Abstract:

The proliferation of data collection in the health, commercial, and economic spheres, brings with it opportunities for extracting new knowledge with concrete policy implications. Examples include individualizing medical practices based on electronic healthcare records, and understanding the implications of job training programs on employment and income. The scientific challenge lies in the fact that standard prediction models such as supervised machine learning are often not enough for decision making from this so-called "observational data": Supervised learning does not take into account causality, nor does it account for the feedback loops that arise when predictions are turned into actions. On the other hand, existing causal-inference methods are not adapted to dealing with the rich and complex data now available, and often focus on populations, as opposed to individual-level effects.

The problem is most closely related to reinforcement learning and bandit problems in machine learning, but with the important property of having no control over experiments and no direct access to the actor's model.

In my talk I will discuss how we apply recent ideas from machine learning to individual-level causal-inference and action. I will introduce a novel generalization bound for estimating individual-level treatment effect, and further show how we use representation learning and deep temporal generative models to create new algorithms geared towards this problem. Finally, I will show experimental results using data from electronic medical records and data from a job training program.