Advances in Causal Representation Learning: Discovery of the Hidden World

Kun Zhang
Carnegie Mellon University & MBZUAI

Can we find the causal direction between two random variables without temporal precedence information? How can we figure out where latent causal variables should be and how they are related? In our daily life and science, people often attempt to answer such causal questions for the purpose of understanding, proper manipulation of systems, and robust prediction under interventions. Accordingly, finding causality and making use of it is an essential problem in scientific discovery and engineering.

Traditional causal discovery approaches [1], such as the PC algorithm and GES, mainly focus on finding causal relations among measured variables, even in the presence of latent confounders (see, e.g., the FCI algorithm). However, in a wide range of real problems, we even do not know what the causal variables are or they are not measurable. That is, measured variables (e.g., image pixels values, insurance claims, and survey responses) are often reflections of the underlying causal variables involved in the generating process, but not the causal variables themselves. For instance, in psychometric studies, the answer scores to questionnaire questions are not directly causally related, but generated by the underlying mental conditions, which might be causally related to each other. Causal representation learning aims to reveal the underlying high-level hidden causal variables, their causal relations, and how they are causally related to the measured variables [2]. It can be seen as a special case of causal discovery.

To achieve reliable causal discovery and causal representation learning, two issues are to be addressed. One is to formulate what footprint or constraints causality leaves in observational data; the other is how to guarantee that the estimated result is consistent with the underlying causal process. Interestingly, the modularity property of a causal system implies properties of minimal changes and independent changes in causal modules, and we show that instantiations of such properties make it possible to recover the underlying causal representations from observational data with identifiability guarantees: under appropriate assumptions, the learned representations are consistent with the underlying causal process up to certain types of indeterminacies.

More specifically, in this talk, we consider various settings corresponding to the three axes of the causal representation learning problem, including whether the observed data are independent and identically distributed (i.i.d.), whether there are parametric constraints (e.g., linear models) on the causal influence, and whether a large number of latent variables are allowed. In each setting, we report to what extent the underlying causal model can be recovered from measured data. For instance, with i.i.d. data, in the linear case, one can uniquely recover the whole causal structure, including causally-related latent variables, in the non-Gaussian case [3]. The conclusion still holds true even if the latent variables do not have any measured variables as indicators [4]. In the Gaussian case, one can recover the equivalence class of the latent hierarchical causal structure from measured data [5], which contains the so-called measurement model as a special case [6].

It is interesting to note that causal presentation learning further nicely benefits from violations of the i.i.d. data assumption: if we have temporal data, it is even possible to recover the underlying latent causal processes from their arbitrary nonlinear mixtures [7]; similarly, nonstationarity or heterogeneity of the data also makes it possible to recover the underlying latent variables from their nonlinear mixtures [8], which has immediately implications in transfer learning and unsupervised data generation or image-to-image translation with identifiability guarantees [9]. Applications of the identifiability theory and developed methods will also be given.

References:

[1] Peter Spirtes, Clark Glymour, Richard Scheines, “*Causation, Prediction, and Search*,” MIT Press, Cambridge, MA, 2nd edition, 2001.

[2] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio, “Toward Causal Representation Learning,” *Proceedings of the IEEE,* 109(5): 612--634, 2021.

[3] Feng Xie, Ruichu Cai, Biwei Huang, Clark Glymour, Zhifeng Hao, Kun Zhang, "Generalized Independent Noise Condition for Estimating Linear Non-Gaussian Latent Variable Causal Graphs," *Conference on Neural Information Processing Systems (NeurIPS)* 2020; https://proceedings.neurips.cc/paper/2020/file/aa475604668730af60a0a87cc92604da-Paper.pdf

[4] Feng Xie, Biwei Huang, Zhengming Chen, Yangbo He, Zhi Geng, Kun Zhang, "Estimation of Linear Non-Gaussian Latent Hierarchical Structure," *Proceedings of International Conference on Machine Learning (ICML)* 2022; https://proceedings.mlr.press/v162/xie22a.html

[5] Biwei Huang*, Charles Low*, Feng Xie, Clark Glymour, Kun Zhang, “Latent Hierarchical Causal Structure Discovery with Rank Constraints,” *Conference on Neural Information Processing Systems (NeurIPS)* 2022; https://openreview.net/pdf?id=lIeuKiTZsLY

[6] Ricardo Silva, Richard Scheine, Clark Glymour, and Peter Spirtes, “Learning the structure of linear latent variable models,” *Journal of Machine Learning Research*, 7(Feb):191–246, 2006; https://www.jmlr.org/papers/v7/silva06a.html

[7] Weiran Yao, Guangyi Chen, Kun Zhang “Temporally Disentangled Representation Learning,” *Conference on Neural Information Processing Systems (NeurIPS)* 2022; https://openreview.net/pdf?id=Vi-sZWNA_Ue

[8] Lingjing Kong, Shaoan Xie, Weiran Yao, Yujia Zheng, Guangyi Chen, Petar Stojanov, Victor Akinwande, Kun Zhang, "Partial disentanglement for domain adaptation," *Proceedings of International Conference on Machine Learning (ICML) 2022;* https://proceedings.mlr.press/v162/kong22a.html

[9] Shaoan Xie, Lingjing Kong, Mingming Gong, Kun Zhang, “Multi-domain image generation and translation with identifiability guarantees”, *Proceedings of International Conference on Learning Representations (ICLR)* 2023; https://openreview.net/pdf?id=U2g8OGONA_V