Lecture 3: Probability, Bayesian inference and Value of Information

Scribe: *Chris Birmingham*

1 Pre-Lecture Notes

1. Quizzes will be one week after each lecture.
2. Presentations will start on Tuesday the 4th of September.
3. Each presentation will be 17 minutes with 10 minutes Q&A

2 Smoothing vs. Filtering

See prior lecture notes for definition of forward and backward filtering.

Smoothing: ”Smoothing is when we take the entire sequence and then we look back at the past at a specific time step, and then we weigh the whole observation sequence to determine the likelihood of that given state at that given time step”

Definition:

$$\begin{aligned}
P(X_k|O_{1:t}) &= \eta b_{k+1:t} f_{1:k} \\
&\quad where \\
f_{1:k} &= P(X_k|O_{1:k}) \\
&\quad and \\
b_{k+1:t} &= P(O_{k+1:t}|X_k) \\
&= \sum_{X_k+1} P(O_{k:t}|X_{k+1}) P(O_{k+2:t}|X_{k+1}) P(X_{k+1}|X_k)
\end{aligned}$$

3 Most Likely Sequence

Similar to dynamic programming

Show:

$$S'^{*}_{X_0->X_t} =< S^{*}_{X_0->X_{t-1}}, S_{X_{t-1}->X_t} >$$
$$P(S'^{*}_{X_0->X_t}) = P(S^{*}_{X_0->X_{t-1}})P(S_{X_{t-1}->X_t})$$

Prove: *by contradiction - assume not most likely path*

$$\begin{aligned} S'^{*}_{X_0->X_t} &=< S'_{X_0->X_{t-1}}, S_{X_{t-1}->X_t} > \\ P(S'^{*}_{X_0->X_t}) &= P(S'_{X_0->X_{t-1}})P(S_{X_{t-1}->X_t}) \\ &< P(S^{*}_{X_0->X_{t-1}})P(S_{X_{t-1}->X_t})P(S^{*}_{X_0->X_t}) \end{aligned}$$

This is the principle behind the Viterbi Algorithm

3.1 Viterbi Algorithm

The Viterbi Algorithm will give us the most likely sequence over the entire set.

The same as process as filtering, but instead of summing over previous states we take the maximum.

$$m_t^{(X_t)} = \eta P(O-t|X_t)max[P(X_t|X_{t-1})m_{t-1}] \tag{1}$$
$$m_t^{(X_1)} = max_{X_1,...,X_{t-1}} P(X_1, ..., X_{t-1}, X_t|O_{1:t}) \tag{2}$$

Why are smoothing and Viterbi different?
Smoothing does point-wise estimates Viterbi considers the joint probabilities, the probabilities of the entire sequence.

Why do smoothing?
Viterbi doesn't say anything about the likelihood of other paths.

4 Model Verification

Does the model match the evidence? What is the probability of the evidence we have?

$$
\begin{aligned}
P(O_{1:t}) &= \sum_{X_t} M(O_t|X_t) \sum_{X_{t-1}} T(X_t|x_{t-1}) P(X_{t-1}|O_{1:t-1}) P(O_{1:t-1}) \\
&= P(O_t|O_{1:t-1}) P(O_{1:t-1}) \\
&= \sum_{X_t} P(O_t|X_t, O_{1:t-1}) \sum_{X_{t-1}} P(X_t|X_{t-1}, O_{1:t-1}) P(X_{t-1}|O_{1:t-1}) P(O_{1:t-1}) \\
&= \sum_{X_t} P(O_t|X_t) \sum_{x_{t-1}} P(X_t|X_{t-1}) P(X_{t-1}|O_{1:t-1}) P(O_{1:t-1})
\end{aligned}
$$

5 Actions

How an agent should act given uncertainty.
Given an option to bet and receive \$4000 with 80% certainty or pass and receive \$3000 with 100% certainty, what is the rational choice?

$$
\begin{aligned}
EV(a) &= \sum_{S'} P(S'|a) V(S') \\
EV(bet) &= .8 * 4000 + .2 * 0 = 3200 \\
EV(bet) &= 1.0 * 3200 = 3000
\end{aligned}
$$

Humans will often not make the rational choice for reasons that are not included in the model. In our scenario the agent will only consider the model and will seek to maximize the expected value.

5.1 HRI Example

The robot is trying to clear the table. If the human does not trust the robot the human can intervene and stop the robot from picking up either the glass or the bottle. The human will not intervene with a probability of 70% for the bottle and with a probability of 20% for the glass. If the human does not intervene for the bottle the robot will get a reward of 5, if the human does not intervene for the glass the robot will get a reward of 10. The robot will get a reward of 0 if the human intervenes. What is the expected value of attempting to get the bottle? The glass? What is the value of asking the human if they will intervene?

$$
\begin{aligned}
EV(Bottle) &= 3.5 \\
EV(Glass) &= 2 \\
EV[bestaction] &= 3.5 \\
EV[a^*] &= max_a \sum_{S'} P(S'|a)V(S')
\end{aligned}
$$

the value of the information is the value of the best action with the information minus the value without that information

$$
\begin{aligned}
VI &= EV[a^*|E] - EV[a^*] \\
&= 4.7 - 3.5 = 1.7
\end{aligned}
$$

$$EV[a^*|E] = EV[a_e^*|intervene - glass]P(intervene - glass) + EV[a_e^*|\neg intervene - glass]P(\neg intervene) \quad = 3.5 *$$

5.2 Markov Decision Process

Components of a Markov Decision Process:

- S: State
- A: Action
- T: Transitions
- R: Reward
- $\Pi : S \Rightarrow A$: Policy

Please refer to Ch 17, pg 646 for the grid world example of sequential decision making.