4.2.3 How to derive Forward Algorithm?

We consider forward algorithm as sum of variable from S_1 to S_T.

$$
\begin{aligned}
P(O_1, O_2, \cdots, O_T) &= \sum_{S_1, \cdots, S_T} P(O_1, O_2, \cdots, O_T, S_1, \cdots, S_T) \\
&= \sum_{S_1, \cdots, S_T} P(S_1) P(S_2|S_1) \prod_{t=3}^{T} P(S_t|S_{t-1}) \prod_{t=1}^{T} P(O_t|S_t) \text{(Markov Assumption)} \\
&= \sum_{S_T} \cdots \sum_{S_2} \sum_{S_1} \underbrace{P(S_1) P(O_1|S_1)}_{\alpha_1^k \text{,k is the choice of} S_1} P(S_2|S_1) P(O_2|S_2) \prod_{t=3}^{T} P(S_t|S_{t-1}) P(O_t|S_t) \\
&= \sum_{S_T} \cdots \sum_{S_2} \underbrace{\left(\sum_{S_1} \alpha_1^k P(S_2|S_1) P(O_2|S_2) \right)}_{\alpha_2^k} \prod_{t=3}^{T} P(S_t|S_{t-1}) P(O_t|S_t) \text{(Recurrence Rela.)}
\end{aligned}
$$

So we know what α is , and can derive the below algorithm:

Algorithm 1 Forward Algorithm For Evaluation Problem

Initialise $\alpha_1^k = P(S_1 = k) P(O_1|S_1 = k)$ for all k
for $t = 2$ to T **do**
    $\alpha_t^k = \sum_i \alpha_{t-1}^i P(S_t = k|S_{t-1} = i) P(O_t|S_t = k)$ for all k
end for
$P(\{O_t\})_{t=1}^T) = \sum_k \alpha_T^k$

Note : we can get the general form of α by

$$
\begin{aligned}
\alpha_t^k &= \sum_{S_{t-1}} \cdots \sum_{S_1} P(S_1) P(O_1|S_1) P(S2|S1) P(O_2|S_2) \cdots P(S_t = k|S_{t-1}) P(O_t|S_t = k) \\
&= \sum_{S_{t-1}} \cdots \sum_{S_1} P(S_1, \cdots, S_{t-1}, S_t = k, O_1, \cdots, O_t) \\
&= P(O_1, \cdots, O_t, S_t = k)
\end{aligned}
$$

4.2.4 Backward Algorithm

Similar to Forward Algorithm, but we can eliminate the variable from S_T to S_1

$$
\begin{aligned}
P(O_1, O_2, \cdots, O_T) &= \sum_{S_1,\cdots,S_T} P(O_1, O_2, \cdots, O_T, S_1, \cdots, S_T) \\
&= \sum_{S_1,\cdots,S_T} P(S_1)P(S_2|S_1)\prod_{t=3}^{T} P(S_t|S_{t-1})\prod_{t=1}^{T} P(O_t|S_t)(\text{Markov Assumption}) \\
&= \sum_{S_1}\cdots\sum_{S_{T-1}}\underbrace{\sum_{S_T}\underbrace{\beta_T^k}_{=1} P(S_T|S_{T-1})P(O_T|S_T)}_{\beta_{T-1}^k}\prod_{t=2}^{T-1} P(S_t|S_{t-1})P(O_t|S_t)\cdot P(S_1)
\end{aligned}
$$

Can we define $\beta_T^k = P(O_T|S_T)$? You can think it yourself.

So we know what β is , and can derive the below algorithm:

Algorithm 2 Backward Algorithm For Evaluation Problem

1: Initialise $\beta_T^k = 1$ for all k
2: **for** $t = T - 1$ to 1 **do**
3: $\quad\beta_t^k = \sum_i \beta_{t+1}^i p(S_{t+1} = i|S_t = k)p(O_{t+1}|S_{t+1} = i)$ for all k
4: **end for**
5: $P(\{O_t\})_{t=1}^T) = \sum_k \beta_1^k p(O_1|S_1 = k)p(S_1 = k)$

Note: we can get the general form of β by

$$
\begin{aligned}
\beta_t^k &= \sum_{S_{t+1}}\cdots\sum_{S_T} P(S_T|S_{T-1})P(O_T|S_T)\cdots P(S_{t+1}|S_t = k)P(O_{t+1}|S_{t+1}) \\
&= \sum_{S_{t+1}}\cdots\sum_{S_T} P(O_{t+1}, \cdots, O_T, S_{t+1}, \cdots, S_T|S_t = k) \\
&= P(O_{t+1}, \cdots, O_T|S_t = k)
\end{aligned}
$$

4.3 Problem 3: Decoding Problem

4.3.1 Forward-Backward Algorithm

Similarly, we show how to derive the formula from the perspective of sum over variable.

In order to find $P(S_t = k, \{O_t\}_{t=1}^T)$:

$$
\begin{aligned}
P(S_t = k, \{O_t\}_{t=1}^T) &= \sum_{S_1} \cdots \sum_{S_{t-1}} \sum_{S_{t+1}} \cdots \sum_{S_T} P(S_1, \cdots, S_{t-1}, S_t = k, S_{t+1}, \cdots, S_T, \{O_t\}_{t=1}^T) \\
&= \sum_{S_1} \cdots \sum_{S_{t-1}} \sum_{S_{t+1}} \cdots \sum_{S_T} \prod_{t=1}^{t-1} P(S_t|S_{t-1}) P(O_t|S_t) \cdot \\
&\quad P(S_t = k|S_{t-1}) P(O_t|S_t = k) || P(S_{t+1}|S_t = k) P(O_{t+1}|S_{t+1}) \prod_{t+2}^{T} P(S_t|S_{t-1}) P(O_t|S_t)
\end{aligned}
$$

We can see that the left part of || is just α_t^k and right part of is just β_t^k

we can easily derive the forward-backward algorithm as below:

Algorithm 3 Forward-Backward Algorithm For Decoding Problem

Initialise $\alpha_1^k = P(S_1 = k)P(O_1|S_1 = k)$ for all k
Initialise $\beta_T^k = 1$ for all k
for $t = T - 1$ to 1 **do**
    $\beta_t^k = \sum_i \beta_{t+1}^i p(S_{t+1} = i|S_t = k) p(O_{t+1}|S_{t+1} = i)$ for all k
end for
for $t = 2$ to T **do**
    $\alpha_t^k = \sum_i \alpha_{t-1}^i P(S_t = k|S_{t-1} = i) P(O_t|S_t = k)$ for all k
end for
$P(S_t = k, \{O_t\}_{t=1}^T) = \alpha_t^k \beta_t^k$
$P(S_t = k|\{O_t\}_{t=1}^T) = \frac{P(S_t = k, \{O_t\}_{t=1}^T)}{P(\{O_t\}_{t=1}^T)} = \frac{\alpha_t^k \beta_t^k}{\sum_i \alpha_t^i \beta_t^i}$

4.3.2 Backward-Forward Algorithm

Then we take a further step to think how to make it backward-forward by changing the order of sum of variable.

From the above expansion, since $P(O_t|S_t = k)$ are independent of variable $S_1, \cdots, S_{t-1}, S_{t+1}, S_T)$ we can leave it out and then the left part can be re-written as :

$$
\begin{aligned}
\text{Left Part} &= \sum_{S_1} \cdots \sum_{S_{t-1}} P(S_1)P(S_2|S_1)\prod_{t=1}^{t-1} P(S_t|S_{t-1})P(O_t|S_t)\cdot P(S_t=k|S_{t-1}) \\
&= \sum_{S_1} \cdots \sum_{S_{t-1}=i} \underbrace{P(S_t=k|S_{t-1})P(O_{t-1}|S_{t-1})}_{\alpha_{t-1}^{i}} P(S_{t-1}|S_{t-2}) \prod_{t=1}^{t-2} P(S_t|S_{t-1})P(O_t|S_t) \\
&= \sum_{S_1} \cdots \sum_{S_{t-2}=i} \underbrace{\sum_{S_{t-1}=j} \alpha_{t-1}^{j} P(S_{t-1}|S_{t-2})P(O_{t-2}|S_{t-2})}_{\alpha_{t-2}^{i}} P(S_{t-2}|S_{t-3}) \prod_{t=1}^{t-3} P(S_t|S_{t-1})P(O_t|S_t)
\end{aligned}
$$

Note: we then get the general form of α by

$$
\begin{aligned}
\alpha_x^j &= \sum_{S_{x+1}} \cdots \sum_{S_{t-1}} P(S_t=k|S_{t-1})P(O_{t-1}|S_{t-1}) \cdots P(S_{x+1}|S_x=j)P(O_x|S_x=j) \\
&= \sum_{S_{x+1}} \cdots \sum_{S_{t-1}} P(O_x,\cdots,O_{t-1},S_{x+1},\cdots,S_t=k|S_x=j) \\
&= P(O_x,\cdots,O_{t-1},S_t=k|S_x=j)
\end{aligned}
$$

Similarly, the right part can be re-written as :

$$
\begin{aligned}
\text{Right Part} &= \sum_{S_T} \cdots \sum_{S_{t+1}} P(S_{t+1}|S_t=k)P(O_{t+1}|S_{t+1}) \prod_{t+2}^{T} P(S_t|S_{t-1})P(O_t|S_t) \\
&= \sum_{S_T} \cdots \sum_{S_{t+1}=i} \underbrace{P(S_{t+1}|S_t=k)P(O_{t+1}|S_{t+1})}_{\beta_{t+1}^{i}} P(S_{t+2}|S_{t+1})P(O_{t+2}|S_{t+2}) \prod_{t+3}^{T} P(S_t|S_{t-1})P(O_t|S_t) \\
&= \sum_{S_T} \cdots \sum_{S_{t+2}=i} \underbrace{\sum_{S_{t+1}=j} \beta_{t+1}^{j}(S_{t+2}|S_{t+1})P(O_{t+2}|S_{t+2})}_{\beta_{t+2}^{i}} P(S_{t+3}|S_{t+2})P(O_{t+3}|S_{t+3}) \cdots
\end{aligned}
$$

Note : we then get the general form of β by

$$
\begin{aligned}
\beta_x^j &= \sum_{S_{x-1}} \cdots \sum_{S_{t+1}} P(S_{t+1}|S_t = k)P(O_{t+1}|S_{t+1}) \cdots P(S_x = j|S_{x-1})P(O_x|S_x = j) \\
&= \sum_{S_{x-1}} \cdots \sum_{S_{t+1}} P(O_{t+1}, \cdots, O_x, S_{t+1}, \cdots, S_x = j|S_t = k) \\
&= P(O_{t+1}, \cdots, O_x, S_x = j|S_t = k)
\end{aligned}
$$

Based on above two parts, we can summarise the following backward-forward algorithm:

Algorithm 4 Backward-Forward Algorithm For Decoding Problem

Initialise $\alpha_{t-1}^i = P(S_t = k|S_{t-1} = i)P(O_{t-1}|S_{t-1} = i)$ for all i
Initialise $\beta_{t+1}^i = P(S_{t+1} = i|S_t = k)P(O_{t+1}|S_{t+1} = i)$ for all i
for $x = t + 2$ to T **do**
    $\beta_x^i = \sum_j \beta_{x-1}^j P(S_x = i|S_{x-1} = j)p(O_x|S_x = i)$
end for
for $x = t - 2$ to 1 **do**
    $\alpha_x^i = \sum_j \alpha_{x+1}^j P(S_{x+1} = j|S_x = i)P(O_x|S_x = i)$
end for
Further Define $\alpha_0 = \sum_i \alpha_1^i P(S_1 = i)$
Further Define $\beta_{T+1} = \sum_i \beta_T^i$
$P(S_t = k, \{O_t\}_{t=1}^T) = \alpha_0 \beta_{T+1} \cdot P(O_t|S_t = k)$
$P(S_t = k|\{O_t\}_{t=1}^T) = \frac{P(S_t=k,\{O_t\}_{t=1}^T)}{P(\{O_t\}_{t=1}^T)} = \frac{\alpha_0\beta_{T+1}\cdot P(O_t|S_t=k)}{\sum_k \alpha_0\beta_{T+1}\cdot P(O_t|S_t=k)}$
For special case: if t = 1: We set $\alpha_0 = P(S_1 = k)$.
If t = T: we set $\beta_{T+1} = 1$.

You can check what's the remaining term to the left/right of the || to see how to derive this special condition.