Cognitive Scaffolding for a Web-Based Adaptive Learning Environment

George Fernandez

School of Computer Science and Information Technology
RMIT University, Melbourne, Australia
gfernandez@rmit.edu.au

Abstract. On-line Web-based learning environments with automated feedback, such as WebLearn [5], present subject questions to the student and evaluate their answers to provide formative and summative assessment. With these tools, formative learning activities such as quizzes and tests are mostly pre-planned, since testing instruments are generated by selecting questions in a pre-specified manner out of question banks created for the purpose. Although this approach has been used with a significant degree of success, the real challenge to support students' learning is to mimic what a human instructor would do when teaching: provide guided learning.

The main difficulty associated with creating such an 'electronic tutor' is to implement the required intelligent dynamic behaviour during learning. That is, at any stage of a student's learning session the system should take into account his/her demonstrated cognitive level to generate the next appropriate formative testing instrument. For students to be able to make the higher-level cognitive contributions as they progress through a session, the system must keep a history of students' answers and must react accordingly. We call here that behaviour *adaptive learning* by *adaptive formative assessment*.

We propose on this paper a strategy to implement an adaptive automated learning system, based on establishing an incremental cognitive path from the lowest to the highest level questions related to a concept. In the research literature this has been often called 'cognitive scaffolding'. For our on-line automated environment, the first hurdle has been how to define the scaffolding and how to implement it from question banks that have not been created for this process. Our approach is embodied in WebTutor, a 'black box' component being developed at RMIT University to work in combination with the generation, presentation and feedback capabilities of the WebLearn system.

1 Introduction

Cognitive scaffolding represents what an instructor does when working with a student "to solve a problem, carry out a task, or achieve a goal which would be beyond his unassisted efforts" [10, pp 90]. It is generally a dynamic process, with the student interacting with the instructor, who attempts to understand from the student responses what the cognitive gaps are, and accordingly provide guided support to progress along the intended learning path. Instructors do this by presenting appropriate examples to reflect on and problems to solve, and demonstrating skills that student can imitate.

W. Zhou et al. (Eds.): ICWL 2003, LNCS 2783, pp. 12–20, 2003.

This typically follows a "from shallower to deeper" approach, as a sequence of steps intended to guide the student to the desired depth of understanding. The best instructors are the ones who, during a session, follow a student's demonstrated progress and adapt the learning activities to promote as much as possible reflection by the student. To do so, they present content to stimulate inquiry in the students, present alternative points of view on a concept, raise points for consideration, and decide on subsequent steps in the instruction. During that process, often students are asked to perform learning activities for which they are unprepared. If that happens, a skilful instructor follows the student's answers and evaluates the shortcomings, and backtracks looking for a place to start again on a firmer footing. The teacher's model of instruction includes a continuous evaluation of what are the difficulties with the problem at hand, and what would be the necessary steps for helping students advance towards their goals.

In on-line teaching and learning there is much less teacher-student face-to-face contact than traditionally, thus changing the emphasis from a teacher-centred to a student-centred approach [3, 8]. The main purpose of teaching is now to properly manage the learning process rather than to transmit information in a clear and organised manner, (a la Level 1 and Level 2 in [9]). Learning environments with automated feedback have been used rather successfully in online learning, albeit mostly for rote learning by focusing on drilling exercises. As discussed above, however, creating an 'electronic tutor' would require engaging students in appropriate self-directed learning activities that foster question, reflection and analysis along an incremental cognitive path. With the support of an appropriate environment, well-structured learning tasks should induce consideration, inquiry and discovery in the students, progressing students through their learning process to the higher levels required for deeper learning (See for example [6, 7]).

However, on-line learning environments today are not capable of adequately supporting learning processes in such a way. Not only they are typically restricted to questions with a given simple format, such as Multiple Choice, Multiple Answer, or Short (Key) Text, but they lack the human instructor's ability to retrace steps and dynamically change the angle of instruction based on what the student seems to have learnt/not learnt up to that stage. We argue that an automated learning system providing formative assessment might, to a certain extent, be able to do that if the system keeps the history of the previous student answers during a session and decides on what testing instrument to generate next based on what the student has already learnt and is still required to achieve.

This issue is naturally related to Computer Adaptive Testing (CAT), where there has been considerable research attention focused on Item Response Theory (IRT) [1]. IRT is attractive because it is based on solid statistical foundations, and because, with the right item bank and variance of examinees, it may be very effective for computer based automated testing. Our interest here is not, however, adaptive testing but adaptive learning by adaptive formative assessment. By this we mean that we intend to endow the learning environment with the capability of guiding students through a learning session where questions are presented as a response to their previous answers in the session, following a strategy resembling a human instructor.

The rest of this paper is organised as follows. Section 2 presents previous research on WebLearn, a Web-based learning environment with which WebTutor is tightly associated. Section 3 presents the basics of Item Response Theory (IRT), a line of research closely related to this paper. Section 4 discusses the conceptual differences between IRT and our line of research. Section 5 presents the proposed strategy to establish the cognitive scaffolding for a given item bank related to a concept or concepts. Section 6 concludes, and presents suggestions for further research.

2 WebLearn, a Web-Based Online Learning Environment

WebLearn is a WWW-based tool that supports self-learning by presenting questions of different types and providing student with automated feedback. The system is easy to use by non-computer experts, it is highly configurable to reflect diverse subject objectives and personal teaching preferences and it can accommodate subjects in many different disciplines. The system supports the teaching of 'WebLearn subjects', divided into modules, each divided into set of learning objectives (e.g. Bloom's Taxonomy of Educational Objectives, [2, 4]). Each module requires questions addressing the learning objectives, compiled into quiz and tests question banks. For formative assessment, WebLearn automatically generates random quizzes from the stated learning objectives, according to the instructor's directions, checks the answers given by the students, and provides immediate feedback. Quiz questions can be Multiple Choice and Multiple Answer (more than one correct answer), Short Text, and a variety of numeric and other types with and without random generation of parameters.
Over the last two years, WebLearn has been working in combination with Maple, a Mathematics symbolic manipulation package. Systems such as Maple provide an environment with which students can interact in mathematical terms, since they include specialised 'engines' that interpret abstract mathematical language. On the other hand, environments such as WebLearn have been designed to present questions to students and analyse their answers against predefined correct answers supplied by the instructors. Our approach combines the generation, presentation and feedback capabilities of WebLearn with the analysis capabilities of the Maple engine. When required, WebLearn automatically generates a formative or summative — a quiz or a test — testing instrument. Students' answers are captured by WebLearn and fed through Maple. The response from Maple is then caught back by WebLearn to be analysed, massaged into an appropriate form, and fed back to the students. This makes possible the correct evaluation of questions with no unique answer, for example, providing proper assessment of any right answer provided by the students. WebLearn treats Maple essentially as a 'black box', making possible to quarantine software changes to either system. The interoperation between Maple and WebLearn offers a wide variety of unique features, including handling of symbolic mathematics in areas such as general calculus, differential equations, Fourier and Laplace transforms, algebra, finite mathematics and geometry.
The development of WebTutor follows the same 'black box' approach.When WebLearn requests the generation of a new testing instrument to present to the students, WebTutor generates the new quiz by inspecting the student's history and deciding on the best way forward. Currently the system uses a very simplistic approach to make this decision, so this research intends to provide a sound strategy to move the student along an incremental cognitive path. Thus, we are developing a formal framework on which to base these decisions, effectively implementing the above mentioned scaffolding. The two main problems we currently face are the

development of appropriately graded question banks, and the provision of a set of criteria and structures to progress students up the cognitive ladder. This paper discusses our progress on the first one of these issues.

3 Item Response Theory

Item response Theory (IRT) was first introduced to provide a formal approach to adaptive testing. The theory establishes how to estimate the unknown 'ability' θ of a student being tested with a test consisting of a number of items (questions). Each of these items measures an aspect of the ability being estimated. Answers to an item are assessed as correct or incorrect; the student receives a score of one for a correct answer, zero otherwise. The main goal of IRT is to determine the true ability of an examinee by studying the probability of a correct response to each individual item in a test. Therefore, the primary interest of IRT is whether an examinee answered each individual item correctly or not, rather than a total test score. The theory considers each examinee to have a numerical ability value θ somewhere on the ability scale. The value of θ is measured on a scale having a midpoint of zero, a unit of measurement of one, and a range from negative to positive infinity.

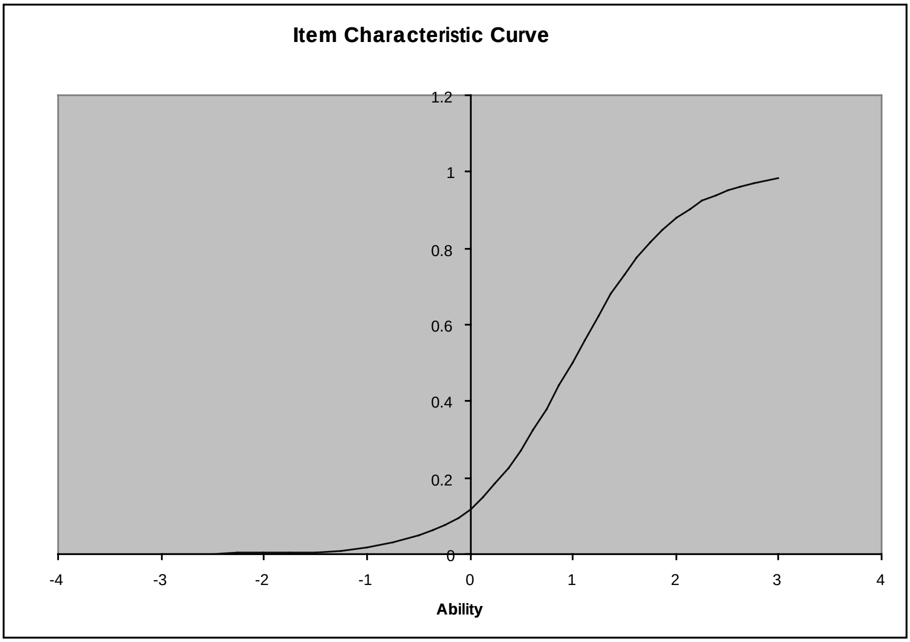

Fig. 1.

Although IRT has been used for free-response items, more often automated tests consist of multiple-choice items. One of the main applications of IRT has been to implement adaptive testing within an automated system by providing a carefully chosen sequence of questions. IRT determines at each step of the testing process which is the best next item to be presented to a given student, provides a stopping condition for the test, and a statistical estimation of the value of θ at the end of the exercise. The fundamental construct of the theory is the Item Characteristic Curve, which for each item represents the probability $P(\theta)$ that an examinee with ability θ will give a correct answer to the item. In the case of a typical test item, this probability will be smaller for examinees of low ability and larger for examinees of high ability (See Figure 1).

The shape of the curve is typically a smooth S, with differences depending on the value of some parameters. The first of these is the difficulty of the item b, determined – somewhat arbitrarily – by the point on the θ axis where the probability $P(\theta)=0.5$. The second parameter is the item discrimination a, which describes to what extent an item discriminates between examinees having abilities below and above, and close to, b. The discrimination parameter is often interpreted as the slope of the curve at abscissa b, although the value is actually $a/4$. There are actually several models in use – with one, two and three parameters – so this model is identified as the two-parameter model.

The 1-parameter model fixes the value of $a=1$, so there is only necessary to determine the difficulty parameter to establish the characteristic curve for the 2-parameter model. The three-parameter model includes the guessing parameter c. Although this last model lacks the mathematical elegance of the one and two parameter model – mainly because it doesn't follow a logistic model – this third parameter c is very important for CAT. In automated testing, it is reasonable to assume that if students don't know the answer to a Multiple Choice question they will attempt to guess it. This is certainly strongly expected in the case of an examination, probably slightly less so in a learning situation. Still, it is clear that the probability of getting a question right is affected by the likelihood of getting it right by pure chance. For example, a question with a true/false answer will have a 'floor' probability of being answered right of .5; even if the examinee knows nothing about the matter he/she would be expected to get it right 50% of the time by chance alone. Naturally, the probability of getting it right will still increase with the value of the ability θ, the higher the value the closer to 1 the probability of getting the question right.

The introduction of the third parameter does not change the basic shape of an item characteristic curve, but it certainly changes some of the values, as follows:

- the new value of the lowest probability becomes c rather than 0;
- the value of b is now the value on the θ axis where $P(\theta)=(1+c)/2$ (the middle point between c and 1);
- the actual value of the slope at b is $a(1-c)/4$.

4 IRT and Adaptive Formative Assessment

Our first step to implement adaptive learning is the definition of the scaffolding. Although there are many similarities between IRT and the requirements for adaptive learning, there are also important differences:

- The value of θ under IRT is loosely defined as the 'ability' of an examinee at the moment of taking the test. This is assumed to embody the knowledge and cognitive capabilities of the examinee at the time of the test. The examinee is not supposed to learn during the examination process. However, adaptive learning as defined here perceives the value of θ to change as learning progresses. Actually, our intention is not to try to determine θ• as adaptive testing tries to do, but to move students along an incremental cognitive path so their value of θ increases on a particular topic.
- The discrimination parameter `a` is very important for adaptive testing, since it indicates the sensitivity of the estimation of θ. Given that the purpose of the examination is to determine the level of θ, a high value of `a` indicates that the item is capable of discriminating between two very close levels of ability within a certain range. For our purposes, though, an exact value of `a` is less relevant, since the intention is not to determine the value of θ but to increase it as a result of the learning process. In practice, however, questions are to be divided into categories, and the discrimination parameter may be used as a decision mechanism to trigger item selections from a higher cognitive category: a correct response to an item close to the category edge with a high value of `a` may indicate that is time to move the student to next category up.
- The difficulty parameter `b` is crucial to our research, specifically to create the cognitive scaffolding based on increasing values of `b`. If the learning system consists of questions with an established level of difficulty, it is possible to progress up the learning path until a certain stopping condition occurs.
- During an examination under IRT, the sequence of items presented to an examinee is determined by selecting, at each step in the procedure, the 'best next item'. Intuitively, this should be:
  - an item with difficulty close to the examinee's θ value, since selecting an item that is too easy or too hard will provide no new information about the value of θ;
  - an item with a high value of `a`, since it is desirable to have an item that is most useful in discriminating between examinees with abilities close to the unknown value θ.

For adaptive learning, however, these considerations are not that important. Students are supposed to learn during a session, and therefore there is no fixed value of θ to estimate in this case. The intention is to present a sequence of items that challenge, but don't discourage, students. We are only interested in a reasonable estimate of the value of θ at any stage of the learning session, to be able to make a decision about when to move the student up the incremental cognitive path.

5 The Cognitive Ladder

In systems such as WebLearn, the question banks have not been developed with cognitive scaffolding in mind, so regardless of how questions have been grouped they need to be re-classified so the system can progress the students up the cognitive ladder. To this end, the items in a question bank must be classified from lower to higher in a chosen cognitive taxonomy. A question may be classified higher than another question in a given taxonomy for different reasons, such as when the higher-level question is perceived as harder, more abstract, or requiring a deeper understanding than the lower-level question. We argue that it is unreasonable to put this classification burden on the instructor, for several reasons:

- Given that a bank used for this purpose may contain thousands of different items, the large number of questions may be too much for an instructor to categorise. If more than one instructor is used, problems of consistency would arise.
- The resulting classification by an instructor would be a very subjective one, and highly dependent of the opinion and experience of the instructor.
- Such a long and demanding task will inevitably result in an inconsistent categorisation, even by one instructor.

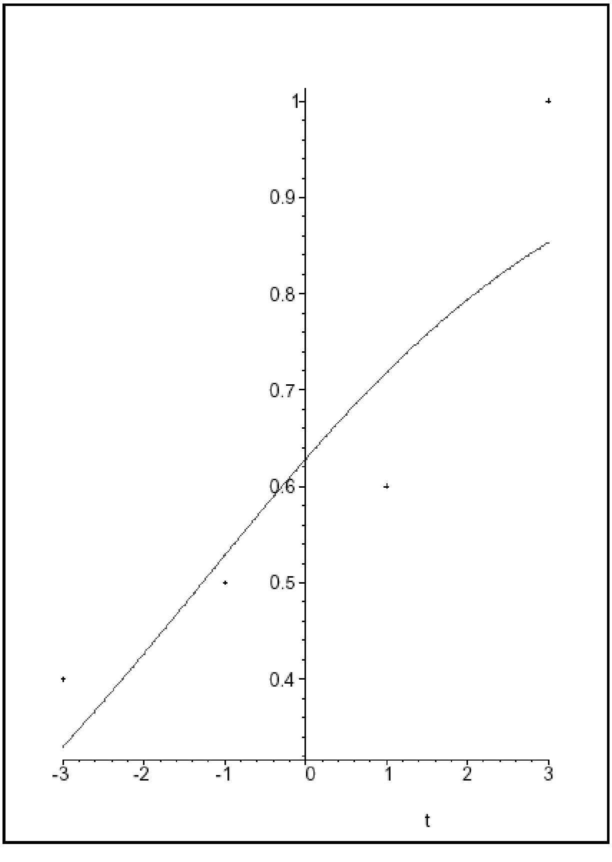

Fig. 2. An item classified, a = 0.411, b = -1.273

Regardless of the taxonomy, the only known invariant is that, given a question in a bank, students of higher ability are more likely to answer the question correctly than students of lower ability. It is possible to use this invariant to automatically classify a whole bank if the abilities of a group of students are known, or estimated by other means. The chosen strategy for this project was to use historic students' results as an estimate of their ability, and to use this information to provide a classification for the whole item bank. It was then possible to establish a correspondence between the estimated difficulty of the students' answers to the question banks for a first year programming subject. Two different, parallel approaches were considered for this project, as follows:

- The results of student tests were considered from the historical records. The unknown ability θ of the student cohort was estimated by their subject and examination results, and divided into categories θ_j. The proportion of correct answers to total number of answers `pj/mj` was then established for students in each category θ_j and used as an estimator of the probability value for the item response curve. This is an estimation of the true probability `P(θj)`, and it is then possible to obtain corresponding pairs `(θj, P(θj))` to fit the characteristic curve and obtain the parameters `a` and `b`. This approach directly classifies automatically all the questions in the bank. Figure 2 depicts one of the characteristic curves obtained.
- A group of five experienced instructors was given a set of 30 questions to categorise into five categories: from 1 (Very Easy) to 5 (Very Hard). This gives a reasonable estimation of the difficulty of the items in the small question sample, with the intention to try to infer from this grading a classification for the whole collection. For this second approach, a Neural Network was then trained with the values obtained from the experts, and made to classify the whole item bank based upon the students' results and the classification by the experts. Once the Neural Network learns the ranking process, any number of questions can be ranked using just their historical information. It was then possible to automatically estimate the value of `b` for each question of the whole bank.

With both approaches, different strategies are being tried on this phase:

- It is possible to consider the final subject results, or only its examination component, as a measure of students' ability. Preliminary results seem to indicate that examination marks are a better indicator than overall subject marks.
- It also remains to be determined whether the experts' opinion is a good estimation of the difficulty of the questions in the sample, so it may be used when there are no historical data available.
- There are several ways of aggregating the experts' opinion, likely to produce different results. Data collection and analysis is progressing in this area.
- There are also several ways of categorising students' abilities θ_j, such as using equally spaced intervals or equal population segments (quartiles, deciles and so on). These are also likely to produce different results.

A complete analysis of the results obtained is currently progressing.

6 Conclusions and Further Research

Preliminary results are very encouraging. The scheme makes possible the automated classification of the items in a question bank to implement the cognitive ladder, even when the bank has not been developed for the purpose. The analysis of the results is currently in progress, in an attempt to establish the best strategies to follow in the near future. Some questions remain unresolved, in terms of the best indicator of the students' ability, whether the experts' opinion and Neural Network strategy provides a good estimator for when there is no historical results available, what is the best way of categorising the experts' opinion, etc. Research is progressing on these issues. After this phase is concluded, the research will attempt to establish an appropriate strategy to progress students up the cognitive ladder.

References

1. Baker, F.: The Basics of Item Response Theory. Second Edition, ERIC Clearinghouse on Assessment and Evaluation, Boston C., & Rudner L. (Eds), USA, ISBN 1-886047-03-0
2. Bloom B.S.: Taxonomy of Educational Objectives, David McKay Company Inc, New York, (1964).
3. Chalmers D. & Fuller A.: Teaching for Learning at University: Theory and Practice. Edith Cowan University, Perth., (1995).
4. Doran M. and Langan D.: A Cognitive-Based Approach to Introductory Computer Science Courses: Lessons Learned. SIGCSE Bulletin. Proceedings of the 26th ACM SIGCSE Technical Symposium. March, Nashville, TN, (1995), pp 218–222.
5. Fernandez G.: WebLearn: A CGI-Based Environment for Interactive Learning. Journal of Interactive Learning Research. Vol 12, No 2, (2001), pp 265–280.
6. John S., Netherwood G., Sudarmo & Fernandez G.: Learning Taxonomy Analyses Of Student-Based Activities Using The Lego Mindstorms System. Proceedings of the 13th AAEE Conference. Canberra, ACT, Australia, September, (2002).
7. Fernandez G., John S. & Netherwood G.: Objective-Based Teaching of Science and Engineering With an On-line Student-Centred Environment. Proceedings of the 12th AAEE Conference. QUT, Brisbane, Australia, pp 332–337, (2001).
8. Laurillard D.: Rethinking University Teaching: A Framework for the Effective Use of Educational Technology. Routledge. London, (1993).
9. Prosser M. & Trigwell K.: Teaching for Learning in Higher Education, Open University Press. Buckingham, (1998)
10. Wood, D., Bruner, J. S., & Ross, G.: The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17, 98–100, (1976).