Task-oriented performance metrics for structure learning

Leonard Henckel

Graphs are an elegant framework to succinctly represent the joint dependencies of a vector of covariates. They can, for example, be used to encode conditional independences, which makes them a useful tool in probabilistic reasoning (Koller and Friedman, 2009). They can also encode qualitative causal relationships and be used for causal reasoning (Pearl, 2009). As a result, the problem of learning a graphical model from observational data, a problem known by many names such as causal discovery, unsupervised learning or structure learning, has received considerable research attention (e.g. Spirtes et al., 2000; Shimizu et al., 2006; Zheng et al., 2018). The resulting algorithms, however, remain inaccurate in applications and practitioners are generally sceptical of them (Reisach et al., 2021). This is in large part because structure learning is simply a very difficult problem, but machine learning has shown that research on difficult problem can progress remarkably quickly if clear performance metrics are available. Structure learning, however, lacks a universally accepted performance metrics in part because graphical models can be used for so many distinct tasks (Gentzel et al., 2019).

In this project, we will try and develop a metric for causal graphical models by centering the core task we use these graphs for: causal inference. Specficially, the metric should capture how well a graph would perform if we used it to infer causal effects. One natural starting point to develop such a metric, is to built on the structural intervention distance by Peters and Bühlmann (2015) by using advances in the literature on causal inference made in recent years. Another, is to revisit causally oriented empirical metrics such as the one proposed by Eigenmann et al. (2020) and investigate their empirical characteristics. While causal inference is the primary focus, a secondary aim is to develop a general framework for task-oriented structure learning metrics.

References

Eigenmann, M., Mukherjee, S., and Maathuis, M. (2020). Evaluation of causal structure learning algorithms via risk estimation. In *Conference on Uncertainty in Artificial Intelligence*, pages 151–160. PMLR.

Gentzel, A., Garant, D., and Jensen, D. (2019). The case for evaluating causal models using interventional measures and empirical data. *Advances in Neural Information Processing Systems*, 32.

Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques.* MIT press.

Pearl, J. (2009). *Causality.* Cambridge University Press, Cambridge, second edition.

Peters, J. and Bühlmann, P. (2015). Structural intervention distance for evaluating causal graphs. *Neural computation*, 27(3):771–799.

Reisach, A., Seiler, C., and Weichwald, S. (2021). Beware of the simulated dag! causal discovery benchmarks may be easy to game. *Advances in Neural Information Processing Systems*, 34:27772–27784.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A. (2006). A linear non-gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.*, 7(Oct):2003–2030.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search.* MIT Press, Cambridge, MA, second edition.

Zheng, X., Aragam, B., Ravikumar, P. K., and Xing, E. P. (2018). Dags with no tears: Continuous optimization for structure learning. *Advances in neural information processing systems*, 31.