Lecture 16 on Decision Theory and Decision Networks

Alice Gao

November 2, 2021

Contents

1 Learning Goals **2**

2 Introduction to Decision Theory **2**

3 Decision Network for Mail Pick-up Robot **2**
- 3.1 Variables . . . 3
- 3.2 Nodes in a decision network . . . 3
- 3.3 Arcs in a decision network . . . 4
- 3.4 The robot's utility function . . . 7

4 Evaluating the Robot Decision Network **8**
- 4.1 Choosing an action . . . 8
- 4.2 Calculating the expected utilities . . . 9
- 4.3 What should the robot do? . . . 11

5 Variable Elimination for a Single-Stage Decision Network **13**

1 Learning Goals

By the end of the lecture, you should be able to

- Model a one-off decision problem by constructing a decision network containing nodes, arcs, conditional probability distributions, and a utility function.
- Determine the optimal policy of a decision network by computing the expected utility of every policy.
- Determine the optimal policy of a decision network by applying the variable elimination algorithm.

2 Introduction to Decision Theory

In our discussion of Bayesian networks, we talked about *reasoning* in an uncertain world. Starting with decision theory, we will talk about *acting* in an uncertain world.

In one sentence, decision theory is probability theory plus utility theory.

Decision theory asks, how should an agent act in an uncertain world? To act, the agent needs to know what they should believe given some evidence. Probability theory answers this. The agent will also have to choose between the many possible decisions available to them. Utility theory answers this; for each possible state in the world, the agent's **utility function** assigns a real number representing the usefulness or desirability of that state.

The principle of maximum expected utility guides decision-making. It states that a rational agent should choose the action that maximizes the agent's expected utility.

This may sound like a blanket solution to AI, but maximizing expected utility is not a trivial task. We've already seen that performing probabilistic inference is NP-hard. As for calculating utility, an agent may not know the utility of a state immediately. They may need to do some kind of search to see which other states can be reached from that state.

In this unit, we will build on our inference-making tools to develop decision-making tools. We will use decision networks: these will combine a Bayesian network with nodes for actions and utilities.

3 Decision Network for Mail Pick-up Robot

Throughout this lesson, I will use a running example to show you how to construct a decision network.

> **Example:** A robot must choose its route to pick up the mail. There is a short route and long route. On the short route, the robot might slip and fall. The can put on pads. Pads won't change the probability of an accident. However, if an accident happens, pads will reduce the damage. Unfortunately, the pads add weight and slow

the robot down. The robot would like to pick up the mail as quickly as possible while minimizing the damage caused by an accident.

What should the robot do?

3.1 Variables

To construct the decision network, we will need both random variables and decision variables (actions).

Random variables represent events that are out of our control. The random variables we will need are:

- A: whether an accident occurs or not.

Decision variables represent events that are in our control. The decision variables we will need are:

- S: whether the robot chooses the short route.
- P: whether the robot puts on pads.

3.2 Nodes in a decision network

We will use three types of nodes in a decision network:

- Chance nodes

  represent random variables (as in Bayesian networks).
- Decision nodes

  represent actions (decision variables).
- The

  Utility node

  represents the agent's utility function on states (happiness in each state).

Both chance nodes and decision nodes can influence the agent's current state. Because the utility node depends on the current state, both chance nodes and decision nodes will influence the utility node.

Problem: Draw the chance nodes, decision nodes, and utility node for the robot decision network.

Solution: We have one chance node for A (whether an accident occurs). We have two decision nodes: one for S (whether the robot chooses the short route) and one for P (whether the robot puts on pads). Finally, we have the utility node.

3.3 Arcs in a decision network

Now that we have the nodes in our decision network, we need to connect them. First, we will consider the chance nodes and decision nodes.

Problem: How do the random variables and the decision variables relate to one another?

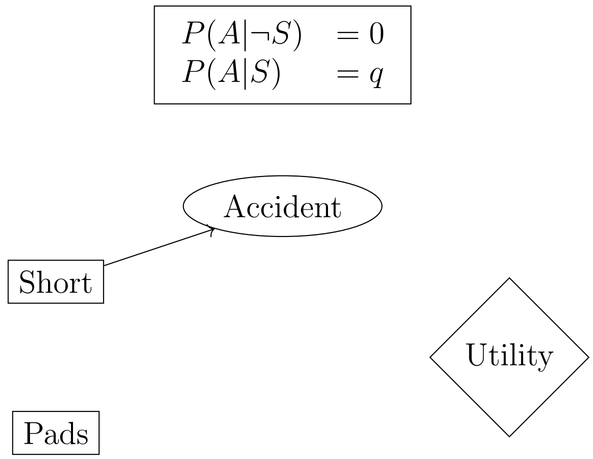
Solution: Firstly, the robot must make both decisions before it can observe whether an accident happens. Depending on the route, there may not even be a non-zero probability for an accident. Also, the robot must choose whether to put on pads before it goes down any route.

In terms of time, the decision nodes come before the chance nodes. Given this, there may be arrows from the decision nodes to the chance nodes, but there cannot be arrows in the other direction. Let's consider which decisions affect the chance node here.

Short does affect Accident here. On the long route, an accident will not occur. On the short route, an accident may occur.

Pads does not affect Accident here. Wearing pads will only affect the severity of damage if an accident occurs.

Based on this analysis, we can connect the nodes in the decision network. Similarly to how we have a conditional probability distribution attached to every node in a Bayesian network, we will also have one such distribution attached to every chance node in the decision network. Note that the chance nodes do not get affected by other chance nodes, but by decision nodes. There is a fixed probability that an accident will occur on the short route, but we do not know what this number is. I'll let this probability be $0 \leq q \leq 1$.



To address the utility node, we need to answer two questions: which variables influence the utility node, and how do these variables influence it?

Problem: Which variables influence the robot's happiness?

(A) P only
(B) S only
(C) A only
(D) Two of (A), (B), and (C)
(E) All of (A), (B), and (C)

A robot must choose its route to pick up the mail. There is a short route and long route. On the short route, the robot might slip and fall. The can put on pads. Pads won't change the probability of an accident. However, if an accident happens, pads will reduce the damage. Unfortunately, the pads add weight and slow the robot down. The robot would like to pick up the mail as quickly as possible while minimizing the damage caused by an accident.

Solution: The route the robot takes will influence the robot's happiness, since the robot would like to pick up the mail as quickly as possible.

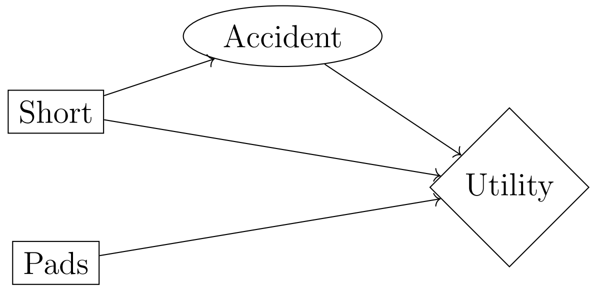
Whether to put on pads or not will also influence the robot's happiness, since the robot would like to minimize the damage caused by any accidents and pads would reduce such damage.

Finally, whether an accident occurs or not will also influence the robot's happiness, since the accident would cause damage.

The correct answer is (E).

Let's update our decision network to reflect these relationships.

$$\begin{aligned} P(A|\neg S) &= 0 \\ P(A|S) &= q \end{aligned}$$



Answering *how* variables affect the utility function is difficult. To help us get started, I'll begin with a clicker question. Then I will give one example of a utility function and reason through the numbers presented.

Problem: When an accident does NOT happen, which of the following is true?

(A) The robot prefers not wearing pads to wearing pads.

(B) The robot prefers the long route over the short route.

(C) Both (A) and (B) are true.

(D) Both (A) and (B) are false.

Solution: For statement (A), recall that we don't need pads to reduce the severity of the damage since there is no damage. Wearing pads also slows down the robot, so the robot would prefer to not wear them. Thus (A) is true.

For statement (B), the short route is better because there is no accident causing damage to the robot. Thus (B) is false.

The correct answer is (A).

We can do similar analyses for other states. I won't cover all of these in depth; instead, let's look at one possible utility function.

3.4 The robot's utility function

Recall a utility function maps states in the world to real numbers representing how useful each state is to the agent. Here is one possibility for the robot, where each state is labeled w_i and the utility function is $U(w_i)$:

	State	$U(w_i)$
$\neg P, \neg S, \neg A$	w_0 slow, no weight	6
$\neg P, \neg S, A$	w_1 impossible	
$\neg P, S, \neg A$	w_2 quick, no weight	10
$\neg P, S, A$	w_3 severe damage	0
$P, \neg S, \neg A$	w_4 slow, extra weight	4
$P, \neg S, A$	w_5 impossible	
$P, S, \neg A$	w_6 quick, extra weight	8
P, S, A	w_7 moderate damage	2

To analyse this utility function, we will consider two cases: when an accident does not happen, and when an accident does happen.

Case 1: An accident does not happen

This corresponds to w_0, w_2, w_4, and w_6.

Problem: When an accident does not happen, does the robot prefer not wearing pads or wearing pads?

Solution: The robot prefers not wearing pads because it can move faster. We expect $U(w_0) > U(w_4)$ and $U(w_2) > U(w_6)$.

Problem: When an accident does not happen, does the robot prefer the short route or the long route?

Solution: The robot prefers the short route because it is faster. We expect $U(w_6) > U(w_4)$ and $U(w_2) > U(w_0)$.

Considering both of these questions should allow us to order the utilities:

$$U(w_2) > U(w_6) > U(w_0) > U(w_4).$$

Case 2: An accident does happen

This corresponds to w_1, w_3, w_5, and w_7.

Problem: When an accident does happen, does the robot prefer the short route or the long route?

Solution: This is a bit of a trick question! If an accident occurs, the robot must have taken the short route. Thus there is no defined utility for w_1 and w_5.

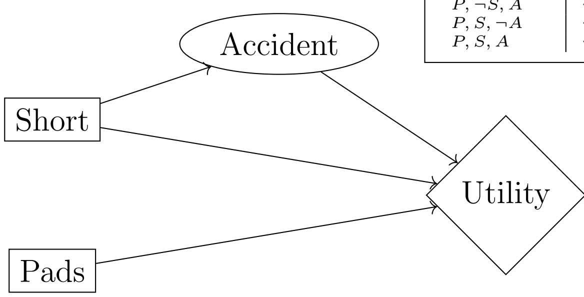
Problem: When an accident does happen, does the robot prefer not wearing pads or wearing pads?

Solution: The robot prefers wearing pads because pads reduce the severity of the damage. We expect $U(w_7) > U(w_3)$.

The complete decision network, including a utility function, is as below:

$$P(A|\neg S) = 0$$
$$P(A|S) = q$$

	State	$U(w_i)$
$\neg P, \neg S, \neg A$	w_0 slow, no weight	6
$\neg P, \neg S, A$	w_1 impossible	
$\neg P, S, \neg A$	w_2 quick, no weight	10
$\neg P, S, A$	w_3 severe damage	0
$P, \neg S, \neg A$	w_4 slow, extra weight	4
$P, \neg S, A$	w_5 impossible	
$P, S, \neg A$	w_6 quick, extra weight	8
P, S, A	w_7 moderate damage	2



4 Evaluating the Robot Decision Network

4.1 Choosing an action

Choosing an action will be similar to the ideas of the variable elimination algorithm for Bayesian networks:

1. Set evidence variables according to current state
2. For each possible value of a decision node:
   (a) Set the decision node to that value
   (b) Calculate the posterior probability for parent nodes of the utility node
   (c) Calculate expected utility for the action
3. Return action with highest expected utility

In step 2, we are essentially iterating over all possible sets of decisions or actions and calculating an expected utility for each set. Step 2(b) refers to the parent *chance* nodes, since decision nodes are fixed. Note that we can only calculate an *expected* utility since the chance nodes introduce uncertainty.

4.2 Calculating the expected utilities

In the robot example, we have two binary decision variables, so we will have four possible actions. For the first action, I will provide some explanation for the derivation. For the other three actions, I encourage you to work through the calculations yourself, but I will still include the solutions.

Problem: What is the agent's expected utility of not wearing pads and choosing the long route?

Solution: The expected utility, which I will call $EU(a)$, is a summation over all possible states where the actions a were taken. In this case, we need $EU(\neg P \wedge \neg S)$, which corresponds to the states w_0 and w_1.

$$\begin{aligned} EU(\neg P \wedge \neg S) &= P(w_0|\neg P \wedge \neg S) * U(w_0) \\ &\quad + P(w_1|\neg P \wedge \neg S) * U(w_1) \\ &= P(\neg P \wedge \neg S \wedge \neg A|\neg P \wedge \neg S) * U(w_0) \\ &\quad + P(\neg P \wedge \neg S \wedge A|\neg P \wedge \neg S) * U(w_1) \end{aligned}$$

We can simplify the conditional expressions since $\neg P \wedge \neg S$ appears in both the query and the evidence.

$$= P(\neg A|\neg P \wedge \neg S) * U(w_0) + P(A|\neg P \wedge \neg S) * U(w_1)$$

P and A are independent, so we can remove $\neg P$ from the evidence as well.

$$= P(\neg A|\neg S) * U(w_0) + P(A|\neg S) * U(w_1)$$

At this point, we can read off the utilities from the table. $U(w_1)$ is undefined, which I've indicated with a dash, but this is not a problem since the probability of observing w_1 must be zero anyway!

$$\begin{aligned} &= (1)(6) + (0)(—) \\ &= 6. \end{aligned}$$

Take some time to answer these next questions yourself. Note that the expressions may be functions of q, since we did not specify the value of q in the decision network.

Problem: What is the agent's expected utility of not wearing pads and choosing the short route?

Solution:

$$\begin{aligned}
EU(\neg P, S) &= P(w_2|\neg P \wedge S) * U(w_2) \\
&\quad + P(w_3|\neg P \wedge S) * U(w_3) \\
&= P(\neg P \wedge S \wedge \neg A|\neg P \wedge S) * U(w_2) \\
&\quad + P(\neg P \wedge S \wedge A|\neg P \wedge S) * U(w_3) \\
&= P(\neg A|\neg P \wedge S) * U(w_2) \\
&\quad + P(A|\neg P \wedge S) * U(w_3) \\
&= P(\neg A|S) * U(w_2) \\
&\quad + P(A|S) * U(w_3) \\
&= (1-q)(10) + (q)(0) \\
&= 10 - 10q.
\end{aligned}$$

Problem: What is the agent's expected utility of wearing pads and choosing the long route?

Solution:

$$\begin{aligned}
EU(P, \neg S) &= P(w_4|P \wedge \neg S) * U(w_4) \\
&\quad + P(w_5|P \wedge \neg S) * U(w_5) \\
&= P(P \wedge \neg S \wedge \neg A|P \wedge \neg S) * U(w_4) \\
&\quad + P(P \wedge \neg S \wedge A|P \wedge \neg S) * U(w_5) \\
&= P(\neg A|P \wedge \neg S) * U(w_4) \\
&\quad + P(A|P \wedge \neg S) * U(w_5) \\
&= P(\neg A|\neg S) * U(w_4) \\
&\quad + P(A|\neg S) * U(w_5) \\
&= (1)(4) + (0)(\text{---}) \\
&= 4.
\end{aligned}$$

Problem: What is the agent's expected utility of wearing pads and choosing the short route?

Solution:

$$\begin{aligned}
EU(P, S) &= P(w_6|P \wedge S) * U(w_6) \\
&\quad + P(w_7|P \wedge S) * U(w_7) \\
&= P(P \wedge S \wedge \neg A|P \wedge S) * U(w_6) \\
&\quad + P(P \wedge S \wedge A|P \wedge S) * U(w_7) \\
&= P(\neg A|P \wedge S) * U(w_6) \\
&\quad + P(A|P \wedge S) * U(w_7) \\
&= P(\neg A|S) * U(w_6) \\
&\quad + P(A|S) * U(w_7) \\
&= (1 - q)(8) + (q)(2) \\
&= 8 - 6q.
\end{aligned}$$

4.3 What should the robot do?

To summarize, here are the four expected utilities:

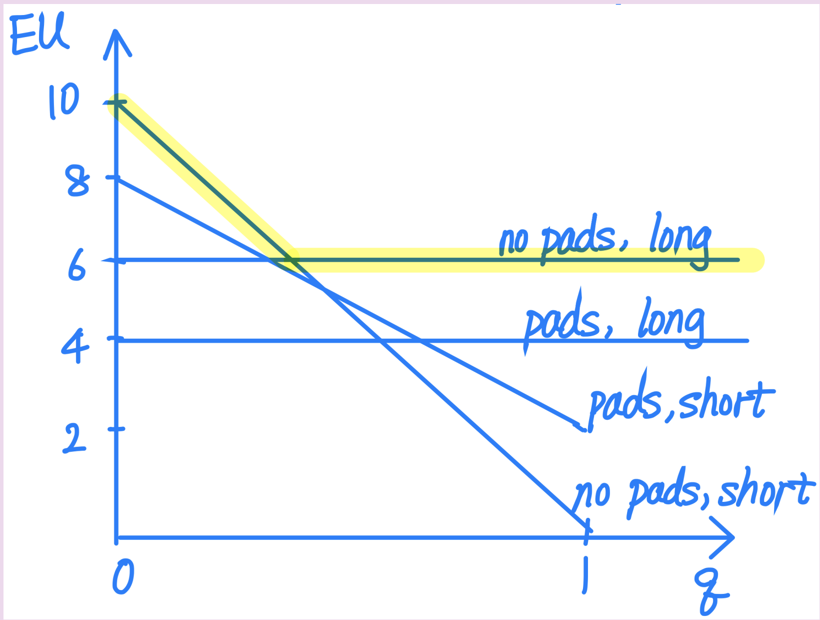
$$\begin{aligned}
EU(\neg P \wedge \neg S) &= 6 \\
EU(\neg P \wedge S) &= 10 - 10q \\
EU(P \wedge \neg S) &= 4 \\
EU(P \wedge S) &= 8 - 6q
\end{aligned}$$

Problem: Should the robot wear pads or not, and should it choose the short route or the long route?

Solution: You might not know how to answer this since we have expressions involving q here. The third option is clearly dominated by the first (which is intuitive — we should not wear pads on the long route because there won't be any chance of damage), but beyond that it is hard to decide what to do.

Fortunately, we can use some math. Since q varies from 0 to 1, we can graph the expected utilities against q and determine the best action as q varies. Try drawing this graph yourself and come to a conclusion on your own before reading on.

Here is the graph:



The robot should choose the set of actions which maximizes expected utility for each q. Looking at the graph, this would be not wearing pads and choosing the short route for some smaller q, or not wearing pads and choosing the long route for larger q. To determine which q these actions correspond to exactly, we can solve for the intersection of the two lines:

$$10 - 10q = 6 \implies q = \frac{2}{5}.$$

In decision theory, we often refer to the solution to this kind of problem as a **policy**. The optimal policy here is:

- if $q \leq \frac{2}{5}$, no pads, short route
- if $q > \frac{2}{5}$, no pads, long route

What can we learn about this optimal policy?

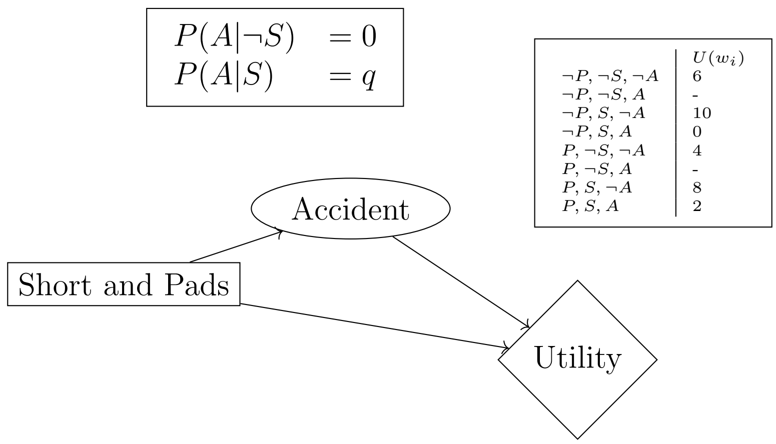
Essentially, it says that we are OK with taking the short route if the chance of an accident is small, but that we should avoid the risk altogether if the chance of an accident is large. This seems pretty intuitive.

Interpreting the choice of pads or no pads is more difficult. Wearing pads might be a good choice on the short route, but this policy decides to never wear pads. Why not? Intuitively, it might be the case that when the chance of an accident is small enough, the burden of wearing pads is too much compared to the potential damage without the pads. The graph reflects this: comparing the lines for choosing the short route, we can see that the expected utility of wearing pads only surpasses the expected utility of not wearing pads for $q > \frac{1}{2}$.

5 Variable Elimination for a Single-Stage Decision Network

In the previous sections, I showed you how to make a decision by calculating the expected utilities by directly calculating conditional probabilities. You can also make a decision using the variable elimination algorithm.

To start, we simplify our decision network by combining the decision nodes into one since the decisions occur independently. Here, we will have one decision node containing the cross product of Short and Pads. Technically, we will have to change the Accident and utility nodes accordingly, but I will omit it for simplicity. The revised network looks like this:



The steps for performing variable elimination in such a single-stage (only one decision node) decision network are:

1. Prune all the nodes that are not ancestors of the utility node.
2. Sum out all chance nodes.
3. For the single remaining factor, return the maximum value and the assignment that gives the maximum value.

Let's carry out this algorithm on our revised network. First, we will define factors for any tables present in the network: these are any conditional probabilities and the utility function.

Accident
$f_1(A, S \wedge P)$:

A	$S \wedge P$	val
t	$S \wedge P$	q
f	$S \wedge P$	$1 - q$
t	$S \wedge \neg P$	q
f	$S \wedge \neg P$	$1 - q$
t	$\neg S \wedge P$	0
f	$\neg S \wedge P$	1
t	$\neg S \wedge \neg P$	0
f	$\neg S \wedge \neg P$	1

Utility function
$u(A, S \wedge P)$:

A	$S \wedge P$	val
t	$S \wedge P$	2
f	$S \wedge P$	8
t	$S \wedge \neg P$	0
f	$S \wedge \neg P$	10
t	$\neg S \wedge P$	−
f	$\neg S \wedge P$	4
t	$\neg S \wedge \neg P$	−
f	$\neg S \wedge \neg P$	6

Then, we will sum out the chance nodes.

Multiply the two factors.

$f_2(A, S \wedge P)$:

A	$S \wedge P$	val
t	$S \wedge P$	$2q$
f	$S \wedge P$	$8 - 8q$
t	$S \wedge \neg P$	0
f	$S \wedge \neg P$	$10 - 10q$
t	$\neg S \wedge P$	0
f	$\neg S \wedge P$	4
t	$\neg S \wedge \neg P$	0
f	$\neg S \wedge \neg P$	6

Sum out A from f_2.

$f_3(S \wedge P)$:

$S \wedge P$	val
$S \wedge P$	$8 - 6q$
$S \wedge \neg P$	$10 - 10q$
$\neg S \wedge P$	4
$\neg S \wedge \neg P$	6

f_3 should look familiar to you—these are exactly the expected utilities we calculated earlier. At this point, we can perform the same analysis as before by varying q.

There are two main reasons for me showing you this second method. One is to show you that we can approach the problem in different ways. The second is to show you that a decision network is not very much different from a Bayesian network, and that the variable elimination algorithm still works for decision networks.