T-79.5102 **Autumn 2008**
Special Course in Computational Logic
Tutorial 9
Solutions

1. Given an ordinary reward function $R(s)$ over states, the transition probability table $T(s, a, s')$, and the discount factor γ, the Bellman equation gives utilities for individual states:

$$U(s) = R(s) + \gamma \max_a \sum_{s'} T(s, a, s')U(s'). \tag{1}$$

(a) To use a reward function $R(s, a)$ that depends also on the action taken, we have to push the reward inside maximisation:

$$U(s) = \max_a (R(s, a) + \gamma \sum_{s'} T(s, a, s')U(s')). \tag{2}$$

Further reorganisation is required, if a reward function $R(s, a, s')$ that depends also on the outcome state is introduced:

$$U(s) = \max_a \sum_{s'} T(s, a, s')(R(s, a, s') + \gamma U(s')). \tag{3}$$

In particular, note that rewards are not subject to discounting.

(b) The question is how (**??**) can be viewed as a special case of (**??**). This is achieved by a reward function

$$R'(s, a) = \sum_{s'} T(s, a, s')R(s, a, s') \tag{4}$$

which is obtained from $R(s, a, s')$ by weighting individual rewards with the respective transition probabilities.

(c) The reward $R'(s)$ for a particular state depends on the action taken in that state, i.e., $R'(s) = R(s, \alpha)$ where

$$\alpha = \arg\max_a (R(s, a) + \gamma \sum_{s'} T(s, a, s')U(s')). \tag{5}$$

2. There are no real actions involved in the problem statement: the owner merely perceives changes in the state of the environment indirectly by observing draught inside or the alarm. Thus we do not need POMDPs to model this domain but prediction techniques from Chapter 15 suffice.

The update of the current set of beliefs consists of two steps:

- In the *estimation* step the current set of beliefs, i.e., the distribution $B(\mathbf{X}_t)$, is used to estimate the state of the world at the next time step $t + 1$. This probability distribution is

$$B'(\mathbf{X}_{t+1}) = \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t)B(\mathbf{x}_t)$$

where $\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{X}_t)$ gives the transition model.

- In the *assessment* step, the observations of the time step $t + 1$, i.e., $\mathbf{e}_{t+1}$ is taken into account. They are incorporated to $B'(\mathbf{X}_{t+1})$ using Bayesian updating:

$$B(\mathbf{X}_{t+1}) = \alpha \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})B'(\mathbf{X}_{t+1})$$

where α scales probabilities so that they sum up to 1.

The transition probabilities given in the problem statement can be nicely represented in a tabular form:

	f_{t+1}	$\neg f_{t+1}$
f_t	0.6	0.4
$\neg f_t$	0.1	0.9

	k_{t+1}	$\neg k_{t+1}$
k_t	0.6	0.4
$\neg k_t$	0.3	0.7

For the sake of simplicity, we assume that F and K are independent, i.e., F_{t+1} depends on F_t and K_{t+1} depends on K_t (in the respective BN representation, we would have arrows $F_t \longrightarrow F_{t+1}$ and $K_t \longrightarrow K_{t+1}$).

We first estimate the state of the system at time $t = 1$ given the probability distribution for $t = 0$, i.e., $B(f_0) = 0.1$ and $B(k_0) = 0.3$:

$$\begin{aligned}
B'(f_1) &= P(f_1 \mid \neg f_0)B(\neg f_0) + P(f_1 \mid f_0)B(f_0) \\
&= 0.1 \cdot 0.9 + 0.6 \cdot 0.1 = 0.15 \\
B'(\neg f_1) &= P(\neg f_1 \mid \neg f_0)B(\neg f_0) + P(\neg f_1 \mid f_0)B(f_0) \\
&= 0.85 = 1 - B'(f_1) \\
B'(k_1) &= P(k_1 \mid \neg k_0)B(\neg k_0) + P(k_1 \mid k_0)B(k_0) \\
&= 0.39 \\
B'(\neg k_1) &= P(\neg k_1 \mid \neg k_0)B(\neg k_0) + P(\neg k_1 \mid k_0)B(k_0) \\
&= 0.61 = 1 - B'(k_1)
\end{aligned}$$

Then at $t = 1$ the cat owner observes that both doors are open — corresponding to evidence $w_1 \wedge a_1$. Next we compute how probable it is that the front door is open at $t = 1$ given the owner's observation w_1:

$$\begin{aligned}
B(f_1) &= \alpha_{f_1} P(w_1 \mid f_1)B'(f_1) \\
&= \alpha_{f_1} \cdot 0.9 \cdot 0.15 = 0.135\alpha_{f_1} \\
B(\neg f_1) &= \alpha_{f_1} P(w_1 \mid \neg f_1)B'(\neg f_1) \\
&= \alpha_{f_1} \cdot 0.1 \cdot 0.85 = 0.085\alpha_{f_1}
\end{aligned}$$

The normalisation constant is α_{f_1} can be computed from the equation

$$\alpha_{f_1} = \tfrac{1}{0.135+0.085} \approx 4.54.$$

In this way, we obtain beliefs $B(f_1) \approx 0.61$ and $B(\neg f_1) \approx 0.39$ and we perform analogous calculations for the kitchen door:

$$\begin{aligned}
B(k_1) &= \alpha_{k_1} P(a_1 \mid k_1)B'(k_1) \\
&= 0.27 \cdot \alpha_{k_1} \\
B(\neg k_1) &= \alpha_{k_1} P(a_1 \mid \neg k_1)B'(\neg k_1) \\
&= 0.06 \cdot \alpha_{k_1} \\
\alpha_{k_1} &= \tfrac{1}{0.27+0.06} \approx 3.03 \\
B(k_1) &\approx 0.82 \\
B(\neg k_1) &\approx 0.18
\end{aligned}$$

This concludes the determination of beliefs for $t = 1$. Now we use exactly the same procedure to compute the beliefs at $t = 2$ when the cat owner observes no draught nor alarm (i.e., pieces of evidence $\neg w_2$ and $\neg a_2$):

$$\begin{aligned}
B'(f_2) &= P(f_2 \mid f_1)B(f_1) + P(f_2 \mid \neg f_1)B(\neg f_1) \\
&= 0.6 \cdot 0.61 + 0.39 \cdot 0.1 \approx 0.41 \\
B'(\neg f_2) &\approx 0.59 \\
B(f_2) &= \alpha_{f_2} P(\neg w_2 \mid f_2)B'(f_2) \\
&= 0.04\alpha_{f_2} \\
B(\neg f_2) &= \alpha_{f_2} P(\neg w_2 \mid \neg f_2)B'(\neg f_2) \\
&= 0.53\alpha_{f_2}
\end{aligned}$$

To perform normalisation, we calculate $\alpha_{f_2} = \frac{1}{0.04+0.53}$ which yields probability values $B(f_2) \approx 0.07$ and $B(\neg f_2) \approx 0.93$. For k_2, we get

$$\begin{aligned} B'(k_2) &= P(k_2 \mid k_1)B(k_1) + P(k_2 \mid \neg k_1)B(\neg k_1) \\ &\approx 0.55 \\ B'(\neg k_2) &\approx 0.45 \\ B(k_2) &= \alpha_{k_2} P(\neg w_2 \mid k_2)B'(k_2) \\ &= 0.16\alpha_{k_2} \\ B(\neg k_2) &= \alpha_{k_2} P(\neg w_2 \mid \neg k_2)B'(\neg k_2) \\ &= 0.41\alpha_{k_2} \end{aligned}$$

Thus $\alpha_{k_2} = \frac{1}{0.16+0.41}$, $B(k_2) \approx 0.28$, and $B(\neg k_2) \approx 0.72$. To answer the question given in the problem statement, we should determine the probability that at least one of the doors is closed at $t = 2$:

$$B(\neg(f_2 \wedge k_2)) = 1 - B(k_2)B(f_2) = 1 - 0.0196 \approx 0.98.$$