2

DoubleMLDeep: Estimation of Causal Effects with Multimodal Data

Martin Spindler[1], Philipp Bach[1], Sven Klaassen[2], Jan Teichert-Kluge[1], Victor Chernozhukov[3]
[1]University of Hamburg, Hamburg, Germany. [2]Economic AI, Hamburg, Germany. [3]MIT, Cambridge, USA

Martin Spindler
Please provide a brief biography for the Presenting author(s)
Martin Spindler is Professor of Data Science, Statistics & Econometrics, at the University of Hamburg. In his research he works on the theory and practical applications of Machine Learning and AI, in particluar Causal Machine Leraning. He is coauthor of the opensource package "DoubleML", one of the leading frameworks in R and Python for causal machine learning. Together with leading international researchers he has founded Economic AI to support companies to develop and apply state-of-the-art solutions with a focus on Causal AI. The methods can be applied for clinical trials and also real world data (RWD).

Philipp Bach
Please provide a brief biography for the Presenting author(s)
--

Sven Klaassen
Please provide a brief biography for the Presenting author(s)
--

Jan Teichert-Kluge
Please provide a brief biography for the Presenting author(s)
--

Victor Chernozhukov
Please provide a brief biography for the Presenting author(s)
--

Single topic, multi-speaker session, Workshop or Single presentation submission

A single presentation/poster

Single presentation or poster submission

This paper explores the use of unstructured, multimodal data, namely text and images, in causal inference and treatment effect estimation. We propose a neural network architecture that is adapted to the double machine learning (DML) framework, specifically the partially linear model. An additional contribution of our paper is a new method to generate a semi-synthetic dataset which can be used to evaluate the performance of causal effect estimation in the presence of text and images as confounders. The proposed methods and architectures are evaluated on the semi-synthetic dataset and compared to standard approaches, highlighting the potential benefit of using text and images directly in causal studies. Our findings have implications for researchers and practitioners in medicine and data science in general who are interested in estimating causal quantities using non-traditional data.