Using Problem Statement Parameters and Ranking Solution Difficulty to Support Personalization

Rômulo C. Silva[1,2], Alexandre I. Direne[2], and Diego Marczal[3]

[1] Western University of Paraná (UNIOESTE)
[2] Federal University of Paraná
[3] Federal Technological University of Paraná
romulocesarsilva@gmail.com
alexd@inf.ufpr.br
dmarczal@gmail.com

Abstract. The work approaches theoretical and implementation issues of a framework aimed at supporting human knowledge acquisition of mathematical concepts. We argue that personalization support can be achieved from problem statement parameters, defined/set during the creation of Learning Objects (LOs) and integrated with the skill level of learners and problem solution difficulty. The last two are formally defined here as algebraic expressions based on fundamental principles derived from extensive consultations with experts in pedagogy and cognition. Our implemented prototype framework, called ADAPTFARMA, includes a collaborative authoring and learning environment that allows short- and long-term interactions. We present our ongoing research about student modeling to support personalization. Finally, we draw conclusions about the suitability of the claims and briefly direct the reader's attention to future research.

Keywords: rating, problem difficulty calibration, Intelligent Tutoring Systems

1 Introduction

The personalization in computer-based learning systems can range from simple student preferences to motivational state detection. Besides, the system is expected to adapt to specific learning needs, including different assessment mechanisms. In Algebra, the student's expertise is usually developed by solving problems that require a set of assessed skills. This is done in both conventional education schools and by applying advanced learning technologies, such as Intelligent Tutoring Systems (ITS). Normally, human teachers detect students' misconceptions when marking tests and exercises. Depending on how much the answer of a question departs from its correct version, two students that missed the same question could be scored different grades for that specific question.

Another aspect that can be used to compose the score is how difficult the question is. The difficulty degree of an exercise can be measured by the number of students that have skipped or made a mistake in that exercise. Thus, a student who finds the correct answer of a question that many missed, probably has more skills than others and the score should reflect that. Conversely, a student who makes a mistake in a question that many were successful to answer, might possess fewer skills. A student error can basically be used as a guideline for two actions: simply to assess the student or to detect misconceptions towards a more effective pedagogical practice. In the latter sense, recently, there has been increasing interest for direct use of errors as a source of teaching material, in order to learn more deeply about the content of the domain and thus develop metacognitive skills [5].

Another desirable aspect in ITS is in predicting or prospecting whether a learner will be able to answer a question correctly or not before it is actually showed to him or her, allowing a more effective personalization support. Usually this kind of feature requires that questions be previously calibrated according to their difficulty and matched to the assessed student's skills.

2 Literature review

Segedy *et al.* [11] propose a taxonomy for adaptive scaffolding in computer-based learning environments, named Suggest-Assert-Modify (SAM). Suggestion scaffolds provide information to learners for the purpose of prompting them to engage in a specific behaviour. Assertion scaffolds communicate information to learners as being true that will be integrated with their current understanding. Modification scaffolds change aspects of the learning task itself.

A manner to support personalization is by implementing algorithms that generate different content sequencing according to a learner's needs. In this sense, Champaign and Cohen propose an algorithm [1] for content sequencing that selects the appropriate learning object to present to a student, based on previous learning experiences of like-minded users. The granularity of sequencing is on the LO level, not exercises or issues. Segal *et al.* [10] propose an algorithm for personalizing educational content in e-learning systems to students. It combines collaborative filtering algorithms with social choice theory. Schatten and Schmidt-Thieme [9] present the Vygotski Policy Sequencer (VPS), based on the concept of Zone of Proximal Development devised by Vygotski. It combines matrix factorization (a method for predicting user rating) with a sequencing policy in order to select at each time step the content according to the predicted score.

Ravi and Sosnovsky [8] propose a calibration method for solution difficulty in ITS based on applying data mining techniques to a student's interaction log. Using the classical bayesian Knowledge Tracing (KT) method [2], the probability that a student has acquired a skill is calculated on the basis of a tentative sequence of exercises for which the soluctions involve a given concept. The logged events are grouped by exercises and classified according to the student's skills.

3 Automatic calculation of rating

Rating systems are frequently used in games to measure the players skills and to rank them. Usually, the rating is a number in a range $[minRank, maxRank]$

such that it is very unlikely that a player falls on the extremes. Inspired by game rating systems and taking the performance of other learners, this study proposes Equation 1 to assess iteratively a student's ability.

The following guidelines were adopted: *(1)* each question is scored a difficulty degree with a Real value in the range [0..10] and the student is rated a number in the range [1..10] to express his or her expertise level in the subject matter; *(2)* the easier the question, the greater the likelyhood that student will answer it correctly (in this case, a student's rating should have just a small increase if he or she enters the correct answer and should have a large decrease in the case of failure); *(3)* students that are successful in the first attempt to solve a question are scored a higher increment in their expertise level compared to those who need several attempts; *(4)* skipped questions are considered wrong.

Consider Equation 1. The details of its parameters are as follows:

$$R_J^q = R_J^{q-1} + Ak_1\alpha(10 - \frac{9T_J^q}{T_{med}^q}) - Ek_2\beta \times 10\frac{T_J^q}{T_{med}^q} \tag{1}$$

- R_J^q: student J's rating after answering question q. $R_J^0 = 5.5$ (initial rating);
- $A = 1$ and $E = 0$ for successful in answering q, otherwise $A = 0$ and $E = 1$;
- T_J^q: number of unsuccessful attempts of student J to answer question q;
- T_{med}^q: median of wrong attempts on question q during classroom time;
- $\alpha = \frac{1}{N_a^q}$ and $\beta = \frac{1}{N_e^q}$ are weight factors to increase and decrease the rating respectively (N_a^q and N_e^q are the number of students that were successful and unsuccessful answering question q, respectively);
- k_1 and k_2: multiplier factors of rating increase and decrease, respectively, calculated by $k_1 = 1 - \frac{R_J^{q-1}}{10}$ and $k_2 = \frac{R_J^{q-1}-1}{10}$.

Although there is no limit to the number of attempts a student can make to answer a question, for calculation purposes, 10 trials is considered the maximum. Factors k_1 and k_2 avoid results of the expression in Equation 1 to reach upper and lower bounds of the range [1..10].

Using only the number of attempts, the difficulty degree of a question q can be defined by Equation 2 and its parameters are as follows:

$$D^q = \frac{\sum_{J=0}^{J=n} T_J^q}{N_e^q + N_a^q} \tag{2}$$

- D^q: difficulty degree of the question q after an exercise session;
- T_J^q: number of unsuccessful attempts of student J to answer question q. If the number of attempts is greater than 10 trials, then 10 is taken as T_J^q;
- N_e^q and N_a^q are the same as in Equation 1

4 The ADAPTFARMA environment

The ADAPTFARMA (Adaptive Authoring Tool for Remediation of errors with Mobile Learning) prototype software tool is a modified version of FARMA [6], an authoring shell for building mathematical learning objects. In ADAPTFARMA, a learning object (LO) consists of a sequence of exercises following their introductory concepts.The introduction is the theoretical part of a LO where concepts are defined through text, images, sounds and videos. The implementation was carried out aiming at its use on the web, either through personal computers or mobile devices.

For each question, the teacher-author must set a reference solution, which is the correct response to the question. ADAPTFARMA allows arithmetic and algebraic expressions to be entered as the reference solution. Under the learner's functioning mode, the tool deals automatically with the equivalence between the learners response and the reference solution.

An important feature of ADAPTFARMA is the capability of backtracking the teacher to the exact context in which the learner made a mistake. It allows the teacher to view a learner's complete interaction with the tool in the chronological order by means of a graphical timeline. In addition, he/she can perform a closer monitoring of problem solutions from other classroom students, as long as system permission is given through the collaboration mechanisms. Likewise, learners can backtrack to the context of any of their right or wrong answers in order to reflect about their own solution steps and find new solution hypotheses. Additionally, on the collaborative side, it is possible for the teacher to carry out a review of students' responses and then provide them with non-automatic feedback, which can be done by exchanging remote messages through the system.

5 Algorithm for exercises sequencing

The ADAPTFARMA environment was designed such that different pedagogical strategies can be used and tested. In this study, we propose an algorithm for sequencing exercises, named Adaptive Sequencing Method (ASM), to be shown in ascending order of difficulty, combined with a mechanism similar to numerical interpolation. We carried out an experiment with 149 highschool students, aging fifteen to seventeen, including pre- and post-tests. The results demonstrate that there has been a significant increase between pre- and post-test scores of students that were subject to ASM (p-value = 0.0037). However, there has been no significant difference in student score gains between ASM-determined and teacher-defined sequencing methods.

A minimal sequence of exercises is defined such that it always begins with the easiest exercise and finishes with the most difficult one. The intermediate-level exercises in the minimal sequence are distributed evenly among the easiest and most difficult exercises such that the number of exercises is $\left\lceil \frac{n}{stepsize} \right\rceil$, where n is the total of exercises and the *stepsize*, set by the LO's author, refers to the number of exercises that may be skipped when the student is successful.

Initially, the algorithm presents the exercises in the minimal sequence order. If the number of attempts in an exercise reaches the average number of attempts obtained in the calibration phase,the next exercise presented to the student is of a mid range difficulty, considering the last exercise correctly answered and the current one. Unlike the calibration phase, the student cannot skip exercises and if he/she continually misses the correct answer, the presentation becomes strictly sequential.

6 Ongoing Research

Our ongoing research related to student modeling, including the learner interaction and context, is based on problem statement parameters and ranking solution difficulty in order to support personalization. During the creation of the LO, the teacher-author sets certain parameters that affect the pedagogical strategy, as follow:

- maximum number of retries (attempts) per question;
- tips for each question;
- remediation rules for each question;
- prerequisites for the solution of the exercise, that can be topics, theoretical pages of the LO itself or other LOs in ADAPTFARMA;
- difficulty degree for each question in the range $[1-10]$, such that $[1-2]$ means very easy, $[3-4]$ means easy, $[5-6]$ means medium, $[7-8]$ means difficult and $[9-10]$ means very difficult;
- exercises sequencing strategy, that can be difficulty-biased, teacher-defined or ASM-determined (presented in the previous section).

The student profile is assembled from the previous parameters. By analysing the tips used and relating them to associated prerequisites, the system can provide feedback to both teacher and student on topics that should be further explored or even recommend other complete LOs to be inspected. In addition, the difficulty degree of the questions and the student rating can be updated after each problem solving session has finished.

7 Conclusion and Future Work

The personalization support in learning systems can include adaptive mechanisms of assessment and generation of different content sequencing. We proposed an automatic rating system that can be used as an additional tool to assess students. Depending on the number of attempts and the difficulty degree of a question, different students can get different scores for the same solution. Also, we proposed an algorithm for sequencing exercises using a formalization of the intuitive notion of difficulty degree combined with a mechanism similar to numerical interpolation. All that was implemented in the ADAPTFARMA environment, a web authoring tool for creating and executing LOs.

Future research concentrates in adding new features to ADAPTFARMA in two ways. Firstly, we are working in a deeper approach to user adaptation that includes more dimensions than just the matching between problem difficulty and student skill. One such new feature will be a function for generating problem statements based on teacher-defined problem template parameters as in [4] and [3]. Secondly, on the interface side, more interaction modes will be available to improve collaboration tasks for monitoring student performance progress.

References

1. John Champaign and Robin Cohen. A Model for Content Sequencing in Intelligent Tutorign Systems Based on the Ecological Approach and Its Validation Through Simulated Students. pages 486–491. Association for the Advancement of Artificial Intelligence (AAAI), 2010.

2. Albert T. Corbett and John R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4):253–278, 1994.
3. R.M. Garcia Rioja, S. Gutierrez Santos, A. Pardo, and C.D. Kloos. A parametric exercise base tutoring system. In *Frontiers in Education, 2003. FIE 2003 33rd Annual*, volume 3, pages S1B_20–S1B_26, Nov 2003.
4. Julio Guerra, Shaghayegh Sahebi, Peter Brusilovsky, and Yu-Ru Lin. The Problem Solving Genome: Analyzing Sequential Patterns of Student Work with Parameterized Exercises. In Pardos Z. Mavrikis M. McLaren B.M. Stamper, J., editor, *Proceedings of the 7th International Conference on Educational Data Mining*, pages 153–160, 2014.
5. Seiji Isotani, Deanne Adams, Richard E. Mayer, Kelley Durkin, Bethany Rittle-Hohnson, and Bruce M. McLaren. Can Erroneous Examples Help Middle-School Students Learn Decimals? volume 6964 of *Lecture Notes in Computer Science*, pages 181–195, Palermo, 2011. Springer Berlin Heidelberg.
6. Diego Marczal and Alexandre Direne. Farma: Uma ferramenta de autoria para objetos de aprendizagem de conceitos matemáticos. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 23, 2012.
7. Niels Pinkwart and Frank Loll. Comparing three approaches to assess the quality of students' solutions. In Darina Dicheva, Riichiro Mizoguchi, and Niels Pinkwart, editors, *AIED 2009 Workshops Proceedings Volume 2, SWEL'09: Ontologies and Social Semantic Web for Intelligent Educational Systems Intelligent Educational Games*, pages 81–85, Jul 2009.
8. Gautham Adithya Ravi and Sergey Sosnovsky. Exercise difficulty Calibration Based on Student Log Mining. In F. Mdritscher, V. Luengo, E. Lai-Chong Law, and U. Hoppe, editors, *Proceedings of DAILE'13: Workshop on Data Analysis and Interpretation for Learning Environments*, Villard-de-Lans (France), Janeiro 2013.
9. Carlotta Schatten and Lars Schmidt-Thieme. Adaptive Content Sequencing without Domain Information. *6th International Conference on Computer based Education*, April 2014.
10. Avi Segal, Ziv Katzir, Kobi Gal, Guy Shani, and Bracha Shapira. EduRank: A Collaborative Filtering Approach to Personalization in E-learning. In Pardos Z. Mavrikis M. McLaren B.M. Stamper, J., editor, *Proceedings of the 7th International Conference on Educational Data Mining*, pages 68–75, 2014.
11. James R. Segedy, Kirk M. Loretz, and Gautam Biswas. Suggest-assert-modify: A taxonomy of adaptive scaffolds in computer-based learning environments. In Gautam Biswas, Roger Azevendo, Valerie Shute, and Susan Bull, editors, *AIED 2013 Workshops Proceedings Volume 2: Scaffolding in Open-Ended Learning Environments (OELEs)*, pages 73–80, Jul 2013.