Been Kim

BIO: Been Kim is a Research Scientist at AI2 and an affiliated faculty in the Department of Computer Science & Engineering at the University of Washington. Her research focuses on interactive and interpretable machine learning models for human-machine collaboration. She received her PhD. from MIT. Prior to her PhD, she worked at the MathWorks as a software engineer.

ABSTRACT: I envision a system that enables successful collaborations between humans and machine learning models by harnessing the relative strength to accomplish what neither can do alone. Machine learning techniques and humans have skills that complement each other --- machine learning techniques are good at computation on data at the lowest level of granularity, whereas people are better at abstracting knowledge from their experience, and transferring the knowledge across domains. The goal of my research is to develop a framework for human-in-the-loop machine learning that enables people to interact effectively with machine learning models to make better decisions using large datasets, without requiring in-depth knowledge about machine learning techniques. In this talk, I present the Bayesian Case Model (BCM), a general framework for Bayesian case-based reasoning (CBR) and prototype classification and clustering. BCM brings the intuitive power of CBR to a Bayesian generative framework. The BCM learns prototypes, the "quintessential" observations that best represent clusters in a dataset, by performing joint inference on cluster labels, prototypes and important features. Simultaneously, BCM pursues sparsity by learning subspaces, the sets of features that play important roles in the characterization of the prototypes. The prototype and subspace representation provides quantitative benefits in interpretability while preserving classification accuracy. Human subject experiments verify statistically significant improvements to participants' understanding when using explanations produced by BCM, compared to those given by prior art. I demonstrate the application of this model for an educational domain in which teachers cluster programming assignments to streamline the grading process.