need to screen out the topics that meet the requirements and store them in a targeted manner. At the same time, we also need to do a good job of classification, so as to improve the effectiveness of saving results [7].

3.5 Genetic Manipulation
In the process of genetic manipulation, we should pass the selected high-quality genes to the next generation. In this process, methods such as selection, gene interleaving, mutation, gene recombination, etc. are often used to obtain higher quality individuals. When selecting the individual application situation, we can also use this method to improve the data information in the question bank, so that the subsequent gene extraction work can proceed smoothly. It should be noted that many unreasonable genes or genes with high fault tolerance will be produced in this process. These contents are all contents that need to be cleaned up in advance in the evaluation stage to improve the reliability of the calculation results [8].

3.6 Generate A New Population
After completing the above operations, start to enter the new population generation work. In this process, first of all, we need to calculate the fitness of the new individual. Secondly, according to the ECP decoding results, the benchmark value for evaluating individual fitness is determined, so as to obtain reasonable and effective population evaluation results. In addition, the evaluation results of individual fitness also need to be backed up in time to complete the generation of new populations on this basis. Finally, for some applied individual content, we need to perform accurate calculations of the results to improve the effectiveness and practical value of the final calculation results [9].
3.7 Get the Expected Calculated Value
In the process of data extension, if the data information has reached the maximum evolution algebra, or the required calculation accuracy, then the individual evolution content has ended at this time, and we can output the completed volume composition. On the contrary, we need to re-execute the above steps to find the optimal solution of the calculation result to satisfy the accuracy and reliability of the calculation result [10].

4. Test Demonstration
In this experiment, it is assumed that there are a total of 1500 test questions in a certain question bank, and the numbered 1-400 are multiple-choice questions. The numbers 401-800 are the blank-filling questions for the test paper; the numbers 801-1200 are the true and false test questions for the test paper; and the numbers 1201-1500 are the big test questions for the test paper. Then use this method to process the test paper. The ideal degree of data information obtained at the initial stage of collation is relatively high, maintaining between 1300-1450. Therefore, the number of iterative processing needs to be reduced. After 325 iterations, the best fitness is obtained. Subsequently, manual inspection of the paper composition effect, which meets the application requirements, shows that the algorithm has high practical value [11].

5. Conclusion
In summary, the intelligent test paper composition algorithm based on the gene expression programming algorithm constructs a mathematical model of the test paper problem and the corresponding fitness function under the premise of given test paper constraints. Then you can quickly calculate the desired result, that is, the test paper that meets the conditions. This can effectively prevent "premature maturity", is simple and quick, and has high practical value.

References:
[1] Ming Weipeng, Ma Guangbin, Zhang Wenyi. Multi-satellite imaging task planning algorithm based on gene expression programming[J]. Journal of the University of Chinese Academy of Sciences, 2020, 37(04): 532-538.
[2] Tan Yan. Working baseline modeling of civil aviation turbofan engine adjustable bleed valve based on gene expression algorithm[J]. Computer Measurement and Control, 2020, 28(06): 165-169+179.
[3] Li Chaoqun, Wei Qingshun. Performance prediction of submersible pumps using gene expression programming[J]. Hydropower Energy Science, 2020, 38(04): 150-153+180.
[4] Han Xiao, Bi Bo, Tang Jinping. Research on Computer Test Paper Generation Algorithm Based on Gene Expression Programming[J]. Computer Technology and Development, 2020, 30(05): 154-159.
[5] Xu Lingling. Research on multi-dimensional computerized multi-stage adaptive test: automatic test paper generation algorithm and its routing rules [D]. Jiangxi Normal University, 2019.
[6] Yang Chunzhe, Chang Hanji. A method of automatic generation of papers for university computer foundation based on genetic algorithm[J]. Modern Electronic Technology, 2018, 41(11): 171-174.
[7] Yan Lina. Research on the application of genetic algorithm in computer automatic test paper generation[J]. Computer programming skills and maintenance, 2018(02): 30-32.
[8] Zhou Wenyings. Research on the application of genetic algorithm in computer automatic test paper generation[J]. Information and Communication, 2017(10):34-35.
[9] Lei Yang, Tian Ying’ai, Li Zhuo, Liu Chengxia. Application and analysis of paper composition algorithm of computer core course process assessment platform[J]. Industry and Information Education, 2016(09): 45-49.
[10] Liu Fengxiang, Du Chunyan, Tian Zhenqing. Computer test paper generation strategy and scoring algorithm design according to the type of question[J]. Journal of Inner Mongolia Normal
University (Natural Science Chinese Edition), 2016, 43(04): 444-447+452.

[11] Tian Zhenqing, Chen Mei. Multi-choice test paper composition strategy and scoring algorithm design [J]. Journal of Inner Mongolia Normal University (Natural Science Chinese Edition), 2016, 42(03): 315-320.