Different approaches to computer adaptive testing applications

Cristina Garrido Morro

1 Introduction

Inferring the student's knowledge level on a subject is a very studied topic and several expert systems have been developed for this purpose. A *Computer adaptive testing* (CAT) application is a system that presents questions to a student and infers the student's knowledge level from his answers [4]. The student's knowledge level is inferred using the answers to the questions previously presented to the student. The difference between CAT and other expert systems that infers the student's knowledge level is that CAT uses the least amount of questions. In addition, the question to be presented is selected carefully taking into account the answers to the previous questions or the student's knowledge level at the moment, so that the question could provide as much information as possible [1].

The challenge of the present article is to show how to implement a computer adaptive testing application following two models: *Item response theory* (IRT) [1] [2] [5] and *Bayesian networks* (BN) [1].

The most used approach is IRT [1]. In this approach the probability that the student answers correctly to a question depends on the student's knowledge level. The higher the student's knowledge level, the bigger the probability of correct response to a question is. Knowing the shape of the probability distribution of the questions, it is possible to predict the user's knowledge level, given the set of responses to the previous questions.

Another approach is a *Bayesian network* (BN) model [1]. A BN is a directed acyclic graph whose nodes represent variables and links represent the dependence relationships between the variables. In our case, X_i is a variable, which can have values *correct* or *incorrect*. It expresses the student's answer to the i^{th} question. $P(X_i)$ is the probability that X_i is correct, i.e. $P(X_i = correct)$. The relationships represent the order in which the items should be mastered. In BN the probability that a student answers correctly to a question is conditioned to the student's responses to the previous questions.

The paper organization is the following. In section 2 you can find a detailed description of the stages to be followed to implement a CAT system. In sections 3 and 4, you can read a description of two models that can be applied to the implementation of CAT systems, IRT and BN, respectively. The final conclusions are drawn in section 5.

2 Computer adaptive testing

There are several expert systems that can estimate the student's knowledge level by analyzing the answers given by him/her to an exam/test. These expert systems require that the student answers the whole test, that is, all the questions in order to get valid information about the student's knowledge level. Computer adaptive testing systems [4] estimate the student's knowledge level using as few questions as possible. CAT systems select which questions must be the next to be answered depending on the student's knowledge level so that the answer to this question gives the biggest amount of information about the student's knowledge level.

While building a CAT system, some stages must be completed in order to get an efficient system. One example of these phases is the following [4]:

1. *Question pool:* Select a set of questions (question pool) related to the subject/theme the students are going to be assesed in.
2. *Pretest of a sample set:* Take a big sample set formed by students, present the test to the sample set and collect the results of the tests.
3. *Parameters of the model:* From the results of the tests, train the model setting up its parameters that will be used in the next stage.
4. *Model implementation:* Implement the process that uses some CAT model explained in the following sections.

The CAT process consist of a loop that estimates the student's knowledge level by the answers to the previous questions and searches for the question that gives the most information according to the student's knowledge level. This question will be the next to be asked to the student. More detailed, we can define the process as in [1] as a sequence of the following steps:

1. Estimate the student's knowledge level by using the responses to the questions that have already been answered by the student. The estimation is done using one of the models that are explained in the following sections. If no questions are answered yet, the student's knowledge level is the average value of the knowledge level of the sample set. Go to step 2.

2. If the student's knowledge level is outside the given threshold boundary then stop the process, otherwise go to step 3.

3. Select the question in the question pool that provides the biggest amount of information according to the student's knowledge level at the moment. The selection of the question depends on the model used for the CAT system. Go to step 4.

4. Show the selected question to the student, wait untill it is answered and go to the step 1.

Note that in order to estimate the student's knowledge level accurately, some other factors should be taken into account. The interface adaption of the CAT system has a direct impact on the user performance [3].

3 IRT model

The IRT model is used to estimate the student's knowledge level and to select an appropiate question according to it [5]. That is, it is applied in the estimation of the student's knowledge level and in the selection of the next question to be presented.

In IRT model the answer given by a student to a question can be probabilistically predicted given the student's knowledge level and the *item charasteristic curve (ICC)* of the question. The ICC of a question expresses the probability that a student answers corretly to the question given the student's knowledge level.

The ICC is defined by n parameters. So, for each question there are different values of the n parameters that define its ICC. The values of these parameters are set up in the stage 3 (Parameters of the model) in the CAT system construction. Several ICCs have been used in CAT systems. For one CAT system only one ICC is needed. The most commonly used ICCs are *two parameters logistic ICC* [1] and *three parameters logistic ICC* [2].

3.1 Two parameters logistic ICC

Two parameters logistic ICC [1] is defined by the following two parameters:

- *Difficulty parameter for question i, β_i,* is the value of the student's knowledge level θ where the probability of a correct answer to the question is equal to the probability of a wrong answer.
- *Discrimination parameter for question i, α_i,* indicates that if the student's knowledge level θ is higher than the difficulty parameter of the question, β_i, then the probability of the student to answer correctly to the question is high.

The ICC is defined as follows [1] :

$$P(X_i = correct|\theta) = \frac{1}{1+e^{-\alpha_i(\theta-b_i)}},$$

where

- θ is the student's knowledge level,
- X_i is the answer (correct or incorrect) to question i,
- α_i is the discrimination parameter for the question i, and
- β_i is the difficulty parameter for the question i.

3.2 Three parameters logistic ICC

Three parameters logistic ICC [2] is defined by the following three parameters:

- *Difficulty parameter for question i, β_i,* is defined as before.
- *Discrimination parameter for question i, α_i,* is defined as before.
- *Guessing parameter for question i, γ_i,* is the probability that a student without any knowledge answers correctly to the question by selecting the answer randomly.

The three parameters logistic ICC is defined as follows [2]:

$$P(X_i = correct|\theta) = \gamma_i + (1-\gamma_i)\tfrac{1}{1+e^{-\alpha_i(\theta-b_i)}},$$

where

- θ is the student's knowledge level,
- X_i is the answer (correct or incorrect) to question i,
- β_i is the difficulty parameter for the question i,
- α_i is the discrimination parameter for the question i, and
- γ_i is the guessing parameter for the question i.

3.3 Estimation of the student's knowledge level θ

Assuming that the questions are independent, the probability that the student answers correctly to questions $1,..,k$ is:

$$P(X|\theta) = \prod_{i=1}^{k} P(X_i|\theta),$$

where

- k is the number of answered questions,
- X is the sequence of answers $X_1, X_2, ..., X_k$, and
- X_i is the answer (correct or incorrect) to question i.

Now, knowing $P(X|\theta)$ and $P(X_i|\theta)$, it is possible to estimate $P(\theta|X)$by using Bayesian rule:

$$P(\theta|X) = \frac{P(\theta)P(X|\theta)}{P(X)}$$

Notice that we get the values $P(\theta)$ and $P(X)$ during the calibration of the model.

The main advantage [1] of IRT CAT systems is that the student's knowledge level estimation is based on the data cases, not on an expert knowledge that previously teaches the system. On the other hand, the most important disadvantage [1] is that the questions can not be divided into concepts.

4 Bayesian CAT model

In a Bayesian model we can compute the probability of an event using the information of the occurrence of previous events. In CAT systems we use a Bayesian model to compute the probability of a correct answer to a question, using the previous answers to the previous questions.

4.1 Bayesian networks

The conditional probability to answer correctly to the question i is [1] :

$$P(X_i|X) = \frac{P(X_i, X_1, ..., X_k)}{P(X_1, ..., X_k)},$$

where

- X_i is the student's response to the question i, and
- X is the vector $X_1, X_2, ..., X_k$ of the student's answer to the previous questions.

Calculating the full joint probability $P(X_1, ..., X_k)$ is impractical. The amount of conditional probabilities grows exponentially while the number of items increases [1] . This problem is solved by using *Bayesian Networks* (BN).

Bayesian networks use only the relevant conditional probabilities that are defined by the structure of the network. The set of relevant conditional probabilities and the structure of the network are learnt from empirical data in the third phase of the CAT construction. In a BN, the probabilities are updated whenever an event occurs.

In BNs, it is assumed that each node X_i, given its parent nodes $parent(X_i)$, is independent from all the other nodes except its children/descendants [1]. The property of BN that let us save conditional probabilities is that the node X_i given its parent's node $parent(X_i)$, is independent from all the other nodes except its children/descendants. This leads to

$$P(X_1, ..., X_k) = \prod_{i=1}^{k} P(X_i|parent(X_i)),$$

where

- $X_1, ..., X_k$ is the set of answers, and
- $parent(X_i)$ is the set of X_i's parent nodes in the network.

4.2 CAT bayesian network structure

In a BN CAT system a node X_i represents the response of a question. Each node is asigned probability $P(X_i)$ that the student answers correctly to the question i. The probabilities are updated every time a new question is answered. Notice that now $P(X_i)$ is not a function of the student's knowledge level θ, but it is a function of the probabilities of its parent nodes.

The relations indicate the order in which the students are expected to master the questions. For example, a link $X_1 \rightarrow X_2$ means that knowing that question 1 was correctly answered, the probability of correct answer to question 2 is increased. Similarly, if question 1 was uncorrectly answered then the probability to answer correctly to question 2 decreases.

The main advantage of BN CAT systems is that the questions can be divided into concepts. The disadvantage is that the BN has to be trained by an expert of the topic, and if this is not possible, the BN would need a lot of data cases to be useful, for example, a few thousand of cases for few dozens of questions [1].

POKS [1] is a new modeling scheme for CAT based on BN but with an additional charasteristic. Now it is assumed that the influence of a parent of a node is independent from the other parents of the node, i.e.

$$P(X|X_{p1},...,X_{pk}) = \prod_{i}^{n} P(X|X_{pi}),$$

where $X_{p1},...,X_{pk}$ are the parents of node X. This assumption allows the system to be trained using less data cases than a ordinary BN CAT system.

5 Conclusion

In this paper we have given an overview of computer adaptive systems. The main approaches to implement CAT systems are IRT model and BN. The main difference between IRT and BN in CAT systems is that IRT computes the probability of a correct answer to a question depending on the student's knowledge level and the answers to previous questions, whereas BN computes the probability of a correct answer to a question taking into account the probability of the answers to previous questions.

After seeing how IRT and BN approaches are used in CAT systems, it is important to know which are the strengths and weaknesses of the methods. The disadvantages of the IRT and BN are focused on how to teach the models from real data and how to organize the contents of the questions.

Fortunately, there is one model that can avoid the disadvantages of IRT and BN CAT systems. POKS is a variant of a BN model that does not need a large amount of data cases to be calibrated and can divide the questions into concepts.

References

[1] Michel C. Desmarais and Xiaoming Pu. A Bayesian student model without hidden nodes and its comparison with item response theory. *International Journal of Artificial Intelligence in Education*, 15, in press.

[2] Eduardo Guzmán and Ricardo Conejo. A model for student knowledge diagnosis through adaptive testing. In *Proceedings of 7th International Conference on Intelligent Tutoring Systems*, pages 12–21, 2004.

[3] Eva Jettmar and Clifford Nass. Adaptive testing: effects on user performance. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–134, 2002.

[4] Gin-Fon Nancy Ju and Alfred Bork. The implementation of an adaptive test on the computer. In *Proceedings of the 5th IEEE International Conference on Advanced Learning Technologies (ICALT)*, pages 822–823, 2005.

[5] Angel Syang and Nell B. Dale. Computerized adaptive testing in computer science: assessing student programming abilities. In *SIGCSE '93: Proceedings of the twenty-fourth SIGCSE technical symposium on Computer science education*, pages 53–56, 1993.