Causal Inference with Video Features as Treatments

Nakamura, Kentaro, Adam Breuer, Michael H. Crespin, Bryce J. Dietrich, and Kosuke Imai

Department of Government, Harvard University, U.S.A.

Abstract

We develop the first statistical methodology for causal inference with video features as treatments. Video is the most engaging content modality on the internet. A central causal question is how audience reactions change in response to treatment features that unfold over the course of a video. Unfortunately, standard causal inference methods are not applicable because confounding features are latent, high-dimensional, and dynamically related to both the treatment sequence and the outcome trajectory. To address these challenges, we first reproduce each video using a deep generative model and leverage the model's internal representations as learned, low-dimensional summaries of video content for causal estimation. We then establish that the average potential-outcome trajectory under dynamic stochastic interventions is nonparametrically identified. Lastly, we propose a consistent and asymptotically normal estimator based on a longitudinal neural network architecture. We empirically validate our approach by constructing a new causal inference benchmark consisting of 10,000 Super Mario Bros. levels played by fixed Mario AI agents, where ground-truth causal effects are known by construction. Finally, we apply our method to television advertisements from the 2020 U.S. presidential campaign and find that increasing the probability of a candidate appearing over time leads to higher average viewer evaluations. With the proposed methodology, researchers can ask which visual features, appearing at which points in a video, influence audience responses, while benchmarking new methods against datasets with known ground-truth causal effects.