International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-7 Issue-4s, November 2018

A Survey on Item Selection Approaches for Computer Based Adaptive Testing

Nandakumar G. S, Viswanandhne S.

***Abstract:** Assessment is an essential part in determining the level of attainment of education. In spite of availability of several formal and informal methods, Computer Based Assessment (CBT) is predominantly used for very large scale assessments. Adaptive Testing has better estimation capabilities if the standard of its items (questions) match the ability of the candidate. Items that are too simple or too difficult give unpredictable reactions and can't give much information about the ability of the student. It is therefore essential to select items from the large pool so that the selected item gives maximum information about the ability of the student. This paper reviews the various methods for the item selection during the computerized adaptive testing.*

***Keywords:** Computer Adaptive Testing, Item selection approaches*.

I.INTRODUCTION

The expanded use of computer innovations has driven educational institutions and recruiters to search for approaches to utilize innovation in testing the students' ability. Earlier, paper and pencil method is used as a standardized assessment method, but in recent days several Computer Based Testing (CBT) methods are widely used. Computer Adaptive Testing (CAT) has several advantages over traditional testing. In traditional testing, often students spend most of their time with questions that are not matching their knowledge and ability. Using CAT, students complete the test with reduced number of questions and those questions are appropriate for their knowledge level. This test is conducted by having an account of the student's performance on each question and selects the next question to be presented with the estimated students' ability. CAT provides a major advantage by attaining the desired level of stability, with minimum number of questions in the estimation of the students' ability.

II.ITEM SELECTION APPROACHES

A. Bayesian Approach

Bayesian Approach is merely not selecting questions based on single point estimation of the students' ability. This approach uses prior information about the students' ability level. A Bayesian approach defines procedure for item selection such as:

Manuscript published on 30 November 2018.
*Correspondence Author(s)
Nandakumar G S, Department of Information Technology, Kumaraguru College of Technology, Coimbatore (Tamil Nadu), India.
Viswanandhne S, Department of Computer Science and Engineering, Sri Krishna College of Engineering &Technology, Coimbatore (Tamil Nadu), India.

i. the first item selection
ii. estimation of ability θ
iii. the next item selection based on estimated ability
iv. test termination

Bayesian adaptive testing procedure selects next question based on the above briefed steps, along with the use of the prior information of the questions.

For selecting items there are two different Bayesian approaches (van der Linden, 1998). One approach is based on "Maximum Posterior-weighted Information (MPI)" and the other based on "Maximum Expected Information (MEI)". Both of these makes use of the students' posterior information θ. (i.e., students' probability in providing the correct answers to the question considering the earlier question attended by the student). So these approaches take expected error as the estimate of θ when it is selected as the maximal information.

III.THE MPI APPROACH

Initially at the start of the assessment, item information function estimated at $\hat{\theta}$ differs with the item information function estimated at θ. [since true ability of the student cannot be estimated at the beginning of the assessment]. For the cases where there are such discrepancies between θ and $\hat{\theta}$, Maximum posterior-weighted information Approach (MPI) approach can used to select next item that is to be administrated. For item selection, MPI approach takes these differences between θ and $\hat{\theta}$into consideration and it follows three steps.

i. Partition the standardized θ into finite intervals R, such that r=1,2,…R (generally R may range between 30 and 50,θranges between -4.0 and +4.0). Let X_r denotes middle of r^{th} interval.
ii. Once the questions previous answers given by others are known and based on the probability of student's θ falling in r^{th} interval, X_r the weight information is calculated.
iii. The total of the weighted information for θ throughout R is evaluated from the following equation.
'$Si = \sum_{r=1}^{R} P(\theta \text{ in } r|u)\ I_i(X_r)$'--------------------- Equation 1

The term **u** refers to the students' response to the earlier questions faced.

$I_i(X_r)$ represents the item information function value evaluated at X_r.

P(θ in r|u) in Equation 1 represents the probability of the student's ability value θ that falls in interval 'r', based on the question students faced given the student's pattern of responses to the previously administered items, which is known as the posterior probability in the Bayesian framework.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

This posterior probability is calculated using

'$P(\theta \text{ in r} \mid u) = \frac{P(u \mid X_r) f(X_r)}{\sum_{r=1}^{R} P(u|X_r) f(X_r)}$'--------------------- Equation 2

where

'f(X_r)' – approximate standard normal density at interval 'r'

'P(u|X_r)'- response pattern of the student having ability level X_r

In the intervals θ, when the 'Items Information Function' is more than the potential of students' capability, then Si will be considerably larger. On the other hand if 'Item Information Function' is less than potential of students' capability then Si will be much smaller. Finally this approach always selects the items that have not been administrated before, which has highest S_i value(Randall D. Penfield,2006).

IV.THE MEI APPROACH

The 'Maximum Expected Information (MEI)' approaches (van der Linden, 1998) uses examinee's posterior distribution of ability θ and thus it considers the expected error in $\hat{\theta}$ during the estimation of the true ability θ.

This approach chooses the question that is likely to bring out maximum information on the test based on all probable responses of the candidate and the student's posterior distribution (Randall D. Penfield, 2006).

In this approach, E_i is computed for those questions that were not faced by the candidate earlier.

'$Ei = \sum_{k=0}^{m} P(u_i = k)\, I_T(\hat{\theta}_k)$' --------------------- Equation 3

where $I_T(\hat{\theta}_k)$ is the test information at $\hat{\theta}_k$.$P(u_i$=k) is the probability of candidate's answer in group 'k' (0<=k<=m) and is calculated using the following equation.

'$P(u_i = k) = \sum_{r=1}^{R} P(\theta \text{ in r} \mid u)\, P_{ik}(X_r)$' --------------------- Equation 4

where $P(\theta \text{ in r}|u)$ is the posterior probability of θ and is calculated using equation 2 and $P_{ik}(X_r)$ is the response for probability k in question i.

Finally, the question with the highest E_i will be chosen for presenting to the candidate[11].

A. Stratified (As) Strategy

Chang & Ying (1999) [6] proposed the following strategy for choosing similar questions in CAT. Here the quality of the assessment in maintained with controlled revealing of questions. Generally questions getting exposed over a period of time and security are of major concerns in managing/operating the test. Also there is a possibility of some questions being presented to students more often than others.

Generally the questions are selected using the maximum Fisher information. With the Fisher Information Function, items with high value of a(item discrimination parameter) gives the maximum information about student's ability. In this A-stratified (AS) strategy, items are stratified into K level in view of the item discrimination parameter (a value). Based on this selecting the question is categorized into 'K' levels. In level 1, questions will have minimum discrimination parameter 'a'. At higher levels, the questions will be with high 'a' values.

Estimation of the student's ability will not be accurate at the initial stage of the assessment. Hence it is better to have questions with lower values of 'a'. During the final stage of the test, it will be more efficient while using higher 'a' values [6].

V.SHADOW TEST APPROACH (STA):

In this approach Van der Linden and Reese (1998)[7], All other item selection approaches selects its item directly from pool of questions. This shadow test approach selects questions from the shadow test which is a test carried out for the entire duration before actual test. The choice of questions is based on the dynamic combination of question from the shadow tests. A shadow test is conducted considering the following aspects: (1) All parameters of actual test (2) Questions faced already by the student being tested and (3) Current capability of the student which is more useful.

Algorithm of STA to select the item:

i. Initialize all the variables of student capability
ii. Construct the initial 'shadow test' satisfying necessary criterion
iii. Given students' ability, select the item that can provide maximum information
iv. Modify the variables in the assessment method
v. Reassemble the shadow test by excluding items that have been already administrated
vi. Repeat from ii till estimating the actual ability of the student.

VI.MAXIMUM PRIORITY INDEX (MPI)

MPI approach uses 'Priority Index' (PI) which is based on the relevance factor for each question in the question bank. This is calculated at every step of choosing the question. Questions with higher PI values are the most preferred questions for presenting to the student. A two-stage question choosing pattern is adopted (Cheng et al., 2007) for to executing this along with flexible content balancing where the limitations are indicated as both lower and upper limits.

$$PI_j = I_j \prod_{K=1}^{K} (w_k f_k)^{c_{jk}}$$

Where

I_j – 'Fisher information of question 'j'

w_k - weight of limitation k.

c_{jk} - Limitation relevancy matrix with (=1if it has limitation, =0 if there is no limitation.

Generally this matrix given by the domain experts [8].

VII.WEIGHTED PENALTY MODEL (WPM)

Chingwei David Shin et all [9] proposed WPM for item selection in CAT. This model strives to optimize content parameters over question classifications while at the same time maintaining question information during question selection.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Also the shortage of questions under some criterion is taken into account [9]. This model works by framing a list of items to the candidates based on a particular question exposure control method or the first one is chosen to be administrated.

VIII. WEIGHTED DEVIATION MODELLING MODEL

A Widely used most heuristics approach for item selection is Weighted deviation model (Buyske, 2005). Algorithm of WDM to select the item:

i. For every question that is not administrated already, the parameters that do not match with the actual is computed.
ii. Weighted sum of all items that are not administrated before are calculated
iii. Questions with the lower value is chosen for presenting to the candidate.

Let ‘J’ be number of questions in the question bank, ‘K’ be the limitations, C be the constraint relevancy matrix, ‘w_k’ is the weight associated with limitation ‘k’, ‘w_I’ represents the weight of test information, ‘l_k’ and ‘u_k’ are the lower bound and upper bound of constraint k. the notations d_{lk} ,d_{uk} and d_I are the ‘deviations’ that may either be the shortage from minimum level or excess from the maximum level. Similarly ‘e_{lk}’ is the excess to ‘l_k’ and ‘e_{uk}’ the deficit from the u_k.
The goal is to minimize the sum of weighted deviations:

$$\sum_{k=1}^{K} \left(w_k \left(d_{l_k} + d_{u_k} \right) + w_\theta d_I \right)$$

such that $d_{lk}, d_{uk}, e_{lk}, e_{uk} \geq 0$, $k = 1,2\ldots.K$

Mansoor Al-A'ali [4] exhibited a description of adaptive testing in view of Item Response Theory (IRT) and experiment with IRT keeping in mind the end goal to assess its applicability and advantages. It additionally presented upgrades for IRT. Questions’ characteristics are estimated based on the least square technique. It is possible to minimize questions to achieve the maximum level of the student where his ability level can be estimated. They have incorporated variables like capability of student, toughness of question, portions covered by faculty in Item Response Theory to more practical and appropriate.

The ultimate aim is to select one question from the existing question bank that is appropriate to test the knowledge of the student. This Adaptive selection of questions approach works as the collection of three methods applied one after the other. Here, each and every method filters the collection of questions at its stage and at the end the most beneficial question remains. Figure z depicts the entire approach for question selection.

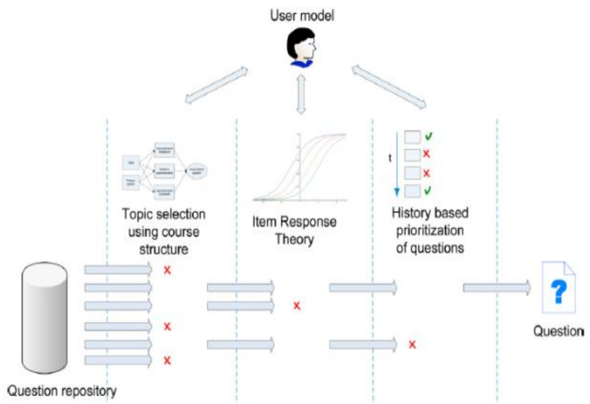

Figure z. Adaptive Question Selection

Dr. V. Natarajan [5] The IRT is becoming more predominant throughout the world in terms effective assessments, setting up test question bank, critical analysis and scoring compared to that of CTT(Classical Test Theory). The main reason for lies in the inter-relationship it possesses between the questions and the parameters of the assessment. This is true for any of the formal tests with any duration in CTT in a hypothetical case.

In CTT, the candidate’s actual assessment marks is total obtained from the test. IRT concentrates more on the individual answers turned to be right or wrong, than the final scores. This is due to the fact that IRT is based on individual questions outcome rather than some gross average assessment score. In CTT the complete score is considered statistically and in IRT it individual values that matter. “An attempt is made to relate the individual item characteristic to the individual’s ability which of course has a different and a definite definition.” [5]

IX.CONCLUSION

Thus the paper presents different approaches for selecting a question in computer based adaptive testing. Each of the approaches has its own method of choosing the question from a large question bank. Every approach has its own advantages and disadvantages. Hence a good selection approach based on the size of the question bank, the candidates’ proficiency level, the duration of the test, etc., should be used in testing such that it chooses the item that retrieves the maximum student’s ability.

REFERENCES

1. Richard M. Luecht and Stephen G. Sireci,” A Review of Models for Computer-Based Testing”, in College Board, ReSearch Report 2011-12
2. Mariana Lilley and Trevor Barker, “An Evaluation of a Computer Adaptive Test in a UK University Context”, University of Hertfordshire, United Kingdom
3. Daniel R. Eignor, Martha L. Stocking, Walter D.Way, Manfred Steffen , “Case Studies In Computer Adaptive test Design through Simulation”, Educational Testing service, NewJersey, November 1993
4. Mansoor Al-A'ali, “Implementation of an Improved Adaptive Testing Theory”, International Forum of Educational Technology & Society (IFETS), ISSN 1436-4522,2007
5. Dr. V. Natarajan, “Basic Principles of IRT And Application to Practical Testing & Assessment”, in IACAT conferences at Amsterdam, Asilomar (US),2008

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

6. Hua-Hua Chang and Zhiliang Ying, “a-Stratified Multistage Computerized Adaptive Testing With b Blocking”, in Sage Publication, Educational Testing Service, Vol. 25 No. 4, December 2001, 333–341
7. Wim J. van der Linden, “Constrained Adaptive Testing with Shadow Tests”, Statisticsfor Social and Behavioral Sciences, DOI 10.1007/978-0-387-85461-8 2
8. Wei He,” Comparison of Four Item-Selection Methods for Severely Constrained CATs”, Northwest Evaluation Association (NWEA),2013
9. Chingwei David Shin, YuehmeiChien, Walter Denny Way, Len Swanson,” Weighted Penalty Model for Content Balancing in CATS”, in Pearson, April 2009.
10. Van der Linden, W. J. (1998). Bayesian item-selection criteria for adaptive testing. Psychometrika, 62, 201–216
11. Randall D. Penfield , Applying Bayesian Item Selection Approaches to Adaptive Tests Using Polytomous Items, 2006

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)