Meeting notes 3 31 2023

Everybody was there.

I. Amber has hybridization up and running for 3 nodes. She says that extension to more nodes shouldn't be difficult.

The next step is to get a 'keep the best' routine that we can use with all three of our heursitics: single-point mutation, random single- and double-point, and hybridization. Dennis has an earlier form of that, which Amber is trying to read and adapt.

II. Patrick and Dennis have both been looking at other literature for potential interface with what we're doing.

A. Patrick said he has been exploring (1) Judea Pearl on Bayesian nets and causal graphs, and (2) literature on learning Bayesian nets from data.

On (2), he's found parallels to our approach;

- Other work involves 'score and search', demanding a scoring function for alternative results. Like ours. Characterized as an optimization problem.

- An exhaustive search can be shown to be NP-hard for "any reasonable scoring method," even if the number of parents in the network is limited to 2.

- Thus people suggest different heuristics "Where there is no guarantee that the ultimate Bayesian net has been found." Sounds familiar.

- A prominent difficulty mentioned with regard to heuristics for networks is to avoid cycles. One approach parallels the 'level' approach we played with: "Any DAG has a topological ordering of the variables such that a vertex's parents must appear earlier in the ordering"

B. Dennis gave an outline of fundamental points in (1):

A basic problem in Pearl is to go from a spreadsheet of merely observed (non-interventional) data in the attempt to (a) find causal connections and (b) therefore predict the results of intervention.

I'll try to get this right…

A spreadsheet of observed data merely shows correlations between variables. But correlations between variables A, B, and C might reflect any of these causal patterns:

$A \longrightarrow B \longrightarrow C$ chain

$A \leftarrow B \rightarrow C$ fork

$A \rightarrow B \leftarrow C$ collider

Can the static data tell you which causal pattern is in play? What Pearl portrays the classical position in statistics as 'correlation can't tell you cause.' His answer is that 'sometimes it can:'

If a chain or a fork is the causal pattern, A and C will be independent given B ($A \perp\!\!\!\perp C \mid B$).

If a collider is the causal pattern, the opposite holds: A and C are independent, but are conditionally dependent given B.

There was discussion of examples of this last pattern:

if A alien abduction and C a stopped clock are both sufficient causes of B Bill's missing the meeting (and the only ones, for purposes of illustration), A and C will be dependent given that Bill missed the meeting: Knowing that his clock didn't stop would raise the probability that he was abducted by aliens. Knowing that his clock did stop 'explain away' or reduce the need to appeal to alien abduction.

If A attractiveness and T talent are each sufficient causes of C celebrity, although A and T are independent in the population, they will be dependent given C. Knowing that someone is a celebrity and is attractive will reduce the probability of T, 'explaining away' or reducing the need to attribute talent.

The conclusion: observational data contains at least partial causal information. Observational data alone can't distinguish between a chain and a fork, however. To do that we need a real intervention or a randomized controlled experiment.

C. Some more. I'll again try to get this right, but Dennis will correct me if I'm wrong:

Pearl's 'do operator' [$P(Y \mid do(X))$ as opposed to just the observational conditional probability $P(Y \mid X)$] is what would happen to Y if one intervened at X.

In a graph "The do-operator erases all the arrows coming into X, and thus prevents information from flowing from X in the noncausal direction." Given that alteration, any effect of X on Y will have to be causal.

Since confounding variables function through noncausal links, the do-operator eliminates confounding. How do you know whether you need to 'deconfound' variables X and Y in a graph, and how to do so if you need to? One way is the 'back-door criterion.'

Trace every path to Y that starts with an arrow into X (and thus is not causal out of X). For example:

$X \leftarrow A \rightarrow B \rightarrow C \leftarrow Y$

If there is a chain or a fork in there, you can 'deconfound' by controlling for the central item of the fork. Thus in this example controlling on A or B would block any back door between X and Y.

If there is a collider in there, as at C, you don't have to do anything. There is no confounder on that path. In fact, if you did control for C you would have opened a back door.

III. We started to discuss how these bits might interface or not interface with what we're doing, but there's lots more thinking to do there.

Both Bayesian nets and Pearl involve probabilities between 0 and 1, not merely the 1 and 0 of activation that we've been dealing with. But we've had that in our sights eventually anyway.

None of this other work involves **temporal** data, emphasized in both Hume and Mill and that has been important in our work. Patrick speculated wildly that temporal data must be exploited in formation of (real) neural connections, suggesting an option for formal modeling in the tradition of genetic algorithms and neural networks, both of which are based on formal borrowings from biology.

IV. For next time, **two weeks** (April 14^{th})

Zhongming is going to think more about the 'space of networks consistent with' incomplete patterns of evidence,

Sophia is going to look over Draft 1,

Amber is going to work toward a 'keep the best' routine that can be used with our other heuristics,

Dennis will help with that, and

Patrick and Dennis are both going to do more exploring on interfaces with Pearl and other potentially relevant literature.