Holcombe Department of Electrical and Computer Engineering Seminar Series

Achieving Causal Fairness in Machine Learning

Yongkai Wu

Ph.D. candidate, Dept. of Computer Science & Computer Engineering, University of Arkansas

Abstract

Fairness in AI systems is receiving increasing attention. Fairness-aware machine learning studies the problem of building machine learning models that are subject to fairness requirements. In this talk, I will present my dissertation research on developing a causality-based framework for measuring discrimination and achieving fairness in classification. In our research, we formulate discrimination based on the causal inference framework where the causal effect is measured from a causal graph and observed data. We propose a unified definition that covers most of previous causality-based fairness notions, namely the path-specific counterfactual fairness (PC fairness). We target an inherent challenge in causal inference, unidentification, which means some causal quantities cannot be uniquely computed from observed data. To overcome this challenge, we propose novel estimation methods to bound the unidentifiable fairness quantities. Then we develop an efficient post-processing method to achieve fairness in unidentifiable counterfactual cases. At the end of my talk, I will briefly introduce my other works dealing with discrimination issues in various machine learning tasks and applications. I will also discuss future research directions on fairness-aware machine learning and FATE (Fairness, Accountability, Transparency, and Explainability) in AI.

Biography of Speaker

Yongkai Wu is a Ph.D. candidate in the Department of Computer Science and Computer Engineering at the University of Arkansas. He received his B.Eng. degree in Electronic Engineering from Tsinghua University, China in 2014. His research interests focus on machine learning, data mining, and artificial intelligence, particularly fairness-aware machine learning and causal inference. His publications have appeared in prestigious conferences including IJCAI, KDD, NeurIPS, WWW, and a premier journal TKDE. He has served as a PC member for several international conferences including AAAI, IJCAI, KDD, PAKDD.