Abstract

Adaptive systems teach and adapt to humans; their promise is to improve education by minimizing the subset of items presented to students while maximizing student outcomes (Cen et al., 2007). In this context, items are questions, problems, or tasks that can be graded individually. The adaptive tutoring community has tacitly adopted conventions for evaluating tutoring systems (Dhanani et al., 2014) by using classification evaluation metrics that assess the student model component—student models are the subsystems that forecast whether a learner will answer the next item correctly.

Unfortunately, it is not clear how intuitive classification metrics are for practitioners with little machine learning background. Moreover, our experiments on real and synthetic data reveal that it is possible to have student models that are very predictive (as measured by traditional classification metrics), yet provide little to no value to the learner. Additionally, when we compare alternative tutoring systems with classification metrics, we discover that they may favor tutoring systems that require higher student effort with no evidence that students are learning more. That is, when comparing two alternative systems, classification metrics may prefer a suboptimal system.

We recently proposed Learner Effort-Outcomes Paradigm (Leopard) for automatic evaluation of adaptive tutoring (González-Brenes & Huang, 2015). Leopard extends on prior work on alternatives to classification evaluation metrics (Lee & Brunskill, 2012). At its core, Leopard quantifies both the effort and outcomes of students in adaptive tutoring. Even though these metrics are novel by itself, our contribution is approximating both without a randomized control trial.

In this talk, we will describe our recently published results on meta-evaluating Leopard and conventional classification metrics. Additionally, we will present preliminary results of framing the value of an educational intervention as multi-objective programming. We argue that human-propelled machine learning, and educational technology in particular, aim to optimize the Pareto boundary of effort and outcomes of humans.

References


González-Brenes, José P. and Huang, Yun. Your model is predictive— but is it useful? theoretical and empirical considerations of a new paradigm for adaptive tutor-