

Student Modeling Applications, Recent Developments & Toolkits [SMART tutorial]

José P. González-Brenes
Pearson
jose.gonzalez-brenes@pearson.com

Michael Yudelson
Carnegie Learning, Inc.
myudelson@carnegielearning.com

Kai-min Chang
Carnegie Mellon University
kkchang@cs.cmu.edu

Yoav Bergner
Educational Testing Services
ybergner@ets.org

Yun Huang
University of Pittsburgh
yuh43@pitt.edu

ABSTRACT

This tutorial is intended for researchers, practitioners and students interested in student modelling. We will cover both the theoretical foundations of student models, and a hands-on approach of how to use existing state-of-the-art toolkits.

LENGTH

Full day

INTENDED OUTCOMES

Participants will be able to understand and apply the selected theory and toolkits for student modeling.

INTENDED AUDIENCE

30 participants with a mix of expertise, and backgrounds (academic, industry).

POTENTIAL FUNDING

Presenters will be either self-funded or funded by their host institution.

DESCRIPTION AND JUSTIFICATION

The educational data mining community is starting to fulfill the promise of using data to improve education. The advancement of the field requires the community to be aware of existing tools and results from student modeling. But with a myriad of student modeling techniques and toolkits available, it is easy to be overwhelmed.

In this tutorial we will cover popular and promising toolkits – and the theory behind them. We will demystify the acronym soup in the educational data mining field (BKT, IRT, 1PL, etc). We will help practitioners and researchers alike to get up to speed on student modeling using the latest technology. We are fortunate enough that the toolkits will be presented by the authors who developed them.

MAIN TOPICS

BNT-SM toolkit. We will cover the foundation of Dynamic Bayes Nets (DBNs), which provide a powerful way to represent and reason about uncertainty in time series data, and are therefore well-suited to model a student's changing knowledge state during skill acquisition. Many general-purpose Bayes net packages have been implemented and distributed; however, constructing DBNs often involves complicated coding effort. To address this issue, we teach how to use a popular extension of the Bayes Net Toolbox (Murphy, 1998) called BNT-SM. BNT-SM is a toolkit designed for the student modeling community. BNT-SM inputs a data set and a compact XML specification of a Bayes net model hypothesized by a researcher to describe causal relationships among student knowledge and observed behavior. BNT-SM generates and executes the code to train and test the model using the Bayes Net Toolbox. BNT-SM allows researchers to easily explore different hypothesis with respect to the knowledge representation in a student model.

FAST (Knowledge Tracing with Features) toolkit. Bayesian Knowledge Tracing (BKT) is an extremely popular method for student modeling. Unfortunately, it does not allow modeling the feature-rich data that is now possible to collect in modern digital learning environments. Because of this, many ad hoc Knowledge Tracing variants have been proposed to model a specific feature of interest. For example, variants have studied the effect of students' individual characteristics, the effect of help in a tutor, and subskills. These ad hoc models are successful for their own specific purpose, but are specified to only model a single specific feature. We present the FAST (Feature Aware Student knowledge Tracing) toolkit, an efficient, novel method that allows integrating general features into Knowledge Tracing. FAST is available online at <http://ml-smores.github.io/fast/>

Bayesian Knowledge Tracing at Scale toolkit. Bayesian Knowledge Tracing (BKT) is an extremely popular method for student modeling. We will present the Toolkit for Bayesian Knowledge Tracing at scale (codenamed *hmmsclbl*). Originally targeted at being used with KDD Cup 2010 data donated by Carnegie Learning, Inc. (millions of records), it has been successfully tested with datasets in the hundred of million records range. *Hmmsclbl* accepts a simple data format file with four columns – success, student, step, skill(s) and can fit and cross-validate models using a number of algorithms. In addition to standard BKT model, *hmmsclbl* is capable of fitting individualized models with factors accounting for student variance. Individualization is done without changing the structure of the model by modifying how the objective function is computed. Open version of the toolkit that fits standard BKT models is available for downloading at <https://github.com/IEDMS/standard-bkt>. The tutorial will cover the public version of the *hmmsclbl* (standard BKT), currently private individualized BKT, as well as several experimental settings, including: estimating using differentiated evolution and context-based parameter multiplexing.

CFIRT toolkit: Collaborative Filtering in the Style of Item Response Theory. We will introduce classic psychometric theory approaches called Item Response Theory. We will present a machine learning approach to multidimensional item response theory (MIRT), showing how collaborative filtering leads to a general class of models that includes many MIRT models. We motivate a class of models which contains many well known psychometric IRT models (e.g. the 1PL, 2PL, M2PL) as well as a large number of models not previously named. The focus will be on compensatory models, although this is not an intrinsic limitation of the method. We will explain how the model space can be systematically searched by ordering the candidate models in terms of their parametric complexity.

PANEL

José P. González-Brenes is a Research Scientist at Pearson. José is the first prize winner of an international data mining competition against 350+ teams for predicting time travel in highways. José likes to study applications of machine learning to education, for example, he has published in both machine learning venues (AISTATS, NIPS) and educational data mining venues (EDM). José received a PhD and a masters degree in Language Technologies from Carnegie Mellon University (USA), an IMBA from National Tsing Hua University (Taiwan) and a BSc in Computer Science from Instituto Tecnológico de Costa Rica (Costa Rica).

Michael Yudelson is a Research Scientist at Carnegie Learning, Inc. His research focuses on educational data mining, big data analytics, knowledge transfer, hierarchical linear models, and hidden markov models. Michael got his Ph.D. from the University of Pittsburgh in 2010 under the supervision of Peter L. Brusilovsky. Michael's doctoral research focused on investigating scalability of user-adaptive hypermedia systems. Early results of this work were awarded the best student paper at the User Modeling 2007 conference. Michael was a postdoctoral fellow at the Human-Computer Interaction Institute of Carnegie Mellon University in 2010-2013. There he worked with Kenneth R. Koedinger and Geoffrey J. Gordon on knowledge transfer and big data analytics using Hidden Markov Models.

Kai-min Chang is a Research Associate at the School of Computer Science in Carnegie Mellon University. His main topic of research is to understand the neural underpinning of semantic knowledge, and knowledge representation in the context of an Intelligent Tutoring System. He has disseminated various machine learning and student modeling toolkits (Yuan et al., 2014; Chang et al, 2006) and coorganized various international workshops on machine learning and brain imaging at leading conferences includes NIPS, ITS, and NAACL.

Yoav Bergner is a Research Scientist in the Center for Advanced Psychometrics at Educational Testing Service. Yoav earned a bachelor's degree in physics from Harvard University and a PhD in theoretical physics from MIT before turning to the applied side and becoming a sculptor and furniture maker for five years. He then turned from shaping the material world to shaping young minds as a NYC public school science and math teacher for the next three years. In the classroom, Yoav developed a pragmatic interest in the potential of digital environments for personalized learning and assessment, so he returned to the research fold, first at the Research on Learning, Assessing and Tutoring Effectively (RELATE) Lab at MIT (just in time for the MOOC explosion!) and now at ETS. Yoav's work bridges educational data mining and psychometric approaches to modeling process data, including online courses and simulation-based tasks, with particular interests in multidimensionality issues, model fit, and collaborative learning and assessment.

Yun Huang is a PhD student in the Intelligent Systems Program, University of Pittsburgh. Her research focus is on applying machine learning techniques to understand student learning. Her advisor is Dr. Peter Brusilovsky, and many of her projects also received advice from Dr. Jose P. González-Brenes. She has been working on online Java Programming Tutoring systems, tackling modeling challenges for complex programming skills, parameterized programming exercises, and cross-content learning environment. One of the model that addresses these challenges is Feature-Aware Student Knowledge Tracing, a general model that allows arbitrary features in Knowledge Tracing, which she developed with Dr. González-Brenes. Their work was nominated for Best Paper Award in 2014 Educational Data Mining Conference. Yun is working towards applying probabilistic methods to construct skill and student models that can take insights from both cognitive theories and student data.

PROPOSED TIMELINE

| | |
|---------------|---|
| 09:00 - 09:45 | Introduction to student modeling, bayesian networks, IRT, Knowledge Tracing, evaluation (TBA) |
| 09:45 - 10:00 | <i>coffee break</i> |
| 10:00 - 11:00 | BNT-SM toolkit for student modeling & brain interfaces (Chang & Xu) |
| 11:00 - 12:00 | FAST framework & toolkit (González-Brenes & Huang) |
| 12:00 - 02:00 | <i>lunch</i> |
| 02:00 - 03:00 | Knowledge tracing at scale (Yudelson) |
| 03:00 - 03:15 | <i>coffee break</i> |
| 03:15 - 04:15 | Collaborative filtering & Item Response Theory (Bergner) |
| 05:15 - 05:30 | Concluding remarks |