Causality in Modern Deep Generative Models

Background

Psychology

Prior expectations affect counterfactual and causal judgements

People often imagine counterfactual changes that would undo a negative outcome, thinking things like “If only he had driven home a little later, he wouldn’t have gotten in a car crash.” Kahneman and Tversky (1981) show that people are sensitive to prior expectations when deciding on an “if only…” counterfactual. They presented stories where they varied whether a character Mr. Jones went home at the usual time via an unusual route or went home at an unusually early time via his usual route. Participants were much more likely to mention counterfactuals where Mr. Jones behaved more like he usually did, choosing the route to counterfactually change when the route had been unusual, and choosing the time to counterfactually change when the time had been unusual.

People use counterfactual simulation to understand “cause”

Gerstenberg et al. (2017) used eyetracking to more directly show that people spontaneously consider counterfactual outcomes while making causal judgements. They tracked eye movements while participants judged whether one billiard ball caused another to go through a gate. When asked for *causal* judgements like this, but not when asked about details of the *outcome* (how close the caused ball came to the edge of the gate), participants’ eye movements traced the counterfactual path that the caused ball would have taken had the causing ball been absent.

Related work

CausalGAN

Kocaoglu et al. (2018) present a method to separate the work of generating images from the work of learning a causal model. They assume that the true causal graph is given, and train one GAN to sample plausible configurations of features from that causal model and another GAN to generate images given different configurations of features. This separation allows them to sample implausible but imaginable images (e.g. women with mustaches) by *intervention* on a labeled variable (e.g. *mustache*) in the causal model, while still being able to sample only plausible images (e.g. men with mustaches) by *conditioning* on the same variable.

References

Gerstenberg, Tobias, Matthew F. Peterson, Noah D. Goodman, David A. Lagnado, and Joshua B. Tenenbaum. 2017. “Eye-Tracking Causality.” *Psychological Science* 28 (12): 1731–44. https://doi.org/10.1177/0956797617713053.

Kahneman, Daniel, and Amos Tversky. 1981. “The Simulation Heuristic.” TR-5. STANFORD UNIV CA DEPT OF PSYCHOLOGY. https://apps.dtic.mil/docs/citations/ADA099504.

Kocaoglu, Murat, Christopher Snyder, Alexandros G. Dimakis, and Sriram Vishwanath. 2018. “CausalGAN: Learning Causal Implicit Generative Models with Adversarial Training.” In *International Conference on Learning Representations*. https://openreview.net/forum?id=BJE-4xW0W.