

NICE TO KNOW

Bayesian Belief Networks - Part 1

Tony Starfield

recorded: March, 2012

Before I start on the next tool that we're going to develop, I would like to philosophize a bit about modeling and what models achieve for us. And I want to develop the idea. I want to concentrate on the word "tool" because I like to think of models as mind tools. What do I mean by this? We have a problem. We might be a little unsure exactly how the system works. The problem might be a bit amorphous. We're trying to think through it. We might be doing that as an individual within a group. We might have thought through it, and we're trying to communicate it. And in all of these cases, it makes a huge difference if one has a model of some sort than if one is just talking about the problem.

The difficulty with just talking about a problem is that it's hard to keep the pieces together in your mind or to keep the pieces together in somebody else's mind. And it's hard to reach closure in thinking about the problem.

So I want to postulate that a model, any model, is better than having no model at all. Having said that, I want to push it a bit and say that some models are much better than others. What do I mean? Conceptual models, drawing a diagram, showing how the different components of a system or of a problem might interact and influence each other is useful. It helps you to talk about the problem. But it's not nearly as satisfying as having a model that allows you to manipulate inputs and see the changes in outputs.

To have a working model, I don't care whether it's numerical. I don't care whether it's on a spreadsheet or not. I don't care whether it is qualitative. But it has to be something whereby you can change the inputs and see what happens to the outputs because that is what really enables you to get your mind around a problem.

So what I want to do in this section is talk about another tool in the toolkit for taking a conceptual model and turning it into a numerical model. We've kind of done that with qualitative modeling. We've done it with various types of spreadsheet modeling. We've done it with frame-based modeling.

Here, I want to look at a situation where the uncertainty is in one's understanding of how one

part of a story influences another part. And so one would start in a situation like this with an influence diagram. So an influence diagram would be your conceptual model.

Let's take an example. Suppose one's looking at a future population of a species, and we're worried about what's going to happen to that species in maybe 50 years' time. And, in particular, we're worried about how development in part of the habitat of the species might influence its future trajectory. We're worried about what role climate change is going to play. And it could be that, along with development, you're going to get interaction between household pets, dogs, cats and so on, and wild species, and this might lead to disease. Maybe at the back of your mind, you've got something like a wolf population.

So one would draw an influence diagram where one would say, "Development, climate change and disease are probably the three things we're going to think about as influencing this population most." And then you'd go one step further in your analysis, and you would say, "Development and climate change are probably both going to influence habitat loss." So one would draw arrows from development and climate change to habitat loss. And disease is probably going to have an impact on the litter size or the production of young in this population. And, if you think about it, climate change could have an impact on that too. So we would draw two more arrows, one from climate change and one from disease leading into litter size. And then one would say, "Habitat loss and litter size, we think, are going to be the major drivers of trying to figure out what's going to happen to the population in future." So we would have an arrow from habitat loss, an arrow from litter size leading into future population.

Now, just drawing a diagram like this sort of defines the terms of the debate. It tells people what we're thinking about, and it tells them where we think changes along the way--development, climate change and so on--are going to impact things that affect the population.

So, as I was saying earlier, a diagram like this is better than no diagram because it helps to clarify what one's talking about. But a diagram like this is not useful, really useful, unless you can start putting some numbers into it.

Well, how would one go about putting numbers into this type of diagram? It turns out there's a tool that's becoming quite popular, which really fits with this type of situation. And it is a tool that deals in probabilities rather than uncertainties, and it is called a "Bayesian Belief Network."

So if I were dealing with a problem like this, step one is to draw the influence diagram. And to show how one develops Bayesian Belief Networks, I'm not going to take the whole of this influence diagram. I'm just going to take a small part of it. Let's look at the arrows that lead from

development and climate change into habitat loss.

And to put quantities into this, I need to define states to describe development, climate change, and habitat loss. So, for example, I might describe development as either controlled or uncontrolled. I might describe climate change as slight or severe over the time period that we're looking at, whatever it is. And I might want to talk about habitat loss being low or high. So that's a start. I now have measures, if you like, for the input variables, which, in this case, are development and climate change, and an output variable, habitat loss.

What I want to do is put some numbers into this, specifically, probabilities. And what I'd do is I create a table. And if you look at this table, you will see that I look at all combinations of development and climate change. So development can be controlled and climate change can be slight. Development can be controlled and climate change can be severe. Development can be uncontrolled; climate change, slight. And development, uncontrolled; climate change, severe.

And as I look at all these combinations, I think you can begin to see that we're very glad we weren't more discerning, that we didn't have three states, for example, for both development and climate change, because then we would have a lot more entries in this table.

And then we ask ourselves how these different combinations of inputs would affect habitat loss. Under what conditions would it be probably low or probably high?

The way we put numbers into this is to imagine that we are 100 percent certain about the inputs. So if I say, "Development is controlled, and climate change is slight," what I mean is, "I'm 100 percent sure that development is controlled, and I'm 100 percent sure that climate change is slight." And I would then ask experts to tell me what the impact of that might be on habitat loss. And quite possibly, they would all agree that habitat loss would definitely be low, under those circumstances. I would then fill in 1.0 for the probability of it being low, and obviously, therefore, 0 for the probability of it being high.

Now what happens if we move onto the next line? We have controlled development, definitely, 100 percent sure, and climate change that we are 100 percent sure is going to be severe. And maybe, under these circumstances, we would get experts to hone in on the idea or to support the numbers that say they reckon that, under those circumstances, there's a 70 percent chance that habitat loss could be low, but there's a 30 percent chance that it could be high. So we feed in those numbers.

And then we go on and fill in the table for the other two combinations.

Now, again, as in, for example, developing a hybrid qualitative model, the numbers in this table are the parameters of our model. And different experts might have different ideas about the probabilities. And if we can get them to compromise, or more or less agree, then we will get one set of numbers. If it turns out that they have really different viewpoints of what might happen, we could actually end up with two models, one populated by the parameters of one group of experts and another populated by the parameters of another group of experts. Either way, we can set up a model like this.

But what do we do with it? Well, the key point is that these are the numbers where we are sure of the inputs. In practice, we're not going to be sure of the inputs. How do we know for sure that development over the next 50 years, for example, is going to be controlled or uncontrolled? So what we want to do in our actual working model is feed in probabilities for the development being controlled or uncontrolled and probabilities for the climate change being slight or severe. And then our model will come back, using these parameters, with a probability that, under those circumstances, habitat loss will be low or high. And the way one develops the mathematics behind this is using Bayesian statistics. And there is software out there that enables you to implement this sort of model.

So once you've put it all together, you are going to be able to manipulate some of the inputs, argue about whether you are less or more certain about what state of inputs you will have in the future, and you are going to be able to see how that affects the probabilities for your key outputs. And that is exactly what I want to do. That is just so much more valuable than drawing an influence diagram.

I want to talk about some of the examples of this and some payoffs from this and some comparisons with other things that we've been doing.

< 00:13:53 END >