If an Algorithm Is Openly Accessible, and No One Can Understand It, Is It Actually Open?

Iris Howley, Williams College, iris@cs.williams.edu

Abstract

As blackbox algorithms play an increasing role in classroom decision-making, calls to "open" these algorithms and explain the inputs and latent variables that determine the decision outcomes grow increasingly louder. However, even systems using open algorithms face similar concerns. Using open algorithms does not mean that stakeholders (e.g., instructors, students, parents, etc.) of the system understand the connection between features in the underlying machine learning model and the outcomes displayed to them. Some algorithms (i.e., multi-level neural networks) are too complex to easily interrogate the decision-making process, but other algorithms (e.g., Bayesian Knowledge Tracing (BKT), classification, etc.) are considerably more comprehendible, but teachers and students still do not understand them. This work asks, what are the ethical implications of providing students and teachers with algorithmic decision-making software they *could* interrogate, but due to lack of knowledge, *cannot* interrogate and how might researchers help bridge that gap?

First steps of this work focus on OARS, an assessment and learning system described in Bassen et al. (2018), that uses BKT to predict student skill mastery. BKT is built on a Hidden Markov Model where student skill mastery is output as either "mastered" or "unmastered." Parameters in this model include probabilities related to existing mastery, slipping (i.e., forgetting), learning, and guessing (Bassen et al., 2018). In relation to deep learning approaches, BKT is rather simple to interrogate for relationships between parameters, however, just because it is possible to understand, does not mean the students and instructors who use BKT-based systems do understand it.

I performed eight semi-structured interviews with instructors discussing their use of educational technology tools in the classroom. 6 of the interviewees used OARS previously, and so could speak directly to their understanding of the underlying BKT model, and how (or why) they trust the skill predictions it produces. In general, interviewees trusted the OARS output without a firm grasp on the underlying decision-making, such as P7, a professor at a community college, who said, "I have not thought much about the specific algorithm because I tend on being more trusting in the algorithm than not."

From initial interviews on this topic, it is apparent that the space of interrogating AIED algorithmic decisions is a multi-dimensional issue. Along the openness algorithm dimension, there are algorithms too difficult to interrogate, algorithms that are complex but explainable, and then there are blackbox commercial algorithms. Along the user motivation dimension, there are users who want to understand the algorithm and have the prior knowledge to do so, users who want to understand but lack the requisite prior knowledge, and users who are less motivated to understand algorithmic processes. User motivation also intersects with system trust, although exactly how is not yet known. Investigation into these issues is necessary to understand when and how intervention should be provided to users of algorithmically enhanced learning (AEL) environments.

As artificial intelligence in education researchers, we must ask, what are the ethical implications of deploying systems that our instructors and students, cannot understand, but could if provided with the appropriate scaffolding? If students and instructors are using our systems to change their approach to learning, should they not first understand, when possible, the outputs their AEL environments are producing? My current project begins to examine how to scaffold the learning of the algorithmic decision-making process for instructors, and how this relates to trust in the algorithmically enhanced learning environment.

References

Bassen, J., Howley, I., Fast, E., Mitchell, J., & Thille, C. (2018, June). OARS: Exploring instructor analytics for online learning. In *Proceedings of the Fifth (2018) ACM Conference on Learning@Scale*.

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. In *Information, Communication, & Society*, *20*(1), 30-44.