Research on Computer Intelligent Test Paper Composition Based on Genetic Algorithm

Liming Wang

College of Foreign Languages, Bohai University, Jinzhou, 121013, China

2860333710@qq.com

Keywords: test paper composition; genetic algorithm; selection; crossover; mutation

Abstract. The test paper composition is an important part of the exam, how to choose an efficient, scientific and rational way, can ensure test papers the greatest extent possible to meet the different needs of users, and random, science, rationality and other characteristics, it is very critical. Aiming at the limitations of the traditional method of generating test paper problem, this paper uses genetic algorithm of generating test paper. Firstly, the mathematical model of the generating test paper algorithmic; then, through a graphical illustration of the basic principles of genetic algorithms; then studied the application of genetic algorithm in intelligent test paper composition and illustrates the main steps of intelligent test paper composition; Finally, genetic manipulation of intelligent test paper composition, continue to carry out individual fitness evaluation, selection, crossover and mutation until the search for the optimal solution or satisfactory solution. Using the genetic algorithms for intelligent test paper composition, it reduces the workload of teachers, to promote scientific and standardized test management has important practical value.

Introduction

The strategy of test paper composition is the core implementation technology of online exam. the method of generating test paper in the existing examination system is divided into the following three categories [1]: a class with manual generating test paper is similar, test database as questions set, by the teacher according each question screening and extract; One is to have an organized papers set as a test database, when generating test paper from there randomly selected one or more sets of test papers for the exam; there is a class of adopting a strategy of generating test papers, the teachers according to their own requirements by means of a computer to automatically test paper generate one or more sets test papers. Comparison of these three methods, the first one is essentially manual generating test paper; the second method of exam papers in a limited number of cases, the lack of flexibility, and prone to repeatedly drawn the same set papers phenomenon; these two methods generating the papers are difficult to meet the needs of users. Therefore, in order to ensure the practicality and reliability for the machine test, have intelligent test paper composition for it.

Most of the intelligent test paper composition program still uses traditional algorithmic, random question method is simple, but has a great time cost, and often due to local satisfy the constraints caused by the failure test paper. Backtracking method applicable to less demand and less question number, if the items number is large, the number of the states type transformation will become huge, time and space complexity are large. Therefore, choosing an efficient, scientific and rational algorithm is the key to automatic test paper. Genetic algorithm (GA) as an intelligent search algorithm with global optimization, parallelism, robustness and simplicity and other characteristics suitable for solving test paper composition, so intelligent test paper composition uses GA.

The Mathematical Model of Test Paper Generating Algorithm

General mathematical model is a mathematical simplification of the real thing. It is often exist in the abstract sense close to the real thing, but it's something essentially different. To describe a real phenomenon can have a variety of ways, such as audio, video, metaphor, rumors and more. In order to describe more scientific, logical, objectivity and repeatability, generally considered to be more people to adopt a stricter language to describe a variety of phenomena, this language is mathematics. Use mathematical language to describe things is called a mathematical model [2].

Mathematical model species

Mathematical model is a powerful tool to study and master the laws of motion of the system, it is the analysis, design, forecast or predict, control the basis of the actual system. Many different types of mathematical models, and there are a variety of different classification methods [3].

(1) Static and dynamic models. Static model is simply more relationship between the each amount of system described is not change with time and change, are generally expressed by algebraic equations. Dynamic model refers to the mathematical expressions describing between the each amount of system with time of the rule changes, usually with differential equations or difference equations to represent.

(2) Distribution parameters and lumped parameter model. Distributed parameter model with all kinds of partial differential equations describing the system's dynamic characteristics, and lumped parameter model is linear or non-linear ordinary differential equations describing the system's dynamic characteristics. In many cases, the distributed parameter model by means of space discretization method, can be simplified as a lumped parameter model with less complex.

(3) Continuous time and discrete time models. In the model the time variable is within a certain time interval of change model called continuous time model, the model described by differential equations in the above types are continuous time models. In dealing with the lumped parameter model, can also be the time variable discretization, the obtained model is called a discrete time model. Discrete time model is described by the differential equation.

(4) Stochastic and deterministic models. In stochastic model the relationship between the variables is given in the form of the probability distribution or statistic value, and the relationship between variables in the deterministic model is determined.

(5) Parametric and non parametric models. the model described by algebraic equations, differential equations, difference equations and transfer functions, and so are the parametric model. Established to determine the each parameters of the parametric model lies in the structure of the known model. Through theoretical analysis is always derived parameters model. Non parametric model is to respond directly or indirectly resulting from the experimental analysis of the actual system, for example, recorded by the experimental system impulse response or step response is non parametric models.

The Construction of Mathematical Model

Generally speaking, when organizing a paper, usually to determine papers perfect score value, item difficulty, the number of questions, the distribution of knowledge point and xamine the ability level and other parameters indicators. Combined with the actual situation in the English exam, Establishing the mathematical model of intelligent test paper composition nine core attributes, a paper constraint into a 9 dimensional vector space:

Questions (question number, scores, the difficulty factor, ability level, knowledge, question types, discrimination, answer time, the frequency of use).

A question is determined by 9 properties, that is, from a 9-dimensional vector $(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9)$ characteristics determine a questions. Therefore, for a set of n test questions papers, can be expressed by matrix $(n \times 9)$:

$$S_g = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{19} \\ a_{21} & a_{22} & \cdots & a_{29} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{n9} \end{bmatrix} \qquad (1)$$

Each attribute in matrix satisfies the following constraint conditions:

(1) Total score papers. $\sum_{i=1}^{n} a_{i2} = P$, P is total score requirements for papers, The default is 100 points, designated by the test paper staff.

(2) The difficulty of test paper. $ND = \sum_{i=1}^{n} a_{i2} a_{i3} \big/ P$, P is total score of test paper.

(3) The ability level. $Z_j = \sum_{i=1}^{n} C_{i4} a_{i4}$, Z_j is scores in the *j-ed* ability level, when $a_{i4} = j$, $C_{i4} = 1$; when $a_{i4} \neq j$, $C_{i4} = 0$. Each level includes the ability to understand, comprehend, grasp, flexible use and so on.

(4) The knowledge point. $\sum_{i=1}^{n} C_{i5} a_{i5} = P_j$, P_j is scores of the *j-ed* knowledge point, set by the user as required. When $a_{i5} = j$, $C_{i5} = 1$; when $a_{i5} \neq j$, $C_{i5} = 0$.

(5) The question types. $\sum_{i=1}^{n} C_{i6} a_{i6} = M_j$, M_j is scores of the *j-ed* question type, set by the user as required. When $a_{i6} = j$, $C_{i6} = 1$; when $a_{i6} \neq j$, $C_{i6} = 0$. At the present stage, CET4 constituted by six question types were listening comprehension, reading comprehension, cloze test, error correction, writing and translation.

(6) The discrimination. $\sum_{i=1}^{n} a_{i2} a_{i7} \big/ P$, P s total score of test paper. The examinee obtained in each test scores from high to low order, divided into high and low groups, on the basis of high and low grouping scoring rate, get the discrimination, the discrimination of test papers is a weighted average of each test questions discrimination.

(7) The answer time。 $\sum_{i=1}^{n} a_{i8} = T$, T is the answer time of papers, that is, the examination time, the default is 120 minutes, specified by the test paper staff.

The Basic Principles of Genetic Algorithms

Genetic algorithm to mimic natural biological evolution mechanisms developed is a random global search and optimization method, it draws on the Darwin's natural selection, survival of the fittest theory of evolution and Mendel's genetics, GA is based on the fitness function (or objective function), through to the individuals in the population genetic operation to achieve the iteration process structure of recombinant individuals within populations, in this process, groups of individuals from generation to generation optimized, and gradually approaching the optimal solution, ultimately, obtain the optimal solution [4]. Also has inherent parallelism, global optimization and fast convergence characteristics, it can automatically acquire and accumulate knowledge about the search space during the searching process, and adaptively control the search process in order to achieve optimal solution [5]. GA is suitable for the processing intelligent test paper composition.
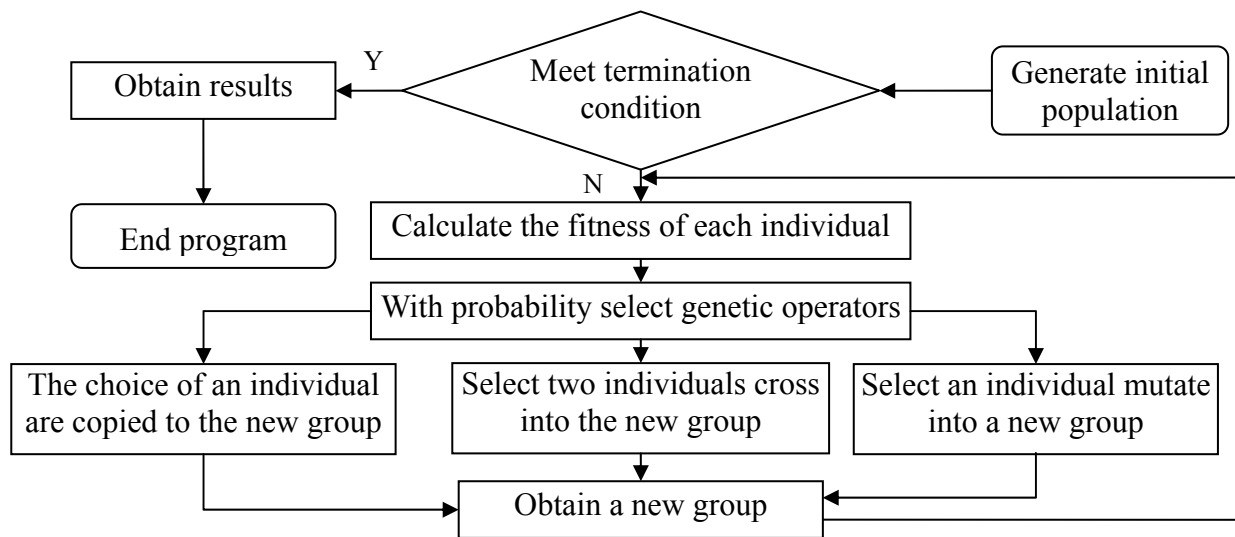
Fig. 1. Genetic algorithm flowchart

The basic principle of the GA is shown in Fig. 1, first, data of issues should be encoded, mapping from external features of problem entities to the bit string, this will generate a random initial population. Search, on the basis of individual fitness function value to select individual, the fitness function value high is excellent individuals, large probability of being selected, in contrast, value low, less probability of being selected [6]. Next, crossover operation randomly paired two chromosomes exchange part of their genes, through mutation, chromosomal gene location will change by a certain probability, to generate the individuals of next generation, to form a new population. The new population out of low fitness individuals, inherit and carry the good genes of forward generation populations, so the whole better than previous generation, more resilient. Thus, after several generations of evolution, e population average fitness is improving constantly, until the problem tends to converge, termination condition is met, the individuals of the last generation population is decoded, get approximate optimal solution.

Application of Genetic Algorithm in Intelligent Test Paper Composition

For intelligent test paper composition problem, the main steps in the application of GA as follows, test paper composition flowchart shown in Fig. 2:
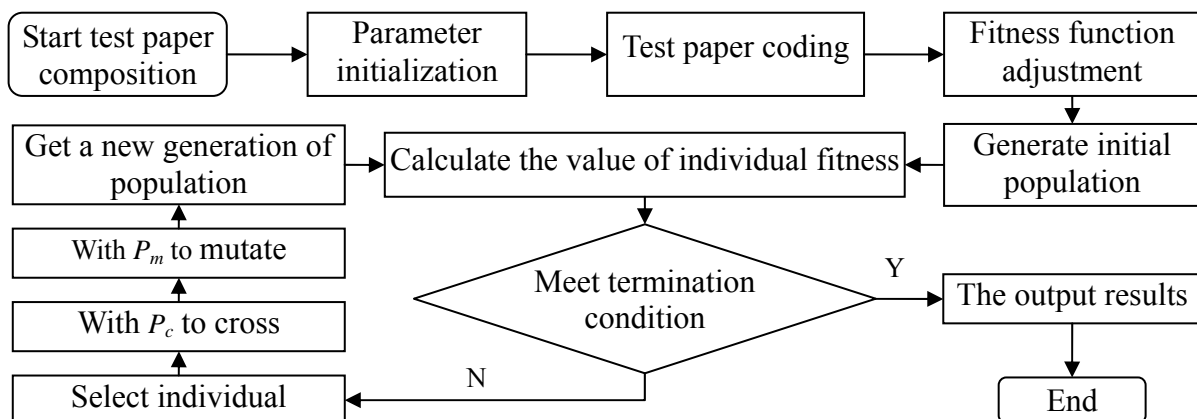

Fig. 2. Test paper composition flowchart

(1) Test paper chromosome coding, in order to avoid GA population chromosome encoding long decline caused by the search efficiency, this article use natural number coding scheme, based on questions type for coding segments, gene number is question number, the process of test paper composition according to the requirements to select question from the question database, the search will end the selected gene consists of a chromosome. Assuming the test database topics to be

selected number of *A*, from *A* which the questions selected to meet user requirements constraints consisting of a *B* test questions papers, selected questions with which number represents, then a papers string corresponding to real number string that is the question number composed by the length of *B*, called chromosomes, thus forming a chromosome encoding.

(2) Population initialization, randomly generate initial population of *M* individuals, the initial population size is usually required by the experience or experiment is given. According to the papers questions categories, the number of each question type, the total score, and difficulty factor of all kinds to generate the initial population. In the test database to generate a random questions number within the scope of questions number corresponding question type, question number and chromosome on this kinds of questions that has been generated for comparison. If repeated, then re-produce, otherwise leave, and repeat the process number for the total of questions. Then calculate the papers total score, if not equal to the total score, regenerate. Otherwise, the calculation of the difficulty factor, compared with the difficulty factor settings, if the difference is within the allowable range, the paper is valid. Otherwise the paper is invalid, regenerate [7].

(3) The fitness function design. Use of the fitness function (evaluation function) is calculated function value for the individual, fitness function used to evaluate the merits index of the individuals, which indicating the search direction; each individual calculated fitness is higher, the greater ability to adapt the environment, be selected probability the larger [8]. In the test paper composition process, it will affect the search speed and the generation of test paper quality, so the design of the fitness should be as simple as possible, due to initial population has considered the question number and other constraints, so the fitness function considers only the knowledge points and the difficulty coefficient which two constraints. Then according to the evolution algebra, it is determined whether a termination condition is satisfied. If so, output, or continue to the next step of genetic manipulation.

Genetic Manipulation of Intelligent Test Paper Composition

After the above operation, genetic operation namely is selection, crossover and mutation.

Selection, reflects the survival of the fittest principle of biological evolution, the purpose is to make the current population fitness individuals genetic better able inherited to the next generation, and enhance the individual's ability to adapt to the next generation. According to selection strategy to select a number of individuals from the population to copy, a selected individuals to produce a new generation of population by using the genetic operators role. Selection strategy is to comply with the principle: the higher the individual's fitness, the greater the probability of being selected.

Crossover, it is in accordance with a certain probability of two individuals (crossover probability *Pc*), starting from a bit by bit swap. First, generate a random number for each individual in the (0, 1), and if the number is less than *Pc*, then select the individual cross, or do not select. The selected individuals were paired randomly, and in accordance with the length *B*, a random cross position *i*, *i* is an integer on [1, *B*-l]. Mutation is individual in accordance with a certain probability (mutation probability *Pm*) reversed [9]. for after crossover, each individuals of the population randomly generates a random number in [0,1], if more than *Pm*, then retain the original characters of the same chromosome; otherwise use the random manner in papers selected an item, get the properties of the questions type and their respective knowledge points, and then randomly from test question database select a question, which is the same kinds of questions and same knowledge point, replacing the individual of mutation.

After the above operation, generates new individuals. Calculating individual fitness, find out the current population in individual fitness the highest and fitness individuals the lowest. If the fitness of the best individual in the current population is even higher than the best individual's fitness to date, places the current population of the best individual as the new date for the best individuals. Replace the individual with the best so far out of the current population of the worst individuals [10]. GA in test paper composition is so constantly in individual fitness evaluation, selection, crossover, mutation, and so on ad infinitum, until the termination condition is met, and the search to the most optimal solution or satisfactory solution.

Conclusion

With the wide application of computer technology in the field of learning, computer aided test has been paid more and more attention. the core content of computer aided test is the intelligent test paper composition, this article carries on the intelligent test paper composition by GA, GA has natural selection, survival of the fittest and simple genetic operation, so that the GA is not subject to its search space restrictive conditions constraints, and does not require additional supporting information, and the search process is not easy to fall into local optima. Can effectively solve the intelligent test paper composition, compared with other methods, it can earlier found to meet the conditions of population, making the test paper meet the diversity also has generated timeliness. Not only reduces the workload of teachers, but also reflects the level of teachers, while effectively checked student' learning effect.

References

[1] L. Zhao, "Research on Intelligent Composing Test Paper Strategy Based on Genetic Algorithm," Harbin Engineering University, 2013.

[2] Baidu Encyclopedia, "Mathematical modeling (branch of Mathematics)," http://baike.baidu.com/link?url=AYDFjW8lvESOXHEDtpgCwo2uTFuO_qIMZCRlwlXU70g4 nHT56aZD2BDMlGI4KkDjxtIUdLZU7LOz4iL_40qgviQ9jiXri5tKlCV-csD-asq, 2015-1-24..

[3] Baidu Encyclopedia, "The mathematical model (Mathematics)," http://baike.baidu.com/link?url=AYDFjW8lvESOXHEDtpgCwo2uTFuO_qIMZCRlwlXU70g4 nHT56aZD2BDMlGI4KkDjxtIUdLZU7LOz4iL_40qgviQ9jiXri5tKlCV-csD-asq, 2015-1-24.

[4] J. Lu, "The Research of Intelligent Test Paper Auto-generation Based on An Improved Genetic Algorithm," Central South University, 2007.

[5] H. Ma, "Research of Intelligent Test Paper Composition System Based on Genetic Algorithm," Shenyang University of Technology, 2009.

[6] F. X. Sun, "Design and Realization of Intelligent Test Paper Auto generating System Based on Genetic Algorithm," Hebei University of Science and Technology, 2013.

[7] D. Y. Xv, "An Intelligent Generating Examination Paper Method Based on Improved Genetic Algorithm," Computer and Digital Engineering, vol. 41, no. 2, pp. 176-178, 2013.

[8] L. Pan, "Application Method Research of Genetic Algorithms in Automatic Generating Test Paper," Northeast Normal University, 2010.

[9] M. Yu, "Application of Automatic Compositing Test Paper Based on Genetic Algorithm," Qingdao Technological University, 2009.

[10] Y. Shang, "The Application and Research of the Paper Composed System Based on Genetic Algorithm," Dalian Maritime University, 2010.