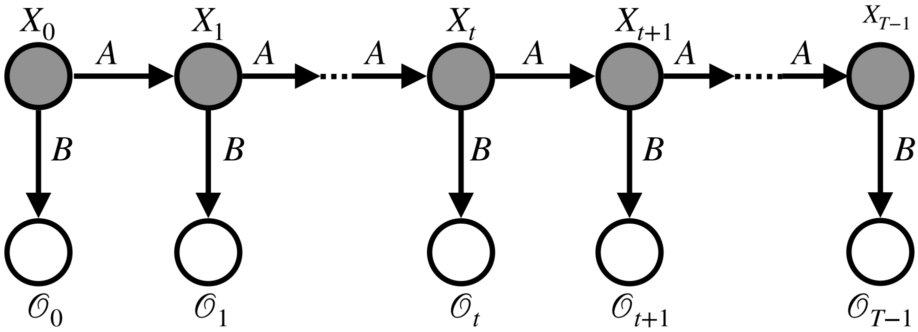
HMM



T： 序列长度 (Length of the observation sequence)
N： 隐藏状态的状态数 (Number of states in the model)
M： 观测状态总数 (Number of observation symbols)
$X = \{X_0, X_1, \cdots, X_{T-1}\}$： 状态序列 (states sequence)
$Q = \{q_0, q_1, \cdots, q_{N-1}\}$： 状态值集合 (distinct states of the Markov process)
$\mathcal{O} = \{\mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_{T-1}\}$： 观测序列 (observation sequence)
$V = \{v_0, v_1, \cdots, v_{M-1}\}$： 观测值集合 (distinct values of observation)

(i) $\lambda = (\pi, A, B)$

π： 初始状态的概率分布 (initial state distribution)，$\{\pi_0, \pi_1, \cdots, \pi_{N-1}\}$ ($\sum_{i=0}^{N-1} \pi_i = 1$)
A： $[a_{ij}]$→转移矩阵 (state transition probabilities)，$a_{ij} = P(X_{t+1} = a_j \mid X_t = q_i)$
B： $[b_j(k)]$→发射矩阵 (observation probabilities matrix)，$b_j(k) = P(\mathcal{O}_t = v_k \mid X_t = q_j)$

(ii) 两个假设 (Two propositions)

① 齐次Markov

$$P(X_{t+1} \mid X_0, X_1, \cdots, X_{T-1}, \mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_{T-1}) = P(X_{t+1} \mid X_t)$$

② 观测独立

$$P(\mathcal{O}_t \mid X_0, X_1, \cdots, X_{T-1}, \mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_{T-1}) = P(\mathcal{O}_t \mid X_t)$$

(iii) 三个问题 (Three problems)

① Evaluation

给定λ，求$P(\mathcal{O} \mid \lambda)$
(evaluate $P(\mathcal{O} \mid \lambda)$ by given λ) → **Forward-Backward**

② Learning

$\lambda_{MLE} = \arg\max_{\lambda} P(\mathcal{O} \mid \lambda)$ → **Baum-Welch**
EM

③ Decoding

$\hat{X} = \arg\max_{X} P(\mathcal{O} \mid \lambda)$ → **Viterbi**

前向算法 (Forward Algorithm)

Define : $\alpha_t(i) = P(\mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_t, X_t = q_i \mid \lambda)$

$\alpha_t(i)$	$= P(\mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_t, X_t = q_i)$	We omit λ
	$= \sum_{j=0}^{N-1} P(\mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_t, X_{t-1} = q_j, X_t = q_i)$	$P(B) = \sum_{B} P(A, B)$
	$= \sum_{j=0}^{N-1} P(\mathcal{O}_0, \cdots, \mathcal{O}_{t-1}, X_{t-1} = q_j) \cdot P(\mathcal{O}_t, X_t = q_i \mid \mathcal{O}_0, \cdots, \mathcal{O}_{t-1}, X_{t-1} = q_j)$	$\sum_{B} P(A, B) = \sum_{B} P(B) \cdot P(A \mid B)$
	$= \sum_{j=0}^{N-1} \alpha_{t-1}(j) \cdot P(\mathcal{O}_t \mid \mathcal{O}_0, \cdots, \mathcal{O}_{t-1}, X_{t-1} = q_j, X_t = q_i) \cdot P(X_t = q_i \mid \mathcal{O}_0, \cdots, \mathcal{O}_{t-1}, X_{t-1} = q_j)$	$P(AB) = P(A \mid B) \cdot P(B)$
	$= \sum_{j=0}^{N-1} \alpha_{t-1}(j) \cdot P(\mathcal{O}_t \mid X_t = q_i) \cdot P(X_t = q_i \mid X_{t-1} = q_j)$	$\mathcal{O}_t$ is only connected with X_t X_t is only connected with X_{t-1}
	$= \sum_{j=0}^{N-1} \alpha_{t-1}(j) \cdot b_i(\mathcal{O}_t) \cdot a_{ji}$	
	$= \left[\sum_{j=0}^{N-1} \alpha_{t-1}(j) \cdot a_{ji} \right] \cdot b_i(\mathcal{O}_t)$	

Then the final probability :

Initialization : $\alpha_0(i) = P(\mathcal{O}_0, X_0 = q_i \mid \lambda) = \pi_i \cdot b_i(\mathcal{O}_0)$

Recursions : $\alpha_t(i) = \left[\sum_{j=0}^{N-1} \alpha_{t-1}(j) \cdot a_{ji} \right] \cdot b_i(\mathcal{O}_t)$

The result :

$$
\begin{aligned}
P(\mathcal{O} \mid \lambda) &= \sum_{i=0}^{N-1} P(\mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_{T-1}, X_{T-1} = q_i \mid \lambda) \\
&= \sum_{i=0}^{N-1} \alpha_{T-1}(i)
\end{aligned}
$$

后向算法 (Backward Algorithm)

Define : $\beta_t(i) = P(\mathcal{O}_{t+1}, \mathcal{O}_{t+2}, \cdots, \mathcal{O}_{T-1} \mid X_t = q_i, \lambda)$

$\beta_t(i) = P(\mathcal{O}_{t+1}, \mathcal{O}_{t+2}, \cdots, \mathcal{O}_{T-1} \mid X_t = q_i)$	We omit λ
$= \sum_{j=0}^{N-1} P(\mathcal{O}_{t+1}, \mathcal{O}_{t+2}, \cdots, \mathcal{O}_{T-1}, X_{t+1} = q_j \mid X_t = q_i)$	$P(A) = \sum_B P(A, B)$
$= \sum_{j=0}^{N-1} P(\mathcal{O}_{t+2}, \cdots, \mathcal{O}_{T-1} \mid \mathcal{O}_{t+1}, X_t = q_i, X_{t+1} = q_j) \cdot P(\mathcal{O}_{t+1}, X_{t+1} = q_j \mid X_t = q_i)$	$\sum_B P(A, B) = \sum_B P(A \mid B) \cdot P(B)$
$= \sum_{j=0}^{N-1} P(\mathcal{O}_{t+2}, \cdots, \mathcal{O}_{T-1} \mid \mathcal{O}_{t+1}, X_t = q_i, X_{t+1} = q_j) \cdot P(\mathcal{O}_{t+1}, X_{t+1} = q_j \mid X_t = q_i)$	Among $\mathcal{O}_{t+1}, X_t = q_i$ and $X_{t+1} = q_j$, $\mathcal{O}_{t+2}, \cdots, \mathcal{O}_{T-1}$ are only connected with $X_{t+1} = q_j$.
$= \sum_{j=0}^{N-1} \beta_{t+1}(j) \cdot P(\mathcal{O}_{t+1} \mid X_{t+1} = q_j, X_t = q_i) \cdot P(X_{t+1} = q_j \mid X_t = q_i)$	$P(AB) = P(A \mid B) \cdot P(B)$
$= \sum_{j=0}^{N-1} \beta_{t+1}(j) \cdot P(\mathcal{O}_{t+1} \mid X_{t+1} = q_j) \cdot P(X_{t+1} = q_j \mid X_t = q_i)$	$\mathcal{O}_{t+1}$ is only connected with X_{t+1} X_{t+1} is only connected with X_t
$= \sum_{j=0}^{N-1} \beta_{t+1}(j) \cdot b_j(\mathcal{O}_{t+1}) \cdot a_{ij}$	

Then the final probability :

Initialization : $\beta_{T-2}(i) = P(\mathcal{O}_{T-1} \mid X_{T-1} = q_i, \lambda) = b_i(\mathcal{O}_{T-1})$

Recursions : $\beta_t(i) = \sum_{j=0}^{N-1} a_{ij} \cdot b_j(\mathcal{O}_{t+1}) \cdot \beta_{t+1}(j)$

The result :

$$
\begin{aligned}
P(\mathcal{O} \mid \lambda) &= \sum_{i=0}^{N-1} P(\mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_{T-1}, X_0 = q_i \mid \lambda) \\
&= \sum_{i=0}^{N-1} P(\mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_{T-1} \mid X_0 = q_i, \lambda) \cdot P(X_0 = q_i \mid \lambda) . \\
&= \sum_{i=0}^{N-1} P(\mathcal{O}_0 \mid \mathcal{O}_1, \cdots, \mathcal{O}_{T-1}, X_0 = q_i, \lambda) \cdot P(\mathcal{O}_1, \cdots, \mathcal{O}_{T-1} \mid X_0 = q_i, \lambda) \cdot \pi_i . \\
&= \sum_{i=0}^{N-1} P(\mathcal{O}_0 \mid X_0 = q_i, \lambda) \cdot P(\mathcal{O}_1, \cdots, \mathcal{O}_{T-1} \mid X_0 = q_i, \lambda) \cdot \pi_i . \\
&= \sum_{i=0}^{N-1} b_i(\mathcal{O}_0) \cdot \beta_0(i) \cdot \pi_i
\end{aligned}
$$

Baum-Welch Algorithm

Define : $\gamma_t(i) = P(X_t = q_i \mid \mathcal{O}, \lambda)$

$$
\begin{aligned}
\gamma_t(i) &= P(X_t = q_i \mid \mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_{T-1}) && \text{We omit } \lambda \\
&= \frac{P(\mathcal{O}_0, \cdots, \mathcal{O}_t, \mathcal{O}_{t+1}, \cdots, \mathcal{O}_{T-1}, X_t = q_i)}{P(\mathcal{O} \mid \lambda)} && P(B \mid A) = \frac{P(AB)}{P(A)} \\
&= \frac{P(\mathcal{O}_0, \cdots, \mathcal{O}_t, X_t = q_i) \cdot P(\mathcal{O}_{t+1}, \cdots, \mathcal{O}_{T-1} \mid \mathcal{O}_0, \cdots, \mathcal{O}_t, X_t = q_i)}{P(\mathcal{O} \mid \lambda)} && P(AB) = P(A) \cdot P(B \mid A) \\
&= \frac{P(\mathcal{O}_0, \cdots, \mathcal{O}_t, X_t = q_i) \cdot P(\mathcal{O}_{t+1}, \cdots, \mathcal{O}_{T-1} \mid X_t = q_i)}{P(\mathcal{O} \mid \lambda)} \\
&= \frac{P(\mathcal{O}_0, \cdots, \mathcal{O}_t, X_t = q_i \mid \lambda) \cdot P(\mathcal{O}_{t+1}, \cdots, \mathcal{O}_{T-1} \mid X_t = q_i, \lambda)}{P(\mathcal{O} \mid \lambda)} && \text{We return back } \lambda \\
&= \frac{\alpha_t(i) \cdot \beta_t(i)}{P(\mathcal{O} \mid \lambda)}
\end{aligned}
$$