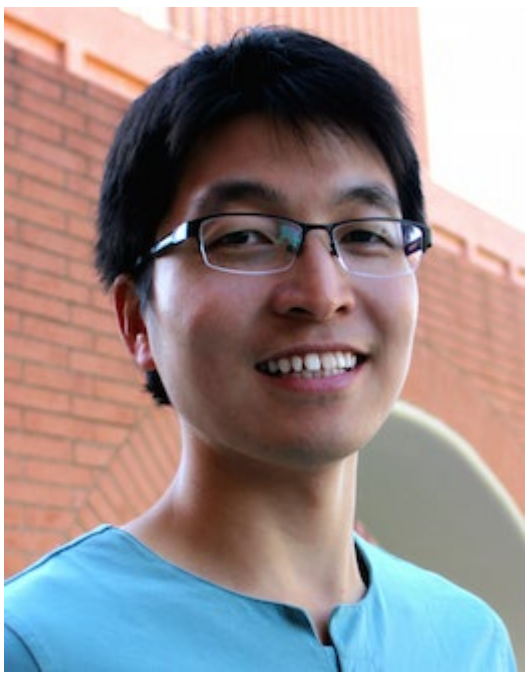
isr PRESENTS

# Haifeng Xu

## Manipulating Learning Algorithms in Strategic Environments



Haifeng Xu is the Alan Batson Assistant Professor in the Department of Computer Science at the University of Virginia. He works broadly on algorithms, game theory and machine learning, with a particular focus on studying how incentives/information/data affect learning and decision making. Prior to UVA, he was a postdoc at Harvard, hosted by Yiling Chen and David Parkes. Haifeng received his PhD in Computer Science from University of Southern California, advised by Shaddin Dughmi and Milind Tambe (now at Harvard). His research was recognized by several awards, including the honorable mention award for both the ACM SIGecom Dissertation Award and the IFAAMAS Victor Lesser Distinguished Dissertation Award, a Google PhD fellowship, the 2016 AAMAS best student paper award, and the 2016 SecMas Workshop best paper award.

There has been significant amount of recent interests in adversarial attacks to machine learning algorithms, particularly deep learning algorithms. In this talk, we pursue a closely related, yet far less explored, theme alone this research agenda, i.e., strategic attacks to learning algorithms. In particular, we consider settings where the learner faces a strategic agent who manipulates the learning algorithm simply to optimize his own utility, as opposed to completely ruining the learner's algorithm in adversarial ML. Such strategic interactions naturally arise in many decision-focused learning tasks including, e.g., learning to set a price for an unknown buyer and learning to defend against an unknown attacker. We describe a general framework to theoretically analyze the attacker's optimal strategic attack, and then instantiate the framework and analysis in two basic scenarios. Finally, we consider how to defend against such strategic attacks and provide formal barriers to the design of optimal defense for the learner.

## Wednesday, February 5
## 2:00pm to 3:00pm
## 4405 Gates Hillman Center

isr institute for SOFTWARE RESEARCH

**School of Computer Science**
5000 Forbes Avenue, 5212 Wean Hall
Pittsburgh, PA 15213-3890

Carnegie Mellon University