STAT 8101: Applied Causality (Spring'22) Week 9: Mar 21, 2022 - Mar 27, 2022

Validating Deep Representations for Interventional Robustness

Han Lin hl3199@columbia.edu MS in Computer Science

- This paper mainly introduces a criteria to evaluate the effectiveness of disentanglement of models (VAEs). Figure 1 and Figure 2 are quite illuminative, which connects the generative factors Z wrt the latent factors produced from VAE models. I think such framework is general enough for practical use.

- However, in most situations, we don't usually have the full dataset like (X, G) as required in Algorithm1, but only the observational data X. For example, when we use VAE models, we will just put into observational data X (e.g. images), and then output latent factors Z. So having additional information about G seems not very practical in many real situations. I guess it's impossible to evaluate the effectiveness of disentanglement without information about G.

- Besides, if G are not confounded by C, then we could just use G to represent Z since we already know the ground truth of generative factors. So I guess this paper will be useful when G are confounded by C, and this evaluation algorithm gives a criteria for the effectiveness of getting disentangled Z from confounded G.

- Another point is that this paper tests VAEs models that are all under the assumptions of independent priors p(z). After Locatello et al. 2019 [1] points out that fully unsupervised learning without any prior information about data/model is impossible, recent works have gradually tried to abandon such assumption and focu on using fully supervised [2] or weak-supervised [3] approach to learn the underlying causal graphs (which is a matrix with rows/cols equal to the number of factors in G to represent their causal relationships). It would also be interesting to see can we derive some other evaluation criteria under this new settings.

[1] Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations
[2] CausalVAE: Disentangled Representation Learning via Neural Structural Causal Models
[3] Disentangled Generative Causal Representation Learning