Work in Progress: Practical Computerized Adaptive Assessment based on Bayesian decision theory

Catherine C. Marinagi and Vassilis G. Kaburlasos
Department of Industrial Informatics, Technological Educational Institution of Kavala, GR-65404 Kavala, Greece \{vkabls,kmairi\}@teikav.edu.gr

Abstract - This work reports on the development of a novel software tool, namely Module for Adaptive Assessment of Students or MAAS for short, for adaptive multi-user Web-based assessment. The framework of Bayesian decision theory has been used for sequential mastery testing to classify students as masters or non-masters based on their responses to adaptively selected test items. MAAS applies a naive Bayesian decision model to adaptive testing, assuming many levels of student performance and taking into account not only right/wrong, but also blank answers. MAAS is embedded in a software platform, namely Platform for Adaptive and Reliable Evaluation of Students or PARES. MAAS is currently in a pilot use both for formative assessment and self-assessment. A preliminary experiment of using MAAS under real world conditions is described.

Index Terms – Bayesian decision theory, Computerized adaptive assessment

I. ESTIMATING PRIOR AND CONDITIONAL PROBABILITIES

Bayesian decision theory has been proposed as a model for adaptive testing. Comparing with Item Response Theory (IRT) model [2] which is widely used in Computer Adaptive Testing (CAT), Bayesian decision theory provides simplicity, lack of assumptions, robustness and computational ease. Rudner [3] has compared adaptive Bayesian decision-theory testing procedures with IRT in terms of classification accuracy, using simulated item response data. Minimum expected cost, a Bayesian item selection procedure was better than the best-case possibility for Item Response Theory.

For each chapter, we assume $J$ levels of student performance, $L_1, \ldots, L_J$. Given the scores of students which have been previously tested on a particular chapter, the prior probabilities $p(L_1), \ldots, p(L_J)$ can be estimated and fixed.

For each item $\tau_n$, the conditional probabilities $p(a_{ng}|L_1), \ldots, p(a_{ng}|L_J)$, can be estimated and fixed, where $a_{ng}$ are student responses in item $\tau_n$. The variable $a_{ng}$ takes three values: ‘R’ for right, ‘W’ for wrong, ‘B’ for blank. These are the three possible student responses to an item $\tau_n$.

For each item $\tau_n$, the probability of the corresponding response $a_{ng}$, can be estimated, where $a_{ng}=\text{‘R’}$, $a_{ng}=\text{‘W’}$ and $a_{ng}=\text{‘B’}$:

$$p(a_{ng}) = \sum_{k=1}^{J} p(a_{ng} | L_k) p(L_k), \quad g=1,\ldots,3$$  \hspace{1cm} (1)

II. ESTIMATING STUDENT STATE VECTOR

Let $S_c$ be the number of students tested on chapter $\kappa_c$, $c \in \{1,\ldots,K\}$ of a particular subject. Each student will correspond to a vector $\text{indS}_c(h)$, $h=1,\ldots,S_c$, of student performance indices. The entry values of a vector $\text{indS}_c(h)$ specify the competence/profile of a student, therefore vector $\text{indS}_c(h)$ will be called student state vector.

We consider that a student $h$, $h=1,\ldots,S_c$, sits for an adaptive test on chapter $\kappa_c$, $c \in \{1,\ldots,K\}$. Initially, a set of $I$ items is randomly generated and the student provides corresponding responses $a_i$, $i=1,\ldots,I$. We apply Bayes
Theorem to estimate the posterior probability that a student \( h \) is in level \( L_k \), \( k=1,\ldots,J \), given the responses \( a_1,\ldots,a_I \) to \( I \) items.

\[
p(L_k | a_1,\ldots,a_I) = \frac{1}{\sum_{j=1}^{I} p(L_j | a_1,\ldots,a_I) \prod_{i=1}^{I} p(a_i | L_j)} \]

For a student \( h \), the state vector of \( \text{indSc}(h) \) in chapter \( \kappa \) of a particular subject is

\[
\text{indSc}(h) = \{ p(L_1 | a_1,\ldots,a_I), \ldots, p(L_J | a_1,\ldots,a_I) \}.
\]

### III. Administering Next Item

Having computed the initial student state vector \( \text{indSc}(h) \), adaptive assessment has to be employed. Now only one item appears on screen at a time, students respond and their state vector is re-evaluated.

Each time a student responds to an item, the current probability of response to any remaining item \( \tau_\text{ng} \) in the Item Bank is estimated using another version of (1), where the current student state vector is incorporated. That is, the posterior probabilities \( p(L_k | a_1,\ldots,a_I) \) are treated as updated prior probabilities \( p(L_k) \). The current probability of a response \( a_\text{ng} \) to an item \( \tau_\text{ng} \), given the responses \( a_1,\ldots,a_I \) to \( I \) items, is

\[
p(a_\text{ng} | L_k) = \sum_{k=1}^{J} p(a_\text{ng} | L_k) p(L_k | a_1,\ldots,a_I), \quad g=1,\ldots,3
\]

where \( a_{ng}=\text{R}, \ a_{ng}=\text{W}, \ a_{ng}=\text{B} \).

In order to select the next item, decision cost is also considered. The \textit{minimum expected cost approach} is applied [3], for \( J \) levels of student performance, \( L_1,\ldots,L_J \). Let \( c_{st} = |L_t-| \) be the cost of deciding that the student’s level is \( L_t \) (observing his/her score) when student’s true level is \( L_t \). The cost for giving a response \( a_\text{ng} \) to each item \( \tau_\text{ng} \) of the item bank is

\[
B_{ng} = \sum_{s=1}^{J} \sum_{t=1}^{J} c_{st} p(L_t | a_\text{ng}) p(L_t | a_\text{ng}), \quad g=1,\ldots,3
\]

where \( a_{ng}=\text{R} \), \( a_{ng}=\text{W} \), \( a_{ng}=\text{B} \).

Using (3), (4) the average cost of each item \( \tau_\text{ng} \) of the item bank, with corresponding responses \( a_\text{ng} \), can be calculated by

\[
\text{Cost}(\tau_\text{ng}) = \sum_{g=1}^{3} p(a_\text{ng}) B_{ng}
\]

The item with the lowest expected cost is selected next.

Another option is the calculation of cost per unit and the random selection of the next item from the appropriate unit.

### Practical Use of MAAS

PARES is a client/server, multi-user Web-based assessment platform developed in Java. Hypertext markup language (HTML) is used for the development of pages including text and figures. PARES includes three modules: the Administrator Module, used to update records of courses, instructors and students, the Instructor module used to update the Item Bank and compose tests, and the Student module used to take tests.

During the academic years 2003-05, previous versions of PARES have been successfully used at a grand scale for formative assessments in a real classroom environment. These versions enable teachers to adapt tests manually. Current version of PARES incorporates MAAS for adaptive student assessment. During the winter 2005-06 semester MAAS was used both for formative assessment and self-assessment in the course \textit{Introduction to Software Engineering}.

The experiment, took place twice during the semester, including self-assessment a week before formative assessment. Data gathered from self-assessment tests were used for calculating prior probabilities of students’ levels, and conditional probabilities for each item. Students had access to the educational resources associated with each answer. The test was executed as follows: Each student initially responded to a basic set of randomly generated items, in a non-adaptive way. MAAS provided an estimation of student’s performance level. Then the test was turned to be adaptive. Items adjusted to student’s level, appeared on screen one-by-one. Student’s level was re-evaluated after any single response. The procedure was repeated until termination criteria were met.

The effectiveness of MAAS will be evaluated by the employment of MAAS in four courses during the spring 2006 semester. Finally, robust statistical evidence will be presented regarding the viability of the proposed technology under real world conditions.

### Conclusions and Future Work

MAAS can help diagnosing students’ preparedness for a subject, tailoring tests to students’ personal level of performance. When used as a self-assessment tool, MAAS can help students to gradually improve their level through the associated feedback of answers and be trained in the use of the assessment tool, before formative assessment. MAAS applies the Bayesian decision theory model to adaptive testing. Not only right/wrong, but also blank answers are considered in calculations and many levels of student performance are assumed instead of two (masters/nonmasters).

MAAS is expected to contribute towards an improvement of the Greek higher education system’s throughput, the latter is the percentage of students passing a course, under a number of constraints, resulting in a better return of investment in education.

### Acknowledgment

This work has been supported in part by the third European framework programme: Operational Programme in Education and Initial Vocational Training II.

### References