Nonlinear Invariance Risk Minimization: A Causal Approach

Han Lin hl3199@columbia.edu MS in Computer Science

- This paper propose an invariant causal representation learning approach that enables out-of-distribution generalization in the nonlinear setting. Compared with prior papers, it assumes non-factorized prior, which allows to capture dependences between latent variables.

- In Phase 1, the formulas are derived by a direct substitution of U with (Y, E). Theorem 1 further states that if the k by k matrix $L = (\lambda(Y1) - \lambda(Y0), \ldots, \lambda(Yk) - \lambda(Y0))$ is invertible, then we actually do not need to know E. I'm still unsure what are the observed E we could gather in real environments.

- Phase 2 aims at finding direct parent of Y. This paper proposes a good idea to find a set of (X_i, X_j) and tests their independence and conditional independence given Y. The paper mentions HSIC and KCI test to achieve these two tests respectively. The paper did not state explicitly how such tests are done (do we need to compare all pairs of X_i, X_j?). Out of curiosity I tried these two tests on some games mage dataset, with number of images around 10X larger than the experiments in HSIC paper, but found it very easy to test as non-independent. Will delve into the reason for this. There are also literatures trying to compare $P(X_i- > Y)$ with $(Y- > X_i)$ directly from observational data [1,2], some of them assumes continuous variables XY and some assumes discrete.

- In Phase 3, after obtaining all direct parent of Y, we can learn classifier w. Such classifier can be estimated using data from any environment e with full support. When a new testing data comes, we could first infer Pa(Y) from O, since p(O—X) is assumed to be invariant across all e. After that we could use w for prediction.

[1] Distribution-free learning of bayesian network structure in continuous domains.
[2] Testing conditional independence using maximal nonlinear conditional correlation.