Optimization of Moodle for Adaptive Testing

Xin-di LAN, Hao-yan YU and Ji-mei LI

College of Information Sciences, Beijing Language and Culture University, China

*Corresponding author

Keywords: Adaptive test, Moodle platform, Item Response Theory, Learning personality characteristic model, Personalized recommendation.

Abstract. In the era of big data, Computerized Adaptive Testing (CAT) with flexibility, easy access and other advantages, has become a new way of examination, but supports for adaptive test on Moodle are still lacking. Moodle, which has been widely used, is a free and open-source platform, and an object-oriented modular development of dynamic learning environment based on the social constructivism theory. The design of this paper applied the Item Response Theory (IRT), the maximum information quantity selection method and the learning personality characteristic model, which can realize the adaptive test on the Moodle. It can track the learner's learning ability and learning process better, and match learning materials with the learner better, in order to improve learning efficiency and effectiveness.

Introduction

CAT has many advantages over traditional exams, and can efficiently record and track learners' test results and learning ability. CAT has greater interactivity as well. A CAT system can give the corresponding test questions according to the learning ability of different students, so as to make a more accurate measurement on the students' learning achievement, which is more flexible than traditional tests.

At present, CAT based on the Item Response Theory (IRT) is the direction of many researches. The focuses of these researches are how to estimate the parameters, as well as the composition policy of test papers, etc.

The main purpose of this paper is to build the adaptive test into Moodle under Item Response Theory (IRT) and parameter estimation method, in order to optimize Moodle platform. We apply the maximum information quantity method in the adaptive test module. A user device carries out a test paper, and the adaptive test module tracks the process of learners’ response. In addition, the learning personality characteristic model is applied to send appropriate learning materials according to the learning personalities, so as to improve users’ satisfaction and learning efficiency.

Literature Review

Moodle

The Modular Object-Oriented Development Learning Environment (Moodle), which is a set of open source network Course Management System (CMS), was originally developed by Dr. Martin Dougiamas, Australia. Moodle is a series of object-oriented modular developmental dynamic learning environment based on social constructivism theory[1]. The basic modules are site management, user management, curriculum management, job module, forum module, chat module, resource module, test modules, and so on. Moodle's interface is simple, but very powerful. It meets the majority of teaching needs, and has good compatibility. Moodle supports various international resource standards such as IMS and SCORM, and has good security performance. Therefore it has been favored by the majority of educators around the world. Dr. Martin Dougiamas released version 1.0 in 2002. For the next 14 years of continuous development and update, the version 3.3.2 has already been released.
According to the data displayed by Moodle official platform in December 2016, a total of 121029 sites from 232 countries are using Moodle. Moodle has become one of the most popular network teaching platforms. In China, by the promotion of Professor Li Zuhou from early 2006, its use is also rapidly expanding[2].

With the research and application of Moodle in depth, teachers began to understand it and try to use it for online course development. At the same time, schools of all levels adopt the platform to build their own network courses. However, there are still some problems in the research and application of Moodle platform[3]. Firstly, the researches do not focus on the auxiliary teaching of Moodle. They mainly focus on the exploration of the practicality of technology, such as curriculum development, management and network applications. However, the study of network-based teaching and learning is not enough. There is little research on the organization and evaluation of learning activities.

Secondly, many researchers only assess Moodle from the macro level, instead of using the platform to carry out network teaching or the exploration of learning model.

Finally, most researchers are limited to the use of existing modules in Moodle, instead of thinking the expansion of its functions, the improvement of its interface, and the stylization of its resource display style.

In summary, it is necessary to develop new features to better adapt to the learner's learning situations on Moodle platform. Therefore, designing adaptive testing on Moodle is of great theoretical value and practical significance.

Adaptive Testing

Computerized Adaptive Testing (CAT) is a new type of test. In a CAT system, different levels of testees can receive a group of adaptive questions matching with their own level. The test can automatically adapt to the specific situation of the testees[4]. In the process of testing, CAT can assess the possible level of the testees based on the testing situation and present the appropriate items, which will more accurately assess the ability of testees[5].

Traditional tests are not targeted and not able to accurately test the real level of the testees. As the adaptive test can solve the problems, now it attracts attention and has significant development in the field of education testing. CAT was first introduced to the United States and has been used in various fields such as the Graduate Record Examination, the Graduate for Management and Administration Test, and the Nurse National Committee License Test[6]. In China, it started late. But in recent years, there have been some adaptive tests and successful applications. For example, the Shanghai TV University has adopted the CAT test design method in the application of computer application capacity test project "VB6.0 program design". In the beginning of the 19 centuries, the CET-4 and CET-6 had also been committed to CAT research and development.

In adaptive testing, the commonly used theory is item response theory.

Item Response Theory and Model. Item Response Theory (IRT) is a modern theory of psychological measurement, which can be used as the theoretical basis for item selection and test preparation. IRT assumes that the subject has a potential trait. In the test, the potential trait generally refers to the potential capacity, and is often estimated by use of the test scores. IRT suggests that the testees' responses and achievements in the test items have a special relationship with their potential traits[7].

Hambleton and Swaminathan defined IRT as follows. In the test, by assessing the characteristics of the testees, the traits or abilities, they could estimate the scores of the testees on these traits (called competency scores) and use these scores to predict the item and answer, and to explain and predict the testees' answers. Traitor ability and item are the core concepts of item response theory. The relationship between the two is the main connotation of IRT[8].

In the process of studying IRT, the earliest practical application is the two-parameter normal oval model proposed by American educational measurement expert Lodd in 1952. However, because the calculation process in the model is too complex, it is not contributed to in-depth study. Subsequently,
Bourne Baum proposed the Logistic model and the related statistical analysis method, which is more computable than the normal oval model. Thus, it becomes the most commonly used model in IRT [9].

Logistic model has three commonly used models—single-parameter model, two-parameter model and three-parameter model [10]. The most commonly used is the three-parameter model and the formula is

\[
p(\theta) = c + (1 - c) \frac{1}{1 + \exp[-2a(\theta - b)]}.
\]  

(1)

In Eq.1, \( \theta \) is the ability of the subject, \( a \) is the degree of discrimination of the test, \( b \) is the difficulty of the test, \( c \) is the guess coefficient of the question, \( p(\theta) \) is the probability that the person with the ability \( \theta \) gives the right the answer to this question and \( D = 1.702 \). When \( c = 0 \), the formula can be transformed into two-parameter model. When \( c = 0 \) and \( a = 1 \), the formula can be transformed into single-parameter model [11].

**Basic Processes of Adaptive Testing.** CAT is based on the level of candidates to adjust the questions. It can adjust the difficulty of the next question according to the previous performance. The more questions the testees do, the more accurately the computer can estimate their ability. Until the pre-set termination conditions are all reached, the test ends and gives the corresponding ability value. In general, the CAT is divided into three parts, which are the starting condition, the subject selection policy and the termination condition [12].

**Part I, Starting Condition.** In this part, we need initializing the testee’s ability and make parameter estimation.

In general, there are three methods to initialize the testee’s ability.

- Select the questions of middle difficulty by CAT as the starting point of the test from the question bank randomly.
- Present the moderately difficult questions based on the previous test records (historical ability value) if the testee has previously participated in the test.
- Select the starting questions by the testee, who determines his own ability value [13].

In the three methods, the second method can only be applied when there is a history test record. The first is suitable for the initial test and simple to operate. The third is relatively free. But the selection is not accurate when the testee does not know his level of competence.

The maximum likelihood estimation method (Birnbaum, 1968) is the most widely used parameter estimation method. It is based on the likelihood function of the testee’s response pattern. The likelihood function is a joint probability that is the testee with ability \( \theta \) to produce the reaction mode \( U \).

The reaction mode with the capacity \( \theta \) and the likelihood function \( \ln L(U|\theta) \) is calculated as follows[14].

\[
L(U|\theta) = \prod_{i=1}^{n} p_i(\theta)^{u_i} q_i(\theta)^{1-u_i}
\]  

(2)

The maximum likelihood function method is to obtain the maximum value of \( \theta \) in the likelihood function \( \ln L(U|\theta) \) to the reaction pattern \( U \) of the testee to a set of questions, and the value of \( \theta \) is the estimated value of the testee’s ability. In general, the maximum likelihood estimate is calculated by the method of calculating extremum of \( \ln L(U|\theta) \).

**Part II, Subject Selection policies.** There are two prevalent subject selection methods that are approximate selection method and the maximum information quantity selection method.

Approximate selection method is that the questions are selected based on the answer of the testee and the difficulty of the questions. It can be applied in the initial stage of CAT. Firstly, the testee will be presented the random selected question or the question of the middle difficulty. If the question is answered correctly, then the computer will select a little more difficult question than the previous question from the question bank. If the answer is wrong, then the computer selects a few less difficult question. The method is simple and easy. But it is primitive since it does not measure the impact of the distinction between questions and guessing answer [15,16].

483
The maximum information quantity selection method is to select the question with the greatest amount of information function based on the current estimated ability of the testee. The problem information function is defined as follow.

\[ I_i(\theta) = \frac{r^2_i(\theta)}{P'_i(\theta)Q'_i(\theta)} \]  

(3)

In Eq.3, \( I_i(\theta) \) represents the amount of information question \( i \) to the testee with the ability \( \theta \). \( P'_i(\theta) \) is the first derivative of the \( P_i(\theta) \) at the point \( \theta \), and \( P_i(\theta) \) is the probability of correct answer of the ability \( \theta \) on the question \( i \), \( Q'_i(\theta) = 1 - P'_i(\theta) \) [17].

This method is the most widely used method in CAT. It has high testing efficiency and can reach required accuracy quickly. According to the definition of the project information function, the greater the degree of discrimination, the closer the difficulty is to the current ability of the testee. When the answer whose degree of guessing is close to 0, the value of the project information function will be the greater. But this will lead to the repetition of the questions that are consistent with this characteristic and too much exposure, which will be unfair to the previous testees [18].

**Part III, Termination Condition.** There are three methods to determine the termination condition.

The first one is fixed-length method. According to the actual test needs, pre-set the number of questions for each testee, that is, the project length. If it meets the previous length of the items, the test will end.

The second is variable-length method. According to the actual needs of the test, set the accuracy of the test in advance. If it meets the predetermined accuracy, the test is over. Determine the difference between the estimated ability value of the adjacent two tests. If the difference is less than or equal to the predetermined accuracy, the test will end.

The last one is comprehensive method. Firstly, set the maximum allowable test length. If the test is within the maximum length and meet the information control method, then it will end. Otherwise, when test reaches the maximum allowable project length, it will end [19].

Fixed-length method does not use complex algorithms and is easy to implement. But some testees will answer needless questions, and some will answer insufficient questions leading to the inaccurate measurement of the ability. Variable-length method is more scientific and accurate, but each step needs to be calculated and the amount of information needs to be saved, which is hard to implement. The third method combines them, which would be more accurate than the fixed-length method and less complex than the variable-method. Although it may not reach the given accuracy, it is a more accurate and reasonable method.

To sum up, there are abundant methods to select subject and terminate adaptive tests, and each method has advantages and disadvantages. We will choose the subject selection policies and the existing test conditions that are suitable to Moodle platform.

**Learning Personality Characteristics Model**

The biggest drawback of traditional teaching mode is ignoring the individual differences of learners. In the era of big data, if we want to meet the learning needs of learners individually, it needs to identify their characteristics accurately. Analyzing learners’ personality traits (including knowledge level, learning style, error/misunderstanding and emotional characteristics) is a prerequisite for adaptive learning. Therefore, we need appropriate techniques to construct a personality characteristics model, which helps to provide accurate and personalized learning services. Methods of constructing personality characteristics model include coverage model, lead plate model, perturbation model, mechanical learning technology, model cognitive-based on theory, constraint-based model, fuzzy logic technology, Bayesian network and semantic network ontology model. Among them, the coverage model and the lead plate model are the most common modeling techniques [20].

The coverage model, proposed by Stansfield, is a method commonly used to describe the knowledge level that users have towards each concept. When using the coverage model method to
construct the learner's knowledge level model, the domain knowledge model represents the knowledge of expert level of a discipline, and the learner model is regarded as a subset of the domain knowledge model. The lead plate model is introduced by Rich into the GRUNDY system to build the user feature model, whose core idea is to cluster all potential users in the ALS according to the specific characteristics. Each group is a lead plate of users. Perturbation models and constraint-based models are modeled on learners’ errors / misunderstandings characteristics. Perturbation model is also known as the deviation model. It is said that learner's knowledge includes not only some knowledge of the domain experts, but also the possible mistaken knowledge of the learners [21].

The learning personality characteristics model is commonly used to provide the appropriate materials for different learners in a CAT system.

**Optimization Policy**

At present, what the Moodle platform mainly applies are traditional test ideas, that is, each set of questions are pre-grouped, and the difficulty of the questions are in a certain proportion. Traditional test ideas are simple, but the number of test papers is limited, which cannot determine the ability of the testees accurately. Therefore, Moodle need to be optimized to become more comprehensive and more authoritative.

This paper proposes some aspects to optimize the Moodle platform, including adaptive testing, test process analysis, and reasonable learning plan on the test result.

**Change Traditional Tests to Adaptive Test**

CAT uses the IRT to build a question bank, and automatically selects questions according to the knowledge level of testees, and finally estimates the learning ability. CAT is different from the general computerized test, because CAT can not only present the items, score the test and feedback automatically, but also can automatically select the most appropriate questions for the testee according to his answers. Therefore, it can make the most appropriate estimation of the testee’s learning ability. CAT mainly applies IRT and maximum likelihood method to make the process random and dynamic according to different composition policy. It can greatly reduce the repeatability of test papers and improve the reliability and validity of the test results.

Our study will develop a question bank for logistics management which has 500 or more questions on Moodle. Allowing for the actual situation of the test, this paper combines the existing composition methods of adaptive test and the subject selection policy. The following is the initially improved policy to fit in this test.

If students' historical test scores are available, the adaptive test module will select items according to the previous record of the ability. In other cases, as a testee is not able to determine objectively his own level of learning ability, it is unscientific to choose the starting point. When a student registers in Moodle, the adaptive test module will set the middle difficult item as the starting point for adaptive test.

In order to make the test more accurate, this paper divides the test into two stages. The first stage is an experimental test, mainly with a set of items to assess the initial ability of the testee. The method is easy to implement and sets the maximum number of questions so that the first stage will not be too long. If the testee is all wrong or all right, then randomly give out an extremely easy or extremely difficult question, so that the first stage can be terminated. Otherwise, the ability level can be estimated based on the maximum likelihood method, and the first stage is terminated.

The second stage is the accurate test. Based on the initial value of the first stage, the maximum likelihood method is used to estimate the value of learning ability in the course of the test. The principle of choosing in this stage is to select the items that provide the greatest amount of information. The amount of information will be increased throughout the test until the specified termination condition is reached.
This paper chooses a comprehensive approach to terminate the test. We set the maximum allowable test length. If the accuracy degree is satisfied within the test length, the test is ended, or the test is ended when the question is the last item. This method is more accurate than the fixed-length method, but it is a more reasonable and easier method to reduce the complexity of the variable-length test.

Focus on Testing Process Analysis

The general problem of computerized testing is that the feedback given after the test is only the score, and the testee does not know where the problem is in his knowledge system. However, in the actual learning process, the feedback on the test is extremely important, since it is an effective guarantee for the testee to find and overcome weaknesses in order to improve grades. Therefore, we propose more information after the adaptive test. The adaptive test module will show testees the questions with wrong answers, the comments on knowledge modules which did not been mastered and the summary of them, so that testees can clearly understand the next step of learning.

In order to achieve this function, the background algorithm has to record the questions with wrong answers from testees in the test process, and store them in the feedback. In addition, it is necessary to design a small program to provide the wrong percentage of the problem, to integrate all the wrong questions in order to analyze the type of knowledge and questions that the testee has not mastered. In this way, the testee is provided with a comprehensive and clear feedback of the test result so that he can better carry out the next stage of learning.

Combine the Learning Personality Characteristic Model with the Feedback

After analyzing the knowledge of the testee, we will provide learners with thorough learning suggestions for the next stage of learning. According to the test results, we will give learners a clear plan of next learning stage, such as learning material and subsequent test. Therefore, we need to complete the process in accordance with the tested knowledge of the learners and their learning characteristics. In this paper, we collect related data from the knowledge level, learning style and error/misunderstanding.

The knowledge level describes the learner's level of knowledge according to all the previous learned information, which is the most direct reflection of the learning effectiveness. This feature can be obtained entirely by the adaptive test part. Through the test results, we can get the wrong part of the test, so that related data of knowledge level can be obtained. The feature of knowledge level can be implemented through establishing coverage model.

Learning style is composed of learners' unique cognitive, emotional and physiological behavior. It is a relatively stable way of learning how learners perceive information, interact with learning environment and reflect it. For example, visual learners prefer image learning materials, but text learners prefer paper lectures and other learning materials. This feature will be implemented using the lead plate model. Moodle uses the Felder-Silverman learning style scale (meditation type and active type, intuition and perception type, speech type and visual type, comprehensive type and sequence type) to obtain learner's learning style. First, if the learner logs into the Moodle platform for the first time, he or she will need to complete the learning style questionnaire based on the Felder-Silverman learning style scale. In the subsequent learning process, by digging into the learner's behavior data, we can summarize the learner's preferred learning materials to obtain the learning style data.

Error and misunderstanding are common in the course of learning. Learners may fail because they do not fully grasp the knowledge points, or because of their tense, anxiety, carelessness, and so on. In order to identify these two types of errors, we also use the log mining method to obtain the learner's behavior data to determine the reason for which he or she is often wrong. This feature can be used through establishing the perturbation model. Moodle platform can record the learner's error type and the corresponding correction policy. After obtaining the learner's behavior data, the adaptive test module will match it with the data recorded in the background to find the appropriate correction method and learning materials.
Summary

The traditional tests on Moodle cannot accurately reflect test results and learning situation, and do not recommend the follow-up test and learning plan. Moreover, the current adaptive test does not pay much attention to the process evaluation and cannot scientifically reflect students’ ability. Therefore, this paper puts forward the optimization for the Moodle test module using CAT. The adaptive test module can reflect the true level of testees objectively and accurately. Combined with the actual situation, the adaptive test module can select and improve the appropriate composition policy from the existing adaptive methods. In addition, taking into account that the Moodle platform has learning records and personality characteristics of learners, this paper integrates the adaptive test results and personality characteristic model and makes an improvement to personalize learners’ learning resources.

This paper designs the Moodle optimization for adaptive test through literature analysis and current situation research. The next step is to implement the optimization and practice it. We will add some new modules, such as adaptive test module and personal learning material module, and integrate these two modules in order to provide more personalized learning environment.

Appendix

Dr. Li Jimei, corresponding author of this paper, is a Professor and the Dean of Department of Information Management, College of Information Sciences, Beijing Language and Culture University, China. Her research interests include information systems.

Xindi Lan and Haoyan Yu, undergraduate students majored in Information Management and Information System, are currently studying in Beijing Language and Culture University.

Acknowledgement

This paper is partially supported by BLCU college project (the Fundamental Research Funds for the Central Universities, NO:16YJ030001, 17YJ0304), by Science Foundation of Beijing Language and Culture University (the Fundamental Research Funds for the Central Universities, No: 17PT01), BLCU distinguished teacher program (No: OTP201607), and Beijing training program (No: 2017100320180).

References


