TDT4171 Ex3

Harald Husum

April 2016

Introduction

This exercise presented me with the task of creating a decision support system for a decision problem of my choosing. Looking for something relevant to my own life, I ended up with the difficult issue of how to choose your path when coming upon a crossroad while hiking in the forest. The idea is simple: You arrive at some local woodland, unfamiliar with the specific area. You are greeted by paths with corresponding signs that indicate length. One seem to be quite short, while the other is significantly longer. You have access to some external knowledge as well. You know what the weather report says, you know whether you have brought food, and you come with a perception of your relative fitness. Which path should you choose to walk down?

Causal model

I ended up modeling this problem as a static Bayesian network. Although it can be argued that this is a dynamic problem (what if you turned back in the face of difficulties?), we assume a decision is stuck with and pursued to the end. This leaves us with one decision at only one point in time, and we have a static model. A Bayesian network seems a good choice here, because we can already see the contours of several random variables, some dependent on others. Utilizing the conditional independence relations of the Bayesian network, we can build a compact model for the domain.

To develop this causal model, we start with the utility node. I have chosen to use a compound utility node. This means we have a final utility calculated from what we can call 'sub utilities'. In our case this is 'risk' and 'enjoyment'. This is based on the understanding that the primary purpose of trekking in woodlands is to achieve some level of enjoyment, and that this enjoyment constitutes a trade-off in terms of the risk one subjects oneself to in order to achieve it.

So what constitutes risk in the context of strolling though shrubbery? Risk should be understood as the danger of dying or getting badly hurt. This can happen in many ways, while trekking. I've chosen to consider three vectors I consider representative: Exhaustion, injury and hypothermia. An increase in any of these means a heightened possibility of not surviving the hike, thus we use them as the basis for our risk function.

Enjoyment is an all together different and more subtle quality than risk. What is it the freedom of the hills offer us, that so many of us enjoy so thoroughly? This is a hard question, almost philosophical in nature, and this is not the right context for answering it. Instead we chose a rather simplistic model for enjoyment. We say enjoyment is some function of the view you get to see from your path, the weather conditions, whether or not you are hungry and how far you end up walking. The latter is in a way an intensifier. If your hike is enjoyable, walking longer will result in more enjoyment. If it is less enjoyable, the opposite is true.

Figure 1: The setting for our problem

Now, having hinted at the presence of a bunch of variables, we will now look at their relations and the assumptions underlining them:

- **Exhaustion** is assumed to be dependent on both the walking length and the fitness of the wanderer. Walking far increases the probability of getting exhausted, as does being in bad shape.
- We deem **injury** to be the result of bad path quality. An uneven surface, full of potholes or rocks, maybe even slippery, is the prime cause of injury. This, however, means injury need also depend on the distance traveled, as the more rough ground you cover, the more danger you get subjected to.
- The **time** spent in the forest is similarly also dependent on path quality and length. It could be tempting to let it depend on shape as well, but in this scenario we assume walking speeds, suggesting the terrain sets the upper limit on speed.
- Getting **cold** and suffering from hypothermia is not so much related to the distance you've covered as to the time you've spent outside. It is also highly unlikely in good weather. We thus propose that cold is dependent on weather and time.
- In the same way, we see **hunger** depending on the duration of your walk and whether or not you chose to bring food along.
- We also include two variables, **perceived shape** and **forecast**. These only depend on physical shape and the actual weather, respectively. They function as windows to otherwise hidden variables.

- Furthermore, **path quality**, **length** and **view** only depend on the choice of path.
- **Physical shape**, **weather** and **food** represent the top of our hierarchy, not depending on anything in our model, along with the **choice** to be taken.

Finally we assume perceived shape, path length, food and, weather forecast to be the only visible variables at the time of choice. The structure of the decision network resulting from these assumptions can be seen in Figure 2.

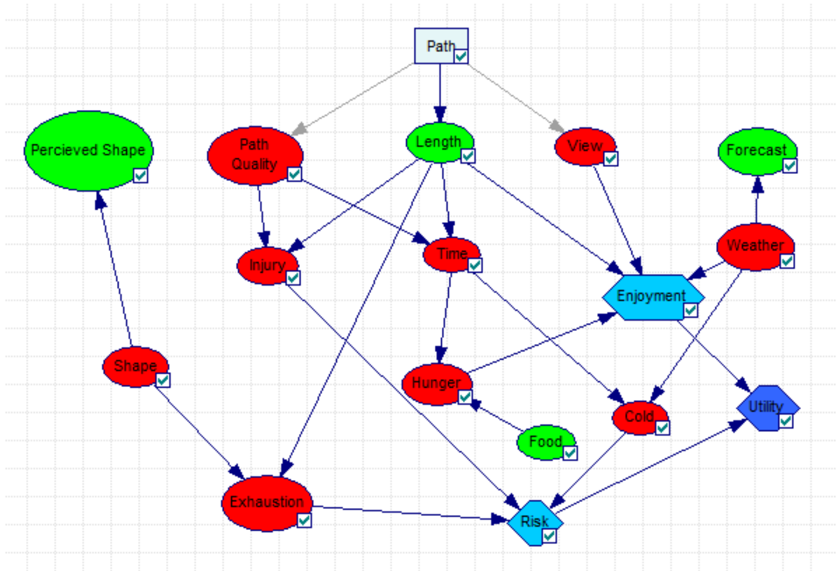

Figure 2: Our resulting decision network. Green nodes are visible, red are hidden, blue is the compound utility function, and grey is the decision.

Probability and utility assignment

It should be mentioned at this point that many of the random variables I have suggested are distinctly continuous in nature. Variables like length, time and weather would probably be best represented over continuous domains. GeNIe does not offer functionality for hybrid Bayesian networks, however, and I don't have the patience for filling out large conditional probability tables by hand. The result is that we will perform quite aggressive discretization on most variables. We allow this, because in this particular domain, I see the potential for a useful and interesting model, despite the coarse categorization.

A selection of quantitative statements

Although the graphical representation and real life intuition indicates path quality and view depend on the choice, in terms of the model, this is not really the case. Both variables go either way 50% of the time, regardless of path choice.

Of both particular amusement and concern, we have the 'perceived shape' variable. I have assigned it values from my own experience as a trekking guide in the mountains of Norway in my youth. As you can see from the assignment, my impression is that people tend to be more optimistic about their physical shape than they have reason to be:

$$P(ps = good|s = good) = 0.95$$

$$P(ps = good|s = bad) = 0.2.$$

Let us also have a look at the choices for the hunger variable. What constitutes hunger can be quite hard to define, and here we rely on a subjective definition, based on a significant level of discomfort. As seen in the table, the probability of hunger never drops to zero. This is supposed to represent the possibility of dropping ones food while hiking, or not bringing enough. We also see that it is fully possible to even long walks without getting too hungry. This is taken from personal experience, so the numbers are a bit fuzzy, but I feel they represent my feeling about food.

Time	Long		Short	
Food	Yes	No	Yes	No
Hunger	0.1	0.7	0.02	0.2

Another distribution of interest is that of injury. It must be assumed to be feasible to gain statistical data on the injury rate per distance on various surfaces. But this is information I have not sought out. I will not deny that my estimates are potentially very hand wavy, but I don't think they are outrageously wrong. This is again based on personal experience - one that I do not wish to elaborate on. The salient part to notice is that we see a lower increase in injury on the longer path than we could have. I assume probability of injury goes somewhat down as one gets used to the terrain.

Length	Long		Short	
Path Quality	Good	Bad	Good	Bad
Injury	0.03	0.1	0.01	0.03

On the additive value function

The utility function I use is in many ways very simple. The final additive value function just reads:

$$Utility = Enjoyment - Risk$$

The lack of weighting in this expression is the result of the subjective nature of both variables. If we had looked at an example involving real world units, like money or time, we could calibrate by weighting the different components. Here we just try to express both components on the same scale instead.

For my enjoyment expression, I chose to assign a high value of 1500 for a long trip in good weather, with nice views and having brought food. The other end of the scale is a long trip in bad weather, with bad views and no food. This gets assigned a measly 9. Other combinations end up somewhere in between, as a I judged them subjectively.

Similarly for the risk expression, a really bad outcome would be to crawl through a forest cold, exhausted and injured. I assign this a 500 badness rating. Should none of this happen, and we find our selves at the bottom of the risk rating, we still assign a 10, to signify the inherent risk of being in a forest. Other values are assigned subjectively as I compare them to the worst outcome.

Model verification and refinement

Having created the network, and having quantified probabilities and utilities, I thought there would still be a lot of room for improvement. I wondered as to whether I had under valued risk in my utility function. Seeing the results of various inputs, however, it seemed the network broadly agreed with my own assessments, and it seems difficult to pinpoint any specific improvement vector. These results can be seen in Table 1.

This does of course not take into assessment the weaknesses in the model previously specified. Making more variables continuous and basing variables on real statistics, could have interesting implications. I assume it would also result in a more useful model.

An interesting variation could be to set a threshold for the utility function, below which a third option comes in to play: neglecting the first choice and going home. This is left out for a reason, however, as I am unsure where this threshold would be reasonable to place for my own part. I leave this as an exercise to the reader.

Table 1: Utility outcomes from the setting of visible variables

Forecast	Sunny				Cloudy				Rainy			
Food	Yes		No		Yes		No		Yes		No	
Perc. Shape	Good	Bad	Good	Bad	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Long	**812**	**782**	**589**	559	**418**	**384**	289	255	93	56	42	5
Short	645	633	581	**569**	355	342	**315**	**302**	**116**	**103**	**95**	**82**