Brummer & Partners MathDataLab Postdoc Proposal

Liam Solus

October 29, 2019

Proposed topic: Probabilistic Graphical Models: Theory and Applications in Causality

Project area: A fundamental goal in artificial intelligence is to develop autonomous systems capable of learning complex cause-effect networks that it can use when making decisions. Modern approaches to this problem typically rely on a *probabilistic graphical model*, which encodes the conditional independence structure of a data-generating distribution via the absence of edges in a graph. Such models are a popular tool for modeling cause-effect systems, and they are now widely used in fields such as economics, epidemiology, genetics, sociology, and the environmental sciences [1]. In order to select a model that best represents the data, it is critical that we have efficient and reliable algorithms for data-driven *model discovery*. Surprisingly, recent advances in the development of such algorithms [2, 5, 6], and in the analysis of their efficiency and reliability [4] have called upon a variety of techniques from areas such as combinatorics, algebra, and discrete geometry. Unfortunately, many of the popular techniques for causal model discovery make the restrictive assumption that all variables in the system are observable. A key problem within the field is to develop efficient and reliable methods for causal model discovery in the *causally insufficient* setting; i.e, where we incorporate the presence of latent confounders.

Project description: In this project, we will use methods from combinatorics, discrete geometry, and algebra to build a theory of causal discovery in the causally insufficient setting. A successful applicant would be invited to pursue directives relating to one or more of the following major goals: (1) develop new methodology for learning causal models incorporating latent confounders, (2) analyze the reliability and complexity of such algorithms, and (3) apply new causal discovery algorithms to data sets in biology and the environmental sciences. New methods for causal discovery in the fully observed cases rely on simplex-type algorithms that depend on the combinatorial geometry of certain *convex polytopes*. A successful applicant could work to generalize this geometry to the causally insufficient setting, develop corresponding simplex-type algorithms, and/or apply such algorithms to data sets from the aforementioned fields. Such methods will require a well-developed theory of *interventional settings* [6] that can arise from experimentation in the presence of latent confounders. Recent results [3] set the stage for describing the complexity of such interventional settings via methods from algebra and discrete geometry.

A postdoc with interests in the theory and application of graphical models to causality would have numerous people to talk to at KTH in both the mathematics and mathematical statistics divisions. An ideal applicant would interact with faculty in both the mathematics and mathematical statistics divisions, thereby strengthening connections between various research groups within the department.

References

[1] J. Pearl. *Causality: models, reasoning and inference.* Vol. 29. Cambridge: MIT press, 2000.

[2] L. Solus, Y. Wang, L. Matejovicova, and C. Uhler. *Consistency guarantees for permutation-based causal inference algorithms.* Submitted to Biometrika. Preprint available at `https://arxiv.org/abs/1702.03530` (2019).

[3] L. Solus. *Interventional Markov equivalence for mixed graph models.* Submitted (2019).

[4] C. Uhler., G. Raskutti, P. Bühlmann, and B. Yu. *Geometry of the faithfulness assumption in causal inference.* The Annals of Statistics 41.2 (2013): 436-463.

[5] Y. Wang, L. Solus, K.D. Yang, and C. Uhler. *Permutation-based causal inference algorithms with interventions.* Advances in Neural Information Processing Systems. 2017.

[6] K. D. Yang, A. Katcoff, and C. Uhler. *Characterizing and learning equivalence classes of causal dags under interventions.* Proceeding of the 2018 International Conference on Machine Learning ICML (2018).