Journal of Scientific Research and Development 2 (5): 47-51, 2015
Available online at www.jsrad.org
ISSN 1115-7569

An intelligent and reliable examination-method using decision tree to classify the examinees in the educational systems

Bahman Arasteh [1,*], Behnam Arasteh [2], Mahsa Ganbarli [1]

[1]Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran
[2]Department of Language Teaching, Tabriz Branch, Islamic Azad University, Tabriz, Iran

Abstract: The classic and paper based testing and examination of examinees traditionally faced with many problems. The main problems of the classic-examination methods are the absence of dynamicity, intelligence and consequently the low reliability. In this paper, we proposed a software-based and intelligent method using decision tree for examination of examinees. This method uses an intelligent selection-algorithm to select appropriate questions from a questions database. In the selection algorithm the level of the examinees are taken into account Information of 80 students in the examinations of three courses English language, advanced programming and Data structures in a computer-engineering faculty is considered as the training and test data-set. The proposed method is evaluated on a real educational system on real examinees; the obtained results illustrated that the proposed method has higher reliability than the classic-examination methods.

Key words: Examination science; Classification; Decision tree; Reliability

1. Introduction

The test and examination science has been traditionally based on the classic methods in many educational systems. Testing and examination at all levels and disciplines traditionally faced with many problems; and consequently the errors of the classic methods prevent to attain the exact examination results (Omirin et al., 2007). The main problem of the classic examination methods is the absence of dynamicity and intelligence (Gonçalves and et al., 2004). Designers of the traditional test-questions design the questions based on their assumptions about the level of students (examinees) and the syllabus and curriculum of the relevant course. If the questions are so hard or so easy then the examination do not give us the exact information about the scientific level of the students. Indeed, an appropriate relation is required between the level of examination and the level of students. The simple and hard questions are not able to determine the exact level of a student (Lilley and et al., 2004). Spending a lot of energy by the examination designers, time limits for responding the questions, the need for staffs to control and correct the questions and low precision of the results are the main issues in the traditional method of examinations and testing in the educational systems (Luecht et al., 2000).

The increasing development in the artificial intelligence (AI), information technologies (IT) and their potential applications in the universities and educational environments are changing the education and examination systems so that the modern and technological learning and examination methods are used instead of classical methods. Similar to other aspects of human life, software and information technology can be used in education and examinations. Using IT and AI can improves the quality of education and examination. Reducing the rate of error in the examinations' results, the cost and the time of examinations are the main advantages of intelligent and software based examinations over the classic examinations. The main shortcomings of the classic examinations which are focused in this study are as follows (Van der Linden and et al. 2000):

- Due to differences in the level of examinees, some of examinees find the examination easy and some other find it hard. Hence, because of constant number of questions (Fixed-item test) and variable levels of examinees, the examination is not able to precisely measure the level of examinees (Wainer and et al., 2000). The uncertainty of the test questions before examination moment and different questions to different students are the main characteristics of the AI based examinations.
- In the classic examinations, we need manpower to control the examination and skilled manpower to correct the questions which will rise the time and cost evenly. Also there exists the error probability in the process of correcting the questions.
- All examinees must present at a fixed time and site to take the exam and the examination cannot be performed individually for each examinee. Indeed

* Corresponding Author.

a classic examination depends on the time and site. Furthermore, the classic examinations have low flexibility.

The main goal of this paper is reducing the above-mentioned shortcomings of the classic examinations. The main research variables which are improved in this paper are as follows:

- Intelligence of the examinations
- Reliability of the examinations
- Adaptability of the examinations

In this paper, we proposed a software based and intelligent method for examination of examinees. The proposed method can be used flexibly in different educational systems on the different examinees and different courses. This method uses an intelligent selection-algorithm to select appropriate questions from a questions database. In the selection algorithm the level of the examinees are taken into account; this feature will improves the reliability of the examination results.

2. Related Works

The computer based examinations can be divided into two categories (Wainer et al, 2000):

- Computer-based Linear Testing
- Computer based adaptive and intelligent Testing

The first type of computer based examinations is similar to classic examinations. On the other hand, second type examinations are intelligent. This type of examination can estimate the scientific level of examinees. The second type is the advanced and newer than the first type. There are many challenges and issues in the second type. Recently, the creation of international association for intelligent and adaptive testing (IACAT) shows the importance and potential applications of this technology. This scientific community by holding different conferences provides good opportunities to study the challenges and problems of student testing. Improving the reliability of estimated results by the computer-based examinations is one of the main challenges of testing and examination systems.

Some of the researchers believe that in the computer based testing the examinees need knowledge and information about the computer system and the examinees may not obtain an acceptable score because of insufficient familiarity with the computer (Omirin, 2007). With regard to this claim, the validity of the computer based examination will be under discussion. On the other hand other researcher rejects this claim and they believe that in the computer based testing the examinees do not require specialty about the computer science. Todays, all students have basic information about the computer and IT and are familiar with the computer and IT.

Item response theory forms the theoretical background and engineering of the computer based and intelligent examinations (Omirin, 2007). This theory includes three steps:

- Selection of questions by the selection-algorithm: the system includes a large set of questions in a database and a selection algorithm. The selection algorithm selects independent and different question for each examinee.
- Initial estimation: the scientific level of examinees is initially estimated with a limited number of questions (about 5 to 10) which is intelligently selected by the selecting algorithm.
- Up-to-dating the initial estimation: After initial estimation, the system selects and presents a series of questions which difficulty of each question depends on the response of the previous question. Indeed, with regard to the responses by the examinee the scientific level will be updated (Wainer et al, 2000).
- Finalizing the test: after exact estimation of scientific level, the test is independently finalized based on the stopping criteria.

The GRE examination is one of the well-known computer-based testing. This testing is performed by the center of US testing (ETS) in about 180 different countries. This test is the pioneer of this technology (Omirin, 2007). Furthermore, since 1998 the computer based version of TOEFL examination is performed with the classic and paper-based version. This technology can be utilized in different educational systems, universities and different courses. Improving the reliability of the obtained results by the computer-based testing is one of the main issues of this technology. In this paper, we propose an intelligent method to improve the reliability of the computer-based testing results using decision tree. The proposed method estimates the scientific level and score of the examinees by means of the limited number of questions and consequently limited time and cost.

3. Proposed Method

As mentioned in the previous sections, the classic testing and examination methods needs more time and cost and at the meantime have low reliability. In the classic examinations all of examinees should be gathered in a same site at the same time and should respond to the same questions. Therefore, the examinees with different levels should respond to a same number of questions within the same time. The same number of questions and the same response time for different examinees will reduce the reliability of test results. For example, if a classic exam with specific questions and specific examinees are repeated several times without changing the examinees, the test results will be different in different circumstances. So, because of constant test questions (Fixed-item test) but variable levels of examinees, the classic tests could not measure the accurate grades of students. While in the intelligent

testing the questions are selected by an algorithm (Selection Algorithm) according to the level of examinee. In the intelligent testing, testing is performed individually for each examinee and each examinee respond to individual questions. The testing time depends on the response time and the examinee does not have to wait to other examinee.

The number of questions is another important point in the classic testing and examinations. The number of questions should be selected in such a way that the obtained scores indicate the exact level of examinees. Another considerable point is that in the classic tests to obtain the perfect score the examinees should answer all questions. For example the examinee with high scientific level should also respond to simple questions. Otherwise, the obtained score do not reflect the exact level of examinee. In the proposed method, the level of each examinee is estimated by a pre-test service. The question-selection algorithm selects a question with regard to the response of the previous question. Hence, in the proposed method the exact level of an examinee is determined with a limited number of questions and limited time and cost. In the proposed method the decision tree is used to determine the level of examinees which is described in the next section.

3.1 Using Decision tree for Examinee classification

Decision tree (DT) is one of powerful and practical tools in machine learning. DT as a data structure can classify a large set of records into smaller subsets. A DT is constructed using a training data set (Barry de Ville and et al., 2006). In the created DT each internal node represents a testing on an attribute. Each internal node focus on a specific attribute and each branch of the node represents a value for the corresponding attribute. The leaf nodes in the DT indicate a specific class and each class corresponds to a scientific level.

One of the main questions in the DT creation is the selection of attributes. To this end, a DT learning algorithm is used. In this study, we used Inducing Decision Trees (ID3) (Barry, 2006). ID3 algorithm is used to create the required DT. In the ID3 algorithm entropy is used to classify the records and this algorithm tries to reduce the height of the tree. One of the main points in the ID3 is to select an appropriate attribute to assign the root of the tree. In order to learning DT, we used a dataset which is gathered from the information of classic examinations of three courses. Information of 80 students in the examinations of three courses English language, advanced programming and Data structures in a computer-engineering faculty is considered as the data set. We select 30% of data set for training and 70% for testing.

To determine the most appropriate attribute to classify the training data, a decision tree based on any attribute is evaluated. Finally the best attribute is selected to testing in the root node. The ID3 algorithm uses a statistical value as Information Gain to classify the training data set. Furthermore, this algorithm uses entropy to classify records. The value of entropy is calculated for each attribute by means of equation 1. In this equation S is a set of attributes. The effectiveness of an attribute in the classification of a data set can be evaluated by a Gain parameter. An attribute with the most gain is selected as the root of the tree. The gain of each attribute is calculated by equation 2. In this equation, A is an attribute in set S. In this study the WEKA software tool is used to induce the DT by the corresponding dataset.

$$Entropy(S) = -\sum_{i=1}^{m} p_i \log_2(p_i) \qquad (1)$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in value(A)} \frac{|Sv|}{|S|} Entropy(Sv) \qquad (2)$$

In order to evaluate the proposed method, we developed a three-layer software tool using the induced DT. The developed software consists of graphical and user-friendly interfaces to interact with the examinees. The software includes a database layer which was developed in SQL Server. The database includes tables to store and retrieve the information about examinees, questions and responses of questions. For example, the table User-test-archive includes the information of previously performed examination in the system. Fig.1 illustrates the entity-relation diagram (ERD) of the database.

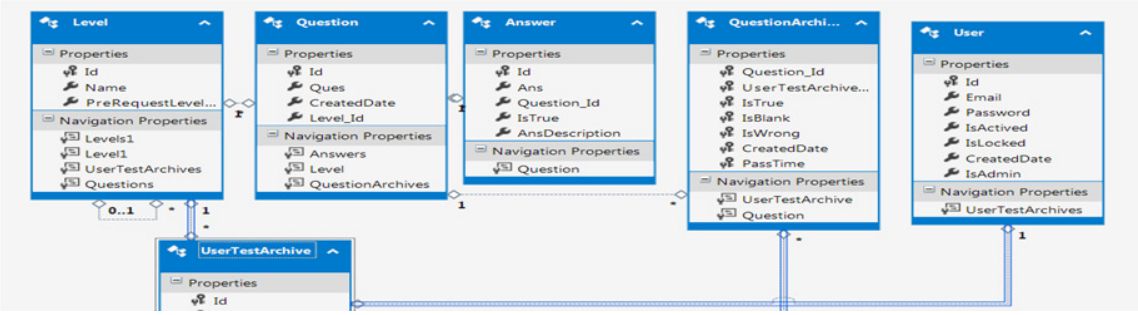

Fig. 1: The ERD of the database

4. Results and Discussion

In order to evaluate the reliability of the proposed method, the developed software based on the proposed method is utilized in a real educational system. This software was tested on the computer-engineering students. This software was used to perform the examinations of two courses (English language and advanced programming). Examination of each course was performed in two types. Firstly the classic version of examination was performed and then the examination was performed by the developed software on the same examinees. With respect to the results of the examinations, each examinee was classified in one the following classes: Excellent, Good, Fair, and Poor.

Each examination in each course was performed on a set of 50 examinees. Evaluating the proposed method on two different courses improves the validity of the obtained results. The obtained results from the classic and computer based examinations was gathered and compared with each other. In order to evaluate the reliability of the results both of examinations (classic and computer based) in each course were repeated three times. The classic examinations were repeated three times with different questions over the same examinees and the results gathered for further analysis. On the other hand the computer-based examinations using proposed method were repeated three times in specific intervals. Indeed, each examinee in each course was examined three times by the classic examinations and three times by the proposed method. Time distance between first and second examinations was two days and between second and third examinations was five days.

Firstly, the obtained results of the three classic examinations were compared with each other. The amount of differences in the results of three examinations in each course illustrates the reliability of the examination results. Table 1 represents the amount of differences in the results of classic examinations in the corresponding courses and Table 2 represents the amount of differences in the results of examinations based on the proposed method.

Table 1: Differences between the results of classic eximantions

	Differences			
	Between First and second examinations	Between First and third examinations	Between second and third examinations	AVG
English language Course	%16	%22	%19	%19
Advanced Programming Course	%11	%28	%17	18,6 %

Table 2: Ddifferences between the results of eximantions based on the proposed method

	Differences			
	Between First and second examinations	Between First and third examinations	Between second and third examinations	AVG
English language Course	%3	%9	%7	6.3 %
Advanced Programming Course	%2	%11	%9	7.3 %

Fig. 2 and 3 shows the obtained results of classic and computer based examinations. The results illustrates that the amount of differences in the proposed method are lower than the classic examinations. If an examination based on the proposed method is repeated on a same examinees leads to a same results. The low difference in the results of the first and other versions (repeated versions) of the computer-based examinations shows a high degree of reliability and validity of the proposed method with regard to the classic methods. Indeed, repeating the computer-based examinations over the same examinees leads to the similar results and the results of the computer based examinations do not depend on the time. With regard to the obtained results which are shown in table 3 the proposed method has higher reliability than the classic-examination methods.

Table 3: Reliability of the proposed method and classic method

	Reliability
classic exminations	%81
proposed Method	%95

5. Conclusion

In this paper, we proposed a software-based and intelligent method using decision tree for examination of examinees. This method uses an intelligent selection-algorithm to select appropriate questions from a questions database. The proposed method was evaluated in a real educational system on different examinees and the obtained results illustrates that the proposed method has higher reliability than the classic-examination methods and can be used in different educational systems. Using other types of learning algorithms in the examination classification is considered as the future work of this study.

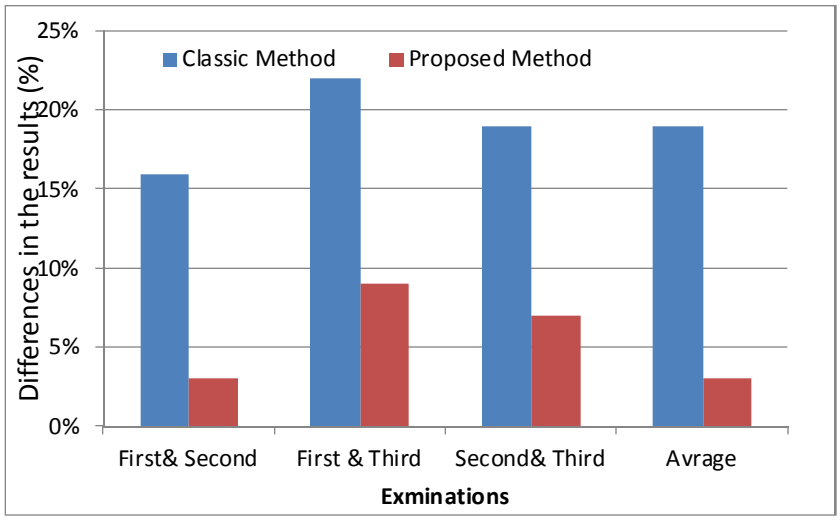

Fig. 2: comparison of the differences in the results of classic and computer-based examinations in engliah language course

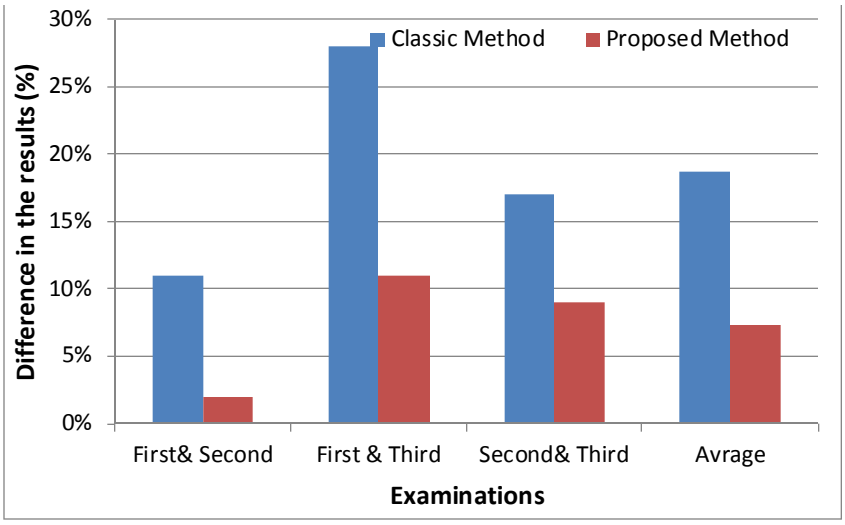

Fig. 3: comparison of the differences in the results of classic and computer-based examinations in adnavced programming course

References

Barry de Ville. (2006). Decision Trees for Business Intelligence and Data Mining. Using SAS enterprise Miner, SAS Institute INC.

Gonçalves, J. P., Aluísio, S. M., de Oliveira, L. H. M., & Oliveira, O. N. (2004). A Learning Environment for English for Academic Purposes Based on Adaptive Tests and Task-Based Systems. Lecture Notes in Computer Science, 3220, 1-11.

Lilley, M., Barker, T. and Britton, C. (2004). The development and evaluation of a software prototype for computer Adaptive testing, Computers and Education, 43(1-2), 109-123.

Luecht, R. M. & Nungester, R. J. (2000). Computer-adaptive sequential testing. In W. J. van der Linden & C. A. W.Glas (eds.). Computerized adaptive testing: Theory and practice. (Pp.117–128). Boston: Kluwer-Nijhof Publishing.

Omirin, M. S. (2007). Difficulty and discriminating indices of three multiple-choice tests using the confidence scoring procedure. Educational Research and Review, 1(2), 14–17.

Van der Linden, W. J., & Glas, C. A. W. (2000). Computerized Adaptive Testing: Theory and Practice, Netherlands: Kluwer Academic Publishers.

Wainer, H, Bradlow, E. T. & Du, Z. (2000). Testlet response theory: An analogue for the 3-PL useful in testlet-based adaptive testing. In W. J. van der Linden & C. A. W.Glas (Eds.). Computer adaptive testing: Theory and practice (pp. 245–269). Boston: Kluwer-Nijhoff Publishing.