Decision Theory: Sequential Decisions

CPSC 322 – Decision Theory 2

Textbook §9.3

Lecture Overview

1. Recap
2. Sequential Decisions

Decision Variables

- Decision variables are like random variables that an agent gets to choose the value of.
- A possible world specifies the value for each decision variable and each random variable.
- For each assignment of values to all decision variables, the measures of the worlds satisfying that assignment sum to 1.
- The probability of a proposition is undefined unless you condition on the values of all decision variables.

Single-Stage decisions

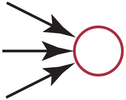
- Given a single decision variable, the agent can choose $D = d_i$ for any $d_i \in dom(D)$.
- The expected utility of decision $D = d_i$ is $\mathbb{E}(U|D = d_i)$.
- An optimal single decision is the decision $D = d_{max}$ whose expected utility is maximal:

$$d_{max} = \arg\max_{d_i \in dom(D)} \mathbb{E}(U|D = d_i).$$

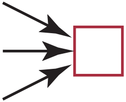
Decision Networks

- A decision network is a graphical representation of a finite sequential decision problem.
- Decision networks extend belief networks to include decision variables and utility.
- A decision network specifies what information is available when the agent has to act.
- A decision network specifies which variables the utility depends on.

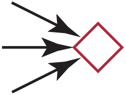
Decision Networks



- A random variable is drawn as an ellipse. Arcs into the node represent probabilistic dependence.



- A decision variable is drawn as an rectangle. Arcs into the node represent information available when the decision is made.



- A value node is drawn as a diamond. Arcs into the node represent values that the value depends on.

Finding the optimal decision

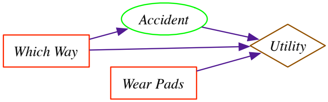
- Suppose the random variables are $X_1, \ldots, X_n$, and utility depends on $X_{i_1}, \ldots, X_{i_k}$

$$\begin{aligned}\mathbb{E}(U|D) &= \sum_{X_{1,\ldots,X_n}} P(X_1, \ldots, X_n|D)U(X_{i_1}, \ldots, X_{i_k}) \\ &= \sum_{X_{1,\ldots,X_n}} \prod_{i=1}^{n} P(X_i|pX_i, D)U(X_{i_1}, \ldots, X_{i_k})\end{aligned}$$

To find the optimal decision:

- Create a factor for each conditional probability and for the utility
- Sum out all of the random variables
- This creates a factor on D that gives the expected utility for each D
- Choose the D with the maximum value in the factor.

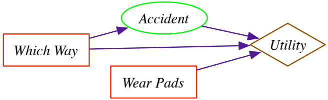
Example Initial Factors



Which Way	Accident	Probability
long	true	0.01
long	false	0.99
short	true	0.2
short	false	0.8

Which Way	Accident	Wear Pads	Utility
long	true	true	30
long	true	false	0
long	false	true	75
long	false	false	80
short	true	true	35
short	true	false	3
short	false	true	95
short	false	false	100

Example Initial Factors



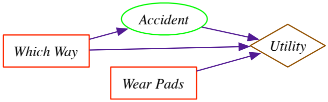
Which Way	Accident	Probability
long	true	0.01
long	false	0.99
short	true	0.2
short	false	0.8

Which Way	Accident	Wear Pads	Utility
long	true	true	30
long	true	false	0
long	false	true	75
long	false	false	80
short	true	true	35
short	true	false	3
short	false	true	95
short	false	false	100

Sum out Accident:

Which Way	Wear pads	Value
long	true	0.01*30+0.99*75=74.55
long	false	0.01*0+0.99*80=79.2
short	true	0.2*35+0.8*95=83
short	false	0.2*3+0.8*100=80.6

Example Initial Factors



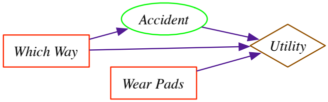
Which Way	Accident	Probability
long	true	0.01
long	false	0.99
short	true	0.2
short	false	0.8

Which Way	Accident	Wear Pads	Utility
long	true	true	30
long	true	false	0
long	false	true	75
long	false	false	80
short	true	true	35
short	true	false	3
short	false	true	95
short	false	false	100

Sum out Accident:

Which Way	Wear pads	Value
long	true	0.01*30+0.99*75=74.55
long	false	0.01*0+0.99*80=79.2
short	true	0.2*35+0.8*95=83
short	false	0.2*3+0.8*100=80.6

Example Initial Factors



Which Way	Accident	Probability
long	true	0.01
long	false	0.99
short	true	0.2
short	false	0.8

Which Way	Accident	Wear Pads	Utility
long	true	true	30
long	true	false	0
long	false	true	75
long	false	false	80
short	true	true	35
short	true	false	3
short	false	true	95
short	false	false	100

Sum out Accident:

Which Way	Wear pads	Value
long	true	0.01*30+0.99*75=74.55
long	false	0.01*0+0.99*80=79.2
short	true	0.2*35+0.8*95=83
short	false	0.2*3+0.8*100=80.6

Thus the optimal policy is to take the short way and wear pads, with an expected utility of 83.

Lecture Overview

1. Recap
2. Sequential Decisions

Sequential Decisions

- An intelligent agent doesn't make a multi-step decision and carry it out without considering revising it based on future information.
- A more typical scenario is where the agent: observes, acts, observes, acts, ...
  - just like your final homework!
- Subsequent actions can depend on what is observed.
  - What is observed depends on previous actions.
- Often the sole reason for carrying out an action is to provide information for future actions.
  - For example: diagnostic tests, spying.

Sequential decision problems

- A sequential decision problem consists of a sequence of decision variables $D_1, \ldots, D_n$.
- Each D_i has an information set of variables pD_i, whose value will be known at the time decision D_i is made.

- What should an agent do?
  - What an agent should do at any time depends on what it will do in the future.
  - What an agent does in the future depends on what it did before.

Policies

- A policy specifies what an agent should do under each circumstance.
- A policy is a sequence $\delta_1, \ldots, \delta_n$ of decision functions

$$\delta_i : dom(pD_i) \rightarrow dom(D_i).$$

  This policy means that when the agent has observed $O \in dom(pD_i)$, it will do $\delta_i(O)$.

Expected Value of a Policy

- Possible world ω satisfies policy δ, written $\omega \models \delta$, if the world assigns the value to each decision node that the policy specifies.
- The expected utility of policy δ is

$$\mathbb{E}(U|\delta) = \sum_{\omega \models \delta} P(\omega)U(\omega)$$

- An optimal policy is one with the highest expected utility:

$$\delta^* \in \arg\max_{\delta} \mathbb{E}(U|\delta).$$