Research on Intelligent Volume Algorithm based on Improved Genetic Annealing Algorithm

Li Tianci\textsuperscript{a}, Lei Guangbo\textsuperscript{b}, Wan Fang\textsuperscript{c,*}, and Shu Yating\textsuperscript{d}

School of Computer Science, Hubei University of Technology, China
\textsuperscript{a}1220096650@qq.com, \textsuperscript{b}64694786@qq.com, \textsuperscript{c}devwaf@qq.com, \textsuperscript{d}1964443049@qq.com

*Corresponding author

Keywords: intelligent volume, genetic algorithm, simulated annealing algorithm, automatic learning, mixed algorithm

Abstract: In order to improve the quality of the traditional exam management system test paper generation, the genetic algorithm of simulated annealing algorithm was put forward, and at last, the automatic group volume algorithm updates the difficulty coefficient. By analyzing the simulation results, the algorithm can improve the efficiency and quality of the volume.

1. Introduction

With the continuous development of Internet technology, various question bank systems are gradually popularized in people's daily study and life. Smart group volume is the core functions of question bank system, and the main problem is how to effectively carry on the data mining on the question bank, so as to quickly produce high-quality test papers to meet the different needs of users.

2. Materials and Methods

Early test questions management system is mainly based on function, priority or random method, or other algorithms search for global optimum at the expense of time, but these methods are not able to achieve the expected goal. As a fast and efficient optimization algorithm, intelligent search algorithm has gradually become a research hotspot of intelligent test paper generation. Genetic algorithm is an algorithm that simulates natural selection and genetic mechanism in biological evolution to solve complex problems at random and optimize operation [1]. Genetic algorithm can find the overall optimal solution in finite time with high probability under certain constraints. However, due to the fact that the actual situation is too complex and the conditions are various, the genetic algorithm is prone to "premature convergence". Moreover, the system can improve the quality of the examination papers by updating the difficulty coefficient of the questions according to the method of intelligent learning. In this paper, the genetic algorithm with strong searching ability, the annealing algorithm with global optimal solution and the intelligent learning research smart volume will be combined.
3. Discussion

In order to investigate the students' mastery of each situation, each exam question set up seven attributes as follows: ① score ② The difficulty coefficient of test questions ③ Questions inclusion of knowledge points ④ Cognitive classification, teaching content requirement level ⑤ Types of questions ⑥ Estimate the time required to answer the questions ⑦ Degree of differentiation.

The key to determining a volume is a matrix of n by 7. Where n is the number of questions contained in the test paper, and the seven attributes of each question are determined by the value of the elements of the matrix S row.

\[
S = \begin{bmatrix}
a_{11} & a_{12} & \ldots & a_{17} \\
a_{21} & a_{22} & \ldots & a_{27} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \ldots & a_{n7}
\end{bmatrix}
\]

Due to these constraints affecting the group of the difficulty of the volume, the system can produce high quality test papers that meet the needs of users if the constraints are controlled accurately through continuous testing and statistics. Then the implementation of the algorithm will be introduced.

3.1. Coding Scheme

In order to overcome the complexity of binary encoding and decoding process, this system adopts the variable length real number coding, and adopts the vector \( x=(x_1, x_2... x_n) \) \( T \) represents the answer to the question as an individual \([2-3]\). If \( x_i \) equals 0 means that question \( i \) was not selected, if \( x_i \) equals 1 means that question \( i \) was selected. If there are \( m \) problems in the question bank, one of the papers needs to pick \( n \) questions, then the vector \( x= (x_1 \ x_2... \ x_n) \) there should be 1 of the number of \( N \) in \( T \) and 0 of the number of \( M-N \). According to the theory of biological evolution, the chromosome representation, each gene on the chromosome represents whether the topic is selected by the way of chromosome expression: 1 indicates that the question is selected, and 0 indicates that the question has not been selected. Each group of solutions is represented by a chromosome.

3.2. Initialization Population

In order to make the search results more diverse, a random method will be chosen to generate a population of \( N \) individuals. The value of \( N \) affects the search efficiency of the algorithm and the optimization of the results. In order to solve the problem that the random results distribution of the initial population is not uniform in genetic algorithm, the classification strategy is adopted when the population is initialized. First of all, the topics are divided into \( A_i \) according to the question type, and then are classified into \( B_i \) according to knowledge points. At last, the subject is randomly selected under the constraints of the coverage of the knowledge points of the test paper, thus the individual sample is formed.

3.3. Fitness Function

Fitness function is an index to measure the quality of individuals, which affects the performance of genetic algorithm in searching the optimal solution. The fitness function selected this time is
shown in Equation (1).

\[ f(x) = \sum_{i=1}^{n} f_i w_i \]  

In Equation (1), \( f \) is the sum of all the indicators of the absolute value of the error and the user's needs, \( f_i \) is the absolute value between the \( i \) index and the error of the user's demand. \( w_i \) is the weight of the \( i \) index in the group, and the \( n \) is the number of indices \([3]\).

### 3.4. Selection

In order to find the optimal solution quickly and accurately, the selection strategy of wheel betting is in direct proportion to the individual selection probability and its adaptability to the remaining individuals after the selection of optimal solutions. The probability formula and the cumulative probability formula of the strategy selection are shown in Equation (2) and (3).

\[ p(i) = \frac{f(i)}{\sum_{j=1}^{n} f(j)} \]  

\[ Q(i) = \sum_{j=1}^{i} P(j) \]  

In Equation (2) and (3), \( P(i) \) is the selection probability of the \( i \)th individual, and \( Q(i) \) is the cumulative selection probability of the \( i \)th individual, and \( f(j) \) is the fitness of the \( j \)th chromosome, and \( n \) is the number of the population. After calculating \( Q(i) \), a probabilistic \( r \) hour is randomly generated \([0,1]\). By comparing the values of \( Q(i) \) and \( r \). If \( Q(i) \) is greater than or equal to \( r \), the chromosomes numbered as \( i \) is selected. If \( r \) is greater than \( Q(i-1) \) and \( r \) is less than \( Q(i) \), the chromosome numbered as \( i \) is selected.

### 3.5. Overlapping

In order to improve the performance of the algorithm, the crossover algorithm is a multi-point crossover method in genetic algorithm, and requires two individuals to cross at the same time in the same problem type section. The modified probability of adaptive mutation is used for crossover probability, and its expression is shown in Equation (4).

\[ P_c = \begin{cases} 
P_{c_1} - \frac{(P_{c_1} - P_{c_2}) - (f_i - f_{avg})}{f_{max} - f_{avg}} & \text{if } f > f_{avg} \\
0 & \text{if } f < f_{avg} 
\end{cases} \]  

In Equation (4), \( P_c \) is a cross probability, \( P_{c_1} = 0.9, P_{c_2} = 0.6 \), and \( f \) are the larger individual fitness in the two cross sections of chromosomes, and \( f_{avg} \) is the average fitness of the individual in the population.

### 3.6. Variation

In order to reduce the complexity of the system algorithm, the mutation algorithm randomly selects one or more individual values and changes their values, and does not repeat the value of the
changed value with those of other positions. The mutation probability is modified by adaptive mutation probability, and its expression is shown in Equation (5).

\[
P_m = \begin{cases} 
    P_{\text{m1}} - \frac{(P_{\text{m1}} - P_{\text{m2}}) - (f_i - f_{\text{avg}})}{f_{\text{max}} - f_{\text{avg}}} & f_i > f_{\text{avg}} \\
    P_{\text{m1}} & f_i < f_{\text{avg}}
\end{cases}
\]

(5)

\(P_m\) is the mutation probability. \(P_{\text{c1}}=0.1, P_{\text{c2}}=0.001\) and \(f_i\) are the fitness of individuals. \(F_{\text{avg}}\) is the average fitness of individuals in the population.

3.7. Simulated Annealing Operation

The simulated annealing algorithm is based on the principle of solid annealing, which gives the search process a certain probability change at a certain time and the final probability tends to zero, thus can effectively avoid the local optimal solution and tend to the global optimal solution. The implementation steps of the simulated annealing algorithm are presented [4-5].

1) The initial temperature is \(t\), the initial minimum temperature is \(T_{\text{min}}\), and the initial solution is \(x_0\);

2) If the temperature ends in the loop, jump directly to the third step. Or a new solution \(x_{\text{new}}\) is generated and the increment \(\Delta x = x_{\text{new}} - x_0\) is calculated. If \(\Delta x < 0\), \(x_{\text{new}}\) is accepted as the new solution, or \(x_{\text{new}}\) will be accepted as a new solution at a certain probability. "Certain probability" is the acceptance criterion of Metropolis, and its formula is Equation (6).

\[
P = \exp\left(-\frac{\Delta E}{KT}\right)
\]

(6)

\(\Delta E\) is the difference of internal energy, \(K\) is constant, and \(T\) is the temperature at that time.

If the termination condition is satisfied, the loop is terminated, otherwise, the \(t\) will gradually decrease, and \(t > T_{\text{min}}\) and jumps to the second step.

3.8. The Difficulty Coefficient of Automatic Learning

According to Bernoulli's law of large numbers in probability theory, the difficulty coefficient of a subject can be determined by a large number of students' probability of correct test. According to this principle, using the average gain and loss calculation method, the difficulty coefficient of the test is defined as \(Q=1-R/T\), in which \(Q\) represents the loss rate of a certain test in a certain test, \(R\) is the number of people who answered the question correctly, and \(T\) is the number of people receiving this test [6]. At the beginning of the test question, the initial value of the difficulty coefficient is decided by the proposition team. After each student's test, the system will automatically correct the test paper and recalculate the difficulty coefficient of each question, replacing the previous difficulty coefficient. With the increasing number of tests, the system will calculate and replace the difficulty coefficient of the test items more frequently, and the difficulty coefficient of the test questions will be more and more close to the actual situation.

4. Results

In order to verify the feasibility of the intelligent test paper algorithm, 500 subjects were extracted from the item bank according to the algorithm implemented by the system, and the following attributes were set. The test paper is full score of 100 points, the average expected score
is 70, the permitted error is 2, 4 types of questions. According to the two distribution of difficulty, the scores of each difficulty level can be calculated as \{12,30,32,19,7\} respectively. The experimental results obtained by the algorithm and random algorithm in this paper are shown in Figure 1, and their fitness function values are compared in Figure 2.

![Figure 1 Algorithm for generating test paper difficulty fraction distribution.](image1)

From Figure 1, it can be seen that compared with the random algorithm, the algorithm introduced in this paper is more in line with the requirements of binomial distribution in terms of the difficulty distribution of test questions in the test papers. From the value of Figure 2, it can be seen that the fitness function value of the algorithm introduced in this paper is much lower than that of the random algorithm. The fitness function shows that the lower the function value is, the better the user’s requirements are, the lower the difference from the target. This algorithm can meet the requirements of the teacher and improve the quality of the examination paper. Generally speaking, compared with the random algorithm, the algorithm introduced in this paper has significantly improved the efficiency and quality of the test paper.

![Figure 2 Comparison of fitness function values.](image2)

5. Conclusion

In this paper, aiming at the premature convergence and slow evolution of traditional intelligent test paper, this paper puts forward an improved algorithm combined with genetic algorithm and simulated annealing algorithm, and the algorithm of improving the difficulty coefficient of test questions is improved by intelligent learning. Through experiment simulation, the algorithm can be applied to intelligent volume, and obviously improve the efficiency and quality of volume.
References


