COMP 424 Artificial Intelligence

Tutorial 4: HMMs, MDPs, and Utility

Questions

Hidden Markov Models

Consider an HMM used by a TA to determine if a student who attends a tutorial understands a topic. A student either understands the material, or is confused. The TA can observe that the student is silent, or asking questions.

The prior probability that a student came to the tutorial already understanding the material is 0.3. At each time step, there is a 40% chance that a confused student will start to understand the material. If the student already understands, there is a 80% chance that they will continue to understand in the next time step. The probability of a student who understands the material asking questions is 0.1. If they don't understand, it's a 50/50 chance whether they ask questions or stay silent.

1. Give the parameters of the model
2. Complete this trellis for the forward algorithm.

	Silent	Asking	Silent	Silent
Understands	0.27	-	0.0732	0.0683
Confused	-	0.132	0.0432	-

3. For the above observation sequence, what is the probability that the student understood the material after the third time step?

Utility

For the summer break, Alice wants to take a trip to Chicago, but she has no car. A plane ticket costs $400, and the advertised travel time is 2 hours. $200 can get her a ticket on an 8 hour train trip. Or, if she's feeling particularly masochistic, Alice can spend 18 hours on a bus for only $50.

Alice is a very busy woman, and time is money. Every hour she spends on the commute costs her another $20 of wasted opportunity, and though the advertised travel times seem reasonable, there's a 40% chance that there will be traffic, which doubles the time for all three modes of travel.

1. Draw a decision graph for this problem.
2. To minimize the cost of her trip, which ticket should Alice buy?
3. How much should Alice be willing to pay a fortune teller to tell her with 100% certainty whether there will be traffic on her commute?
4. Now suppose Alice can also choose between a few dates for her trip. Tickets will be more expensive for earlier dates, but the later dates are on holiday weekends, when there may be more traffic on all forms of transportation. Draw the decision graph that reflects this change.

Markov Decision Processes

Consider the following (very) simple grid. There are three states, and two possible actions an agent can take, left and right. Transitions are stochastic; an action succeeds with probability 0.7, fails (no movement) with probability 0.2, and results in moving in the opposite direction with probability 0.1. If a direction is blocked, the agent does not move. Taking any action in a state results in the reward for that state, i.e. $R(s_2, \mathit{Left}) = R(s_2, \mathit{Right}) = -10$.

s_1	s_2 -10	s_3 +100

1. Assuming the initial policy is $\pi^0(s) = \mathit{Left}\ \forall s \in S$, and initial $V_0(s) = 0\ \forall s \in S$, perform 4 iterations of policy evaluation with discount factor $\gamma = 0.9$.
2. Given this estimate for V^{π_0}, what policy π' will an iteration of policy improvement give?
3. Suppose the episode ends when the agent reaches the rightmost state. How will this change the policy iteration?

COMP 424 Artificial Intelligence

Tutorial 4: HMMs, MDPs, and Utility

Solutions

Hidden Markov Models

1. Give the parameters of the model

Initial Probabilities $\pi_i = Pr(X_1 = i)$	
Und	0.3
Conf	0.7

Transition Probabilities $a_{ij} = P(X_{t+1} = j \mid X_t = i)$		
	Und	Conf
Und	0.8	0.2
Conf	0.4	0.6

Emission Probabilities $b_{ik} = P(E_t = k \mid X_t = i)$		
	Ask	Sil
Und	0.1	0.9
Conf	0.5	0.5

2. Complete this trellis for the forward algorithm.

	Silent	Asking	Silent	Silent
Understands	0.27	**0.0356**	0.0732	0.0683
Confused	**0.35**	0.132	0.0432	**0.0203**

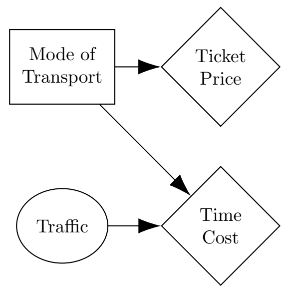
$$\begin{aligned}
\alpha_2(1) &= \pi_2 \cdot b_2(2) \\
&= 0.7 \cdot 0.5 \\
&= 0.35 \\
\alpha_1(2) &= b_1(1)(\alpha_1(1) \cdot a_{11} + \alpha_2(1) \cdot a_{21}) \\
&= 0.1 \cdot (0.27 \cdot 0.8 + 0.35 \cdot 0.4) \\
&= 0.0356 \\
\alpha_2(4) &= b_2(2)(\alpha_1(3) \cdot a_{12} + \alpha_2(3) \cdot a_{22}) \\
&= 0.5 \cdot (0.0732 \cdot 0.2 + 0.0432 \cdot 0.6) \\
&= 0.02028
\end{aligned}$$

3. For the above observation sequence, what is the probability that the student understood the material after the third time step?

$$\begin{aligned}
Pr(X_3 = \text{Und} \mid \mathbf{E}_{1:3},\ \theta) &= \frac{Pr(X_3 = \text{Und},\ \mathbf{E}_{1:3} \mid \theta)}{Pr(\mathbf{E}_{1:3})} \\
&= \frac{\alpha_1(3)}{\alpha_1(3) + \alpha_2(3)} \\
&= \frac{0.0732}{0.0732 + 0.0432} \\
&= 0.6289
\end{aligned}$$

Utility

1. Draw a decision graph for this problem.



2. To minimize the cost of her trip, which ticket should Alice buy?

$$\begin{aligned} EU(Plane) &= -400 - 20 \times (\ 2 \times 0.6 + \ \ 4 \times 0.4) = -456 \\ EU(Train) &= -200 - 20 \times (\ 8 \times 0.6 + 16 \times 0.4) = -424 \leftarrow \\ EU(Bus) &= \underbrace{-\ 50}_{\text{Ticket}} - \underbrace{20 \times (18 \times 0.6 + 36 \times 0.4)}_{\text{Time}} = -554 \end{aligned}$$

Alice should buy the train ticket, with a maximum expected utility of -424.

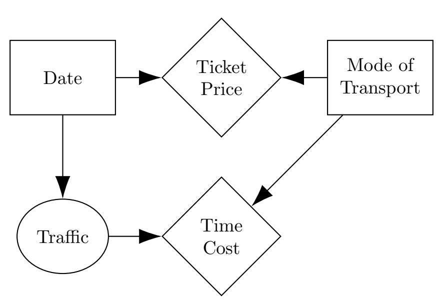
3. How much should Alice be willing to pay a fortune teller to tell her with 100% certainty whether there will be traffic on her commute?

$$\left.\begin{aligned} EU(Plane|Tr=0) \quad &= -400 - \ \ 40 = -440 \\ EU(Train|Tr=0) \quad &= -200 - 160 = -360 \\ EU(Bus|Tr=0) \quad &= -\ \ 50 - 360 = -410 \end{aligned}\right\} \quad a^*_{Tr=0} = Train,\ EU(a^*_{Tr=0}|Tr=0) = -360$$

$$\left.\begin{aligned} EU(Plane|Tr=1) \quad &= -400 - \ \ 80 = -480 \\ EU(Train|Tr=1) \quad &= -200 - 320 = -520 \\ EU(Bus|Tr=1) \quad &= -\ \ 50 - 720 = -770 \end{aligned}\right\} \quad a^*_{Tr=1} = Plane,\ EU(a^*_{Tr=1}|Tr=1) = -480$$

$$\begin{aligned} VPI(X) &= \left(\sum_x P(Tr=x) EU(a^*_{Tr=x}|Tr=x)\right) - EU(a^*) \\ &= P(Tr=0)EU(a^*_{Tr=0}|Tr=0) + P(Tr=1|E)EU(a^*_{Tr=1}|Tr=1) - EU(a^*) \\ &= (-360 \times 0.6 + -480 \times 0.4) - (-424) = 16 \end{aligned}$$

4. Now suppose Alice can also choose between a few dates for her trip. Tickets will be more expensive for earlier dates, but the later dates are on holiday weekends, when there may be more traffic on all forms of transportation. Draw the decision graph that reflects this change.



Markov Decision Processes

1. Assuming the initial policy is $\pi^0(s) = \textit{Left}\ \forall s \in S$, and initial $V(s) = 0\ \forall s \in S$, perform 4 iterations of policy evaluation with discount factor $\gamma = 0.9$.

$$V_{k+1}(s) = R(s, \pi(s)) + \gamma \sum_{s' \in S} T(s, \pi(s), s') V_k(s')$$

s_1	s_2	s_3
0	0	0

$$\begin{aligned} V_1(s_1) &= \quad 0 + 0.9(0.7 \times 0 + 0.2 \times 0 + 0.1 \times 0) = 0 \\ V_1(s_2) &= -10 + 0.9(0.7 \times 0 + 0.2 \times 0 + 0.1 \times 0) = -10 \\ V_1(s_3) &= 100 + 0.9(0.7 \times 0 + 0.2 \times 0 + 0.1 \times 0) = 100 \end{aligned}$$

s_1	s_2	s_3
0	-10	100

$$\begin{aligned} V_2(s_1) &= \quad 0 + 0.9(0.7 \times 0 + 0.2 \times 0 + 0.1 \times -10) = -0.9 \\ V_2(s_2) &= -10 + 0.9(0.7 \times 0 + 0.2 \times -10 + 0.1 \times 100) = -2.8 \\ V_2(s_3) &= 100 + 0.9(0.7 \times -10 + 0.2 \times 100 + 0.1 \times 100) = 120.7 \end{aligned}$$

s_1	s_2	s_3
-0.9	-2.8	120.7

$$\begin{aligned} V_3(s_1) &= \quad 0 + 0.9(0.7 \times -0.9 + 0.2 \times -0.9 + 0.1 \times -2.8) = -0.98 \\ V_3(s_2) &= -10 + 0.9(0.7 \times -0.9 + 0.2 \times -2.8 + 0.1 \times 120.7) = -0.21 \\ V_3(s_3) &= 100 + 0.9(0.7 \times -2.8 + 0.2 \times 120.7 + 0.1 \times 120.7) = 130.83 \end{aligned}$$

s_1	s_2	s_3
-0.98	-0.21	130.83

$$\begin{aligned} V_4(s_1) &= \quad 0 + 0.9(0.7 \times -0.98 + 0.2 \times -0.98 + 0.1 \times -0.21) = -0.81 \\ V_4(s_2) &= -10 + 0.9(0.7 \times -0.98 + 0.2 \times -0.21 + 0.1 \times 130.83) = 1.12 \\ V_4(s_3) &= 100 + 0.9(0.7 \times -0.21 + 0.2 \times 130.83 + 0.1 \times 130.83) = 135.19 \end{aligned}$$

s_1	s_2	s_3
-0.81	1.12	135.19

2. Given this estimate for V^{π_0}, what policy π' will an iteration of policy improvement give?
   $\pi'(s) = \textit{Right}\ \forall s \in S$.

3. Suppose the episode ends when the agent reaches the rightmost state. How will this change the policy iteration?
   $V_k(s_3)$ will continue to be 100 at each step, exerting less positive influence on s_2. After policy evaluation converges, greedy policy improvement will give a policy π' with $\pi'(s_1) = \textit{Left}$.