USING BAYESIAN NETWORKS AND PARAMETERIZED QUESTIONS IN INDEPENDENT STUDY

L. Descalço[1], Paula Carvalho[1], J.P. Cruz[1], Paula Oliveira[1], Dina Seabra[2]

[1]Departamento de Matemática, Universidade de Aveiro (PORTUGAL)
[2]Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro (PORTUGAL)

Abstract

The teaching paradigm is changing from a traditional model of teachers as suppliers of knowledge and toward a model of teachers as advisers who carefully observe students, identify their learning needs, and help them in their independent study.

In this new paradigm, computers and communication technology can be effective, not only as means for knowledge transmission, but also as tools for automatically providing feedback and diagnosis in the learning process. We present an approach integrating parameterized questions from two computer systems (Megua and PmatE), combined with a Web application (Siacua) implementing a Bayesian user model, using already many hundreds of questions from each of the two systems.

Our approach allows the students a certain level of independence and provides some orientation in their independent study, by giving feedback about answers to questions and also about the general progress in the study subject. This progress is shown in the form of progress bars computed by Bayesian networks where nodes represent "concepts" and knowledge evidences.

Teachers use Megua for creating and organizing their own database of (parameterized) questions, make them available for students, and for seeing the progress of each student or class in every topic being studied.

Keywords: Bayesian, networks, mathematic, b-learning, parametrized, problems.

1 INTRODUCTION

In the classroom, learning is largely conditioned by teacher directives, which is dependent on his experience. In this world of very fast change, students demand a learning environment compatible with this change: fast learning and logical and consistent analysis of situations, preferably related to real live problems. Nowadays, students are exposed to communication and information technology from young age, which makes them experts on this technology, sometimes better than their teachers. For these students, the learning process is substantially different of what it was for the previous generation, which includes their parents and teachers.

Literacy is expanded from a static understanding of written texts and participation in classes of knowledge transmission, to a context where student skills are challenged, demanding their active participation. Information world minimizes distances and, with such a huge amount of Web pages, videos, and severable other learning materials so easily accessible, traditional knowledge transmission classes are obsolete.

New learning conceptions, present in the new official mathematics education programs, assume learning to be an active process were students are capable of controlling and assessing their learning. For that to be effective, it is necessary to develop good learning environments, allowing students to apply concepts, solve problems, discuss ideas, and have immediate feedback about their progress and their skills development.

Active learning requires the students to engage in learning activities and thing about the work they are developing. The failure in learning is the result of exclusion in participation.

Intelligent tutor systems (ITS), computer systems capable of adapting to each student, can contribute to the learning process in a selective individualized manner. These systems store information about the student, collected from the interaction of the student with the system, from which a behaviour pattern is created, and use this information in adaptation or for giving some feedback to the student allowing his self-diagnosis.

The too high number of students per class, situation typically found in higher education in Portugal, makes it difficult to adopt an approach were the student work is taken as the centre of the learning process. The use of ITS in this context, is helpful for the students to build their own study system, and the appropriate environment outside the classroom, still using all support and learning material built up for that purpose.

We present an approach integrating parameterized questions from two computer systems (Megua and PmatE), combined with a Web application (Siacua) implementing a Bayesian user model, using already many hundreds of questions from each of the two systems.

2 PARAMETERIZED QUESTIONS

A typical limitation for the effective use of ITS in the learning process is the limited number of available questions in the several topics in study. Fortunately, we do not have this difficulty in the University of Aveiro. We can use the questions from PmatE ([1], [2]), a project with 25 years of existence, containing now many hundreds of parameterized questions in several areas of knowledge, presently with 1252 parameterized questions available for mathematics. Since the questions are parameterized, which means there are parameters in the questions that are instantiated in runtime by the computer, we have in fact many thousands of different questions available for use in ITS.

An instantiated question from PmatE consists of an initial text followed by four statements about that text, randomly chosen from a set of four or more specified in, what we call the Question Generator Model (QGM), or Model for short. Hence, the student can say which of the four statements are true or false, and up to four evidences are supplied to the ITS system. Fig. 1 is an example of a question generated from a QGM. In this example, the centre of ellipses and their semi-axes are randomly chosen from a set of possibilities specified in the QGM. The parameters are used to generate the initial text and statements, but also the pictures, which makes it possible to obtain a rich set of different questions from a single QGM, using PmatE technology.

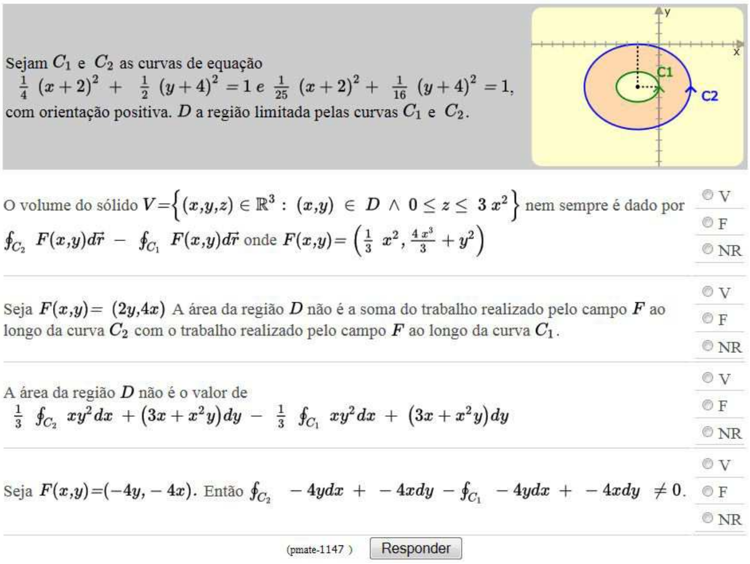

Figure 1: A question from PmatE

The QGM from PmatE have been used for many years in diagnosis tests ([3], [4]) and in the famous yearly competitions, with about twelve thousand participants. We are now reusing these questions in a different way, including them in our ITS designed to help students in their independent learning.

Another project for creating parameterized questions is Megua (Mathematics Exercise Generator, University of Aveiro; [5], [6], [7]). It includes a package for Sage Notebook, allowing the user to create multiple choice questions, with detailed answer, and sending them immediately to a Web application. Figure 2 presents an example of one of these questions. This new project has three main advantages, compared with PmatE: it is open source, the questions include a detailed answer and the teacher can create the questions and make them available without requiring a programmer. The only inconvenient on this system is that it requires the teacher to use LaTeX, HTML and Sage/Megua for creating a QGM, although only very basic features are needed. The number of questions for mathematics we are using from Megua is growing, and we have already 3068 questions generated form 358 different QGMs.

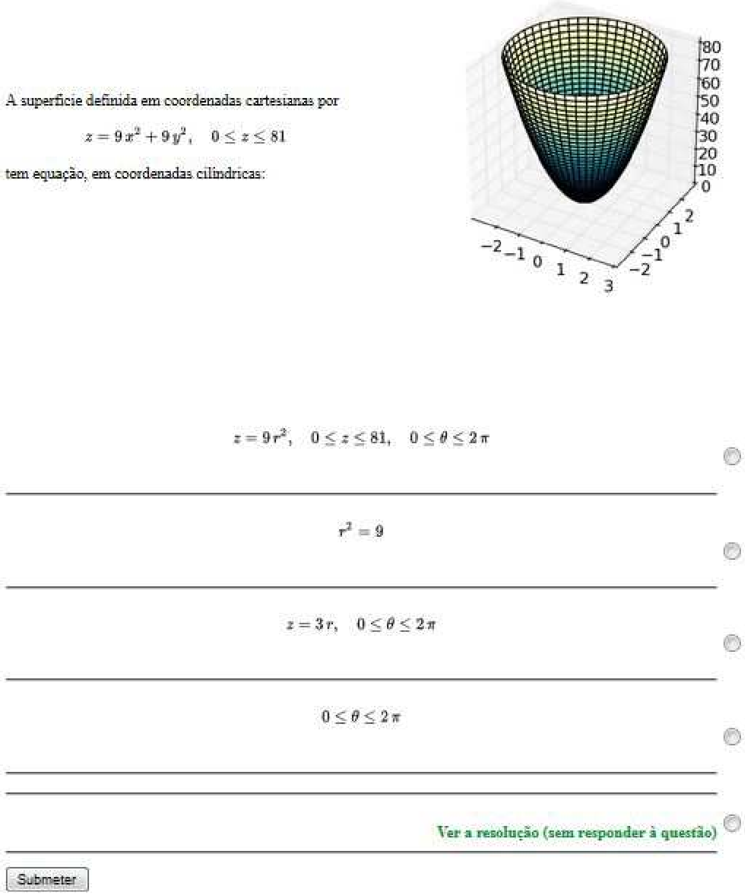

Figure 2: A question from Megua

There are many advantages on using parameterized questions, instead of static questions: different students can answer different questions generated by the same Model, and so very similar, what is useful for assessment, when students are seated side by side; the students can answer similar questions where parameters have different values, which makes clear what is essential in the question and not dependent on the parameter values; the teacher can think on a general model for a question instead of reusing questions from previous exams doing small changes, what is a repetitive and time consuming task teachers have to do.

3 MEGUA AND PARAMETERIZED QUESTIONS

Project Megua first goal ([5], [6]) is the development of a software package, with the same name, for authoring and sharing databases of parametrized mathematical exercises. These exercises include detailed parametrized answers, using LaTeX typesetting and HTML+MathJAX ([7]) as a source for problems both to be printed or used in the Web (currently only through Siacua Web application).

Although this project is in its early days it has been intensively used at the University of Aveiro by teachers of mathematical courses with a large number of students from all scientific areas. Inspiration came from projects [1], [8] and [9].

The chosen platform for this project is Sage Mathematics ([10]), an environment for mathematics ready for recent Ubuntu GNU/Linux distributions or for Windows® using a virtual machine and a web browser. This software can also be used in a server as a small "cloud" system using its companion, the Sage Notebook (now being replaced by Sage Math Cloud; [11]). To install Megua package on Sage Mathematics one can follow [12].

Megua package features took advantages of the Sage Notebook. Some of them are:

- An author or group of authors can develop their own database of problems (Megua package).

- Editing cycle: the author can create and change the problem, and see immediately the result, just pressing shift and enter (Megua package).
- This can be done using a “small cloud” meaning that users can edit and share problems in the web without having to install software on the computer (Sage Notebook).
- If an HTML+MathJAX exercise is being produced it can be sent to the Siacua system. If it is a LaTeX exercise it can be incorporated in a larger LaTeX file (exams or working books) or extracted to some document.
- Figure production from several sources: LaTeX and TikZ ([13]), HTML and related stuff, Sage Mathematics plots. Graphics from the Python Matplotlib library ([14]) or R ([15]) have been used and are under development.

An exercise has the following structure:

- **Summary.** A summary is a textual informative description.
- **Siacua Parameters.** Parameters that will be sent to Siacua together with the exercise: concepts, weights, slip, guess, difficulty level and discrimination.
- **Question text.** A problem text is written using in LaTeX typesetting or HTML with a LaTeX subset for MathJAX browser rendering. A parameters is a sequence of letters (typically ending in a number) that will be replaced by a value, or something else.
- **Options.** Several statements about the text where the first is true and all the others are false. Some of these statements are then randomly chosen and ordered in Siacua.
- **Answer.** A detailed parametrized answer describing a good answer for the posed problem. For the web, this part should also contain several multiple choice answers bases on the defined problem parameters.
- **Code.** Python/Sage Math sequence of rules that attribute values or text to parameters.

4 SIACUA WEB APPLICATION

With the availability of the QGMs from PmatE it is natural to try to develop computer learning systems making use of them. We have already developed other Web applications using QGMs ([16]) and we have now implemented a Web application, called Siacua (which stands for "Computer Interactive Learning System, University of Aveiro", in Portuguese; [17]). This application implements a Bayesian user model, together with an algorithm for knowledge propagation. Each student has an associated Bayesian network, where each node represents a topic, although we sometimes call it a concept node to distinguish from evidence nodes. Before the student starts answering questions, his Bayesian network contains only concepts, it is the conceptual map of the subject being studied. Each time the student answers a question, a new node is added to the network, an evidence node, connected with the concepts it involves, and the student knowledge is propagated. The user model we use has been proposed and tested with simulated students in [18] and [19] and has been tested with real students in [20] and [21].

Siacua is an open system in the sense that the student can see his progress, which is shown in the form of progress bars. The interaction is minimal, consisting of answering true/false questions form PmatE, multiple choice questions from Megua and selecting questions. The selection of questions can be made by clicking the corresponding progress bar, what randomly selects a question related with the corresponding concept, or by introducing the question number. These question numbers are only shown to the student in the moment he is answering the question. This allows the student to identify the question for answering again or interacting with teachers and colleagues.

We aim to infer the cognitive state of the student, from the collected data, diagnosing his state in the learning process. Since the data is incomplete and sometimes not correct, this is a step difficult to implement in the interactive learning system. Nevertheless, this application provides a very simple way of presenting questions to the student, and giving some feedback. The student knows immediately if his answer is right or wrong, can see the detailed solution and has information about his general progress in the course, given by the progress bars, as illustrated in Figure 3.

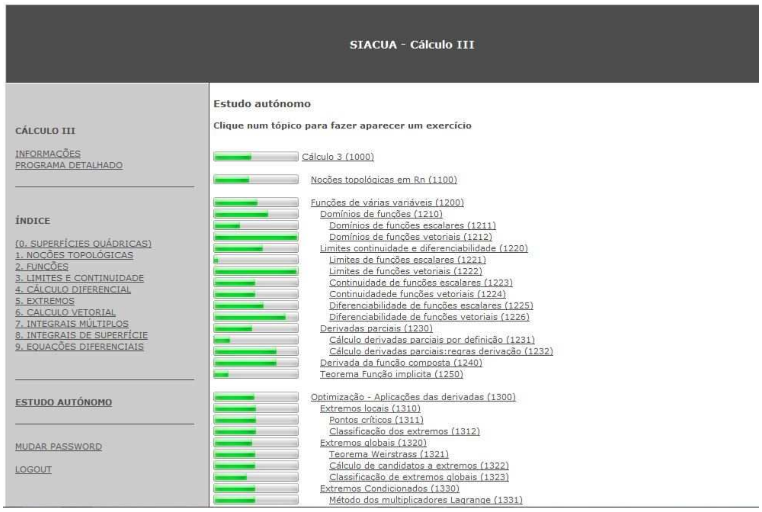

Figure 3: Progress in Siacua

5 SYSTEMS INTEGRATION

Questions from PmatE are available, through a Web service, for use in Siacua. We have selected some of these many questions and classified them using our concept maps. Technically, the Web application sends a request to a Web service from PmatE, sending the question id, and the Web service generates the question using the corresponding QGM. We note that this is done in runtime. The parameters are randomly generated when the request is made.

Questions from Megua are previously generated. This means that for a given QGM, a set of instances is created and stored in Siacua database, before they can be used. This option has advantages and disadvantages. The main disadvantage is that the number of questions generated by a given Model is fixed. Nevertheless, this number can be high, with some cost on memory use in the database, and in practice the behavioural is the same as with runtime generated questions. An obvious advantage is speed, since question instanced are already available without the need of computation. Another advantage is that the teacher can choose some question instances with better parameters, producing better looking and more interesting questions than those obtained allowing the computer to randomly choose any parameters.

A question can be created using Megua, inside Sage Notebook and sent to Siacua, being immediately available online. A teacher can experiment inside Sage Notebook, creating a question, and selecting several instances with convenient question parameters. Then he has to associate the question to concepts from the concept map of the course, and define the parameters for the Bayesian network. Finally the teacher can send several instances of the questions to Siacua. The parameters for the Bayesian network are: a list of concepts and the corresponding weights, parameters "slip" (probability of giving a wrong answer, knowing all concepts it involves), "guess" (probability of giving a correct answer without knowing any concept it involves), "level" (difficulty level) and "discr" (discrimination factor). Hence, when a question is created, the parameters for the Bayesian network have to be set and sent to Siacua together with the question itself.

We note that Siacua is used as a complementary tool in the learning process, not supposed to replace the official space of any course, which can be contained in Moodle, or somewhere else. This approach allows some flexibility in the design of Siacua contents. The program and exercises in "courses" contained in this application are not exactly the same as those in some course but instead, are supposed to be general and consensual enough to be useful for several similar courses.

6 BAYESIAN STUDENT MODEL

Bayesian Networks (BN) are directed acyclic graphs with probabilities in the nodes ([22], [23]). We use the graph to represent the conceptual map of the course, with each node representing a topic (or concept). The probabilities in the BN are used to compute knowledge of the student in each topic.

The model we use is defined in [19] and we have implemented it using Genie & Smile ([24], [25]).

The conditional probabilities associated with node Q, representing a question, are computed using the parameters "slip", "guess", "level", "discr" and the weights of concepts in the question. In general we have

$$P(Q=1 \mid \wedge_{C\in\Omega} C=1) = 1 - slip$$

and

$$P(Q=1 \mid \bigwedge_{C\in\Omega} C=0) = guess$$

where Ω is the set of concepts associated with question Q.

We have to set also the other conditioned probabilities, for the cases where the students knows some concepts related to the question but not others. To computed these probabilities we use the G function from IRT (item response theory)

$$G(x) = 1 - \frac{(1-c)(1+e^{-1.7ab})}{1+e^{1.7a(x-b)}}$$

where *a* is the value of "discr", *b* the value of "level", *c* is the value of "guess". The graph of G is given by Figure 4.

We have *G(0)=c* and

$$\lim_{x\to\infty} G(x) = 1.$$

Let x^* be such that $G(x^*)$=1-slip. The conditional probabilities are given in general by

$$P\left(Q=1 \mid \bigwedge_{C\in\Omega_1} C=1 \wedge \bigwedge_{C\in\Omega_0} C=0\right) = G\left((\sum_{C\in\Omega_1} w_C) \times x^*\right)$$

where

$$\Omega = \Omega_1 \cup \Omega_0$$

is the set of concepts associated with question Q and w_C denotes the weight of concept C in question Q.

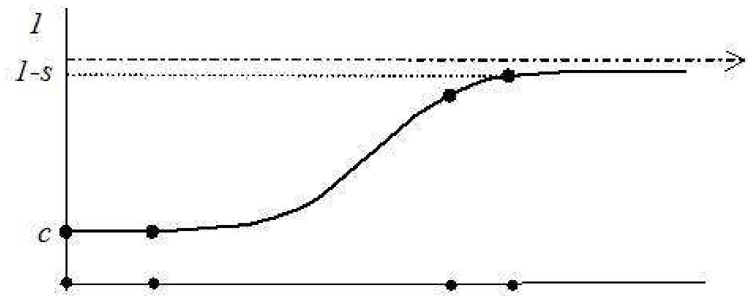

Figure 4: Function G

Using this model, all conditional probabilities can be computed from these small set of parameters. Having set the probabilities we can compute the beliefs after obtaining an evidence. This is done simply by marginalizing, factorizing and using the Bayes rule.

In general, the first time a student uses the application, his BN is the concept map. After providing an answer to a new question, a new evidence node is added to the network, connect to the related concepts, the conditional probabilities are computed using the Bayesian model parameters, and the evidence in the binary node is set to 1 if the answer is correct or 0 if it is wrong. If the student has already answered the question before, then we just update the evidence in the node already present in the graph. In both cases, after updating the network, the beliefs are propagated and the knowledge of the student, represent by the progress bars, is updated using the new beliefs in the BN.

Using this model, each time a student answers a question, not only the progress bars of the topics related to the question are updated but also all other progress bars above, up to the root of the tree, representing the course itself. The BN is used as a robust and theoretically well founded tool for doing this computation.

7 AKNOWLEDGMENTS

This work was supported by CIDMA ("Center for Research & Development in Mathematics and Applications'") and FCT ("FCT- Fundação para a Ciência e a Tecnologia") through project UID/MAT/04106 /2013.

REFERENCES

[1] PmatE: Projeto Matemática Ensino (2015). http://pmate.ua.pt.

[2] Oliveira, M. Paula; Silva, Sabrina Vieira (2006). An overview of PmatE: developing software for all degrees of teaching. Proceedings of the International Conference in Mathematics Sciences and Sciences Education, June 11-14, University of Aveiro.

[3] Anjo, António Batel, M. Oliveira, J. Pinto (2006). Tdmat 9,12 - A diagnostic test in Mathematics for 9th and 12th grades. Proceedings of International Conference in Mathematics Science and Science Education.

[4] Anjo, António Batel; Rui G. Isidro; Paula Oliveira; Sónia Pais; Joaquim Sousa Pinto (2007) TDmat - Mathematics Diagnosis Evaluation Test for Engineering Sciences Students. International Journal of Mathematical Education in Science and Technology, v38 n3, ISSN: 0020-739X, pp 283-299.

[5] Cruz, Pedro; Oliveira, Paula; Seabra, Dina (2012). Exercise templates with Sage. Tbilisi Mathematical Journal, 5(2), pp. 37-44.

[6] MEGUA package for parameterized exercises in Sage Mathematics (2015). http://cms.ua.pt/megua/.

[7] Mathjax (2015). http://www.mathjax.org/.

[8] Zeileis, Achim; Umlauf, Nikolaus; Leisch, Friedrich (2012). Flexible Generation of E-Learning Exams in R: Moodle Quizzes, OLAT Assessments, and Beyond. Working Papers from Faculty of Economics and Statistics, University of Innsbruck.

[9] Almeida, José João; et. al. E-studo: sistemas de ensino com geração automática de exercícios. http://passarola.di.uminho.pt/.

[10] Sage Mathematics Software (2015). http://www.sagemath.org.

[11] Sage Math Cloud (2015). http://cloud.sagemath.org.

[12] MEGUA Code (2015). http://code.google.com/p/megua/.

[13] Till Tantau. TikZ and PGF (2015). http://sourceforge.net/projects/pgf/.

[14] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment, Computing In Science & Engineering, 9:3, pp. 90-95.

[15] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/.

[16] Aires, Maribel; Descalço, L. (2006). A computer aided learn by example system, proceedings of the conference ICMSE 2006, International Conference in Mathematics, Sciences and Science Education, University of Aveiro.

[17] Siacua: Interactive Computer Learning System, University of Aveiro (2015). http://siacua.web.ua.pt.

[18] Millán, Eva; Pérez-de-la-Cruz, José-Luis (2002). A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation. User Model. User-Adapt. Interact. 12(2-3), pp. 281-330.

[19] Millán, Eva (2000). Thesis: Bayesian system for student modeling. AI Commun. 13(4), pp. 277-278.

[20] Castillo, Gladys; Descalço, L.; Diogo, Sandra; Millán, Eva; Oliveira, Paula; Anjo, Batel (2010). Computerized evaluation and diagnosis of student´s knowledge based on Bayesian Networks, Sustaining Tel: From Innovation to Learning and Practice, Lecture Notes in Computer Science, Volume 6383/2010, pp. 494-49.

[21] Castillo, Gladys; Millán, Eva; Descalço, L.; Oliveira, Paula; Diogo, Sandra (2013). Using bayesian networks to improve knowledge assessment. Computers & Education 60(1), pp. 436-447.

[22] Jensen, Finn V.; Nielsen, Thomas D. (2007). Bayesian Networks and Decision Graphs. Information Science and Statistics. Springer.

[23] Koski, Timo; Noble, John M. (2009). Bayesian Networks, An Introduction. Wiley Series in Probability and Statistics. Wiley.

[24] Druzdzel, Marek J. (1999) SMILE: Structural Modeling, Inference, and Learning Engine and GeNIe: A development environment for graphical decision-theoretic models (Intelligent Systems Demonstration). In Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99), pp. 902-903, AAAI Press/The MIT Press, Menlo Park, CA.

[25] Genie & Smile (2015), Decision Systems Laboratory of the University of Pittsburgh. https://dslpitt.org/genie/.