Review of: "Causality in Machine Learning: Innovating Model Generalization through Inference of Causal Relationships from Observational Data"

Federico Nutarelli[1]

1 Institute for Advanced Studies Lucca

Potential competing interests: No potential competing interests to declare.

The present article undertakes an effort to consolidate current techniques in causal inference with the aim of extending the applicability of machine learning methodologies. This endeavor is pursued through a comprehensive examination of score-based methodologies and causal graph-based techniques. While the article demonstrates commendable writing, it falls short in elucidating certain acronyms in advance, potentially causing confusion among readers. Furthermore, there is room for improvement in the exposition of specific methodologies, particularly those discussed in the literature review, such as CausalVAE, CausalRNN, and others. In a broader context, considering the article's objective of reviewing causality methodologies, it would be advisable to include an introductory paragraph that clarifies the concepts of Causal Discovery Methods and Causal Representation Learning.

More significantly, I did not come across any references in the article to the works of Athey and Imbens or Chernuzokov, who have introduced a multitude of causal machine learning techniques, including but not limited to Double Debiased ML, Generalized Random Forests, Causal Matrix Completion, and, more in general, they introduced the concept of Heterogeneous treatment effects, among others. It is imperative for the authors to acknowledge this body of literature and provide a clear delineation of how their methodology diverges from the aforementioned stream, along with an elucidation of the enhancements it offers.

Finally, there is a notable absence of a comprehensive technical exposition of their methodology, discussions concerning its asymptotic validity, the construction of confidence intervals, and a more in-depth examination of the results. Beyond a mathematical elucidation of the Data Generating Process (DGP) and the model generated by the causal machine learning algorithms, it is crucial to incorporate conventional performance metrics, such as at least Root Mean Square Error (RMSE), while explaining how enhanced predictive performance contributes to the revelation of causal effects. For instance, in Chernuzokov's work from 2018, a heightened performance of the machine learning algorithm played a pivotal role in uncovering the significance of p-scores.

Overall I would invite the authors to re-submit the article for the mentioned reasons.