Notes on Pearl & Russell 2000: Bayesian Networks

Gray Stanton and Austin Ellingworth

October 28, 2020

Discussion Qs

- A shower thought from the1995 paper from last week: the paper states the adding arcs can impede but never help identification. To me, this suggest that the creator of the network ”chooses” (which in practice is true) when to add an arc and when not to. But in a perfect world, if we know an arc exists, I think it should be included. It may hurt identificiation, but if we know an arc exists and choose not to include it, is what we are ”identifying” really true in actuality?
- Is Figure 1 really a ”causal” relationship? Would never say ”dang! fall just caused me to slip.”
- When setting up DAGs for anaylsis, in practice, how is it decided what is included and which directions arrows go? Could see potential relationships where directionality is iffy.
- To what type of practical problems are Bayesian networks uniquely suited?
- The authors makes a distinction between general Bayesian networks and causal networks. When would we want to use a Bayesian network that is not causal?
- A final note: Pearl must come from somewhere warm. There is mention of rain making pavements slippery but not ice or snow.

Uses for Bayesian Networks

- Probability models based on DAGs arise in many places; among AI/cognitive science, they are called Bayesian networks.
- Core feature is bidirectional inference.
- Bayesian network represents the joint distribution of all propositional variables (nodes). Edges represent informational/causal dependencies, quantified by conditional probabilities between node and it’s parents. Lack of edge indicates conditional independence assumption.

- Bayesian network represents model of the world, not of reasoning about the world. Reasoning about the world via the model can be done by propagating information. Allows for prediction, abduction, explaining away.

- Global semantics of a Bayesian network is that the joint distribution factors as
$$P(X_1, \dots, X_n) = \prod_i P(X_i|pa_i).$$

- Also have equivalent local semantics, where network specifies that each variable is independent of all non-descendants given it's parents. Useful for construction, as selecting direct causes as parents will satisfy independence conditions.

- Reasoning from evidence on Bayesian networks can be done, but is NP-hard in general.

- Also can encode changes over time in dynamic bayesian networks, with a new copy of each state variable for each timestep. More expressive than other state-based methods like hidden Markov models or Kalman filters

- For fixed network architecture, the conditional probabilities $P(X_i|pa_i)$ can be learned iteratively via EM or gradient based methods. Also possible to learn network architecture.

- Distinction between general Bayesian networks and causal Bayesian networks. The latter can express the result of an intervention, and requires that the parents of each node are precisely it's direct causes.

- To process counterfactual information, replace conditional probability with more general functional form, and process info on "twin network".

- Causal discovery is apparently possible, namely identifying causal structures in raw data without using controlled experiments.

- Bayesian networks can be used to model "plain beliefs" using non-standard probabilities

- Bayesian networks could be used as model of cognition for propositional reasoning under uncertainty.