

Paper: Towards Causal Representation Learning

Han Lin

hl3199@columbia.edu

MS in Computer Science

This paper gives a very comprehensive (but not very detailed) overview of many aspects of causal representation learning. The following are some points I found inspiring:

Section II talks about the relationship of causal modeling with respect to physical and statistical modeling, it's great to put them together for comparison. From table 1, physical models represent the underlying mechanism of the system which is robust to distribution shifts and could answer counterfactual questions. On the opposite, statistical models are based only on data. Causal models are in between. As sample size increases, we could imagine statistical model to be able to describe the underlying physical models with decreasing error. However, when distribution shift happens or we need to answer counterfactual questions, the collected data from the original distribution might not be able to tell us such information. For last two columns of this table of structural causal and causal graphical, I'm wondering are there works that could give some insight/progress as potential answer.

In Section IV, Figure 1 illustrates interventions of causal models via a set of probability distributions, and introduces the concept of ICM.

The concept of SMS gives some idea about the relationship with causal models and adversarial attacks, and points out the importance of disentangled factorization. I read section VI part(a), but feels like I did not understand well. It would be better if we could discuss more about this section in class.

Section VII covers a lot about causal models in different ML methods. It's interesting to see the contradiction between SSL and ICM. The authors also mention the anticausal direction, and introduce the measure from Daniusis's paper is interesting, and it seems that we could use such method to predict the gains we could have by SSL.

For the reinforcement learning section (even though I think it's better to put it under the category of causal reinforcement learning), I also feel that counterfactuals could probably be used more in RL, especially multi-agent RL (MARL) problems. Previous research in MARL mainly defines counterfactual just as the divergence between marginal distributions and conditional distributions, which does not focus on the underlying causal models very much. I think this might still be a good direction for future research.