INFO8006 Introduction to Artificial Intelligence

Exercises 5: Reasoning over time

Learning outcomes

At the end of this exercise session you should be able to:

- Define a Markov Model
- Define what is Filtering/Prediction/Smoothing/Most Likely Explanation in general and how it is done in the context of Markov Model
- Explain the simplified matrix representation of HMM (Hidden Markov Model).

Exercise 1: Umbrella World (AIMA, Ex 15.2)

In this exercise, we examine what happens to the probabilities in the umbrella world in the limit of long time sequences.

R_{t-1}	$P(r_t \mid R_{t-1})$
T	0.7
F	0.3

R_t	$P(u_t \mid R_t)$
T	0.9
F	0.2

Table 1: Transition and Observation probability tables.

1. Suppose we observe an unending sequence of days on which the umbrella appears. Show that, as the days go by, the probability of rain on the current day increases monotonically toward a fixed point. Calculate this fixed point.
   For all t, we have the filtering formula

$$P(R_t|u_{1:t}) = \alpha P(u_t|R_t) \sum_{R_{t-1}} P(R_{t-1}|u_{1:t-1}) P(R_t|R_{t-1}) \tag{1}$$
$$= \alpha \langle 0.9, 0.2 \rangle \left(P(r_{t-1}|u_{1:t-1}) \langle 0.7, 0.3 \rangle + (1 - P(r_{t-1}|u_{1:t-1})) \langle 0.3, 0.7 \rangle \right) \tag{2}$$
$$= \alpha \langle 0.9 \times (0.4 P(r_{t-1}|u_{1:t-1}) + 0.3), 0.2 \times (-0.4 P(r_{t-1}|u_{1:t-1}) + 0.7) \rangle \tag{3}$$
$$= \frac{\langle 0.36 P(r_{t-1}|u_{1:t-1}) + 0.27, -0.08 P(r_{t-1}|u_{1:t-1}) + 0.14 \rangle}{.41 + 0.28 P(r_{t-1}|u_{1:t-1})} \tag{4}$$

   Let $x_t = P(r_t|u_{1:t})$, then

$$x_t - x_{t-1} = \frac{0.36 x_{t-1} + 0.27}{0.41 + 0.28 x_{t-1}} - x_{t-1} \tag{5}$$
$$> 0.36 x_{t-1} + 0.27 - 0.41 x_{t-1} - 0.28 x_{t-1}^2 \tag{6}$$
$$= -0.28 x_{t-1}^2 - 0.05 x_{t-1} + 0.27 \tag{7}$$
$$> 0, \quad \forall x_{t-1} \in]-1.08, 0.9[\tag{8}$$

   where the two bounds are obtained by solving the second order equation as $x_{1,2}^0 = \frac{0.05 \pm \sqrt{0.05^2 + 4 \times 0.27 \times 0.28}}{-2 \times 0.28}$ and the upper bound gives the value of the fixed point. Careful students will notice that the demonstration is not complete, indeed to be rigorous we should also prove that $|x_t - x_{t-1}| < |x_{t-1} - x_{t-2}|$, this is let as an exercise for the motivated students.

2. Now consider forecasting further and further into the future, given just the first two umbrella observations. Compute the exact value of this fixed point. We easily find:

$$P(R_{2+k}|U_1, U_2) = \sum_{R_{2+k-1}} P(R_{2+k}|R_{2+k-1}) P(R_{2+k-1}|U_1, U_2) \tag{9}$$
$$= \langle 0.7, 0.3 \rangle P(r_{2+k-1}|U_1, U_2) + \langle 0.3, 0.7 \rangle (1 - P(r_{2+k-1}|U_1, U_2)). \tag{10}$$

   The fixed point is computed by assuming $P(R_{2+k}|U_1, U_2) = P(R_{2+k-1}|U_1, U_2) = \rho$:

$$\rho = 0.7\rho + 0.3 - 0.3\rho \tag{11}$$
$$\rho = \frac{0.3}{0.6} = 0.5 \tag{12}$$

Exercise 2: The coin

You are in a room containing a table, on this table are placed 3 very precious biased coins (named A, B and C). Suddenly another person enters the room and takes the coins. He throws a coin 4 times and then tells you the following information: First he tells you that he drew the first coin uniformly at random and did not throw it. Then, to select the next coin for each following throw, he either kept the same coin with a probability 2/3 or replaced it by another coin with equal probabilities. When you entered the room you inspected the coins and noticed that the coins A, B and C have a head probability of 80%, 50% and 20% respectively. The result of the throws are head, head, tail and head. If you answer right to his questions he will give you the coins. His four questions are as follows:

1. Provide a hidden Markov model that describes the sequence of throws.

   - The hidden state at time t is $X_t \in D_{X_t} = \{A, B, C\} = \{1, 2, 3\}$, it represents the coin chosen.
   - The evidence at time t is $E_t \in D_{E_t} = \{Head, Tail\} = \{1, 2\}$, it represents the result of the throw.
   - The prior matrix
   $$f_0 = P(X_0) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}^T$$
   .
   - The transition matrix
   $$T = P(X_t|X_{t-1}) = \begin{bmatrix} \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{bmatrix} \quad \text{where} \quad T_{ij} = P(X_t = j|X_{t-1} = i)$$
   .
   - The sensor matrix
   $$B = P(E_t|X_t) = \begin{bmatrix} 0.8 & 0.2 \\ 0.5 & 0.5 \\ 0.2 & 0.8 \end{bmatrix} \quad \text{where} \quad B_{ij} = P(E_t = j|X_t = i)$$
   .

2. What are the probabilities of the last coin given the sequence of evidence ? Let's first define the observation matrices:

   $$O_1 = O_2 = O_4 = \begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} \quad O_3 = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.8 \end{bmatrix}$$

   We apply a forward pass to filter and to derive the following equalities:

   $$\begin{aligned} f_1 &= P(X_1|e_1) = \alpha_1 P(X_1)P(e_1|X_1) \\ &= \alpha_1 \sum_{X_0} P(X_0)P(X_1|X_0)P(e_1|X_1) = \alpha_1 O_1 T^T f_0 \end{aligned} \tag{13}$$
   $$\begin{aligned} f_2 &= P(X_2|e_{1:2}) = \alpha_2 \sum_{X_1} P(X_{1:2}, e_2|e_1) \\ &= \alpha_2 \sum_{X_1} P(e_2|X_2, X_1, e_1)P(X_2|X_1, e_1)P(X_1|e_1) = \alpha_2 O_2 T^T f_1 \end{aligned} \tag{14}$$
   $$f_3 = P(X_3|e_{1:3}) = \alpha_3 O_3 T^T f_2 \tag{15}$$
   $$f_4 = P(X_4|e_{1:4}) = \alpha_4 O_4 T^T f_3 \approx \begin{bmatrix} 0.472 & 0.374 & 0.154 \end{bmatrix}^T \tag{16}$$
   .

3. What are the probabilities of the first coin chosen given the sequence evidence? And of the first coin thrown? We apply

a backward pass to smooth and to derive the following equalities:

$$
\begin{aligned}
b_4 &= P(e_4|X_3) = \sum_{X_4} P(e_4, X_4|X_3) \\
&= \sum_{X_4} P(e_4|X_3, X_4)P(X_4|X_3) = \sum_{X_4} P(e_4|X_4)P(X_4|X_3) \\
&= TO_4 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} && (17) \\
b_3 &= P(e_{3:4}|X_2) = \sum_{X_3} P(e_{3:4}, X_3|X_2) \\
&= \sum_{X_3} P(e_{3:4}|X_2, X_3)P(X_3|X_2) = \sum_{X_3} P(e_4|X_3)P(e_3|X_3, e_4)P(X_3|X_2) \\
&= TO_3 b_4 && (18) \\
b_2 &= P(e_{2:4}|X_1) = \sum_{X_2} P(e_{2:4}, X_2|X_1) \\
&= \sum_{X_2} P(e_{2:4}|X_1, X_2)P(X_2|X_1) = \sum_{X_2} P(e_2|X_2, e_{3:4})P(e_{3:4}|X_2)P(X_2|X_1) \\
&= TO_2 b_3 && (19) \\
b_1 &= P(e_{1:4}|X_0) = TO_1 b_2 && (20)
\end{aligned}
$$

The probability of the first coin given the evidence selected is:

$$P(X_0|e_{1:4}) = \alpha P(X_0, e_{1:4}) = \alpha f_0 \times b_1 \approx \begin{bmatrix} 0.457 & 0.331 & 0.212 \end{bmatrix}^T \tag{21}$$

The probability of the first coin thrown given the evidence selected is:

$$
\begin{aligned}
P(X_1|e_{1:4}) &= \alpha P(X_1, e_{2:4}|e_1) = \alpha P(X_1|e_1)P(e_{2:4}|X_1, e_1) \\
&= \alpha f_1 \times b_2 \approx \begin{bmatrix} 0.580 & 0.329 & 0.091 \end{bmatrix}^T && (22)
\end{aligned}
$$

4. What is the most likely sequence of coins thrown? The most likely sequence given the evidence is:

$$x_{1:4}^{ML} = \arg\max_{x_{1:4}} P(x_{1:4}|e_{1:4}) \tag{23}$$

$$\tag{24}$$

. It can be computed efficiently by using the Viterbi algorithm. Let us define the vector $m_i \in \mathbb{R}_+^3$ such that $m_i(j)$ gives the probability of the most likely path to the ith state with value j. It can be computed recursively with the following equations:

$$
\begin{aligned}
m_1 &= P(X_1|e_1) = f_1 && (25) \\
m_t &= \max_{x_{1:t-1}} P(X_t, x_{1:t-1}|e_{1:t}) && (26) \\
&= \max_{x_{1:t-1}} \alpha P(X_t, x_{1:t-1}, e_t|e_{1:t-1}) && (27) \\
&= \max_{x_{1:t-1}} \alpha P(e_t|X_t, x_{1:t-1}, e_{1:t-1})P(X_t|x_{t-1}, x_{1:t-2}, e_{1:t-1})P(x_{1:t-1}|e_{1:t-1}) && (28) \\
&= \alpha P(e_t|X_t) \max_{x_{t-1}} P(X_t|x_{t-1}) \max_{x_{1:t-2}} P(x_{1:t-1}|e_{1:t-1}) && (29) \\
&= \alpha P(e_t|X_t) \max_{x_{t-1}} P(X_t|x_{t-1}) m_{t-1}(x_{t-1}) && (30)
\end{aligned}
$$

⋆ Exercise 3: September 2019 (AIMA, Ex: 15.13 + 15.14)

A professor wants to know if students are getting enough sleep. Each day, the professor observes whether the students sleep in class, and whether they have red eyes. The professor has the following domain theory:

- The prior probability of getting enough sleep, with no observations, is 0.7.
- The probability of getting enough sleep on night t is 0.8 given that the student got enough sleep the previous night, and 0.3 if not.
- The probability of having red eyes is 0.2 if the student got enough sleep, and 0.7 if not.

- The probability of sleeping in class is 0.1 if the student got enough sleep, and 0.3 if not.

Answer the following questions:

1. Formulate this information as a dynamic Bayesian network that the professor could use to filter or predict from a sequence of observations. Provide the conditional probability tables.

2. Then reformulate the dynamic Bayesian network as a hidden Markov model that has only a single observation variable. Give the complete probability tables for the model.

3. For the evidence values e_1 = 'not red eyes, not sleeping in class', e_2 = 'red eyes, not sleeping in class' and e_3 = 'red eyes, sleeping in class' compute the following conditional probability distributions:

   (a) $P(\text{EnoughSleep}_t | e_{1:t})$ for $t = 1, 2, 3$.

   (b) $P(\text{EnoughSleep}_t | e_{1:3})$ for $t = 1, 2, 3$.

Supplementary materials

Berkeley Handout 6