GENERATION OF INTELLIGENT ASSESSMENT SHEET WITH COMPOUND CRITERIA IN ADAPTIVE LEARNING SYSTEM

R. Kavitha*

A. Vijaya**

D. Saraswathi***

Abstract –

The computer assisted testing has proven to be an efficient and effective way to evaluate students' learning status such that proper tutoring strategies can be adopted to improve their learning performance. In a testing system, the quality of the test items will considerably affect the accuracy of a test; therefore, several measures have been proposed to represent the quality of each test item, such as degree of difficulty and discrimination. However, the quality of a test not only depends on the quality of the item bank, but also relates to the way the assessment sheet is constructed. Selection of appropriate test items is important when constructing a assessment sheet that meets compound assessment requirements, such as expected difficulty degree, expected discrimination degree, number of the test items, and the specified distribution of concept weights. In this paper, a compound criteria assessment sheet generating problem is formulated and an intelligent approach is proposed to cope with the problem. The new approach utilizes the techniques of item response theory, fuzzy logic, enabling construction of a good quality assessment sheet in accordance with specified requirements. Some evaluated results show that the new approach is capable of achieving better performance than previously used test-sheet-generating methods.

Key Words — Computer assisted testing, Assessment sheet generation, Intelligent Testing, Intelligent Tutoring.

* Dept. of MCA, AIMIT, St. Aloysius College (Autonomous), Mangalore, Karnataka, India

** Dept. of Computer Science, Govt. Arts College (Autonomous), Salem, Tamilnadu, India

*** Dept. of Computer Science, KSR College of Arts and Science, Tiruchengode, Tamilnadu, India

Introduction

Advancement of information and communication technologies has paved the way for innovation in the education system. By constructive approach of teaching, that is education has to be learner centered and learning occurs in a cognitive manner in learners' mind by means of past experiences gained and active learning that is "learning by doing in nature." Many adaptive learning and intelligent testing systems have been proposed to offer learners the customized courses. Various computer-assisted application platforms have been built, such as intelligent tutoring systems (ITSs) distance learning systems (DLSs) and computerized adaptive testing (CAT) systems [1]–[3], These systems, which are now regarded as parts of the e-learning systems [4], [5], have facilitated the traditional teaching learning- evaluation methods, thus making education more flexible and diverse. Tests are generally the most common and effective way in evaluating a learner's knowledge or ability. Traditionally, teachers or examiners need to take days or even weeks to compose a test, but the test cannot always satisfy the need in discriminating the learners' knowledge, and the attributes such as the test completion time and the difficulty degree of a test are hard to be controlled. Computer-based tests have been proven to be more effective and efficient than traditional paper-and-pencil tests due to several reasons:

First, the test sheets can be composed dynamically based on the practical requirements; second, more test items can be presented in multimedia styles; third, the student testing portfolio can be recorded and analyzed to improve their learning performance. With user interactivity and adaptability, computer-based assessment expands testing possibilities beyond the limitations of traditional paper-and pencil tests. Therefore, how to progress an efficient learning process is a critical issue. The difficulty of test depends on the examinee's performance and level of ability. It is focused because; each item affects a student's overall success throughout the test in terms of difficulty. In modern education, computer-assisted testing systems are promising in generating tests more efficiently and effectively for evaluating a person's skill. Compared to the traditional paper-and-pencil media, computer-assisted testing platforms are more favored by students. In addition, personalized assessments tailored to each student are the developing trends [3]–[5]. Personalized CAT systems select an appropriate question from the question bank based on the examinee's answer to the previous question. Ho and Yen [6] also showed that the platforms used by the examinees, such as a PC or a personal digital assistant (PDA), did not affect the performance of the CAT. The need for interactive and personalized tutoring environments has encouraged the development of computer-assisted-instruction (CAI) systems which are able to record the learning status of each student and provide adaptive subject materials and practice drills. Therefore, it is very important to precisely determine the learning status of each student so that proper tutoring strategies can be applied accordingly [5], [6]. A high-quality test is the major criterion for determining the learning status of students.

In [6], researchers propounded a "Knowledge Based Computer Assisted Instruction System", which can change the numeric component of items when the test is in progress, preventing students from memorizing the answers. Another branch of relevant researches is Computerized Adaptive Testing (CAT), which applies various prediction methodologies to shorten the length of the test participation time without loses of precision [7]. In conjunction with traditional multiple-choice, fill-in-the-blank, and short essay type questions, In [6], they suggested that Web-based instruction could also allow students' progress to be evaluated by participation in group discussions and portfolio development. Furthermore, in [8], they suggested that designers of Web-based instruction systems could create facilities to allow students to submit comments on courseware design and delivery. However, the quality of a test is not only dependent upon the quality of the item bank, but also the way in which the test sheet has been constructed. Further, it is important to select appropriate test items when constructing a test sheet that meets multiple assessment requirements, such as average difficulty degree, average discrimination degree, number of test items, and the specified distribution of concept weights. A well-constructed test sheet not only helps evaluation of the learning status of the students, but also facilitates improved diagnosis of any problems within the learning process.

In this paper, the compound criteria, assessment sheet generating problem is formulated, and an intelligent approach is proposed in constructing a realistic test sheet, according to the specified requirements. A testing and diagnostic system based on the approach is also presented. Experimental results have shown that the approach can achieve better performance than other previously used methods.

Background

In computer based tests, randomized presentation of items is automatically programmed into testing software to present different items to the test takers. The downside of such randomization is that it prevents planned sequencing of items. Randomizing items does not accommodate a test user or a constructor who wishes to ensure that items progressively become tougher. It may unfairly increase test anxiety for some of the candidates. Increased anxiety at any stage during the test for whatever reason is likely to have a negative effect on that person's performance for the remainder of the test [1]. Instead of giving each examinee the same fixed test, CAT item selection adapts to the ability level of individual examinees. In [5], they proposed an automatic leveling system for e-learning examination pool using entropy measure. The questions were leveled based on the response given by the greater part of learners with similar background. In order to assess the capacity of each question or task to distinguish between those who know and those who do not, the trial group of candidates should possess a range of knowledge from those with good knowledge to those lacking it [6].

Although many computer-assisted testing systems have been proposed, few of them have addressed the problem of systematically composing test sheets for multiple assessment requirements [2], [7]. Most of the existing systems construct a test sheet by manually or randomly selecting test items from their item banks. Such manual or random test item selecting strategies are inefficient and usually are not able to simultaneously meet multiple assessment requirements. Some previous investigations attempted to employ a dynamic programming algorithm to find an optimal composition of the test items [6]. As the time complexity of the dynamic programming algorithm is exponential in terms of the size of input data, the required execution time will become unacceptably long if the number of candidate test items is large.

In a testing system, the quality of the test items will significantly affect the accuracy of the test; therefore, several measures have been proposed to represent the quality of each test item, e.g., degree of difficulty and discrimination. These measures can be derived and updated, according to the statistical results of each test. Especially in a network-based testing system, the test results are recorded and analyzed for updating the degree of difficulty and discrimination, ensuring the quality of the item bank [8].

From the literature, it is very well seen that, there is a need of adaptive assessment with intelligence to satisfy compound criteria through some new enrichments.

Test sheet generating approach

Question Attributes in a Test

A test of n ($n \geq 1$) question(s) should be generated. Each question has several attributes, as a unique item id, difficulty degree, discrimination degree and the weight of concept(s) that the question involves.

Item difficulty is used to find out how each item affects a student's overall success throughout the test in terms of difficulty. Because it is being tried to group questions according to their difficulty level. Normally, item difficulty is scaled in a range from 0.00 to 1.00. Actually, it is inversely proportional to the number of correct answers of each question. This means that if any question has the least amount of correct answer is the hardest question in test. Hence, Item Difficulty can be calculated as,

$$ID = MSCA / SCAE$$

where,

ID is Item Difficulty,

MSCA is Minimum Sum of Correct Answers, SCAE is Sum of Correct Answers of Each Question.

Item discrimination degree indicates a question's ability to discriminate between the students who know the knowledge and those who do not. Generally, it is computed by ranking the students according to the total score. Then based on Kelley's "27% of sample" [26], select the 27% upper scoring students and the 27% lower scoring students in terms of the total score. The value of a discrimination degree ranges in [−1.00, 1.00].

Item Discrimination = (Up / U) – (Lp / L)

Where,

Up= Number of high performers with question right

Lp = Number of low performers with question right

U = Number of high performers

L = Number of Low performers

The higher the discrimination degree, the better the question does in evaluating the students' knowledge. A discrimination degree that is no smaller than 0.3 is usually regarded as acceptable. If the discrimination degree is smaller than zero, the question is not suitable for the test and should be deleted.

As questions are used for assessing whether the student has grasped the concept(s), each question is related with one or more concept(s). Suppose *M* concepts are checked in the test. Using 0 to 4 representation scheme, the relations between concepts and questions can be formulated as,

0: Test item has no relationship with that concept

1: Test item has weak relationship with that concept

2: Test item is related to that concept

3: Test item has high relationship with that concept

4: Test item has very high relationship with that concept.

Finding Expected Degree of Difficulty

Since the learning status of the students may vary significantly, a test item may be easy for one student, yet may be difficult to others. The instructors should consider the learning status of each student when determining the difficulty degree for each test item. This approach employs fuzzy logic theory to determine the difficulty levels of test items, according to the learning status of each student. The corresponding fuzzy rules are defined as follows;

IF the score is high THEN the learner's level is high.

IF the score is medium THEN the learner's level is medium.

IF the score is low THEN the learner's level is low.

After determining the learner's level, the difficulty levels of test items can be obtained by means of the membership functions for the learner's level versus difficulty. If one assumes that the degrees of membership for the learner's level's are x_1, x_2 and x_3 for High, Medium, and Low respectively, the final degree of difficulty can be calculated by,

Expected Degree of Difficulty = $(\mu_{\mathbf{High}} (x_1) * x_1 + \mu_{\mathbf{Medium}} (x_2) * x_2 + \mu_{\mathbf{Low}} (x_3) * x_3) / (x_1 + x_2 + x_3)$

Formulate test sheet

In an item bank, a subset of n candidate test items Q1, Q2, . . .,Qn will be selected for composing a test sheet. The model proposed here considers different compound assessment requirements. Assume there are 'n' items in the item pool and 'm' concepts to be dealt with. The measures which are in need of assessment are as follows:

(1) Decision variables xi : $1 \le i \le n$, xi is 1 if test item i is selected; 0, otherwise.

(2) Degree of Difficulty of an item di : $1 \le i \le n$

(3) Degree of discrimination of item ei : $1 \le i \le n$

(4) Concept involved cj : $1 \le j \le m$

(5) Degree of association between an item Qi and concept Cj r_{ij} : $1 \le i \le n$ and $1 \le j \le m$

(6) Lower Bound on Expected Concept Relevance CLj : $1 \le j \le m$

(7) Upper Bound on Expected Concept Relevance CHj : $1 \leq j \leq m$

To generate test sheets for multiple requirements, this paper proposes a new approach. Without loss of generality, one can assume each test item contains the following information:

- a measure of its difficulty level;
- a measure of its discrimination value;
- a weight to represent the relationship between each test item and each concept.

Based on the assumptions, the compound criteria test-generating problem can be described as follows:

(1) The number of step is the number of test items to be included in the test sheet, and in each stage a decision is made to select a test item from the item bank.

(2) At each stage, there are several possible states which represent the difficulty level, the discrimination level, and the distribution of concept weights, based on the currently selected test items.

(3) Since a test item is selected in each stage, the difficulty level, the discrimination level, and the the distribution of concept weights are recomputed, and the new values are used as the starting values of the next stage.

(4) The final difficulty level of the test sheet should be closer to the desired values under the constraint, so that the weights of the concepts satisfy the specified distribution.

Minimize Z = $\sum_{i=1}^{n} (|s_i - s_g|)$

Where

$$s_i = \frac{(s_{i-1} * i - 1) + d_i}{i}$$

s_i is the degree of difficulty of the test sheet if a test item Qi is selected

s_g is the expected degree of difficulty for a student

Subject to,

$\sum_{i=1}^{n} r_{ij} \geq CL_j$ and $\sum_{i=1}^{n} r_{ij} \leq CH_j$ where j=1,2, …, m

$\sum_{i=1}^{n} d_i x_i \, / \, \sum_{i=1}^{n} x_i \geq 0.5$

The above approach tries to generate a test sheet which is having difficulty degree closer to the expected difficult level. Also, only the items having maximum discrimination values are taken. Concurrently the degree of association between the item and the concept is checked with the given range of association.

Hence forth, this approach provides an intelligent test sheet with more quality.

Evaluation and Discussion

The performance of the proposed approach has been evaluated through an experiment completed for 5 cases which specify various degrees of difficulty and discrimination with different similarity thresholds. The proposed approach is compared with the Random Item Selection method and with the objective requirements.

The random selection program generates the test sheet by selecting test items randomly to meet the constraints of number of test items.

TABLE I. COMPARISON OF DEGREE OF DIFFICULTY OF DIFFERENT APPROACHES

Case No.	Expected Difficulty	Random Selection	Intelligent approach
1	65	75.5	63
2	78	67	79.25
3	91	87	90
4	63	68.25	60.1
5	40	31	42.5

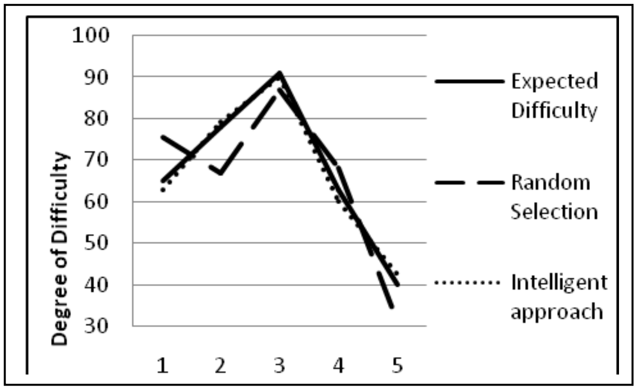

Figure 1: Degree of Difficulty of different approaches

From the above table and figure, it can be apparently seen that assessment sheets with near expected difficulty degrees can be obtained than by the random selection approach.

Conclusion and Future Work

The approach of using computer-assisted testing systems to release teachers from the burden of composing tests and improve the assessment quality of tests is significant and promising in modern education. The multiple criteria test-sheet-generating problem is formulated, and an intelligent approach is proposed to generate test sheets that meet multiple assessment requirements. The question attributes in a question bank are adaptively adjusted, always reflecting students' learning states. From some experimental results, the approach achieves desirable performance under considerations of difficulty,

Several other AI or optimization based technologies and heuristic algorithms could be exercised to develop more efficient test sheet generating approaches for very large item banks. The combination of intelligence and personalization is the future direction, which will be addressed in the forthcoming work.

References

[1] C. Chou, "Constructing a computer-assisted testing and evaluation system on the world wide web-the CATES experience," *IEEE Trans. Educ.*, vol. 43, no. 3, pp. 266–272, Aug. 2000.

[2] S. Piramuthu, "Knowledge-based web-enabled agents and intelligent tutoring systems," *IEEE Trans. Educ.*, vol. 48, no. 4, pp. 750–756, Nov. 2005.

[3] E. Guzm´an and R. Conejo, "Self-assessment in a feasible, adaptive webbased testing system," *IEEE Trans. Educ.*, vol. 48, no. 4, pp. 688–695, Nov. 2005.

[4] A. Kavˇciˇc, "Fuzzy user modeling for adaptation in educational hypermedia," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 4, pp. 439–449, Nov. 2004.

[5] G.-J. Hwang, B. M. T. Lin, H.-H. Tseng, and T.-L. Lin, "On the development of a computer-assisted testing system with genetic test sheet generating approach," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.,*vol. 35, no. 4, pp. 590–594, Nov. 2005.

[6] Xiao-Min Hu, Jun Zhang, "An Intelligent Testing System Embedded with an Ant-Colony optimization based Test composition Method", *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 39, no. 6, pp. 659–669, Nov. 2009.

[7] Marks, A. M., & Cronje, j. C. "Randomised items in computer-based tests: Russian roulette in assessment?". Journal of *educationaltechnology & society*, *11* (4), 2008.

[8] Betül Erdoğdu, "Computer based testing evaluation of question classification for Computer Adaptive testing", A Master Thesis (2009).