

Bayesian nets

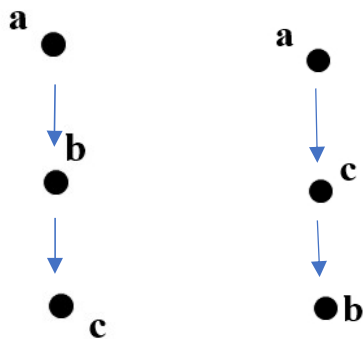
Meeting notes 11/16/2022

Amber, Dennis, Zhongming and Patrick were physically present.

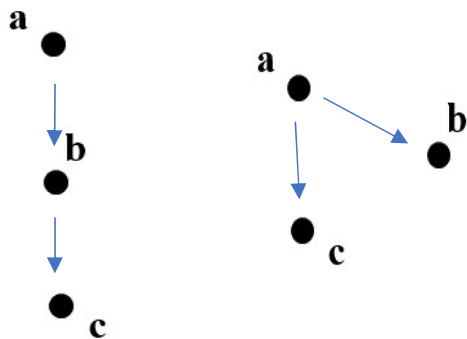
1. Since last week, Amber has worked out the evidence matrix idea, encoding timing, and has programmed the generation of an evidence matrix (convertible to a string) for any network. Yes, I know, very impressive.

That will allow us to (1) generate an evidence matrix from the ‘world’ and (2) generate an evidence matrix (or string) for each of our candidate representations, then allowing us to score how close different representations are in terms of how well their output matches the world.

She also programmed a ‘scoring’ routine in terms of the number of digits at which evidence strings matched. One initially counterintuitive result: these two networks, though they look structurally similar:



score as farther apart than these two networks:



Amber also had the explanation: The bottom two share everything from a to b, leaving the a to c as an area of mismatch. The top two come out wrong on much more.

2. That work sets us up for the last stage in this exploration: the genetic algorithm. We start with a random set of networks, score those in terms of how well their evidence matrix matches that of the ‘world’, pick the most successful and mutate those for a new population, score those and so forth... until, hopefully, we get a match.

Dennis has worked out the genetic algorithm for both one-point mutation and two-point mutation. Now we just have to put that together with Amber’s bits. The plan is to have that in hand by our next meeting, after Thanksgiving on November 30th.

3. There was discussion of how the best way to do the genetic algorithm might be.

One way—indicated above: mutate a random spot (or two random spots) of one of the chosen ‘bests’ at a point.

Another way—mutate only spots on which there wasn’t a match.

The second way might cut down the exploration space, which is always a problem for genetic algorithms. On the other hand, it seems it might get stuck on local maxima, unable to break out of a sub-optimal pattern.

A third possibility—use the second way until it gets stuck, at which point it tries random mutations.

It is an interesting question which one of these would be most effective. Just doing some runs with the different forms would be of interest.

While we’re at it: Exploration of the sample space could be expected to be more efficient, though this would require memory, if the program didn’t try combinations that it had already seen and rejected.

These alternatives are probably the next step for Dennis and Amber in programming.

4. Patrick said that he was also exploring alternatives to a genetic algorithm in network-variation-exploration on the representation end. Genetic algorithms are well established and well regarded for this kind of thing, but they do run into the explosion problem of enormous sample spaces with increasing numbers of nodes.

His claim: that although genetic algorithms are based on a biological analogy, it may be the wrong biological analogy. If we think of our representations as little brains trying to figure out the world, we don’t figure out what happened on Tuesday through millions of years of evolution. Somehow neurons wire and rewire in individual heads on the fly.

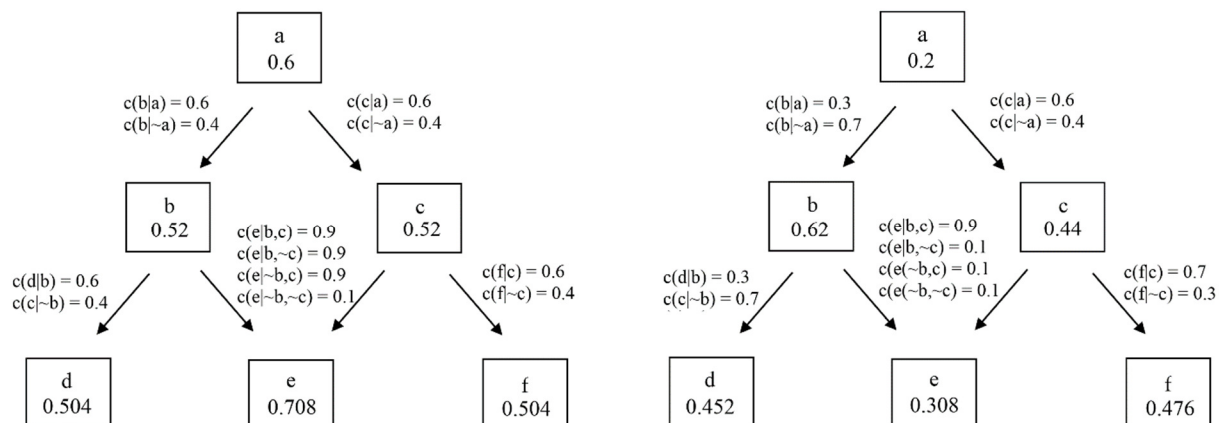
Are there formal models that are more like that? Patrick is still looking. A great deal of the neurobiology isn't yet understood, though people do seem to agree that neurogenesis involves (1) 'ebullient' generation of lots of neurons, (2) migration to needed parts of the brain, and (3) 'pruning' down to the really effective ones.

That would suggest something like starting with a 'complete' network on the representation side and using our evidence matrix to prune it...

5. Zhongming then led us through his beginning work on convergence in Bayesian nets. The idea is that this will hook up with what we're doing if and when we can scale this model up to Bayesian networks.

The basic question: We're trying to figure out how a representation can 'glomp onto' a world in terms of structure. Now suppose that part of the problem is taken care of, but we still have rival theories as Bayesian networks with the same structure but with (a) different credences at nodes and (b) different conditional probabilities between nodes.

Two rival theories might look like this:



One thing that is known about Bayesian updating—and is thought to be very much in its favor—is that if you have what is essentially one node with two credences:



and if you hit both of these with the same string of evidence, you are guaranteed that these credences will eventually converge. Basically, the evidence string is bound to 'swamp' the initial credences.

What would be nice is to extend that result to full networks: That if you subjected the nodes of the two networks above to the same string of evidence-at-the-same-nodes, that those networks would eventually converge on both credences and conditional probabilities between them.

If we think of the Bayesian networks as scientific theories or pictures of the world, that result would say: Given long enough strings of identical evidence, those scientific theories will converge even if they start out with very different credence and conditional probability values.

Proving that, or bits of it, is Zhongming's 'mission impossible' task. Your mission, should you choose to accept it...

In his presentation, Zhongming demonstrated how credences would change down a chain from changes in a top or 'root' node, and demonstrated both algebraically and with a program how you would end up with convergence in that case (and often very quickly).

He also showed the complications that come in with 'collisions' between two nodes.

6. In order to go further, what will be required is a more general demonstration that (1) includes changes at nodes other than the root, (2) calculation of posterior conditional probabilities as well as posterior credences at nodes, and (3) a treatment of what appear to be the three major structures in any Boolean network: chains, branches, and collisions.

Patrick said he thought that if we could prove this for simple 3-node cases of each—chains, branches, and collisions—that could tell us that it would hold for any more complicated networks made out of these pieces.

The next step will be for Patrick to give Zhongming some of the algebra for re-calculating conditional probabilities, and relevant parts of papers and book chapters.

All good stuff. Next meeting Wednesday November 11th, same time and place.

Thanks,
Patrick