A tool for students' grouping in classroom

Sergio A. A. Freitas, Rita C. Silva, Edna D. Canedo
Gama Faculty - FGA
University of Brasilia - UnB
Brasilia, Brazil
sergiofreitas@unb.br,ritasilva@unb.br,ednacanedo@unb.br

Tiago Franklin R. Lucena
Centro Universitário de Maringa - Unicesumar
Maringa, Brazil
tiagofranklin@gmail.com

***Abstract*— this paper aims to present a students' clustering method based on their backward knowledge of a specific domain. The categorization process uses the tracking information of students' answers in multiple choice questions (MCQ). We used an enhanced MCQ representation in which all the choice items (right or wrong) are linked to an existing conceptual map. We track the students´ answers and propose a method to identify each student´s profile on a certain subject and then cluster all the students' similar profiles. The resulting grouping map indicates how much the students learned about concepts on the knowledge domain.**

Keywords—Concept maps; Multiple choice questions; Student profile clustering

I. INTRODUCTION

Nowadays, the use of informatics tools in education process contributes to the rise of various objects and virtual learning environments. The uses of connected computers in the educational processes are many, as for: creating simulation, to interact with pedagogical objects, interactive social rooms and, especially, as a virtual learning environment (VLE). In a VLE, there is a large set of tools for instructors and students, for example: forum, chat rooms, lessons and questions.

Often the instructors use some kind of questions as assessments [1]: short answer, essay, true/false, numerical and multiple choices (MCQ). Such questions are commonly used as a process documenting and measuring the learned knowledge [2]**.** The results of the questions can be used to help students in their specific knowledge of some content [3]. The number of wrong answers enables the instructor to plan reinforcing classes for the current content.

Shaffer [4] proposed a scoring system used in the interactive games for education. Though it is a general proposal, it neither points out any mapping of conceptual mapping nor treats partial learning. Santos [5] proposed a teaching-learning environment to individualize student assistance. They create an interactive learning environment that clusters students based on the interaction, but lacks to associate the users groups based on their expected knowledge (i.e. a conceptual map).

For the backward knowledge, Freitas et al [3] used ontologies to extend multiple choice questions. Their proposal links right and wrong options to an existing conceptual map, and enables the instructor to evaluate the reason a specific student wrongly/correctly answered a multiple choice questionnaire.

By using Freitas et al' proposal, we track the students´ answers and propose a method to identify each student´s profile on a certain subject and then cluster all the students' similar profiles.

By extending their proposal, we can group students that need attention in classes and the ones that do not. This enables teachers to boost their classes with a specific student group on specific contents.

This paper is structured as follows: in section II we present the original Smart Quiz Framework, in section III we propose the extension of the SQ to support grouping, in section IV we evaluate some results and, finally, in section V we made some conclusions.

II. THE SMART QUIZZES FRAMEWORK

In their proposal, Freitas et al [3] called their tool as Smart Quizzes (SQ). Smart Quizzes are a series of Multiple Choice Questions designed for the Engineering´s courses of University of Brasilia. The SQ matches both the mapping of complex concepts and partial learning in the same framework.

The methodology to create SQ is divided in three parts: (1) the creation of a conceptual map for the lectures and Multiple Choice Questions (MCQ) where both the correct and incorrect options point to the conceptual map, (2) the identification of student's partial learning based on his/her answers, and finally (3) the construction of student's profile based on the conceptual map.

A. Conceptual map

With the aid of an instructor it is constructed an ontology diagram [6] for a given lecture. The idea is the mapping of the most important concepts in each lecture. The instructor determines what is important and what is not. The resulting map represents the knowledge expressed by linked concepts to be learned by the students.

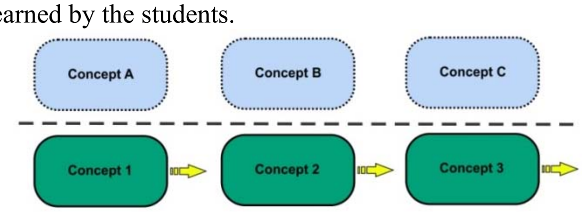

Figure 1: A conceptual mapping example.

Figure 1 is an example of linearity placed on a concept model of a lecture. The repository covers concepts that are directly related to a specific subject. It consists of concepts studied in others courses, thus they need not to be directly treated in the current classes. However, the students should know them in order to answer an MCQ.

B. Partial learning identification

Based on the conceptual map presented in Figure 1 for each subject directly related a sequence of MCQ is elaborated. In general, the MCQ have five choices and the student needs to identify the letter of the choice that best answers a question or completes a statement. The questions are classified by levels: level 1 - easy; level 2 – medium and level 3- hard. The SQ applied on a given domain considers questions at all levels. The method to classify a question as level 1, 2 or 3 depends on the judgment of the instructor.

An MCQ refers to a specific topic. The student has understood some aspects of a concept, even though his final choice is not 100% correct [7,8]. So, no statement is considered completely wrong and the student knowledge is always evaluated in the positive sense. According to this method, SQ infers which concepts or part of a concept is not completely understood by the students. This step can help in the diagnosis of a student, as it is possible to understand where the misunderstood concepts are *located* [9].

Finally, the idea is if there are five primary concepts in the statements, not all in a single option. Hence, if letter b) suggests a statement considering a correct identification of two of these concepts, one of them is not correct.

Thus, according to the method proposed, if the student chooses letter b), it is inferred that he can identify two of these concepts, but cannot solve one of them correctly.

C. The student's profile

The linearity of the conceptual map is due to the way that the student is directed to a set of questions: the student will be able to continue in the lecture if reaches a level of knowledge previously established to a given concept. Likewise, the student will miss grades in all previous related concepts when the answer is wrong.

The cumulative score for a student in a particular concept is computed in his Smart Quiz Pot – SQ pot – for a given concept. The students accumulate points on a concept by pointing a question correctly, or even mistakenly mark an item that is related to the concept of this "pot".

If the student selects an alternative in which there is, incorrectly, the application of a concept, he/she will lose points in corresponding SQ pots. If the alternative is correct, he/she will earn points in the main concept of the MCQ.

The map is based on the proposition that all content and issues are involved in the MCQ. The ultimate goal is to recognize the areas in which the student can be considered as approved in the lecture.

It is the knowledge of the instructor about the subject that it is used to map the sub-issues involved in each lecture. He must accompany the submission and activation of the lists of exercises where each student must select the correct alternative from among several others.

The mapping of each concept as a SQ pot aids the tracking and representation process (Figure 2). The SQ pots are related to each other and are related to a set of questions divided in three different levels. The student will fill up a given pot that represents a specific concept.

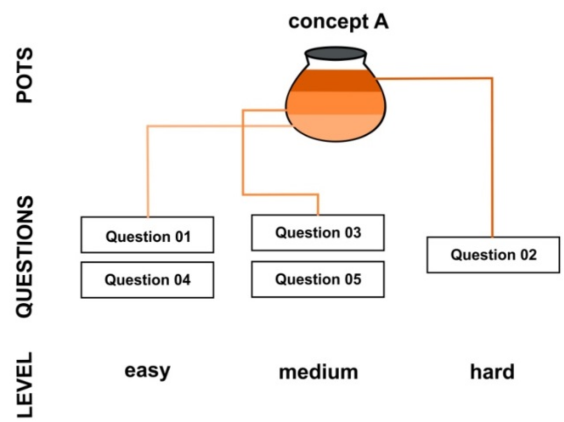

Figure 2: Representation for a conceptual SQ pot.

The representation of knowledge by adopting this strategy of establishing a metaphor with pots and functionality of storing substances meets the need to predict how the system simulates the process of knowledge acquisition. The pots show graphically what might be called the space of concepts.

III. THE GROUPING METHOD

Originally, the SQ framework is directed to identify a student' profile and use such information to direct her learning processes.

Based on Freitas' proposal [3] we extended the students' profile identification in order to enable it to cluster students based on similar fail groups. A similar fail group is a set of students' profiles in which SQ pots do not increase, or even decrease, when a set of MCQ are answered.

Consider $C = \{c_1, c_2, \dots, c_i, \dots, c_m\}$ as the set of concepts that must be learned by some course. The set is organized as ontology. For each concept c_i on the ontology, there is a set $Qc_i = \{q_1, q_2, \dots, q_i, \dots, q_n\}$ of multiple choice questions of c_i as their main scope.

Now consider the internal structure of every question q_i as a tuple $Tc_i = < c_i, mode, choices >$, where c_i is the main concept addressed by the question, $mode$ indicates the question level (easy, medium or hard) and $choices$ are the set of possible options: $choices = \{ch_1, ch_2, \dots, ch_k, \dots, ch_l\}$, and each individual choice ch_k is a tuple $< \frac{correct}{incorrect}, wc_1, \dots, wc_t >$. If a choice ch_k is correct then the main question concept is validated. Otherwise a wrong option is chose and $wch_1, \dots, wch_t$ indicates the misunderstood background concepts.

Using such representation, there are two paths when a student S_i answers a question q_i: (1) if the student choses the correct option then the main concept for q_i is reinforced. Its corresponding conceptual SQ pot is increased by a value. (2) Otherwise, a wrong option is chosen. Misunderstood background concepts $(wc_1, \dots, wc_t)$ interfere in the understanding of the question q_i. The corresponding SQ pots for each wc_l is then decreased by a value.

Now consider all changes on the SQ pots of a student S_i after the application of a series of MCQs. The result is a set of tuples $changes_{Si} =< (c_m, value1), (c_n, value2), \dots >$ where $c_n$are the concepts that are changed after the MCQs answers and $value$ indicates how much the corresponding SQ pot is increased or decreased.

The set of $changes_S$ for all students are the classroom map (TCM). TCM is the base to calculate the groups of students. The heuristic is defined as:

1. Transverses TCM and calculates the set of newest learned concepts NLC. $NLC =< (S_1, C_1, V_1), (S_1, C_2, V_2), \dots, (S_2, C_1, V_3), \dots >$, indicating how much $(V_1, V_2, \dots, V_j)$ the students $S_1, S_2, \dots, S_i$ learned about concepts $C_1, C_2, \dots, C_k$ after answering a series of MCQs.
2. Calculates the percent mean value MvC_i for each concept C_i of TCM. The set of all MvC_i (MvC) indicates the base line value for each concept C_i for a given classroom.
3. Use MvC to filter TCM in two lists:
   a. The list of ill-learned concepts (ILC) are those composed of all tuples in which the MvC_i for a given concept C_i is less than MvC . Now, sum the individual students' values for those concepts. The resulting values lesser than zero indicates the group of students.
   b. The rest of TCM will compose the list of well-learned concepts (WLC).
4. Now, ranks in reverse order both the ILC and WLC based on the number of concepts occurrences. The result classification indicates how much the concepts are (mis)understood by the class.
5. In ILC, the best ranked line indicates the worse understood concept. The list of all students associated on it will compound a group. The instructor can also build others groups based on the second ranked, third ranked and so on.
6. In WLC, the best ranked lines indicate the best understood concepts.

As a direct application of the resulting ILC groups is its uses in determining concepts and group of students that need reinforcing classes. This enables the instructor to give specific booster classes. Note that those classes are very objective because the instructor knows the ill-learned concept. As a second application, the instructor can analyze the ILC list to identify related concepts. Those related concepts could be used to restructure lessons. Finally, the ILC list of all concepts at the end of a course could be used to identify miss concepts.

Finally, the list WLC can be used as an indicator of which concepts are well understood and the students that are learning them.

IV. EVALUATION

As a work in progress, we tested the grouping method on Computer Organization and Architecture (COA) course. COA is obligatory for Software Engineering and optional Electronic Engineering.

We used the introductory part of COA: introduction to Boolean functions, logical gates and applications. It was tested a series of 10 MCQs on 32 students. Those 10 MCQs maps the following 8 concepts:

C1 – Memory
C2 – Flip-flops
C3 – Adders
C4 – Boolean function
C5 – Truth table
C6 – Counters
C7 – Shifters
C8 – Logic gates

The MCQ have five options and were applied using Moodle. Both correct and incorrect options have only one concept associated. The resulting data is exported as an excel table. The data is then treated: the options are converted to associated concepts and the results as computed in excel.

Table 1 illustrates part of the collect data:

Table 1: Collected data

	Question - Q1	
	C1, C2, C3, C4, C5	
	C3	
Stud. ID	**Concept**	**Value**
1	C1	-0.5
2	C3	1.0
3	C4	-0.5
4	C1	-0.5
5	C3	1.0
6	C3	1.0
7	C5	-0.5

Each student (Stud. ID) answer a question Q1. The concepts C1, C2, C3, C4 and C5 are, in this order, mapped on options a), b), c), d) and e). The correct option is third one, associated with the concept C3. For each correct answer, the student wins 1.0 point on his corresponding SQ pot for the associated concept (e.g. C3). For each wrong answer, the student *wins* -0.5 point on his SQ pot associated to the concept.

For each question it was calculated the MvC for all concepts at the question. Also it is calculated the mean percent of each concept at the same question. An example of the result for question Q1 is presented at Table 2.

Table 2: Example of McV and concepts' values calculus.

MvC for Q1	0,015625
C1	-0,08
C2	-0,03
C3	0,344
C4	-0,06
C5	-0,16

The MvC for Q1 is the arithmetic average the values attributed for each student's answer. The value for each concept C is the percent mean value.

The arithmetic average of all MvC will calculate the global MvC. The sum of each concept C for all questions Q will calculate the global influence of such concept on the answer of all questions. Table 3 presents the resulting values:

Table 3: Results for questions and concepts.

MvC	0,24063
C1	0,89063
C2	0,42188
C3	0,1875
C4	-0,0156
C5	-0,2656
C6	0,46875
C7	0,875
C8	-0,1563

The MvC value is the reference value. All values for concepts must be compared to it, as shown in table 4.

Table 4: Calculated ILC and WLC.

ILC	**C5**	-0,26563
	C8	-0,15625
	C4	-0,01563
	C3	0,18750
WLC	**C2**	0,42187
	C6	0,46875
	C7	0,87500
	C1	0,89062

Observe that values below the MvC will form the Ill-learned Concepts list (ILC) and values above will form the Well-Learned Concepts List. Also note the C5 (Truth table) is the worst learned concept. It influences the bad results. Of course C8, C4 and C3 also have a bad influence.

Finally, considering only C5, we calculate its influences at all questions to all students. Part of the results is shown in table 5.

Table 5: The result students' grouping.

S ID	C5
5	0,0
6	0,0
7	-0,5
8	-0,5
9	-0,5
10	0,0
11	-1,0

Each values below 0 (zero) indicates a bad influence of concept C5 at the student (red marked). The set of all marked students is the group that indicates which students do not learned concept C5. The instructor can now help such students.

V. Conclusion

Preliminary results showed that the mapping list is compatible with the real students' profiles. It was observed on assessments that the identified profile gives a good picture of each student at a given moment. Also the grouping methodology provides a very good report about the learning behavior of a specific group of students. It enables the instructor to map the whole class.

The resulting groups indicate two clusters: the students that need attention (ILC) and the ones that do not (WLC), enabling teacher to boost their classes to a specific student group in specific contents.

Finally, the list of well-formed students (WLC) can be used as a clue to other learning methodologies such as peer tutoring [10] and learning by teaching [11].

References

[1] Ángel Fidalgo-Blanco, María Luisa Sein-Echaluce, and Francisco J. García-Peñalvo, "Using Learning Analytics to improve teamwork assessment," *Computers in Human Behavior*, vol. 47, pp. 149–156, 2015.

[2] Jay Lemke, Robert Lecusay, Michael Cole, and Vera Michalchik, *Documenting and Assessing Learning in Informal and Media-Rich Environments*. Cambridge, Massachusetts: The MIT Press, 2015.

[3] Sergio A. A. Freitas et al., "Smart Quizzes in the Engineering Education," in *49th Hawaii International Conference on System Sciences (HICSS)*, Koloa - HI - EUA, 2016, p. 66-74.

[4] David W. Shaffer, *How Computer Games help Children Learn*. New York, USA: Palgrave Macmillan, 2006.

[5] R. Santos et al., "Teaching-Learning Environment Tool to Promote Individualized Student Assistance," in *In proceedings of the 15th International Conference Computational Science and Its Applications*, vol. Lecture Notes in Computer Science, 9155, 2015, pp. 143-155.

[6] T. Gruber, "Ontology," in *Encyclopedia of Database System*, Ling Liu and M. Tamer Özsu, Eds. USA: Springer-Verlag, 2008.

[7] C. Boutilier, "Abduction to plausible causes: an event-based model of belief update," *Artificial Intelligence*, vol. 94, no. 9, pp. 143-166, 1996.

[8] P. Gärdenfors and H. Rott, "Belief revision," in *Handbook of Logic in Artificial Intelligence and Logic Programming*.: Oxford University Press, 1995, vol. 4, pp. 35-132

[9] Golemati, Maria et. al., "Creating an Ontology for the User Profile: Method and Applications," in *Proceedings of the First IEEE International Conference on Research Challenges in Information Science (RCIS)*, Morocco, 2007, pp. 407-412.

[10] Liesje de Backer, "Promoting University Students' Metacognitive Regulation through Peer Learning: The Potential of Reciprocal Peer Tutoring," *Higher Education*, vol. 70, pp. 469-486, 2015.

[11] Sergio A. A. Freitas, W. C. M. P. Silva, and G. Marsicano, "Using an Active Learning Environment to Increase Students Engagement," in *29th IEEE Conference on Software Engineering Education and Training (CSEET 2016)*, Dallas - TX, EUA, 2016.