Facilitating score and causal inference trees for large observational studies. (English)

Zbl 1433.68382

Summary: Assessing treatment effects in observational studies is a multifaceted problem that not only involves heterogeneous mechanisms of how the treatment or cause is exposed to subjects, known as propensity, but also differential causal effects across sub-populations. We introduce a concept termed the facilitating score to account for both the confounding and interacting impacts of covariates on the treatment effect. Several approaches for estimating the facilitating score are discussed. In particular, we put forward a machine learning method, called causal inference tree (CIT), to provide a piecewise constant approximation of the facilitating score. With interpretable rules, CIT splits data in such a way that both the propensity and the treatment effect become more homogeneous within each resultant partition. Causal inference at different levels can be made on the basis of CIT. Together with an aggregated grouping procedure, CIT stratifies data into strata where causal effects can be conveniently assessed within each. Besides, a feasible way of predicting individual causal effects (ICE) is made available by aggregating ensemble CIT models. Both the stratified results and the estimated ICE provide an assessment of heterogeneity of causal effects and can be integrated for estimating the average causal effect (ACE). Mean square consistency of CIT is also established. We evaluate the performance of proposed methods with simulations and illustrate their use with the NSW data in [R. H. Dehejia and S. Wahba, “Causal effects in nonexperimental studies: re-evaluating the evaluation of training programs”, J. Am. Stat. Assoc. 94, No. 448, 1053–1062 (1999; doi:10.1080/01621459.1999.10473858)] where the objective is to assess the impact of a labor training program, the National Supported Work (NSW) demonstration, on post-intervention earnings.

MSC:
- 68T05 Learning and adaptive systems in artificial intelligence
- 62G05 Nonparametric estimation
- 62G10 Nonparametric hypothesis testing
- 62P10 Applications of statistics to biology and medical sciences; meta analysis

Keywords:
- CART: causal inference; confounding; interaction; observational study; personalized medicine; recursive partitioning

Software:
- MatchIt

Full Text: [Link]