Abstract: Deep Counterfactual Prediction using Instrumental Variables

Jason Hartford
University of British Columbia
jasonhar@cs.ubc.ca

Greg Lewis
Microsoft Research & NBER
glewis@microsoft.com

Kevin Leyton-Brown
University of British Columbia
kevinlb@cs.ubc.ca

Matt Taddy
Microsoft Research & University of Chicago
taddy@microsoft.com

Supervised machine learning (ML) provides a myriad of effective methods for solving prediction tasks. In these tasks, the learning algorithm is trained and validated to do a good job predicting the outcome for future examples from the same data generating process. However, decision makers (and automated decision systems) look to the data to model the effects of a *policy change*. Such changes imply that the future relationship between inputs and outcomes will be different from what is in the training data. In such cases, ML algorithms will do a poor job of predicting the many potential futures associated with each policy option.

For example, optimal pricing requires predicting sales under changes to prices, a doctor needs to know how a patient will respond to various treatment options, and advertisers want to identify ads that cause sales. To accurately answer such *counterfactual* questions it is necessary to model the structural (or causal) relationship between policy (i.e., treatment) and outcome variables. Randomized control ('AB') trials are the gold standard for establishing causal relationships, but conducting such trials is often impractical or excessively expensive. Observational data, by contrast, is abundant.

The instrumental variables (IV) framework is a general class of methods for using observational data to establish causal relationships. It has a long history, especially in economics [e.g., Wright, 1928, Reiersøl., 1945]. The idea is to use sets of variables that only affect treatment assignment and not the outcome variable—so-called *instruments*—to consistently estimate the causal treatment effect. The framework is most straightforward in the case of an imperfect experiment. Consider a scenario where one of the inputs to treatment assignment has been randomized, but where other influences are potentially endogenous: they are dependent on unobserved variables that influence the outcome. For example, in a medical trial we might have a treatment that is made available to a random sample of patients. However, only a portion of those patients actually take the treatment (perhaps because it causes discomfort). In this scenario, the random availability of treatment is our instrument and an IV analysis is used to infer the causal treatment effect in the face of selective partial adherence.

This paper provides a recipe for combining ML algorithms to solve for causal effects in the presence of instrumental variables. We show that a flexible IV specification resolves into two prediction tasks that can be solved with deep neural nets: a first-stage network for treatment prediction and a second-stage network whose loss function involves integration over the conditional treatment distribution. This *Deep IV* framework imposes some specific structure on the stochastic gradient descent routine used for training, but it is general enough that we can take advantage of off-the-shelf ML capabilities and avoid extensive algorithm customization. We outline how to obtain out-of-sample causal validation in order to avoid over-fit and describe schemes for both Bayesian and frequentist inference. The result is a modular and scalable framework for reliable causal inference from observational data.

References

O. Reiersøl. Confluence analysis by means of instrumental sets of variables. *Arkiv för Matematik, Astronomi och Fysik*, 32a(4):1–119, 1945.

P. G. Wright. *The Tariff on Animal and Vegetable Oils*. Macmillan, 1928.