ADAPTIVE ASSESSMENT BASED ON DECISION TREES AND DECISION RULES

Irena Nančovska Šerbec, Alenka Žerovnik and Jože Rugelj
Faculty of Education, University of Ljubljana, Ljubljana, Slovenia

Keywords: Adaptive assessment, Formative assessment, Machine learning, Decision tree, Decision rules.

Abstract: In the e-learning environment we use adaptive assessment based on machine learning models called decision trees and decision rules. Adaptation of testing procedure relies on performance, current knowledge of test participants, on the goals of educators and on the properties of knowledge shown by participants. The paper presents sequential process of adaptive assessment where human educator or intelligent tutoring system uses different adaptive rules, based on machine learning models, to make formative assessments.

1 INTRODUCTION

Educational assessment is the process of evaluation and documenting, usually in measurable terms, knowledge, skills, attitudes, and beliefs. In the e-learning environment we need assessment that is, besides being valid and reliable, quick and automated. The last two properties could be achieved by means of adaptive assessment. E-assessment involves the use of a computer to support assessment which happens in the case of web-based assessment tools.

Computerized adaptive testing (CAT) systems are computer supported tests that adapt to the student's knowledge level and use a shorter number of queries tailored to his individual characteristics (Wainer, 2000). The existing computerized adaptive testing (CAT) systems base their adaptation mainly on the learner's performance using statistical models, which can be considered as restrictive from a pedagogical viewpoint. As it will be discussed later, some novel adaptation approaches for testing, which seem more pedagogically promising, have already been suggested (Nirmalakhandan, 2007; Lazarinis, Green, & Pearson, 2009, 2010; Sitthisak, Gilbert, & Davis, 2008).

Many research communities work on adaptability of learning objects in the e-learning environment. Their work is based on different learning theories and practices, cognitive psychological research and teaching strategies. Adaptability is the main property of intelligent tutoring systems (ITS) that provide direct customized instruction or feedback to students. They can be classified as an intersection of education, psychology (cognitive science, developmental psychology) and computer science (artificial intelligence, multimedia, Internet) (Woolf, 2009).

The paper presents prospective software tool for adaptive testing based on machine learning models, such as decision tree (Nančovska Šerbec et al, 2006, 2008) and decision rules. The tool could be used in the blended learning scenario for formative and self-assessment, after the topic is taught by the educator (see Figure 2). Adaptation of the assessment to individual student relies on the following factors:

- current knowledge of the student,
- goals of the educator/student,
- properties of the knowledge absorbed by the participants

From the e-learning point of view, data mining or artificial intelligence applications in e-learning could be divided into the following categories (Castro at al, 2007):

1. Applications dealing with the assessment of students' learning performance.
2. Applications that provide course adaptation and learning recommendations based on the students' learning behaviour.
3. Approaches dealing with the evaluation of learning material and educational Web-based courses.

Nančovska Šerbec I., Žerovnik A. and Rugelj J..
ADAPTIVE ASSESSMENT BASED ON DECISION TREES AND DECISION RULES.
DOI: 10.5220/0003521104730479
In *Proceedings of the 3rd International Conference on Computer Supported Education* (ATTeL-2011), pages 473-479
ISBN: 978-989-8425-50-8

4. Applications that involve feedback to both teachers and students of e-learning courses, based on the students' learning behaviour.
5. Developments for the detection of atypical students' learning behaviour.

Our system could be classified into categories 1 and 4. Machine learning methodologies decision trees and decision rules are data mining methods, which means that our research could be ?related? to the field of *educational data mining* (EDM) (Baker & Yacef, 2009; Romero & Ventura, 2010).

In the next section we will present the architecture of the proposed tool for adaptive assessment.

2 ADAPTIVE ASSESSMENT TOOLS

A Computer Adaptive Test (CAT) can be defined as a test, administered by a computer where the presentation of each item (exercise, question, task) and the decision to finish the test are dynamically adapted according to the answers of the examinees. CAT tools are used mainly as skill meters presenting the overall learner's score on a subject and a pass/fail indication. More specifically, test items dynamically adjust to student's performance level, and as a result, tests are shorter and test scores tend to be more accurate (Lazarinis, Green, & Pearson, 2009).

In this paper a framework for creating adaptive tests is presented (see Figure 1),. The framework incorporate rules (based on machine learning models) which allow personalized assessments. The Web tool implementing the framework supports the sequence of adaptive exercises (Gouli et al,, 2002).

In the following subsections we describe the main parts of adaptive testing tool, for modelling the adaptive engine. We continue with machine learning models which are used to generate if-then rules for adaptation.

2.1 Components of the Adaptive Testing Tools

Typically, adaptive e-learning tools consist of four parts that work closely together (see Figure 1, Lazarinis, Green, & Pearson, 2009). The *domain model* maintains the topics, concepts and other fragments that are used in creation? of lessons. *User (student) models* contain information about the learners that varies from demographics (name, address, etc) to their current knowledge and to learning style and preferences.

The *adaptation model* is a part in the adaptive multimedia tool. Model consists of collection of rules that define how the adaptation must be performed. The rules are used for updating the user model through the generated relationships between the concept and the existing learner knowledge. The final part, the adaptive engine, performs the actual adaptation. The adaptation model describes the conditions and the actions on which the presentation of the information is based and the adaptive engine implements these rules. In the presented adaptive system the domain model consists of the topics and the testing items that are adaptively presented to the learners. The user model component of the exemplar adaptive learning tool corresponds to the *learner profile* module of our adaptive testing tool. The adaptation model consists of a set of customizable if-then-else rules concerning performance, knowledge and the goals of the test participants (Lazarinis, Green, & Pearson, 2009).

In the paper, we are concerned only with the design of the *adaptation model* and not with other components of adaptive testing tool. The *adaptation engine* selects the current exercise by following the rules defined in the adaptation model.

2.2 Decision Tree

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences.

Decision trees are generated or induced by using datasets that consist of validated statically generated tests with wide set of exercises. Through the process of building of decision trees we capture the knowledge structure of the statically tested students. Process of assessment is routed by decision tree structure and it is ended by student's classification by reaching the tree leaf.

The procedure of generating decision tree from training set is called induction of the tree. We start with blank tree and whole set of training objects. Then on every step with the help of heuristic evaluation function we choose an attribute (exercise), which wasn't used jet (Witten & Frank, 2005). If there is not enough training objects or if the data contains missing values, then it usually leads to overfitting. The result of that are large trees with a lot of unimportant branches. That's why it is important where to stop growing of the tree by using the procedures for pruning unimportant branches. In most of the algorithms today we can find a mechanism called n-fold cross validation, which is one of the mechanisms for preventing the overfitting of the model to the training set). The final model is built on a whole training set.

The algorithm for tree induction, which selects the attributes (problems) in the nodes, utilizes the high dependency among attributes (problems).

2.3 Decision Rules

Decision rules are tree-like structures that specify class membership based on a hierarchical sequence of (contingent) decisions. For generation decision rules we used RIpple-DOwn Rule learner . It generates a default rule first and then the exceptions for the default rule with the least (weighted) error rate. Then it generates the "best" exceptions for each exception and iterates until pure. Thus it performs a tree-like expansion of exceptions. The exceptions are a set of rules that predict classes other than the default (Weka, 2010).

3 EXPERIMENTAL SETTINGS

We suppose blended learning scenario. In the following subsection we will explain the activities taken in order to design the adaptation model.

3.1 Diagram of Activities

Diagram of activities taken to model the system, is given in the figure 3. By these activities we gather information about students, their knowledge, learning styles, preferences, etc. In the paper we are concerned with activities labelled in grey boxes. Student (learner profile) and domain modelling will be part of further research.

The first step of designing of the adaptation model is presentation of thematic unit by the educator After the presentation of the thematic unit by a professor (in our case elementary mathematics “Expressions”, “Introduction into programming with programming language Pascal” or “Common knowledge about European Union” (Nančovska Šerbec at al. 2006, 2008), we carry out static (not-adaptive) web-testing of students with wider set of questions, e.g. 20-30 questions or exercises. Each exercise is randomly selected from the set of items of the selected topic. We use items scored correct (associated with number of points) or incorrect. Questions of type: “choose correct answer” or “fulfil the answer” are solved for each domain.

For example, the domain of EU consists of web-inquire results. Static tests were sent to students and teaching staff on a program study of computer science at the Faculty of Education. Solved tests were anonymously stored in MySQL database. We have collected 120 instances. The test contained 20 questions, each of them marked with 5 points. The values of the attributes were numerical (except the attribute class which was descriptive) and they presented achieved points of an individual student for each individual question. The maximum number of points for the test was 100. Attribute Success was the rating of student into three classes as regards the points achieved on the test.

For machine learning modelling we use the open source software tool Weka: tree-induction algorithm J4.8 and RIpple-DOwn Rule learner RIDOR (Witten, Frank, 2005)

After the generation of decision trees/rules we implement the adaptive model based on the tree structure. Table 1 presents the classification accuracy of the tree and the rules. We can see that the classification accuracy of decision tree is better. The system asks questions sequentially, one after another. In the background, algorithm follows the structure of decision tree. The question in the root of a tree is given to all students. (see Figure 4) After they answer particular question, the algorithm chooses the next one regarding to the correctness of the current answer. The testing is finished after the leaf of decision tree/decision rule is reached. It means that student’s knowledge is successfully rated.

A set of decision rules built on the EU domain is given In the Figure 2. Although their classification accuracy is worse compared to decision trees, they represent interesting pedagogical paradigm for assessment because we suppose that all participants belong to predefined class, e.g. excellent. For example, the last subgroup of rules in the Figure 2 can be interpreted in following manner: the test participant knowledge about EU is excellent, but he doesn’t know the number of languages used in EU. In that case his knowledge is sufficient. If the participant doesn’t know the correct answer to the following questions: the number of languages used in EU, which country doesn’t belong to EU, the meaning of EU logo and the country name where the Danube rises, we can classify his knowledge as not sufficient. The adaptive system based on decision rules uses four questions to make decision whether the participant knowledge is not sufficient. To diagnose the excellent knowledge on the EU topic, the system uses three questions: the year of

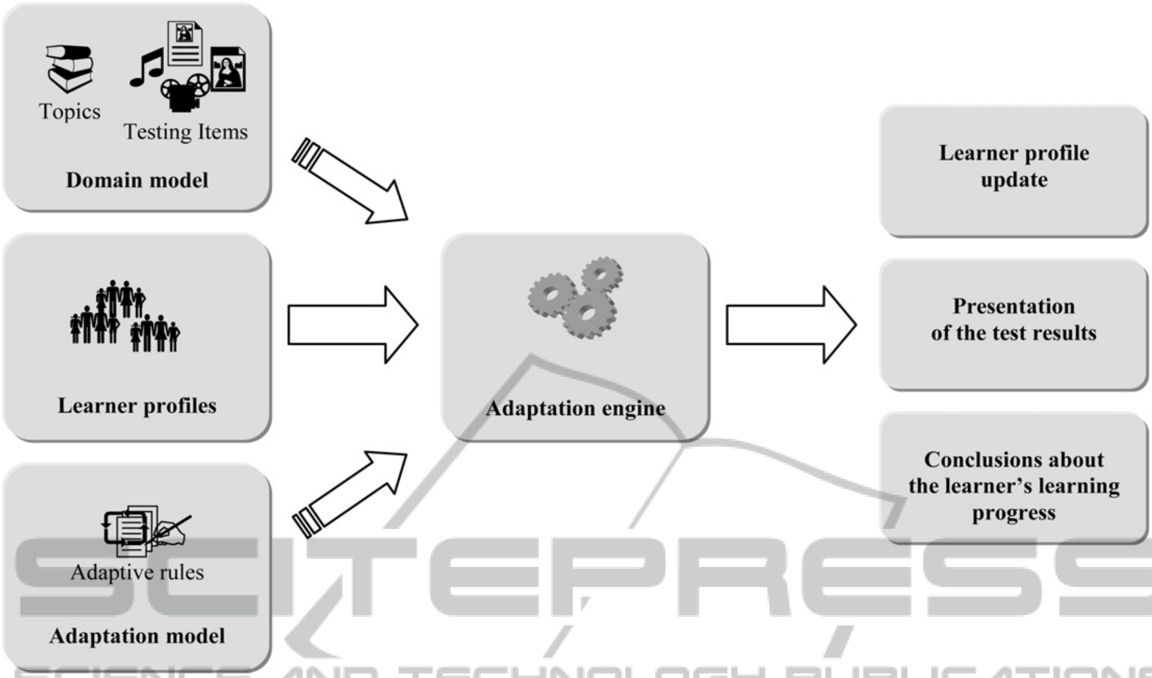


Figure 1: Main components of the adaptive testing tool (Lazarinis, Green, & Pearson, 2009; 2010).

Table 1: Classification accuracy (in %) of decision tree on each domain (10-fold cross validation/whole training set).

Domain	**Decision tree** accuracy (cross val./whole set)	**Decision rules** accuracy (cross val./whole set)
Expressions (math.)	76/100	73/83
Programming	95/98	82/90
EU knowledge	88/96	85/95

```
class = excellent  (120.0/117.0)
   Except (foundation <= 2.5) => class = sufficient (60.0/0.0) [31.0/0.0]
            Except (contry_not_EU <= 2.5) and (not_EU_inst <= 2.5) and
                  (ECTS <= 2.5) => class = not_sufficient  (3.0/0.0) [1.0/0.0]
   Except (mean_of_EU_logo <= 2.5) => class = suff. (12.0/0.0)[5.0/0.0]
          Except (contry_not_EU <= 2.5) and (not_EU_inst <= 2.5) =>
                                             class = not_suff.  (3.0/0.0) [2.0/1.0]
   Except (no_laguages_EU <= 2.5) => class = suff. (3.0/0.0)
                                                              [2.0/0.0]
            Except (contry_not_EU <= 2.5) and (meaning_of_EU_logo <= 2.5)
            and (Danube_rise_country <= 2.5) => class = not_suff.  (2.0/0.0)
                                                                  [2.0/1.0]
```

Figure 2: Decision rules for EU domain.

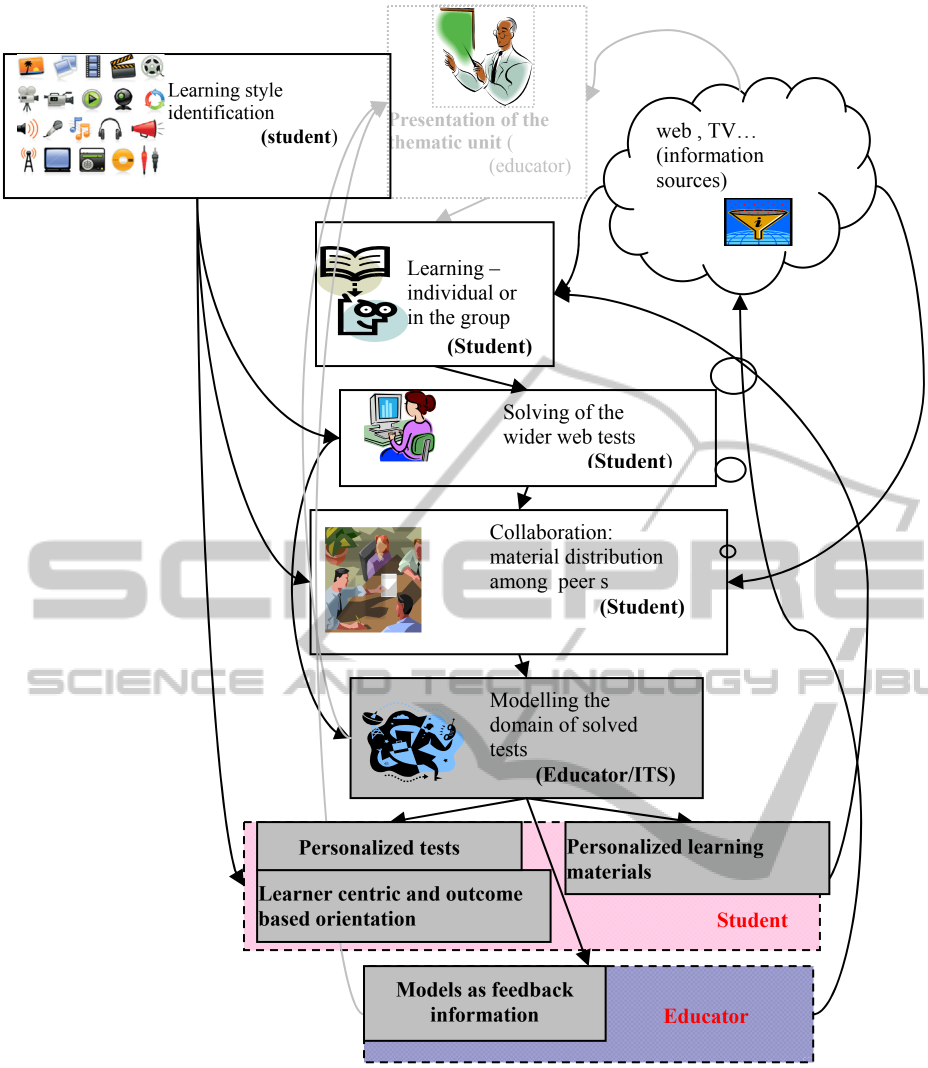
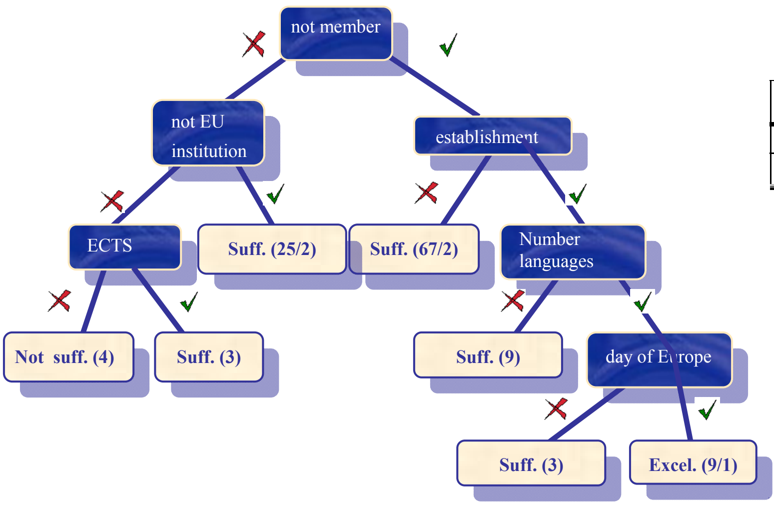


Figure 3: Diagram of activities.



Mark	Meaning
✓	Correct answer
✗	Incorrect answer

Leaf's mark	Joint num. of class. instances/incorr. class. instances
Excel.	9/1
Suff.	103/4
Not suff.	4

Figure 4: Decision tree on the domain of EU.

foundation, the meaning of the EU logo and the number of EU languages.

If we need a quick, adaptive testing without realization of wider, static test, we can use models related to population of student with similar learning properties. Because testing with wider static test is expensive, it could be carried out on a representative pattern of students.

4 PEDAGOGICAL VALUE OF MODELS

Assessment is a part of the developmental process of learning and is related to the accomplishment of learning outcomes. Recently, the main goal of assessment has shifted away from content-based evaluation to intended learning outcome-based evaluation. As a result, by means of assessment the focus has shifted towards the identification of learned capability instead of learned content. This change is associated with changes in the method of assessment.

Self-assessment is a crucial component of learning. Questions should be appropriate to the learner's level of knowledge based on the concept of a hierarchy of knowledge and their cognitive ability in order to use questioning more effectively as a pedagogical strategy (Sitthisak, Gilbert, & Davis, 2008).

Our models for adaptive assessment could be used as a part of formative assessment tool. With quick, adaptive self-assessment individual student can follow his progress and gain the feedback information about the success of his learning.

Other possible purpose of such a tree could be directing students in the learning process with the purpose to improve his/her learning outcome. Students classified in the left leaves in the tree structure could be directed to adopt topics which distinguish them from the right or better classified colleagues.

Another benefit for a teacher is an overview on the accomplished learning outcomes in the group., he can easily find out the main topics which need to be explained again From the decision tree, because a great deal of students did not accomplish the learning goals on that area. A teacher needs to recognize if a misunderstanding is due to lack of student's knowledge or due to difficulty of the exercise.

A tree or a group of decision rules also gives us the information about difficulty level of specific exercises or assignments for specific group. It is interesting to compare those results from the tree and the learning outcomes levels from the predefined curricula. With respect of levels of learning outcomes, teacher can recognize the main areas where students need more explanation or additional exercise.

We can see that models are beneficial from different points of view. They are allowing teachers to see their students as individuals and also give teachers the information about a whole group of learners or individual classrooms. There is also a benefit from the student's point of view. Why should the student take a long and exhausting test when he can achieve the same feedback with just a few questions and exercises? The main benefit for the student is that a tree or the applied decision rules are not allowing to ask a question or to offer an exercise with the same learning goals that he has already proven to be accomplished before.

A test will begin with an exercise that is on the root of the tree. This is the exercise with the learning goal that has been proven to be the most decisive. The next exercise will depend on whether the answer was true or false. This will be repeated until it reaches the leaf of the tree or it comes to a decision. Decision can be whether a grade or a descriptive rating. In this way we can test a group of students and allow each student to have a personalized test to achieve the feedback as quickly and as accurate as possible.

5 CONCLUSIONS

We modelled the students' knowledge captured in the tests on knowledge about the selected thematic units. Decision trees and decision rules are interesting models not only because they are predictors of learning outcomes but because of the transparency of their structure. As knowledge structures, they are interesting for teachers as feedback information about the learning outcomes of their students.

For students, they are interesting as a paradigm for adaptive formative assessment or self-assessment where the adaptation of the assessments relies on factors such as the knowledge, educational background, goals, preferences and performance of the learners.

One limitation of the tree-based approach is that the question/exercise/problem related to the topic in the root of the tree is asked all students. Another limitation is that the tree should be built off-line, before the adaptive assessment, and its structure doesn't adapt to the knowledge shown through the current adaptive assessments. Other limitation is that the students should answer all the questions in the nodes on the path from the root to the leaf, without possibility to omit some of them. Weakness of these models is that their classification accuracy is in average around 90%. Wrong classification can be especially problematic in the cases of successful students with low self-confidence. Besides this, we can not predict how the rating will influence on the motivation of students.

The adaptive assessment tool is in its initial testing phase and a lot of improvements are needed.

REFERENCES

Albert, D. and Mori, T. (2001) *Contributions of cognitive psychology to the future of e-learning. Bulletin of the Graduate School of Education,* Hiroshima University, Part I (Learning and Curriculum Development), 50, p. 25-34.

Baker R. & Yacef K. (2010). *The State of Educational Data Mining in 2009*: A Review and Future Visions. Journal of Educational Data Mining, Volume 1, Issue 1 1: 3– 17.

Castro F., Vellido A., Nebot A. and Mugica F. (2007) *Applying Data Mining Techniques to e-Learning Problems*, In Studies in Computational Intelligence (SCI) 62, 2007, p. 183–221.

Colace, F., & De Santo, M. (2010). *Ontology for E-Learning: A Bayesian Approach*. IEEE Transactions on Education, 53(2), 223-233

Giannoukos, I., Lykourentzou, I., & Mpardis, G. (2010). *An Adaptive Mechanism for Author-Reviewer Matching in Online Peer Assessment*, 109-126.

Gouli, E., Papanikolaou, K., & Grigoriadou, M. (2002). *Personalizing Assessment in Adaptive Educational Hypermedia Systems. Assessment,* (ii), 153-163.

Lazarinis, F., Green, S., & Pearson, E. (2009). *Focusing on content reusability and interoperability in a personalized hypermedia assessment tool. Multimedia Tools and Applications, 4*7(2), 257-278. doi: 10.1007/s 11042-009-0322-8.

Lazarinis, F., Green, S., & Pearson, E. (2010). *Creating personalized assessments based on learner knowledge and objectives in a hypermedia* Web testing application. Computers & Education, 55(4), 1732-1743. Elsevier Ltd. doi: 10.1016/j.compedu.2010.07. 019.

Nančovska Šerbec I., Žerovnik A. and Rugelj J. *Machine Learning Algorithms Used for Validation of the Student Knowledge*, MIPRO 2006, May 22-26, 2006, Opatija, Croatia. Vol. 4, Computers in education, p. 95-100.

Nančovska Šerbec I., Žerovnik A. and Rugelj J.(2008) *Adaptive assessment based on machine learning technology*. V: AUER, Michael E. (ur.). International Conference Interactive Computer Aided Learning ICL, September 24- 26 2008, Villach, Austria_. The future of learning - globalizing in education. Kassel: University Press, cop.

Nirmalakhandan N. (2007) *Computerized adaptive tutorials to improve and assess problem-solving skills,* Computers & Education 49, 2007, p.1321–1329.

Romero C. and Ventura S.. *Educational Data Mining: A Review of the State-of-the-Art.* IEEE Tansactions on Systems, Man and Cybernetics, part C: Applications and Reviews, 40(6), 601-618, 2010.

Sitthisak, O., Gilbert, L., & Davis, H. (2008). *An evaluation of pedagogically informed parameterised questions for self-assessment*. Learning, Media and Technology, 33(3), 235-248. doi: 10.1080/174398 80802324210.

Triantafillou, E., Georgiadou, E., & Economides, A. (2008). *The design and evaluation of a computerized adaptive test on mobile devices*. Computers & Education, 50(4), 1319-1330. doi: 10.1016/j.compedu. 2006.12.005.

Weka 3 - Data Mining with Open Source Machine Learning Software, The University of Waikato: <http: //www.cs.waikato.ac.nz/ml/weka/>,downloaded in 2010.

Witten I. H. and Frank E. (2005) *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, Morgan Kaufmann, San Francisco.

Woolf, B. P. (2009). Building Intelligent Interactive Tutors for revolutionizing e-learning. Pragmatics.