Abstract:

Recent interest in online education, such as Massively Open Online Courses and Intelligent Tutoring Systems, allows collecting large amounts of data from students solving items at different levels of proficiency over time. Existing approaches for inferring students’ knowledge from data require a cognitive model – a mapping between the tutor problems and the set of skills they require. This is a very expensive requirement, since it depends on expert domain knowledge. The success of previous methods in using student performance data to construct this mapping automatically has been limited in that they cannot handle data collected over time, or that they require expensive expert domain knowledge. This dissertation studies how to model students’ time varying knowledge, without requiring such knowledge. We introduce four novel methods:

Dynamic Cognitive Tracing: an easily implemented prototype that jointly estimates cognitive and student models.

Automatic Knowledge Tracing: a method to discover first a mapping of items to skills and then a student model separately.

Topical Hidden Markov Model: a fully Bayesian method to discover both cognitive and student models jointly.

ItemSim: a method to bias the estimation of Topical Hidden Markov Model to discover more interpretable cognitive models by grouping similar items together.

We evaluate our methods on synthetic and real student data collected from students interacting with two commercial tutoring systems. Our automatic methods require much less human work and can achieve significantly higher accuracy at predicting future student performance than the models handcrafted by experts.