Evaluation of online assignments and quizzes using Bayesian Networks

Yamna Ettarres[1,2]

[1] Higher School of Digital Economy, University of Manouba
[2] Larodec Laboratory, University of Tunis

Abstract. This work presents a novel approach to evaluate the relevance of online quizzes and assignments on the success of students in an online exam, using Bayesian Networks BNs. The case under consideration is the module Preparation to Computer Science and Internet Certificate PC2i, learned in licenses' first year, using the online platform Moodle, in Tunisia [6]. The main problem to be faced is what pedagogical activities are to be used in order to prepare students to succeed their C2i certification exam. BNs are used to automatically model relations between licenses' names, assignments, quizzes and results of students in an online exam.

Keywords: e-learning, Bayesian Networks, evaluation of quizzes, evaluation of assignments.

1 Introduction

With e-learning, the learner must be in the core of the learning process; trained according to his needs using interesting, suitable and value-added content in the most efficient way. Design, conception and evaluation of the impact of online resources and activities on the learning process are very important. Always online activities are presented to learners without assessing their impact and interest in relation to concepts having to be mastered. It is therefore difficult to determine relevant activities. This work presents a novel approach to evaluate online activities, such as quizzes and assignments, in the module Preparation to Computer Science and Internet Certificate PC2i, using Bayesian Networks BNs. Data used for an automatic learning of BNs is obtained from reports generated by Moodle platform. The result is an automatic construction of BNs representing relations between license's name, assignments, quizzes and student performance in an online supervised exam. Using inference, relevant assignments and quizzes are determined.

E. Popescu et al. (eds.), *Innovations in Smart Learning*,
Lecture Notes in Educational Technology, DOI 10.1007/978-981-10-2419-1_7

2 PC2i in Tunisia

Computer Science and Internet Certificate C2i is a certificate attesting development, enhancement, validation and mastering of Information and Communication Technologies ICT. The module Preparation to C2i, PC2i is mandatory and taught in the license's first year, using Moodle platform [6]. PC2i enables preparation of students to obtain the certificate C2i level one, which is managed by the Virtual University of Tunis VUT [3]. The C2i certification exam contains 2 parts. In the theoretical part, multiple choice questions are presented. The practical part deals with mastering Web research, word processing, spreadsheets and presentations. Success of students in this exam may be based on the quality of online quizzes and assignments developed by the C2i team. We have already developed an intelligent tool to accomplish a classification of students, to facilitate their selection for the certification exam [3].

3 Bayesian Networks

3.1 Presentation

e-learning problems and processes can be represented using uncertain and causal relations [4]. For this reason, our approach is based on the use of Bayesian Networks BNs [7]. These Directed Acyclic Graphs DAGs are largely accepted to be interesting tools to model uncertain reasoning and enable a faithful representation of causal relations among a set of random variables [7]. A set of mutually exclusive states and a Conditional Probability Distribution CPD are related to each variable X_i. In a BN, each node is independent from the others, giving its parents. The joint probability distribution JPD is represented by the following formula:

$$p(X_1,..,X_n) = \prod_{i=1}^{n} p(X_i | parents(X_i)) \quad (1)$$

When evidences are available, probabilities related to different nodes can be calculated using inference algorithms based on the Bayes rule.

3.2 Construction of a BN

In this work, the construction of the BN is based on the use of data structured in a database. We have used Bayesian Network tools in Java BNJ [1] to automatically construct the BN. The K2 algorithm [2] is used, where the expert defines an order or a hierarchy among variables. Previous variables, to a defined variable are tested if they can be its parents. Multiple orderings are given by the domain expert in order to reduce K2's sensitivity to nodes ordering. The fully connected and most suitable BN is then adopted.

4 The automatic evaluation of online assignments and quizzes approach

Data recuperated from Moodle platform, is represented by 714 records, for the academic year 2008-2009. Students are from the Higher Institute of Computer Science and Management- University of Kairouan HICSM-UK.

4.1 Data recuperation

The data used concerns identification of students, their marks in quizzes, assignments and online exam. The Grader Report presented in the Administration block in Moodle is used. It contains students ID, first Name, last name, email and obtained marks in quizzes, assignments and exam. Data is exported using Excel spreadsheets. The Participant block in Moodle is used to recuperate the list of participants classified by licenses' groups. Data are stored into two spreadsheets. In the file students.xls, we have student' ID, first and last names, license's name and marks in the exam. In the file activities.xls, we have the identification of students and their marks in quizzes and assignments.

4.2 Data storage

Data is gathered and stored in a database using a recuperation process from XLS files. Only one table is used. Its attributes are students' ID, first and last names, licenses' names, marks in quizzes, assignments and online exam.

Treatment of missing data. Missing values may be observed in data gathered. It concerns marks affected to assignments, when they are not corrected by tutors. The nominal value (most frequent one) or the average value may replace numerical missed values. The average mark in the license is chosen to replace missing values.

This choice is motivated by the fact that students affected to different licenses have different profiles.

Discretization process of data. The discretization process is the process of portioning continuous data into different classes. The number of intervals and their boundaries are to be determined. Discretization is the fact of finding a compromise between consistency, simplicity and classification accuracy [6], [7]. Three classes are adopted, represented by these intervals: low ([0..10[), medium ([10..14[) and good ([14..20]).

Retained criteria for evaluation: Nodes of the BN. Retained criteria are license name, quizzes, assignments and online supervised exams.

4.3 Automatic construction of the BN

Automatic learning of a BN's structure and parameters from a database is the automatic construction of BNs. We have chosen to use the K2 algorithm [2]. This greedy search and score-based algorithm, is developed by Cooper and Herskovits [2]. It is based on the maximization of the structure's probability given data. The domain expert has to specify variables' order. Used in multiple real problems, K2 is able to construct in a rapid manner BNs with a moderate complexity[2].

5 Developed Web application using BNJ classes: Automatic Evaluation of PC2i assignments and quizzes

Bayesian Network tools in Java BNJ [1] are used in our work. We have developed a C2i Automatic Evaluation System C2iAES for an automatic evaluation of assignments and quizzes. It is a PHP5 Web application. Evidence about the online exam is introduced to evaluate the impact on quizzes and assignments. Data is sent from the web interface to the developed inference class in Java using an XML file, Observed.xml. This file is parsed using the Application Programming Interface API Java Document Object Model JDOM. Java classes are developed to handle data, create the BN and to recuperate the result of the evaluation process.

6 Use of the developed Web application C2iAES

Three main menus are presented in the developed Web application C2iAES. The first one deals with the creation of a project and the database. The second one, presents evaluation using an inference process. The third one enables loading of an existing project.

At first, the user specifies the name of the project. A folder is then created to store used and generated files. The user introduces data using two XLS files: students and activities. In the next step, a database having the same name as the project, is

then automatically created and filled with data retrieved from XLS files. An automatic treatment of data is also accomplished, for instance to replace missing values or to eliminate students having no assignments and no quizzes.
Using the second menu named Activities evaluation; BNs for evaluation are constructed automatically using the K2 algorithm and different ordering versions specified by a domain expert. BNs are presented to the user in order to select the most suitable one. The BN chosen, which best explains causal relations between variables, must be a completely connected one, while being a convincing model to the evaluation problem.
The evaluation of activities is then accomplished by the introduction of data in the BN as evidence gathered. A web interface is used for this issue. The final result is an XML file presenting evaluation results.
An existing project can be easily loaded using the third menu. The user is able to choose the BN and to accomplish a new evaluation of activities, after the introduction of new evidence.

Test results. The BN chosen is presented in Fig. 1. Quizzes are represented by nodes T1,..,T6. Assignments are represented by nodes D1,..,D10. The online exam is represented by the node DS1.

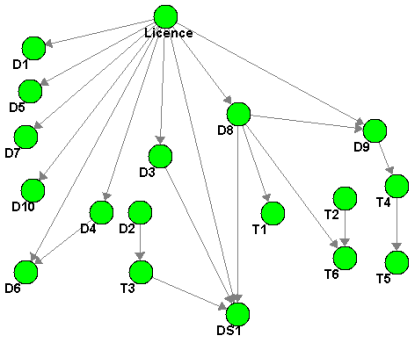

Fig. 1. BN's Structure

The user introduces evidence to the BN using a Web interface to indicate that results in the exam are good, for instance. The result of the evaluation is presented in Fig. 2. It is the XML file, Observed.xml. To obtain a good result in the online exam, the most interesting quizzes are T1, T2 and T6. Interesting assignments are D1, D2, D3, D4, D5, D6 and D8. The impact of the license is also important specially related to computer science SI.

Activities Evaluation

NODE	STATES AND PROBABILITIES
D1	1. P(medium) = 0.0 2. P(Good) = 1.0 3. P(Low) = 0.0
D10	1. P(medium) = 1.0 2. P(Good) = 0.0 3. P(Low) = 0.0
D2	1. P(medium) = 0.026501766784452 3 2. P(Good) = 0.9381625441696113 3. P(Low) = 0.03533568904593639

Fig. 2. Evaluation Results

7 Conclusion

The impact of online assignments and quizzes on the success of learners should be well studied and evaluated. In this paper, we present our proposed evaluation methodology of quizzes and assignments. BNs are used to obtain a reliable model of the relation between licenses' name, assignments, quizzes and results in an online exam. The evaluation approach is important for the construction of a good- quality online course. Students will be concentrated on interesting activities. Similar data in any module can be used and our approach can be easily applied. As perspectives, we propose the evaluation of the content of other types of courses and activities, such as the use of workshop activities and peer assessments in XML and Web development courses. We propose also the integration of this application under Moodle platform.

References

[1] Bayesian Network tools in Java (BNJ), http://bnj.sourceforge.net/

[2] Cooper, G. F., Herskovits, E. A Bayesian method for the induction of probabilistic networks from data. Machine Learning, vol 9, pp. 309-347 (1992).

[3] Ettarres, Y. The national strategy of teaching C2i in Tunisia. International Conference on Public Sector eTraining in the Arabian Gulf, April 11-14, 2010, Manama, Bahrain (2010).

[4] Ettarres, Y., Mellouli, K., Wuillemin, P-H. A Multicriteria Bayesian Intelligent Tutoring System MBITS. 10th IEEE International Conference on Intelligent Systems Design and Applications, Nov 29, 2010 - Dec 1, 2010, Cairo, Egypt (2010).

[5] Ettarres, Y., Mellouli, K. Automatic classification of students for C2i certification exam. IEEE Conference EDUCON Education Engineering 2011. Learning Environments and Ecosystems in Engineering Education (EDUCON 2011), April 4 - 6, 2011, Amman, Jordan (2011).

[6] Liu, H., Hussain, F., Lim, C., Dash, M. Discretization: An Enabling Technique. Data Mining and Knowledge Discovery, pp. 393-423, (2002).

[7] Pearl. J. Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann Publishers, Inc, San Francisco, USA, (1988).