CS 470

Hidden Markov Model - An Example

Daqing Yi

1 Hidden Markov Model

There are two states for hidden random variable x, which are S_1 and S_2. There are three states for observable variable e, which are R, W and B.

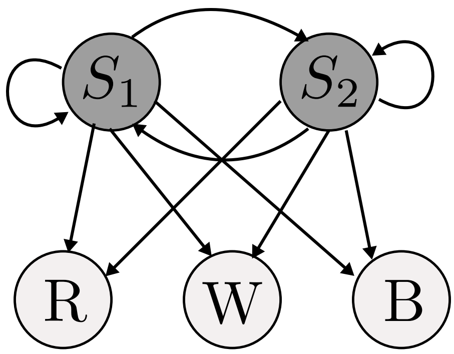

Figure 1: Hidden Markov Model

We have the prior probability defined in Table 1a, the transition probability defined in Table 1b, and the observation probability defined in Table 1c.

	$P(x)$
S_1	0.8
S_2	0.2

(a) Prior probability

	$P(S_1 \mid x)$	$P(S_1 \mid x)$
S_1	0.6	0.4
S_2	0.3	0.7

(b) Transition probability

	$P(R \mid x)$	$P(R \mid x)$	$P(R \mid x)$
S_1	0.3	0.4	0.3
S_2	0.4	0.3	0.3

(c) Observation probability

Table 1: Probability table.

2 Forward Algorithm

Forward algorithm estimates $\boldsymbol{f}_{1:t} = \boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$. For simplification, we have $\boldsymbol{f}_{1:t} = \alpha \boldsymbol{\alpha}_t$. By ignoring the normalization factor, we can calculate only $\boldsymbol{\alpha}_t$. Normalizing $\boldsymbol{\alpha}_t$ gets $\boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$.

Algorithm 1 Forward Algorithm

$\boldsymbol{\alpha}_1 = \boldsymbol{P}(\boldsymbol{e}_1 \mid \boldsymbol{x}_1) P(\boldsymbol{x}_1)$
for $t = 2 : T$ **do**
    $\boldsymbol{\alpha}_t = \boldsymbol{P}(\boldsymbol{e}_t \mid \boldsymbol{x}_t) \sum_{\boldsymbol{x}_{t-1}} P(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) \boldsymbol{\alpha}_{t-1}$
end for

Let's have a sequence of observation "**RWBB**". We will split the calculation into 4 steps.

1. $t = 1$, we have the observation of **R**.

$$\alpha_1(S_1) = P(x_1 = S_1) P(e_1 = R \mid x_1 = S_1) = 0.8 \times 0.3 = 0.24$$
$$\alpha_1(S_2) = P(x_1 = S_2) P(e_1 = R \mid x_1 = S_2) = 0.2 \times 0.4 = 0.08$$

2. $t = 2$, we have the observation of **W**.

$$\begin{aligned}\alpha_2(S_1) =& P(e_2 = W \mid x_2 = S_1) \left(P(x_2 = S_1 \mid x_1 = S_1)\alpha_1(S_1) + P(x_2 = S_1 \mid x_1 = S_2)\alpha_1(S_2)\right)\\ =& 0.4 \times (0.6 \times 0.24 + 0.3 \times 0.08) = 0.067\end{aligned}$$

$$\begin{aligned}\alpha_2(S_2) =& P(e_2 = W \mid x_2 = S_2) \left(P(x_2 = S_2 \mid x_1 = S_1)\alpha_1(S_1) + P(x_2 = S_2 \mid x_1 = S_2)\alpha_1(S_2)\right)\\ =& 0.3 \times (0.4 \times 0.24 + 0.7 \times 0.08) = 0.046\end{aligned}$$

3. $t = 3$, we have the observation of **B**.

$$\begin{aligned}\alpha_3(S_1) =& P(e_3 = W \mid x_3 = S_1) \left(P(x_3 = S_1 \mid x_2 = S_1)\alpha_3(S_1) + P(x_3 = S_1 \mid x_2 = S_2)\alpha_3(S_2)\right)\\ =& 0.3 \times (0.6 \times 0.067 + 0.3 \times 0.046) = 0.0162\end{aligned}$$

$$\begin{aligned}\alpha_3(S_2) =& P(e_3 = W \mid x_3 = S_2) \left(P(x_3 = S_2 \mid x_2 = S_1)\alpha_3(S_1) + P(x_3 = S_2 \mid x_2 = S_2)\alpha_3(S_2)\right)\\ =& 0.3 \times (0.4 \times 0.067 + 0.7 \times 0.046) = 0.0177\end{aligned}$$

4. $t = 4$, we have the observation of **B**.

$$\begin{aligned}\alpha_4(S_1) =& P(e_4 = B \mid x_4 = S_1) \left(P(x_4 = S_1 \mid x_3 = S_1)\alpha_1(S_1) + P(x_4 = S_1 \mid x_3 = S_2)\alpha_3(S_2)\right)\\ =& 0.3 \times (0.6 \times 0.0162 + 0.3 \times 0.0177) = 0.0045\end{aligned}$$

$$\begin{aligned}\alpha_4(S_2) =& P(e_4 = B \mid x_4 = S_2) \left(P(x_4 = S_2 \mid x_3 = S_1)\alpha_1(S_1) + P(x_4 = S_2 \mid x_3 = S_2)\alpha_3(S_2)\right)\\ =& 0.3 \times (0.4 \times 0.0152 + 0.7 \times 0.0177) = 0.00566\end{aligned}$$

3 BACKWARD ALGORITHM

Backward algorithm estimates $\boldsymbol{b}_{k+1:t} = \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$. For simplification, we have $\boldsymbol{b}_{k+1:t} = \alpha\boldsymbol{\beta}_{k+1:t}$. By ignoring the normalization factor, we can calculate only $\boldsymbol{\beta}_{k+1:t}$. Normalizing $\boldsymbol{\beta}_{k+1:t}$ gets $\boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$. If we start from $k = t$, $\boldsymbol{\beta}_{t+1:t} = \boldsymbol{P}(\Phi \mid \boldsymbol{X}_t)$

Algorithm 2 Backward Algorithm

```
β_{t+1:t} = 1
for k = t − 1 : 0 do
    β_{k+1:t} = Σ_{x_{k+1}} P(e_{k+1} | x_{k+1}) P(x_{k+1} | x_k) β_{k+2:t}
end for
```

Let's have a sequence of observation "**RWBB**". Let $t = 4$, we will split the calculation into 5 steps.

1. $k = 4$, we have $\boldsymbol{e}_{t+1:t} = \Phi$.

$$\beta_{5:4}(S_1) = 1$$
$$\beta_{5:4}(S_2) = 1$$

2. $k = 3$, we have the observation of **B**.

$$\begin{aligned}\beta_{4:4}(S_1) =& P(e_4 = B \mid x_3 = S_1)\\ =& P(e_4 = B \mid x_4 = S_1)P(x_4 = S_1 \mid x_3 = S_1)\beta_{5:4}(S_1)\\ &+ P(e_4 = B \mid x_4 = S_2)P(x_4 = S_2 \mid x_3 = S_1)\beta_{5:4}(S_2)\\ =& 0.3 \times 0.6 \times 1 + 0.3 \times 0.4 \times 1 = 0.3\end{aligned}$$

$$\begin{aligned}\beta_{4:4}(S_2) =& P(e_4 = B \mid x_3 = S_2)\\ =& P(e_4 = B \mid x_4 = S_1)P(x_4 = S_1 \mid x_3 = S_2)\beta_{5:4}(S_1)\\ &+ P(e_4 = B \mid x_4 = S_2)P(x_4 = S_2 \mid x_3 = S_2)\beta_{5:4}(S_2)\\ =& 0.3 \times 0.3 \times 1 + 0.3 \times 0.7 \times 1 = 0.3\end{aligned}$$

3. $k = 2$, we have the observation of **BB**.

$$\begin{aligned}\beta_{3:4}(S_1) =& P(e_3 = B, e_4 = B \mid x_2 = S_1)\\ =& P(e_3 = B \mid x_3 = S_1)P(x_3 = S_1 \mid x_2 = S_1)\beta_{4:4}(S_1)\\ &+ P(e_3 = B \mid x_3 = S_2)P(x_3 = S_2 \mid x_2 = S_1)\beta_{4:4}(S_2)\\ =& 0.3 \times 0.6 \times 0.3 + 0.3 \times 0.4 \times 0.3 = 0.09\end{aligned}$$

$$\begin{aligned}\beta_{3:4}(S_2) =& P(e_3 = B, e_4 = B \mid x_2 = S_2)\\ =& P(e_3 = B \mid x_3 = S_1)P(x_3 = S_1 \mid x_2 = S_2)\beta_{4:4}(S_1)\\ &+ P(e_3 = B \mid x_3 = S_2)P(x_3 = S_2 \mid x_2 = S_2)\beta_{4:4}(S_2)\\ =& 0.3 \times 0.3 \times 0.3 + 0.3 \times 0.7 \times 0.3 = 0.09\end{aligned}$$

4. $k = 1$, we have the observation of **WBB**.

$$\begin{aligned}
\beta_{2:4}(S_1) =& P(e_2 = W, e_3 = B, e_4 = B \mid x_1 = S_1) \\
=& P(e_2 = W \mid x_2 = S_1) P(x_2 = S_1 \mid x_1 = S_1) \beta_{3:4}(S_1) \\
& + P(e_2 = W \mid x_2 = S_2) P(x_2 = S_2 \mid x_1 = S_1) \beta_{3:4}(S_2) \\
=& 0.4 \times 0.6 \times 0.09 + 0.3 \times 0.4 \times 0.09 = 0.0324
\end{aligned}$$

$$\begin{aligned}
\beta_{2:4}(S_2) =& P(e_2 = W, e_3 = B, e_4 = B \mid x_1 = S_2) \\
=& P(e_2 = B \mid x_2 = S_1) P(x_2 = S_1 \mid x_1 = S_2) \beta_{3:4}(S_1) \\
& + P(e_2 = B \mid x_2 = S_2) P(x_2 = S_2 \mid x_1 = S_2) \beta_{3:4}(S_2) \\
=& 0.4 \times 0.3 \times 0.09 + 0.3 \times 0.7 \times 0.09 = 0.0297
\end{aligned}$$

5. $k = 0$, we have the observation of **RWBB**.

$$\begin{aligned}
\beta_{1:4}(S_1) =& P(e_1 = R, e_2 = W, e_3 = B, e_4 = B \mid x_0 = S_1) \\
=& P(e_1 = R \mid x_1 = S_1) P(x_1 = S_1 \mid x_0 = S_1) \beta_{2:4}(S_1) \\
& + P(e_1 = R \mid x_1 = S_2) P(x_1 = S_2 \mid x_0 = S_1) \beta_{2:4}(S_2) \\
=& 0.3 \times 0.6 \times 0.0324 + 0.4 \times 0.4 \times 0.0297 = 0.010584
\end{aligned}$$

$$\begin{aligned}
\beta_{1:4}(S_2) =& P(e_1 = R, e_2 = W, e_3 = B, e_4 = B \mid x_0 = S_2) \\
=& P(e_1 = R \mid x_1 = S_1) P(x_1 = S_1 \mid x_0 = S_2) \beta_{2:4}(S_1) \\
& + P(e_1 = R \mid x_1 = S_2) P(x_1 = S_2 \mid x_0 = S_2) \beta_{2:4}(S_2) \\
=& 0.3 \times 0.3 \times 0.0324 + 0.4 \times 0.7 \times 0.0297 = 0.011232
\end{aligned}$$

4 Viterbi Algorithm

The Viterbi Algorithm recursively estimates $\boldsymbol{m}_{1:t} = \max_{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{t-1}} P(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{t-1}, \boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$ and use $\arg\max_{x_t} \boldsymbol{m}_{1:t}$ to estimate states sequentially. By ignoring the normalization factor α, we use $\boldsymbol{m}_{1:t} = \alpha \boldsymbol{m}_t$ in calculation, because we get same result in maximization calculation.

Algorithm 3 Viterbi Algorithm

```
m_1 = P(e_1 | x_1) P(x_1)
x_1^* = argmax_{x_1} m_1
for t = 2 : T do
    m_t = P(e_t | x_t) max_{x_{t-1}} [P(x_t | x_{t-1}) m_t]
    x_t^* = argmax_{x_t} m_t
end for
```

Let's have a sequence of observation "**RWBB**". We will split the calculation into 4 steps.

1. $t = 1$, we have the observation of **R**.

$$m_1(S_1) = P(e_1 = R \mid x_1 = S_1) P(x_1 = S_1) = 0.8 \times 0.3 = 0.24$$

$$m_1(S_2) = P(e_1 = R \mid x_1 = S_2)P(x_1 = S_2) = 0.2 \times 0.4 = 0.08$$

$$\hat{x}_1^* = \arg\max_{x_1}(m_1(S_1), m_1(S_2)) = S_1$$

2. $t = 2$, we have the observation of **W**.

$$\begin{aligned} m_2(S_1) =& P(e_2 = W \mid x_2 = S_1)\max_{x_1}(P(x_2 = S_1 \mid x_1 = S_1)m_1(S_1), P(x_2 = S_1 \mid x_1 = S_2)m_1(S_2)) \\ =& 0.4 \times \max(0.6 \times 0.24, 0.3 \times 0.08) = 0.4 \times 0.6 \times 0.24 = 0.0576 \end{aligned}$$

$$\begin{aligned} m_2(S_2) =& P(e_2 = W \mid x_2 = S_2)\max_{x_1}(P(x_2 = S_2 \mid x_1 = S_1)m_1(S_1), P(x_2 = S_2 \mid x_1 = S_2)m_1(S_2)) \\ =& 0.3 \times \max(0.4 \times 0.24, 0.7 \times 0.08) = 0.4 \times 0.4 \times 0.24 = 0.0288 \end{aligned}$$

$$\hat{x}_2^* = \arg\max_{x_2}(m_2(S_1), m_2(S_2)) = S_1$$

3. $t = 3$, we have the observation of **B**.

$$\begin{aligned} m_3(S_1) =& P(e_3 = B \mid x_3 = S_1)\max_{x_2}(P(x_3 = S_1 \mid x_2 = S_1)m_2(S_1), P(x_3 = S_1 \mid x_2 = S_2)m_2(S_2)) \\ =& 0.3 \times \max(0.6 \times 0.0576, 0.3 \times 0.0288) = 0.3 \times 0.6 \times 0.0576 = 0.0104 \end{aligned}$$

$$\begin{aligned} m_3(S_2) =& P(e_3 = B \mid x_3 = S_2)\max_{x_2}(P(x_3 = S_2 \mid x_2 = S_1)m_2(S_1), P(x_3 = S_2 \mid x_2 = S_2)m_2(S_2)) \\ =& 0.3 \times \max(0.4 \times 0.0576, 0.7 \times 0.0288) = 0.4 \times 0.4 \times 0.0576 = 0.0069 \end{aligned}$$

$$\hat{x}_3^* = \arg\max_{x_3}(m_3(S_1), m_3(S_2)) = S_1$$

4. $t = 4$, we have the observation of **B**.

$$\begin{aligned} m_4(S_1) =& P(e_4 = W \mid x_4 = S_1)\max_{x_3}(P(x_4 = S_1 \mid x_3 = S_1)m_3(S_1), P(x_4 = S_1 \mid x_3 = S_2)m_3(S_2)) \\ =& 0.3 \times \max(0.6 \times 0.0104, 0.3 \times 0.0069) = 0.3 \times 0.6 \times 0.904 = 0.00187 \end{aligned}$$

$$\begin{aligned} m_4(S_2) =& P(e_4 = W \mid x_4 = S_2)\max_{x_3}(P(x_4 = S_2 \mid x_3 = S_1)m_3(S_1), P(x_4 = S_2 \mid x_3 = S_2)m_3(S_2)) \\ =& 0.3 \times \max(0.4 \times 0.0104, 0.7 \times 0.0069) = 0.3 \times 0.4 \times 0.0104 = 0.00125 \end{aligned}$$

$$\hat{x}_4^* = \arg\max_{x_4}(m_4(S_1), m_4(S_2)) = S_1$$