A Theoretical Framework for Knowledge Level Assessment in Intelligent Tutoring Systems Using Machine Learning and Adaptive Testing

Rangasubramanian K,
Student, Department of Electronics and Communication,
Sri Sivasubramaniya Nadar College of Engineering,
Kalavakkam, Tamil Nadu, India
krangasubramanian@gmail.com

Abstract

Machine learning, in general and knowledge level assessment, in particular are gaining momentum these days in the academic and industrial forums due to their immense scope and myriad applications. Ground breaking research has been carried out in the area of student modeling, performance prediction, difficulties faced by a student in a learning environment and so on. This paper attempts to propose only a theoretical framework wherein an intelligent tutoring system would be able to pinpoint a candidate’s knowledge level in a particular topic using various attributes such as question difficulty, time for solving and correctness of answers.

Keywords: Machine learning, Intelligent Tutoring System, Adaptive Testing, Knowledge Level Assessment.

1. Introduction

Intelligent Tutoring Systems (ITSs) are computer programs that enable one-to-one tutoring (R. Baker, A. T. Corbett, and V. Aleven, 2008) with the primary focus of imbibing knowledge of the subject and then testing, correcting and improving the abilities of the student in that subject (Henrique Seffrin, Ig I. Bittencourt, Seiji Isotani, and Patricia A. Jaques, 2016).

Knowledge sharing occurs through content shared to the student from and by the ITS. Herein lies the advantage of the ITS. The student can learn at his/her own pace and also can refer the content again and again in case a concern/doubt occurs. Moreover, each student has his/her own space to learn, unlike a classroom where many students are put together under one teacher, thereby reducing the time invested in one student by the teacher.

Knowledge gained in a topic must not be confused with exam preparedness or grasping power of the student, even though there is only a thin line of difference between the two.

A student may mug up the topic before an exam to score marks, but may not be able to apply it critically and intuitively. Therefore, he/she is exam-prepared but that does not become a correct indicator for his/her knowledge level. A similar analogy can be used to say that grasping power, even if it is quantitatively analyzed, does not become a just indicator for a student’s knowledge level. Therefore, knowledge level assessment can be done only by analyzing how the student is able to use his/her knowledge to solve problems that genuinely test the same.

2. Related Work

As stated earlier, a lot of research has been conducted combining the fields of education, e-learning and machine learning. This has been done to a large extent by student modeling and user activity analysis.

A. The existence of association rules among the topics covered in a course (S. Valenti and A. Cucchiarelli, 2003) was done by analyzing answers to Multiple Choice Questions (MCQs) by students and then analyzing the answers using different scores for different outcomes. Quinlan’s C4.5 was used for decision trees and classification. While the model did predict relations and similarities between two similar topics with high confidence factors, it failed to distinguish
between topics which were totally non-relatable when the confidence factor was high for both the topics.

B. Student evaluation and prediction of students’ final grade in e-learning (Tahira Mahboob, Sadaf Irfan, and Aysha Karamat, 2016) was done on the basis of previous assessments. The model was very similar to the Hidden Markov Model (HMM) and was an attempt to overcome the limitations of other similar models. Quinlan’s C4.5-J48, Naïve Bayes and Random Forest were the algorithms applied on the data, out of which Random Forest depicted that 100% instances were correctly classified. However, this paper does not say anything significant about the student’s knowledge level.

C. Student’s learning achievement, its quantification and ranking the students based on their learning achievement (Shyi-Ming Chen and Ting-Kuei Li, 2010) was done by simply generating the importance degrees of attributes of questions. The paper was an attempt to overcome the limitations in the paper proposed by Bai and Chen (2008) on the same topic. It established the simpler and easier method of using matrices over that of fuzzy logic and even arrived at the same results. However, it does not say anything concrete about knowledge level assessment.

D. Students’ algebraic knowledge has been modeled using Dynamic Bayesian Networks (DBNs) (Henrique Seffrin, Ig I. Bittencourt, Seiji Isotani, and Patricia A. Jaques, 2016). The researchers have used concept maps to arrive at relations between topics and then, depending on the interdependencies of the topics, have evaluated the student’s knowledge in a particular topic using DBNs. However, it has not given a generalized solution.

E. Analyzing user experience, incorporating concept maps and generating activity reports have been proposed as the key players in an e-learning system (Nazeeh Ghatasheh, 2015) in a bid to assess the knowledge level. However, even though the model gave correct results to a large extent, concept maps cannot be relied on due to ambiguity and tediousness involved, thus resulting in wrong assumptions as was the problem in the paper proposed by Valenti and Cucchiarelli (2003).

F. Learner knowledge level calculation and concept weight estimation using concept maps and neural networks (Ahmad Kardan and Negin Razavi, 2014) have been attempted. In the end, the total score of the student is calculated and based on that, concept weight estimation is done using neural networks. However, the paper is not able to pinpoint a particular knowledge level of the learner. Moreover, the use of concept maps will result in complications associated with [3] and [7].

G. Student knowledge level estimation by using Bayesian networks and adaptive assessment (Samad Kardan and Ahmad Kardan, 2009) has been one of the most popular research topics of late. This is very effective in assessing student’s knowledge level. However, it does not take into account some finer aspects of assessment like time for answering and also does not explain clearly the type and order of questions.

H. Adaptive testing/assessment has become very popular (Yang Chen, 2013; M. Birenbaum, 1994; C.E. Dowling, 1996; Di Challis, 2005; I. Hatzilygeroudis, C. Koutsojannis, C. Papavlasopoulos, and J. Prentzas, 2006). Also known as Computer Adaptive Testing (CAT), it adjusts the level of questions to the learner’s level that it estimates from previously answered questions. However, the accuracy of this is questionable. Due to the time limited nature of real world tests, the questions that are put forth earlier will have a greater impact on the score of the learner. Moreover, the CAT system does not account for the possibility of the chance of intelligent guessing by the learner. Moreover, when the difficulty level of successive questions is not different, the system will stop the assessment, resulting in a shorter assessment, from which no concrete and just information can be obtained. These problems have been solved to a significant extent in the model proposed by Samad Kardan and Ahmad Kardan (2009).

3. Proposed Framework

Literature reviewed above suggests that adaptive testing is by far the most efficient, effective and compatible method for testing. It does not follow the one-size-for-all logic of other testing models. Moreover, it is currently one of the most compatible forms of testing given the requirements of the current world. This research is aimed at improving on its methods and overcoming its limitations.
3.1. Time taken for solving

A student who knows the formula to answer a question may readily apply it to the problem. However, one who does not know it will attempt to derive it and then apply it. This will naturally take more time. Therefore, proposition

*P1: Time taken for solving the question is a key player in determining the knowledge level of the student.*

3.2. Order of questioning

Intelligent guessing is a major problem in any assessment. There is always a chance that a student could arrive at the correct answer by means of good guesswork. Naturally, the fact that he/she answered correctly by means of intelligent guesswork cannot contribute to the outcome of the knowledge assessment process. So, if a student answers a question of a new knowledge level correctly, a completely different question of the same knowledge level should be put forth, before moving onto higher concepts and knowledge levels. It is even better if the same concept can be tested in different manners, depending on time invested. Therefore, proposition

*P2: The order of questioning is very important in order to eliminate the chances of intelligent guessing.*

3.3. Application oriented questions

Knowledge does not refer just to the knowing of facts/formulas but also applying them aptly and creatively to the problem at hand. Knowledge is related to how much a student has understood in a topic. Therefore, proposition

*P3: The questions should be more application oriented to really test the knowledge level of the student.*

3.4. Implementation of Adaptive Testing Model (ATM)

The maximum knowledge level to be tested in a topic should be fixed. Thereafter, questions should be uploaded in sufficient number that can effectively nullify the problems arising due to guessing and slipping by appropriate use of DBNs. This should be done by putting forth secondary questions after the primary questions unconditionally. Tertiary questions should be put forth depending on the confidence level exhibited by DBNs. The time taken for solving and the difficulty level of the question should be taken into account to ascertain the same. When confidence level is high, the ITS can move to the next knowledge level. The assessment can stop after the candidate achieves a high confidence level in the topmost knowledge level.

4. Discussions

The framework has been structured with the sole purpose of improving on, and at the same time, overcoming the challenges of, the models proposed earlier using adaptive testing. Machine learning, especially the intelligent use of DBNs along with adaptive testing can produce very accurate results in this field, provided the factors proposed are taken into account.

Putting forth a question of the same difficulty level when the candidate has failed to answer the previous one correctly is not difficult to understand when one considers elementary probability. The chronological order of the steps mentioned depending on the binaries of the response initiated by the user are simply effective ways of making sure of the candidate’s knowledge level and eliminating the chances of intelligent guesswork.

Involving machine learning and DBNs in this process would be highly effective, especially while taking into consideration the time taken for solving, difficulty of questions and that the order of questioning itself is used to reinforce upon the former two. This could possibly be done by a few algorithms like regression, classification, principal component analysis, to name a few.

The approach considered here has only been qualitative. Quantitative modeling and figuring out the best algorithms for the same that give a high performance rate could be considered as future areas of research in this field.
5. Conclusion

Machine learning has a high significance and regard in today’s world and this is only going to increase. In the field of learning and education, adaptive testing has emerged as one of the unique and effective ways for assessments. It is therefore, only natural that the combination of machine learning and adaptive testing would be highly efficient to get the desired results. Learning and education are very two important words today. There is no doubt that such technologies and improvements in the same would improve the experience of learning for all. As stated earlier, the approach here is only qualitative and no effort has been taken to empirically verify or prove the effectiveness of the model. This could be considered for further research.

6. References