Introduction to student modeling and Bayesian Knowledge Tracing

Francesc Martori
IQS Universitat Ramon Llull
ASISTEMBE
Barcelona, Spain
francesc.martori@iqs.edu

Jordi Cuadros
IQS Universitat Ramon Llull
ASISTEMBE
Barcelona, Spain
jordi.cuadros@iqs.edu

Lucinio González-Sabaté
IQS Universitat Ramon Llull
ASISTEMBE
Barcelona, Spain
lucinio.gonzalez@iqs.edu

***Abstract* — The use of integrated intelligent or cognitive tutoring systems that support the learning process of our students requires tools that allow these tutors to make decisions based on the students' state of knowledge. Therefore it is very important to have tools to model the students' learning process as student modeling is one of the key factors that affect automated tutoring systems in making instructional decisions. This communication is an introduction to the techniques and methods for student modeling, being Bayesian Knowledge Tracing (BKT) the most well-known among them.**

Keywords* — *student modeling, BKT, tutors, knowledge component

I. Student Modeling

In the last years, adaptive educational software has been entering into classical learning environments, especially in the United States [1]. This software is intended to detect and fit to individual differences in student knowledge, engagement, and motivation, in order to choose the curricular materials and methods of presentation best suited for each student. This adaptation needs in turn an accurate assessment of, at least, the student's knowledge. The area of study covering the set of tools and techniques to achieve this assessment is known as student modeling [2].

A student model is the base for personalization in computer based educational applications. Self [3] pointed out that student modeling is a process devoted to represent several cognitive issues such as analyzing the student's performance, isolating the underlying misconceptions, representing students' goals and plans, identifying prior and acquired knowledge, maintaining an episodic memory, and describing personality characteristics.

From all different aspects a student model can focus on, knowledge is the most intriguing construct. A variety of mechanisms were invented for the purpose of estimating student knowledge, such as tests, quizzes and exams among all others. The underlying idea is that the best estimation of student knowledge is obtainable by observing student performance [4], but while student performance is observable, student knowledge remains latent. Bayesian Networks are a common tool to address this latency and Bayesian Knowledge Tracing (BKT) is the most important application of them so far.

II. Bayesian Knowledge Tracing

BKT [5] is a student model used to infer a student's knowledge given their history of responses to problems, which it can use to predict future performance. Using students' responses to questions, which are tagged with the skills that the instructor wants the students to learn, the model tells the probability a student has mastered a skill.

A skill or knowledge component is a description of a mental structure or process that a learner uses, alone or in combination with other knowledge components, to accomplish steps in a task or a problem. A full description and taxonomy of knowledge components can be found in [6].

BKT is a two state Hidden Markov Model, these states being the one in which the student knows a given skill, and the one where the student does not. The "knowledge" state is absorbent, implying that the student will not forget the skill once it is learned. To calculate the probability that a student knows the skill given their performance history, BKT uses four probabilities:

- L_0, the probability a student knows the skill before attempting the first problem,
- T, the probability a student, who does not currently know the skill, will know it after the next practice opportunity, that is the transition probability at each practice opportunity,
- G, the probability a student will answer a question correctly despite not knowing the skill,
- S, the probability a student will answer a question incorrectly despite knowing the skill.

According to this model, knowledge affects performance (mediated by the guess and slip rates), and knowledge at one time step affects knowledge at the next time step, but no further.

Then, if a student is in the "no knowledge" state at time *t*, then the probability he will be in the "knowledge" state at time *t*+1 is *T*. The probability that a student has mastered a skill can be calculated using (1), (2) and (3).

$$P(L_{t-1}|Correct_t) = \frac{P(L_{t-1}) \cdot (1-S)}{P(L_{t-1}) \cdot (1-S) + (1-P(L_{t-1})) \cdot G} \quad (1)$$

$$P(L_{t-1}|Incorrect_t) = \frac{P(L_{t-1}) \cdot S}{P(L_{t-1}) \cdot S + (1-P(L_{t-1})) \cdot (1-G)} \quad (2)$$

$$P(L_t) = P(L_{t-1}|Action_t) + (1 - P(L_{t-1}|Action_t)) \cdot T \quad (3)$$

Usually, a separate BKT model is fit for each skill and only the first attempt at each question is taken for each student, as it is the attempt containing the most information about the student's knowledge.

This probability can be used to estimate the likelihood that a given student will provide a correct answer in any future attempt, as shown in (4)

$$C_{i,t+1} = P(L_t) \cdot (1 - S) + (1 - P(L_t)) \cdot G \quad (4)$$

III. Student Modeling Aplications

Student modeling is nowadays being used in both research and real learning contexts. Some of the most relevant learning contexts where student modeling is being applied are the following:

- ASSISTments: (https://www.assistments.org/) This intelligent tutor developed by the Worcester Polytechnic Institute is used by more than 600 teachers from 42 American states and 14 countries and their students solved 10^6 problems in 2015.
- LearnLab and PSLC Datashop: (http://learnlab.org/) It is managed by Carnegie Mellon University and University of Pittsburgh. Funded by the National Science Foundation, they leverage cognitive theory and computational modeling to identify the instructional conditions that cause robust student learning.
- Open Learning Initiative (OLI): (http://oli.cmu.edu/) (http://oli.stanford.edu/) OLI is a grant-funded organization that offers innovative online courses to anyone who wants to learn or teach. It was founded at Carnegie Mellon University and it landed in Stanford University in 2013.

References

[1] Koedinger, K. R., Corbett, A. T., 2006. Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 61-78). New York: Cambridge University

[2] Vanlehn, K., 1988. Student Modeling. In M.C. Polson & J.J. Richardson (Eds.) Foundations of Intelligent Tutoring Systems. London, UK: Psychology Press

[3] Self, J. A. 1990. Bypassing the intractable problem of student modelling. In C. Frasson & G. Gauthier (Eds.), Intelligent tutoring systems: At the crossroads of AI and education (pp. 107–123). Norwood, NJ: Ablex.

[4] Gong, Y. 2014. Student modeling in Intelligent Tutoring Systems. Worcester Polytechnic Institute.

[5] Corbett, A. T. and Anderson, J. R. 1995. Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction, 4(4), 253-278. http://dx.doi.org/10.1007/BF01099821

[6] Koedinger, K. R., Corbett, A. T. and Perfetti, C. 2012. The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. Cognitive Science, 36, 757 - 798. DOI: 10.1111/j.1551-6709.2012.01245.x