Intensified Leveling and Assessment System in Intelligent Tutoring using Decision Tree and Item Response Theory

R. Kavitha
Assistant Professor, Department of MCA,
Aloysius Institute of Management and Information Technology (AIMIT),
Mangalore, Karnataka, India.
kavitha.rajamanii@gmail.com

Abstract

Due to the rapid growth of information and communication technology, the education environment has been enriched and become more diversified. Hence computer based assessment came as a prevalent method of administering the tests. Randomization of test items here may produce unfair effect on test takers which is unproductive in the outcome of the test. There is a need to develop the Intelligent Tutoring System that assigns intelligent question depending on the student’s response in the testing session. It will be more productive when the questions are assigned based on the learner ability in the early stage itself. So, this study focus on building up a framework to automatically assign intelligent question based on the learner ability not only all through the session, but also on entry. At end, automatic up gradation of the question levels can be done dynamically. The questions are classified based on item difficulty using Item Response Theory model. The learners are classified based on their level by Decision Tree using ID3 algorithm. Thereby the objective of Intelligent Tutoring System is achieved by using adaptability and intelligence in testing.

1. Introduction

Computers and electronic technology today offer myriad ways to enrich educational assessment both in the classroom and in large-scale testing situations. With dynamic visuals, sound and user interactivity as well as adaptability to individual test-takers and near real-time score reporting, computer-based assessment vastly expands testing possibilities beyond the limitations of traditional paper-and-pencil tests. This is referred as the Computer Based Testing (CBT). 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.” Each student has their own learning style, and this yields a result that each student’s performance in learning cannot be assessed and evaluated in a unique and simple way. Also it cannot be evaluated only by measuring the test results depending on the number of right and wrong answers. Therefore, how to progress an efficient learning process is a critical issue. For this, testing must be intelligent. It must behave as if a teacher asks questions to a student in real class environment. If student cannot answer the question, teacher must ask easier question about similar subject and if student answer this time, then, again more difficult question must be asked. Computer Assisted Assessment (CAA) is “the networked nature of approach, which provides for distribution of formative or summative assessment directly to multiple client computers with little or no additional hardware or software installation needed apart from standard access to the Web.” The regard is the testing part and the concern is to develop an intelligent testing application that will produce intelligent questions depending on the student’s responses and performance during testing session. This kind of testing is called as Computer Adaptive Testing (CAT).

CAT is a methodology of testing which adapts to the examinee’s level. CAT selects questions in order to maximize the performance of examinee by observing the past success throughout the test. Therefore, 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. This is how the Intelligent Tutoring System (ITS) can be developed. It is a system that provides direct or indirect customized instruction or feedback to learners whilst performing a task. In this regard, the question levels must be determined somehow. In this study, item difficulties of questions were estimated using item responses (Answers of students to each question) by Item Response Theory model. At the same time, the learner ability can be predicted by learning through the decision tree which
was built using academic attributes by ID3 algorithm using training data set. So, at the commencement of the test, the intelligent question is assigned to the learners based on their ability. Also, by the end, based on the overall responses in that particular session, the question levels can be revitalized as such.

The remainder of the paper is organized as follows. The second section is devoted to the literature survey behind this study. Third section deals with the methodology of how well the framework is built. The final section provides the conclusion and enhancements.

2. Literature Survey

There are lots of academicals and commercial work done on computer based testing applications. The need of speed, time flexibility, low-cost, fair scoring and besides the unceasingly increasing information technology makes the computer based testing applications essential. In recent years many researches has been done on this issue and below is the historical evolution.

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 [4].

In a research study [2], it was proved that randomization is ensuring the test security, yet progressively allowing items to become more difficult as the test items are presented to each test-taker, will prevent occurrence of the item randomization effect. Instead of giving each examinee the same fixed test, CAT item selection adapts to the ability level of individual examinees. After each response, the examinee’s ability estimate is updated and the subsequent item is selected to have optimal properties at the new estimate.

Dynamic assessment (DA) [10] has been advocated as an interactive approach to conducting assessments to students in the learning systems as it can differentiate student proficiency at the finer grained level. Different from traditional assessment, DA uses the amount and nature of the assistance that students receive which is normally not available in traditional practice test situations as a way to judge the extent of student knowledge limitations.

In [3], Feng et al compared dynamic assessment against a tough contrast case where students are doing assessment all the time in order to evaluate efficiency and accuracy of dynamic assessment in a tutoring system.

The development of item response theory (IRT) in the middle of the last century has provided a sound psychometric footing for CAT. It is a modern test theory and is currently an area of active research. The key feature of IRT is its modeling of response behavior with distinct parameters for the examinee’s ability and the characteristics of the items. The need to upgrade from ordinary CBT to CAT is well concentrated and illustrated [9].

Asking an easy question to a high ability student would not provide true information about his/her ability even the answer is correct. Likewise, a difficult question answered wrongly by a less successful student would not show the real ability level of the student. By selecting and administering questions that match the individual student’s estimated level of ability, questions that present low value information can be avoided [7]. Low performance student might be disappointed and confused, high performance students might be bored and tired of questions with inappropriate levels of difficulty. So it can be stated that in addition to increasing efficiency, CAT also increase the level of interaction and motivation of the student.

S.C. Cheng et al [6] 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.

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

3. Methodology

Computer Based Testing has thousands of questions in a question pool but the difficulty levels are not determined. By intellectual question classification, Instructors will be able to assign different questions to different students. This kind of instructional methodology can develop education quality and efficiency. Since, the Computer Adaptive Testing pose questions based on the ability of learner, the item difficulty of the question has to be found out initially. And these questions are composed of text formatted and they are multiple-choice questions which are commonly used in electronic testing environments. All available tests in CAT are assigning the question with medium level difficulty to the learner first. After that, based on the response successive questions depending on their capability level is posed. This study involves finding the learning level or capability of the learner first. Hence, their entry into the test itself is based on
individual ability. This conform the effectiveness of Intelligent Tutoring Systems.

The next subsection gives an insight on data gathering. The subsection 3.2 gives an approach on how the questions are classified based on the item difficulty using IRT model. The subsection 3.3 deals with the classification of learners using ID3 algorithm. For this study, three classes of learners were developed and the decision tree is constructed. When the new test taker comes, based on the decision tree, the level of that learner can be easily found and then the specific question that outfits the level is assigned absolutely. At the end of the test session, both the decision tree and also the question level can be revised to keep it up to date.

The framework proposed for this study is as follows:

![Diagram](image)

**Figure. 1 Proposed structure of Intensified Leveling and Assessment in ITS**

### 3.1. Data Gathering

The Data set introduced here consists of the Third Class Test taken by 46 students of MCA course during the second semester for the subject Object Oriented Analysis and C++ in AIMIT. There were 25 multiple choice questions, some with 4 choices and some others with 2 choices as True or False type. Questions were delivered via a CBT on Moodle Learning Course Management System. Also, the programming ability, pre test and assignment proficiency was taken for the same students, same course to develop the decision tree for classifying the learners.

### 3.2. Question Classification using IRT

The reason for classifying the questions is to enhance the computer based system to adaptive mode with intelligence in tutoring. All items in the test item pool range in varied difficulty values. If an examinee gets an item right, an item having a greater difficulty is selected from the item pool and delivered to the examinee as the next question of the test. If he/she gets it wrong, then an item having a smaller difficulty is selected from the item pool and delivered to the examinee as the next question of the test.

The initial nominal question levels should be determined. Then the item difficulties can be found and they can be placed into exact nominal question levels. For this study, five levels of difficulties of items are considered as Very Easy, Easy, Medium, Hard and Very Hard. The nominal levels can be consigned as follows.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Question Level</th>
<th>Numerical Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Very Easy</td>
<td>-1.0</td>
</tr>
<tr>
<td>2</td>
<td>Easy</td>
<td>-0.5</td>
</tr>
<tr>
<td>3</td>
<td>Medium</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>Hard</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>Very Hard</td>
<td>1.0</td>
</tr>
</tbody>
</table>

IRT encompasses any model "relating the probability of an examinee's response to a test item to an underlying ability." Here, the true score is defined on the latent trait of interest rather than on the test, as is the case in classical test theory. Item difficulty can be determined by using IRT approach with one parameter which uses the formula below.

\[
ID = \frac{MSCA}{SCAE}
\]

Where,

- ID = item difficulty
- MSCA = Minimum Sum of Correct Answers
- SCAE = Sum of Correct Answers of Each Question

<table>
<thead>
<tr>
<th>Q No</th>
<th>SCAE</th>
<th>Item Difficulty</th>
<th>Nominal Question Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>0.83</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>37</td>
<td>0.27</td>
<td>-0.5</td>
</tr>
</tbody>
</table>
According to the data set, MSCA is 10. The 21st question is the least correctly answered question and hence has the maximum item difficulty of 1.0.

For the items in Table 2, the mean and standard deviation is calculated using ID values and it is used to place items into related nominal question level.

Mean = 3.46
Standard Deviation = 1.2

According to these mean and standard deviation values, a scale, shown in Figure 2, is designed for placing items into nominal question levels. And all items are placed into related nominal question level as shown in Table 2.

3.3. Finding the Learner ability through Decision Tree

The same set of learners was taken into this study for developing the decision tree. They were classified into 3 labels as Perceptive, Average and Naïve based on their performance in that test. A decision tree is developed using ID3 algorithm for easy prediction of level of a new test taker. Decision tree induction is the most prevalent method in supervised learning from class-labeled training tuples.

The algorithm Iterative Dichotomiser (ID3) is used in this study to develop decision tree. Information gain is used as the attribute selection measure which gives the “information content” of measure. The attributes used here are programming ability, assignment proficiency, and pre-test. The programming ability is considered with two levels as Excellent and Fair. The Assignment proficiency is represented as Good, Average and Poor levels and pre-test with high and low. The decision tree developed is as below.

![Figure 3. Classification of learners by Decision Tree](image)

Here, C1 represents class of Perceptive learners; C2 represents Average learners and C3 for Naïve learners. From the above tree, it is well understood that, those learners who have good assignment writing proficiency are perceptive and average are medium level learners. Also, poor assignment writers can be average learners when they have the excellent programming ability. The tree developed by ID3 was subjected to post-pruning method and the above tree is framed.

3.4. Intelligent Question Assigning by ITS

When a new test taker comes, based on the attributes for decision tree, the level of that specific learner can be predicted by passing over the decision tree. Based on the level, the precise difficulty level of question that matches him/her is taken from the intelligent examination item pool. The methodology followed here is,

<table>
<thead>
<tr>
<th>Learner Level Vs Intelligent Question Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of Learner</td>
</tr>
<tr>
<td>Perceptive</td>
</tr>
<tr>
<td>Average</td>
</tr>
<tr>
<td>Naïve</td>
</tr>
</tbody>
</table>
If the learner succeeds by giving correct response, then the question from next greater level of difficulty is posed. Or else in case of wrong response, the question from next lower level of difficulty is posed. At end, both the decision tree and item difficulty level are automatically updated to meet the need of future test takers.

4. Conclusion and Further Enhancements

There is a great need in the education area to monitor test results on a large scale as well as to identify questions that are most likely to be benefited by student according to the knowledge level of the student. The applications of item response theory modeling help this issue. Item banking allows for the development of computerized adaptive tests that reduce respondent burden and increases reliable measurement by using a methodology that targets in on a respondent’s true score. This study proposes an intensified leveling system using IRT and Decision Tree. The level of new learners can be predicted using the decision tree which was built over by training set of data. The items in the pool are leveled on difficulty using IRT. Therefore, the system poses the intelligent question from the pool based on the predicted level of ability. Since the test is based on their learning ability, test fairness is maintained. Also effective learning is achieved without any compromise which is the objective of Intelligent Tutoring System.

Other type of item analysis model can be tested to give better performance since many techniques are available in IRT with different parameters. This process is easy to test since all the data has already been made nominal and is ready to be classified. The system can be designed in such a way that it can accept other attributes to develop the decision tree which can make it a generic one.

Reference

[1] Jiawei Han and Micheline Kamber, Data mining concepts and techniques -, second edition.
[3] Mingyu Feng, Neil Heffernan, Can we get better assessment from a tutoring system compared to traditional paper testing? Can we have our cake (better assessment) and eat it too (student learning during the test)?,