Meeting notes 6 24 2023

Amber and Patrick physically present (once Amber got in), Dennis and Sophia skyping in.

Patrick said he had sort of ‘lost his philosophical bearings’ in the project, and had been researching some machine learning literature, wondering about the extent to which we were reinventing the wheel and what philosophical lessons we wanted to pursue.

I. With regard to the machine learning literature:

The task we’re after—finding causal structure on the basis of incoming evidence—is called ‘causal discovery’ in the literature, and is distinguished from (though uses certain tools from) the Judea Pearl effort. In Clark Glymour’s characterizations:

“In traditional causality research, algorithms for identification of causal effects, or inferences about the effects of interventions, when the causal relations are completely or partially known, address a different class of problems; see Pearl (2000)…”

In Patrick’s outline, there are 3 traditions of causal discovery that may be of interest to us:

A. The constraint-based PC tradition (named PC for Peter Spirtes and Clark Glymour, both in philosophy at Carnegie Mellon). Here’s an image from Glymour, Kun Zhang and Spirtes, “Review of Causal Discovery Methods Based on Graphical Models,” *Frontiers in Genetics* 2019:

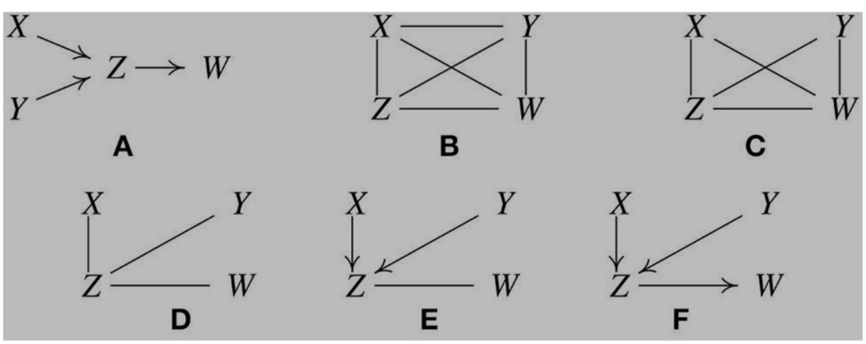

FIGURE 1 | Illustration of how the PC algorithm works. **(A)** Original true causal graph. **(B)** PC starts with a fully-connected undirected graph. **(C)** The $X - Y$ edge is removed because $X \perp\!\!\!\perp Y$. **(D)** The $X - W$ and $Y - W$ edges are removed because $X \perp\!\!\!\perp W \mid Z$ and $Y \perp\!\!\!\perp W \mid Z$. **(E)** After finding v-structures. **(F)** After orientation propagation.

The outline of elimination progressively builds on eliminating links between variables which are independent given larger sets of other variables. There are then a set of orienting rules in order to make the links directed depending on what was conditional on what in elimination. Details in Glymour, Zhang, and Spirtes.

This is a ‘constraint-based’ approach because it relies on the constraint of independence considerations from elsewhere. It isn’t guaranteed to give the specific ‘ground truth,’ but with certain assumptions is guaranteed to give the ‘Markov Equivalence Class,’ in which the evidence isn’t sufficient to distinguish graphs with links in different directions.

We’ve talked about that idea too, with an eye to what partial evidence can tell us.

Although this doesn’t look like a picture of how science proceeds (starting from all possible links between everything and eliminating them), it is interesting that neuron formation in the brain starts with a proliferation and later pruning.

B. The GES tradition (Chickering). This differs from PC in that it starts with no links rather than a complete graph, and uses a score-based method of keeping changes that increase a score, rather than independence measures. We’ve been working with that kind of score-based system throughout.

GES proceeds in two steps:

1. Add links and keep them if they increase the score (in one form, try multiple links and keep those that increase the score the most).

When the score can no longer be improved that way:

2. Remove links which increase the score.

“In the large sample limit,” Glymour says, “the two algorithms converge on the same Markov Equivalence Class under assumptions are nearly the same. On finite samples, the algorithms may give different results…”

This too is interesting, in that it starts with proliferation and then pruning. But because of the two separate steps it doesn’t seem a great match for how science proceeds.

C. Continuous updating

NOTEARS “was revolutionary because it was the first algorithm to frame structure learning as a purely continuous optimization problem” (Molak’s blog)

This sounds the most like ours, and ‘continuous optimization problem’ in terms of a score sounds a lot like how science proceeds. But we don’t have a nice Glymour-type outline (it’s more recent). Two of the cited articles will be sent with these notes, but Patrick says he can’t understand them yet.

Sophia is at a conference in which it might be appropriate to ask other people whether they could give us a lead on them.

II. Patrick’s easy way in to all this is via the blog:

https://towardsdatascience.com/beyond-the-basics-level-up-your-causal-discovery-skills-in-python-now-2023-cabe0b938715

The conclusion of that blog is direction to gCastle, a Python library that supposedly has versions of all of this. "…the biggest, most complete, and most up-to-date list of causal discovery algorithms you can find in any open-source causal Python package."

Dennis is going to see if he can access it and check it out.

III. A major question is where we want to go from here—are there philosophical questions for which this work might be suggestive, for example?

A couple things floating in the discussion:

The Markov equivalence idea, and theories that will be 'empirically equivalent,' particularly given limited data. That may be a nice step on realism and anti-realism in science.

The idea of path-dependence, as in the 'finite sample' note on GES. Can what order the evidence comes in influence what theory you converge on?

Sophia also brought up the issue of a scientific parallel to 'strengthening the antecedent.' Sometimes we have cases where the existence of information p makes it very likely that q… but if p and r, it becomes very **un**likely that q. It seems like this should feed into questions of partial evidence and path dependence in how the evidence comes in.

IV. We'll plan on meeting in a couple weeks—July 10th or thereabouts. We'll see whether in person or virtual is most convenient.