Intelligent Testing Composition System

Xiangran Du[1, a], Shujin Wu[1,b] and Yulin He[2,c]

[1] Department of Information Engineering, Tianjin Maritime College, Tianjin, China

[2] College of Computer Science & Software Engineering, University of Shenzhen Shenzhen, Guangdong Province, China

[a] duxiangran1226@126.com, [b] 56616353@qq.com, [c] heyulin2010@126.com

Abstract. The function of generating paper is a core part for Computer examination system, which has huge influence on the quality of the examination system. The process of generating paper is that some examination questions are selected from paper database according to the composing evaluation function satisfying the users' demands. In this paper, the methods of intelligent composing evaluation function are presented, which can play promotion for improving the testing system. The intelligent evaluation function not only save much time and human resource, but also improve the accuracy of the evaluation function by self-learning ability without the help of the educational experts.

Introduction

Computer examination system is an important part of computer education while the emergence of it not only improves the efficiency of designing papers and the fairness of the exam but also avoids spending huge manpower and resources like traditional examination. A complete set of examination system mainly includes examination bank, examination paper library, examination paper generating modeling and test paper analysis module. Examination paper generation module is the most important part of the examination system. The main function of generating paper modeling is guided by evaluation function to extract test set rapidly and accurately which satisfies users' requirements by using generating paper algorithm and it's a key part to determine whether the test paper is reasonable or not.

Background and relative researches

The intelligence of the computer examination system can make the examination system adjust the content and difficulty of the exam dynamically according to the actual situation of the candidates [2]. For the moment, the intelligence of the examination systems mainly focuses on generating paper algorithm and neglected evaluation function of the system. For instance, WANG Feng-rui achieved Intelligent test sheet composition research based on differential evolution algorithm in 2009 [3]; WANG You-ren and his research team from Nanjing University of Aeronautics and Astronautics put forward Mathematical model and algorithm of intelligent test paper auto-generation system from item pool in 2014 [4], and YUAN Gui-xia presented Modeling and Research on Computer Composing Test Paper [5], etc. Research in evaluation function of the computer examination system is still in the initial stage.

The evaluation function of computer examination system evaluates examination paper according to the questions and the relevant parameters of the examination [6]. It is almost impossible to extract an ideal testing sheet if the evaluation function is inaccurate and the test results can't reflect the true level of candidates. The parameters of the testing paper are given by educational expert in traditional examination system, so it is difficult to evaluate every piece of test sheet and the papers generated in this way will be influenced by subjective consciousness easily without the help of a strong educational team. Optimizing evaluation function by using particle swarm optimization is proposed in this paper. Examination system will improve every question's attribution of the test bank continuously without

artificial intervention to guide generating paper algorithm to extract testing paper better by using evaluation function.

Mathematical Model on Evaluation Function

Generating test paper is a multiple objective and multiple constraint optimization problems and its mathematical description is that assuming that there are n candidate questions named $Q_1, Q_2... Q_n$ in a test bank, paper generation requires to find m constraint conditions marked $C_1, C_2, ... Cm$, in the test bank that satisfies the value of the objective function to be the maximum. Establishing a (math) model of generating paper modeling involves the objective function of test paper, some relevant attributes of test paper, the constraint or range of the attributes and variable declaration, etc.

The target of generating test paper is the test set which is extracted from the question bank to satisfy the constraint condition. These conditions were designated by lots of people including item types, difficulty level, test points, examination time, score and so on. The formula of generating test paper (T) including reliability (R), validity (V), difficulty (Dif) and discrimination (Dis) is given as follows.

$$T = w \quad R + w \quad * V + w \quad Dif + w \quad *Dis$$
(1)

The w_i represents weight parameter and means the importance degree of every component. The ranges of four weights are from 0 to 1 and these weights satisfy constraint Eq. 2.

$$w \quad w \quad + w \quad + w \quad = 1$$
(2)

Reliability mainly inspect what extent can the test paper truly reflect on the students' level and it also can be regarded as the credibility of the test paper. Common method to test the reliability is re-testing method. If the specific value is between 0.9 and 1, the reliability of the test paper is in a reasonable range. Otherwise, it is unreasonable.

Validity determines the level of the test paper content sort with the requirement of the syllabus. The calculation formula is represented by Eq. 3.

$$V = \sum_{i=1}^{N_t} \sum_{j=1}^{N_n} v_{ij} \cdot x_{ij} / TN$$
(3)

Among it, V_{ij} and X_{ij} respectively represent whether the j-th question of the i-th questions satisfies the requirement of the syllabus or not and whether they are included in the paper. TN represents the total number of questions in the test paper. The range of validity is from 0 to 1.When it tend towards 0 means the coincidence rate is low and when it tend towards 1 means the coincidence rate is high. Validity value is given by the setter of the examination and it is between 0.8 and 0.95.

Difficulty means the difficulty of the test paper and it is often represented by the degree of difficulty and the product of test score. Common difficulty of the test paper includes 5 levels which are easy, simple, normal, difficult and harder and its correlation coefficient can be respectively represented by d_1, d_2, d_3 and d_4. Difficulty can be represented by the quotient of the difficulty sum and the total score of papers, its calculation formula is represented by Eq. 4.

$$Dif = (\sum_{i=1}^{N_t} \sum_{j=1}^{N_n} (d_{ij} \cdot s_{ij} \cdot x_{ij})) / TS$$
(4)

Discrimination Degree shows how the test paper distinguishes the practical capacity of the students and it's the sum of all the exam questions' discrimination in the test paper. The calculation formula is given as follows.

$$Dis = \sum_{i=1}^{N_t} \sum_{j=1}^{N_n} f_{ij} \cdot x_{ij}$$
(5)

Among it, F_{ij} represents the discrimination degree of j-th question of the i-th questions. Normally, discrimination degree is divided into 3 grades and the k grade question's percentage F_{Pk} in test paper can be calculated by following formula.

$$FP_k = (\sum_{i=1}^{N_t}\sum_{j=1}^{N_{is}} f_{ij}' \cdot x_{ij})/TS \qquad (k=1,2,3)$$

(6)

The percentage of different question discrimination grade in test paper can be set as manual setting or examination system's default setting.

Intelligent Evaluation Function

The test paper evaluation modeling in the examination system is limited by (the above-mentioned) four factors to evaluate the pros and cons of the test paper through objective function. The evaluation system involves a large number of parameters. Generally speaking, the educational experts summarize through their own experience the initialization of these parameters, which is static in most cases. Once these parameters are set, they will not change. Intelligent evaluation modeling can automatically adjust the corresponding parameters by artificial intelligent optimization algorithm and continuous learning.

Particle swarm optimization is one of the efficient optimization algorithms, and it simulates the bird swarm's prey in order to achieve the goal optimization. Research on particle swarm optimization started late, and has been successfully applied to many practical items through its own characteristics and become a new research hotspot on optimization problems in recent years [8].

When particle swarm optimization is optimized, each particle represents a set of parameters that need to be optimized and expressed by three-dimensional vectors p = [p_{111}, p_{112}, p_{113},..., p_{NSP}]. Parameter N is the number of the examination type in a testing paper, the S is the number of the examination question for every examination type and the P is the attribute for every examining question. Particle swarm which is composed of many particles represents solution space set of optimization problems. The ultimate goal of particle swarm optimization is to find out the most suitable particle and regard it as the optimal solution of the optimization problems. Each particle in the particle swarm includes two parts: position information and optimization speed information. Position information is represented by n-dimensional vector and shows the parameter that needs to be optimized. Corresponding to the test system, position information represents the parameter of constraint condition of test paper composition. The velocity of each particle is corresponding to the corresponding position information, and it is applied to adjust the optimization speed of each particle and avoid the situation that the corresponding attribute falls into the local extreme value. Position information and speed information of particle swarm are updated through the following formula.

S(t+1)=w*S(t)+C1*rand*(LBest-S(t))+C2*rand*(GLBest–S(t))   (7)

L(t+1) = L(t) + S(t+1)   (8)

Among it, S(t+1) and L(t+1) respectively represent position information and speed information of generation t+1. LBest is the global extreme value of each particle and the optimal value of position information that the particle is in the updating process. GLBest is the position information of the optimal particle in the particle swarm. Generally speaking, the speed and direction controlled and optimized by Parameter C1, C2 and rand are given by experiment.

Experiment and Discussion

The construction of the evaluation function in intelligent examination system has been explicitly shown. An examining database is explained and an experiment on the optimization of the examination system is presented in this part.

The experiment accomplishes on the fundamental application of computer technology mainly studied the computer basic knowledge and the ability of using the office-software. The constitution of the examining question database shows in the table 1.

Table 1. Constitution of Examining Question Database

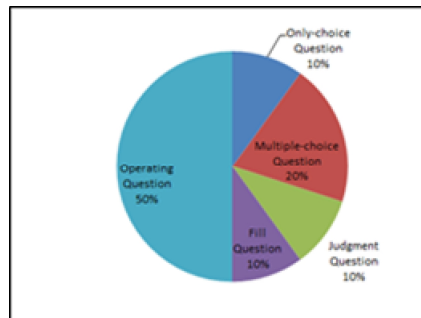|	Only-choice Questions	Multiple-choice Questions	Judgment Questions	Fill Questions	Operating Questions
Number	186	172	178	182	166
Score	2	4	2	2	10

Fig. 1 Construction of Testing Type

The Fig 1 shows the types of question in an examination paper. There are five testing question types composing a testing paper, including Only-choice Question, Multiple-choice Question, Judgment Question, Fill Question and Operating Question. Operating Question accounts for half of the testing paper and the Multiple-choice question is the second largest section, taking up 1/5. The other questions occupy same proportions.

To evaluate the optimizing performance, an experiment will be implemented to verify the performances of the intelligent evaluation function in the testing system. Some experiments have verified the superior performance of the particle swarm optimization when the parameters C_1 equals to C_2 and equals to 2 than others. In this experiment, the parameters are designed by $C_1 = C_2 = 2$ and rand=0.3 or $C_1 = C_2 = 2$ and rand=0.5 and the optimization results shown in the Fig 1 and Fig 2 respectively.

For showing the optimizing performances with different parameters clearly, the results before 1000 generation appear in the Fig 1 and Fig 2 instead of the show whole generation.
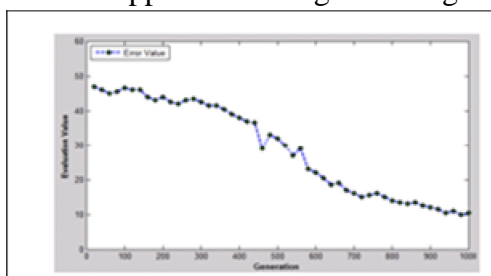

Fig. 2 The results with C1=C2=2 and rand=0.3   Fig. 3 The results with C1=C2=2 and rand=0.5
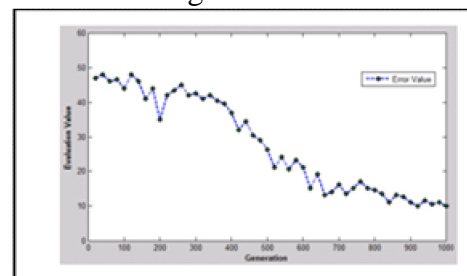
Both of the line fluctuated between 0 and 50 of the evaluation value and the trend of error value declines gradually which means errors can be reduced efficiently by using the two different parameters from the chart. It is obvious that the error value of the Fig. 1 fluctuated wider than another one. It declined slowly at the beginning then dropped sharply at generation from 400 to 100 and finally it remained decreasing slowly to 10 of the evaluation value. From the change reflected in the table, we can see that the second diagram is more stable than the first one.

As shown in the two curve diagrams above, we can find that the changing proportion of generation and evaluation value by analyzing error value.

References

[1] Du Xiang-ran, Wu Shu-jin, Wang Wen-qing. Analysis on Algorithm of Computerized Exam Paper Generation. Journal of Tianjin Vocational Institutes, No.1, 2015. pp. 156-161.

[2] Wim J. Van Der Linden, Ellen Boekkooi-Timminga, A Maximin Model for Test Design with Practical Constraints, Psychometrika, Vol. 54, No.2, 1989, 237-247.

[3] C. Chou, Constructing a Computer-assisted Testing and Evaluation System on the World Wide Web the CATES experience, IEEE Transactions on Education. Vol. 43, 2000, pp. 266-272

[4] T Hong Duan, Wei Zhao, Gaige Wang and Xuehua Feng, Test-Sheet Composition Using Analytic Hierarchy Process and Hybrid Metaheuristic Algorithm TS/BBO. Hindawi Publishing Corporation Mathematical Problems in Engineering. Vol.7, 2012, pp. 1-22.

[5] Gwo-Jen Hwang, Bertrand M.T. Lin, Tsung-Liang Lin. An effective approach for test-sheet composition with large-scale item banks [J]. Computers & Education. Vol. 46, 2006, pp. 122-139.

[6] Gui-Xia Yuan, Modeling and Research on Computer Composing Test Paper Intelligently System. Computer Simulation, Vol. 11, 2011, pp. 370-373.

[7] Ren-Jie Wu, Study on Intelligently Composing Test Paper Based on Ant Colony Optimization. Computer Simulation, Vol.8, 2011, pp. 380-384.

[8] Yin, P.-Y., Chang, K.-C., Hwang, G.-J., Hwang, G.-H., & Chan Y, A Particle Swarm Optimization Approach to Composing Serial Test Sheets for Multiple Assessment Criteria, Educational Technology & Society, Vol. 9, 2006, pp. 3-15.

[9] Xiangran Du, Min Zhang, Xizhao Wang, Self-optimizing evaluation function for Chinese-chess, Hybrid Information Technology [J], Vol.7 No.4, 2014, pp .163-172..

[10] Grabe, M. and Sigler, E. Studying online: Evaluation of an online study environment. Computer & Education. 2002, Vol.38, pp. 375–383.

[11] Huang, G.-J. Hwang, Chu, H.-C., Yin, P.-Y., Lin, J.-Y, An Innovative Parallel Test Sheet Composition Approach to Meet Multiple Assessment Criteria for National Tests. Computer & Education. 2008, Vol.3, pp. 1058–1072.