TDT4171 Artificial Intelligence Methods
Assigment 3

Odd André Owren

February 2020

Introduction

In this exercise, I was tasked with creating a decision support system for a decision problem of my own choice. Since this is supposed to be a system to help out in my day-to-day life, I decided to choose a decision problem that is relevant for exactly this. As a pretty active student, I enjoy working out a lot, especially in a gym with weights.

It is pretty normal to divide into 4 main muscle groups; chest, shoulders, back and legs when talking about working out with weights. Now, the next step is to create a model that uses a different set of variables to make a decision. There are two variables available to me at the time of decision, which is the predicted number of visitors at the gym and what the plan is to work out. With a goal of highest possible strength and muscle gain and the lowest risk for injury, which muscles should I work out today?

Causal model

The model that I ended up with is a static Bayesian network. The reason for a static model rather than a dynamic is purely because the only variable that really could impact our choice is whether or not the gym becomes full while working out and I would have to work out something else than planned or go home early. Overall, the amount of people at the gym stays approximately constant while working out.

When developing a causal model, it is common to begin with the utility, and "backtrack" from here. I have gone for a compound utility node which gives a utility based on two other utilities; risk of injury and strength gains. This allows me to model for the primary reason for me to work out, strength gains, while also accounting for injury risk for a given workout.

To quantify risk and strength gain, I have used 2 variables for each. To create a utility for risk I have assumed that this depends on exhaustion and technique during the workout. For strength gain this is assumed to depend on volume and preparation. Volume and exercise has been joined together as one variable, as these are closely related, and modelling these as individual variables would increase the uncertainty of the model.

Now, over to the other variables throughout the model. The variables and their relations, as well as assumptions, is listed below.

- **Exhaustion / Volume** is a variable that heavily effect the utility function for strength gain. Is is assumed that this variable depends on time, number of exercises and preparation. Longer workouts leads to lower volume, the same does fewer exercises. Something not as obvious is that a bad preparation leads to the need to push yourself harder, and a higher exhaustion and volume.

- **Technique** is a variable describing the form during a lift, which is pretty important for not getting injured. It is assumed that this variable is dependent on how much equipment is available and exhaustion. Improvising exercises due to low amount of equipment gives a higher probability of bad technique, and the same goes for high exhaustion.

- **Time** is a variable that I assume is dependent on the amount of available equipment, as well as the number of exercises and whether or not a partner joins the workout. Less available equipment, more exercises and a partner joining all makes it so that the workout takes more time.

- **Workout-partner and Preparation** are both variables that is assumed to be dependent on which workout is planned. Since there is some bias as to which muscle group is preferred by the workout partner and which muscle group is the easiest to prepare for, this has to be accounted for. As an example, a planned workout for chest increases the probability that a partner joins in, as well as the preparation is more likely to be good.

- **Available equipment and expected visitors** are both variables that is assumed to depend on the actual number of visitors at the gym. The actual number of visitors is a hidden variable, and the expected visitors variable is then a window to the hidden variable of number of visitors.

- **Number of exercises and planned workout** are the last two variables. These two variables are the defining ones of what workout is chosen. These variables can be described as "Given that a chest workout is chosen, what is the probability of planned workout being chest? How many exercises is it probable that is chosen in the workout?". This then creates the basis for the rest of the network as described above.

As presented in the introduction, there are only two visible variables at the time of decision, which is the expected amount of visitors and the planned workout for the day. At last, the whole network structure is pictured in figure 1

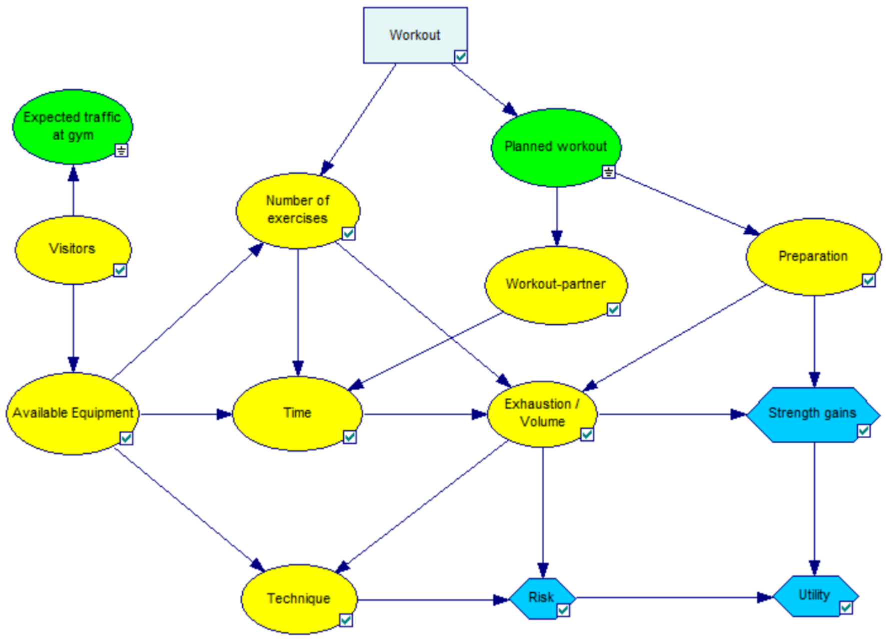

Figure 1: Bayesian network for the decision problem described above. Visible variables are green, hidden variables are yellow. Utility nodes are blue.

Probability assignment

Now, it is time to assign probability values to the model. Since there are a lot of hidden variables, a lot of this model depends on uncertainties that we need to account for. Most of these uncertainties are quantified from experience through a couple of years of working out. This also adds the need to refine the model over time, as the current decision model has pretty raw and biased probabilities.

One of the problems I encountered while modelling was that I wanted to model for the previous workout as to not repeat the same workout multiple days in a row. This turned into a bit of a hassle as this kind of model is complicated for representing a quantification of something along the lines of "If I worked out chest yesterday, what is the probability of working out chest today?". Instead, I accounted for this through the observable variable of planned workout. This way I am quantifying the probability of the chosen muscle group being the one that was planned, i.e. "Given that we choose to work out chest, what is the probability that we planned to work out chest?". This is a probability distribution that could be used as an expectancy for the choice, as the variables elsewhere in the model is what could make us change our choice of workout to what is not planned. Of course, sticking to a plan is important, so this distribution is pretty heavy towards sticking to the choice sticking to plan. The distribution table for Planned workout is as shown in table 1.

Workout → Planned ↓	Chest	Back	Shoulders	Legs
Chest	0.98	0.02	0.03	0
Back	0.01	0.97	0.01	0.03
Shoulders	0.01	0	0.95	0.01
Legs	0	0.01	0.01	0.96

Table 1: Probability distribution of planned workout, given a chosen workout

One problem with the assignment of probabilities is that there are in some places multiple levels of hidden variables, e.g. technique which is dependent on 3 layers of hidden variables through Number of exercises, Time and Exhaustion. This causes the uncertainty that comes from probabilities to be magnified through each layer. This makes the refinement of the model more difficult, as small changes to each layer here might make big changes in the resulting utility.

The probabilities for technique depends as mentioned on the amount of available equipment and exhaustion. I have also assumed that amount of exhaustion is more important than available equipment in whether or not the technique is good. The probability table for this is seen in figure 2 below.

Available Equipment	A lot			Some			Little		
Exhaustion / Volume	High	Medium	Low	High	Medium	Low	High	Medium	Low
Good	0.4	0.6	0.8	0.36	0.54	0.72	0.3	0.42	0.6
Bad	0.6	0.4	0.2	0.64	0.46	0.28	0.7	0.58	0.4

Table 2: Probabilities for good or bad technique given available equipment and Exhaustion

Utility assignment

The utility function is as mentioned above a compound utility derived from risk and strength gain. This is simply a subtraction model with

$$Utility = \text{Strength gains} - \mu \cdot \text{Risk}$$

where μ is a ratio between strength gains and risk, which can be adjusted to give a good model of how these utilities compare to each other when deciding a total utility. I chose to use a constant to adjust the risk, since the utilities of both risk and strength gains are given a value between 0 and 1500, as to keep these to a common standard. These can of course be adjusted, but I wanted to model the maximum and minimum in both cases as 1500 and 0 respectively. A workout with high volume and good preparation gives a strength utility of 1500, while low volume and bad preparation gets a 0, as you could just as well stay home at that point. Of course, a good workout also leads to exhaustion, and if this is combined with improvised exercises from lack of equipment, the risk is at 1500. The same case as previously holds here as well, if you don't exhaust yourself at all and you can use all the equipment as you wish, there are basically no risk with working out.

With my biased assignment of utility, I decided to put μ to 0.8, as I prioritise strength gains over potential risks, as stupid as that might turn out to be. As they say; No pain, no gain.

Verification and refinement

As far as I am concerned, this model is far from perfect. One of the problems is that the resulting utilities from each choice is too similar, which can be linked to the fact there are not enough variables that really distinguish the different choices that are available.

Although, with some refinement of the model, one could use it to make choices, but the utilities of each choice makes it hard to really be sure that the choice is obvious. If I set the evidence of expected visitors to the gym to "Many" and planned workout to "Chest", the different utilities ranges from 334.19 to 331.83. That is a measly 2.36 range with 4 different choices in-between. This makes it pretty obvious that the model is in need of refinement to be of further use, and it could also be an idea to either add more distinguishing variables or clarify the variables a bit more.

The rounded utilities for each visible variable is shown in table 3 below. From this table it is clear to see that the domain for this model might be too limited and indistinguishable between the choices, and therefore hard to model in a way that makes any sense. There seems to be a bias towards working out the chest as well. This might be a problem that comes from the fact that the creator of the model had a bit of a bias when assigning probabilities. This bias also changes when adjusting μ as defined previously for risk. This pushes the bias towards another muscle group, as the risk utility has a bias against chest as chest has a higher probability of high volume. For further refinement of the probabilities in the model, it would be useful with a stronger dataset that contains more precise data of output for some scenarios. At the moment the model is based on empirical observations, which leads to a great uncertainty in the model, and this combined with a difficult domain leads to a model that is some distance from being perfect.

Planned	Chest			Back			Shoulders			Legs		
Exp. visitors	Many	Some	Few	Many	Some	Few	Many	Some	Few	Many	Some	Few
Chest	334	354	362	304	324	332	182	203	211	205	225	234
Back	333	353	361	303	323	332	181	202	211	204	225	233
Shoulders	332	352	361	302	323	331	180	201	210	203	224	233
Legs	333	353	361	303	323	331	181	202	210	204	224	233

Table 3: Utilities for each combination of visible variables