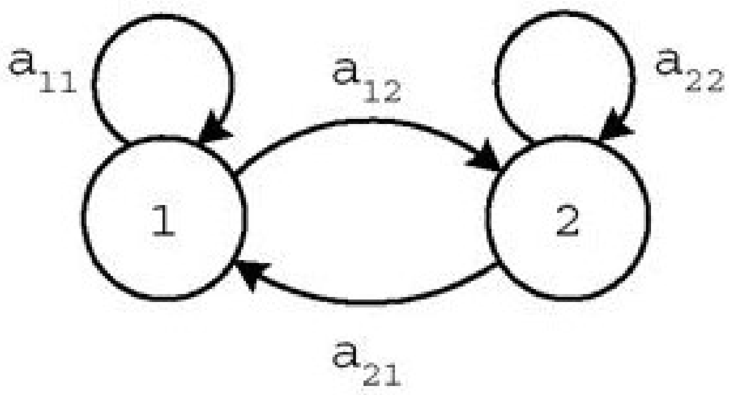
Hidden Markov Model (HMM)

Tutorials

Hidden Markov Model



$\lambda = \{\mathbf{A}, \mathbf{B}, \pi\}$

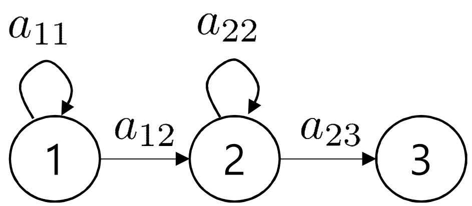
where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \; B = \begin{bmatrix} b_1(v_1) & b_2(v_1) \\ b_1(v_2) & b_2(v_2) \\ \vdots & \vdots \\ b_1(v_k) & b_2(v_k) \\ \vdots & \vdots \\ b_1(v_M) & b_2(v_M) \end{bmatrix}$$

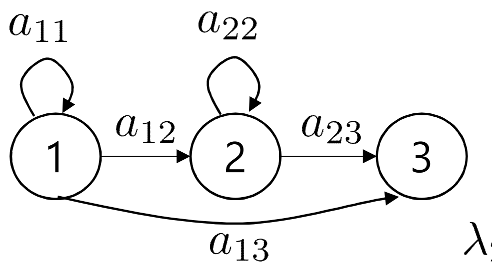
$$\pi = \{\pi_1, \pi_2\}$$

1. $\mathbf{N}$: number of states in the model
   - states, $\mathbf{s} = \{s_1, s_2, \ldots, s_N\}$
   - states at the time $t, q_t \in \mathbf{s}$
2. $\mathbf{M}$: number of discrete observation symbols
   - observation, $\mathbf{v} = \{v_1, v_2, \ldots, v_M\}$
   - observation at time $t, o_t \in \mathbf{v}$
3. $\mathbf{A} = \{a_{ij}\}$: state transition probability distribution
   - $a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i) \quad 1 \le i, j \le N$
4. $\mathbf{B} = \{b_j(k)\}$: observation symbol probability distribution
   - $b_j(k) = P(O_t = v_k \mid q_t = s_j)$, $1 \le j \le N, 1 \le k \le M$
5. $\pi = \{\pi_i\}$:initial state distribution
   - $\pi_i = P(q_1 = s_i)$,$1 \le i \le N$

HMM : Two HMMs prepared for classification



$$\lambda_1 = \{\mathbf{A}_1, \mathbf{B}_1, \boldsymbol{\pi}_1\}$$



$$\lambda_2 = \{\mathbf{A}_2, \mathbf{B}_2, \boldsymbol{\pi}_2\}$$

- state transition probability distribution

$$A_1 = \begin{bmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & a_{23} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.7 & 0.3 \\ 0 & 0 & 0 \end{bmatrix}$$

- observation symbol probability distribution

$$B_1 = \begin{bmatrix} b_1(v_1) & b_2(v_1) & b_3(v_1) \\ b_1(v_2) & b_2(v_2) & b_3(v_2) \\ b_1(v_3) & b_2(v_3) & b_3(v_3) \\ b_1(v_4) & b_2(v_4) & b_3(v_4) \\ b_1(v_5) & b_2(v_5) & b_3(v_5) \\ b_1(v_6) & b_2(v_6) & b_3(v_6) \end{bmatrix} = \begin{bmatrix} 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/2 & 1/2 \end{bmatrix}$$

- initial state distribution

$$\boldsymbol{\pi}_1 = [\pi_1, \pi_2, \pi_3] = [1, 0, 0]$$

- state transition probability distribution

$$A_2 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0 & 0.6 & 0.4 \\ 0 & 0 & 0 \end{bmatrix}$$

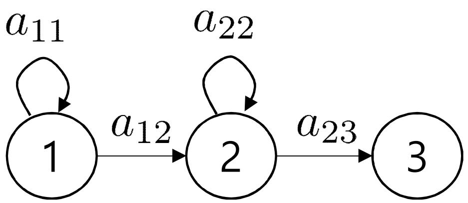
- observation symbol probability distribution

$$B_2 = \begin{bmatrix} b_1(v_1) & b_2(v_1) & b_3(v_1) \\ b_1(v_2) & b_2(v_2) & b_3(v_2) \\ b_1(v_3) & b_2(v_3) & b_3(v_3) \\ b_1(v_4) & b_2(v_4) & b_3(v_4) \\ b_1(v_5) & b_2(v_5) & b_3(v_5) \\ b_1(v_6) & b_2(v_6) & b_3(v_6) \end{bmatrix} = \begin{bmatrix} 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/2 & 1/2 \end{bmatrix}$$

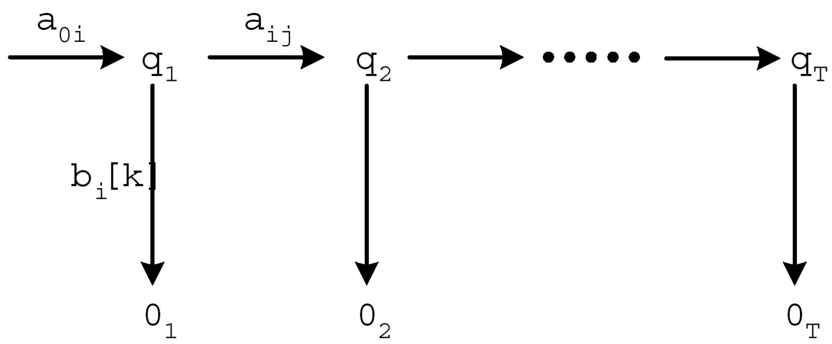
- initial state distribution

$$\boldsymbol{\pi}_2 = [\pi_1, \pi_2, \pi_3] = [1, 0, 0]$$

HMM : Two HMMs prepared for classification



$$\lambda_1 = \{\mathbf{A}_1, \mathbf{B}_1, \boldsymbol{\pi}_1\}$$



- state transition probability distribution

$$A_1 = \begin{bmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & a_{23} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.7 & 0.3 \\ 0 & 0 & 0 \end{bmatrix}$$

- observation symbol probability distribution

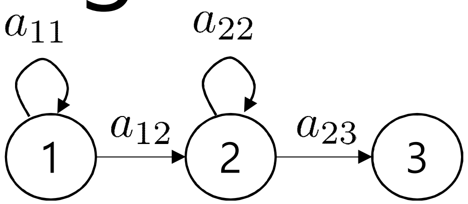
$$B_1 = \begin{bmatrix} b_1(v_1) & b_2(v_1) & b_3(v_1) \\ b_1(v_2) & b_2(v_2) & b_3(v_2) \\ b_1(v_3) & b_2(v_3) & b_3(v_3) \\ b_1(v_4) & b_2(v_4) & b_3(v_4) \\ b_1(v_5) & b_2(v_5) & b_3(v_5) \\ b_1(v_6) & b_2(v_6) & b_3(v_6) \end{bmatrix} = \begin{bmatrix} 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/2 & 1/2 \end{bmatrix}$$

- initial state distribution

$$\boldsymbol{\pi}_1 = [\pi_1, \pi_2, \pi_3] = [1, 0, 0]$$

1. Choose an initial state $q_1 = s_i$ based on the initial state distribution $\boldsymbol{\pi}$.
2. For $t = 1$ to T:
   - Choose $o_t = v_k$ according to the symbol probability distribution in state $s_i \Longrightarrow b_i(k)$
   - Transition to a new state $q_{t+1} = s_j$ according to the state transition probability distribution for state $t+1$
3. Increment t by 1, return to step 2 if $t \leq T$;else, terminate.

The generation:



$$\lambda_1 = \{\mathbf{A}_1, \mathbf{B}_1, \boldsymbol{\pi}_1\}$$

- state transition probability distribution

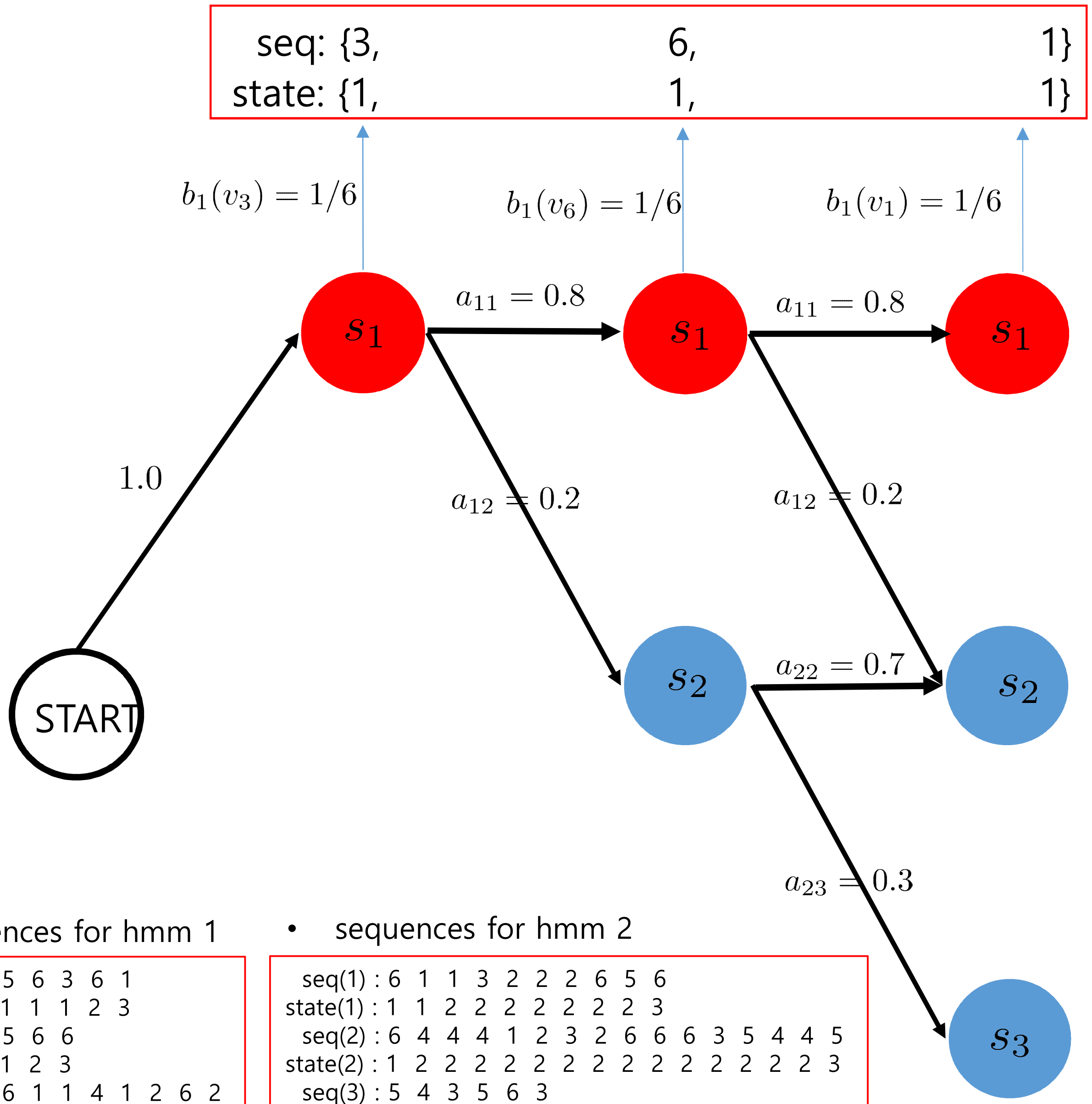
$$A_1 = \begin{bmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & a_{23} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.7 & 0.3 \\ 0 & 0 & 0 \end{bmatrix}$$

- observation symbol probability distribution

$$B_1 = \begin{bmatrix} b_1(v_1) & b_2(v_1) & b_3(v_1) \\ b_1(v_2) & b_2(v_2) & b_3(v_2) \\ b_1(v_3) & b_2(v_3) & b_3(v_3) \\ b_1(v_4) & b_2(v_4) & b_3(v_4) \\ b_1(v_5) & b_2(v_5) & b_3(v_5) \\ b_1(v_6) & b_2(v_6) & b_3(v_6) \end{bmatrix} = \begin{bmatrix} 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/2 & 1/2 \end{bmatrix}$$

- initial state distribution

$$\boldsymbol{\pi}_1 = [\pi_1, \pi_2, \pi_3] = [1, 0, 0]$$



- sequences for hmm 1

seq(1) : 5 5 6 3 6 1
state(1) : 1 1 1 1 2 3
seq(2) : 1 5 6 6
state(2) : 1 1 2 3
seq(3) : 3 6 1 1 4 1 2 6 2
state(3) : 1 1 1 1 1 1 1 2 3

- sequences for hmm 2

seq(1) : 6 1 1 3 2 2 2 6 5 6
state(1) : 1 1 2 2 2 2 2 2 2 3
seq(2) : 6 4 4 4 1 2 3 2 6 6 6 3 5 4 4 5
state(2) : 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3
seq(3) : 5 4 3 5 6 3
state(3) : 1 1 1 1 1 3

Three Basic HMM Problems

1. **Scoring**: Given an observation sequence $\mathrm{O} = \{o_1, o_2, \ldots, o_T\}$ and a model $\lambda = \{\mathbf{A}, \mathbf{B}, \pi\}$, how do we compute $P(\mathrm{O}|\lambda)$ the probability of the observation sequence?

➔ The Forward-Backward Algorithm.

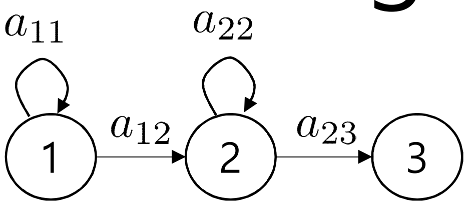
2. **Matching**: Given an observation sequence $\mathrm{O} = \{o_1, \ldots, o_T\}$, how do we choose a state sequence $\mathbf{Q} = \{q_1, q_2, \ldots, q_T\}$ which is optimum in some sense?

➔ The Viterbi Algorithm

3. **Training**: How do we adjust the model parameters $\lambda = \{\mathbf{A}, \mathbf{B}, \pi\}$ to maximize $P(\mathrm{O}|\lambda)$?

➔ The Baum-Welch Reestimation Procedures

The Scoring: Forward



$$\lambda_1 = \{\mathbf{A}_1, \mathbf{B}_1, \boldsymbol{\pi}_1\}$$

- state transition probability distribution

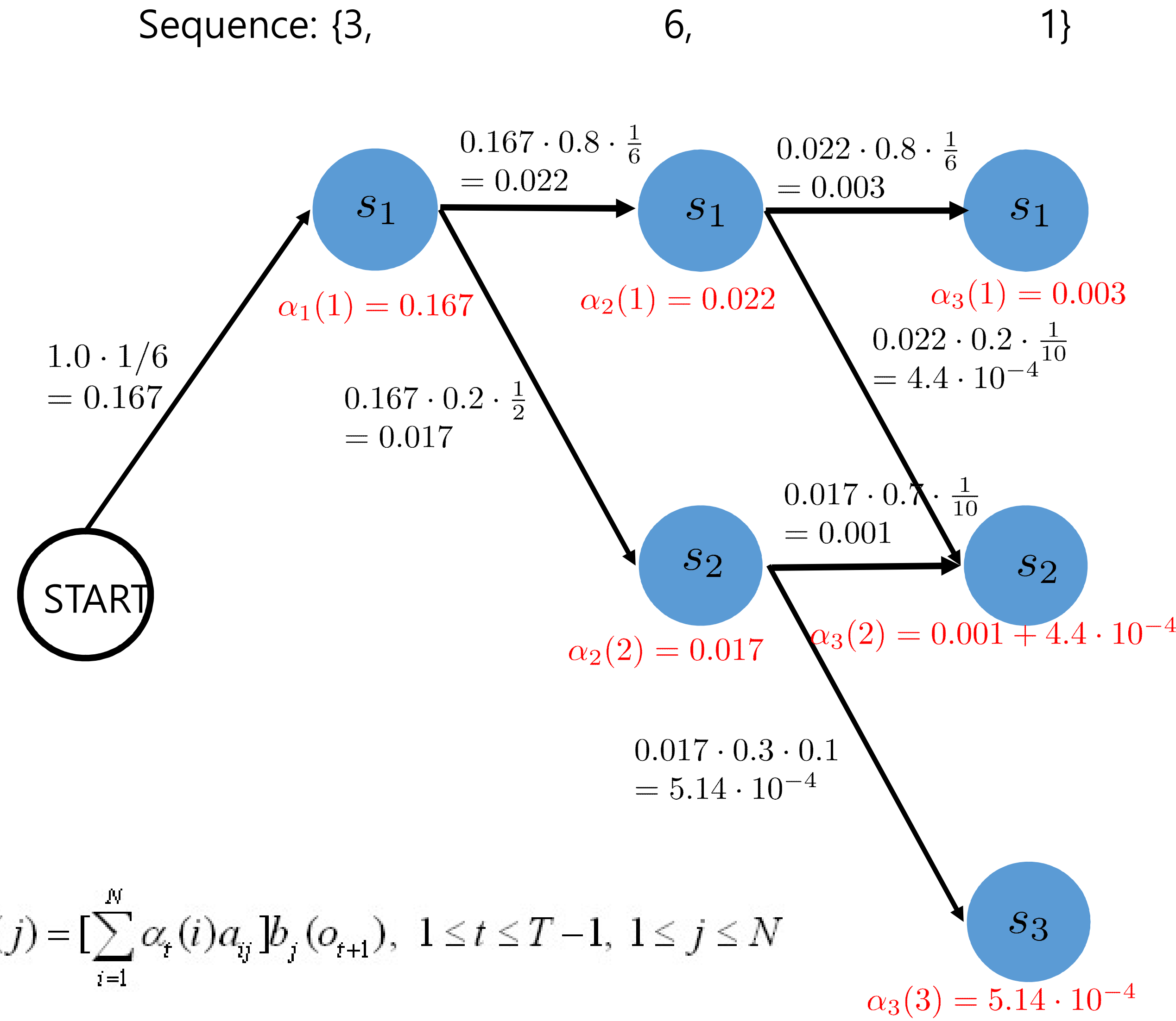
$$A_1 = \begin{bmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & a_{23} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.7 & 0.3 \\ 0 & 0 & 0 \end{bmatrix}$$

- observation symbol probability distribution

$$B_1 = \begin{bmatrix} b_1(v_1) & b_2(v_1) & b_3(v_1) \\ b_1(v_2) & b_2(v_2) & b_3(v_2) \\ b_1(v_3) & b_2(v_3) & b_3(v_3) \\ b_1(v_4) & b_2(v_4) & b_3(v_4) \\ b_1(v_5) & b_2(v_5) & b_3(v_5) \\ b_1(v_6) & b_2(v_6) & b_3(v_6) \end{bmatrix} = \begin{bmatrix} 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/2 & 1/2 \end{bmatrix}$$

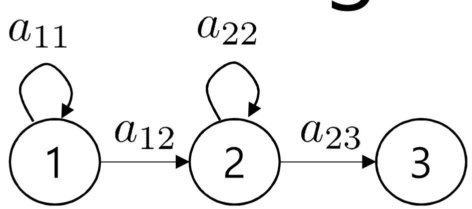
- initial state distribution

$$\boldsymbol{\pi}_1 = [\pi_1, \pi_2, \pi_3] = [1, 0, 0]$$



$$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(o_{t+1}), \ 1 \le t \le T-1, \ 1 \le j \le N$$

The Scoring: Backward



$$\lambda_1 = \{\mathbf{A}_1, \mathbf{B}_1, \boldsymbol{\pi}_1\}$$

- state transition probability distribution

$$A_1 = \begin{bmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & a_{23} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.7 & 0.3 \\ 0 & 0 & 0 \end{bmatrix}$$

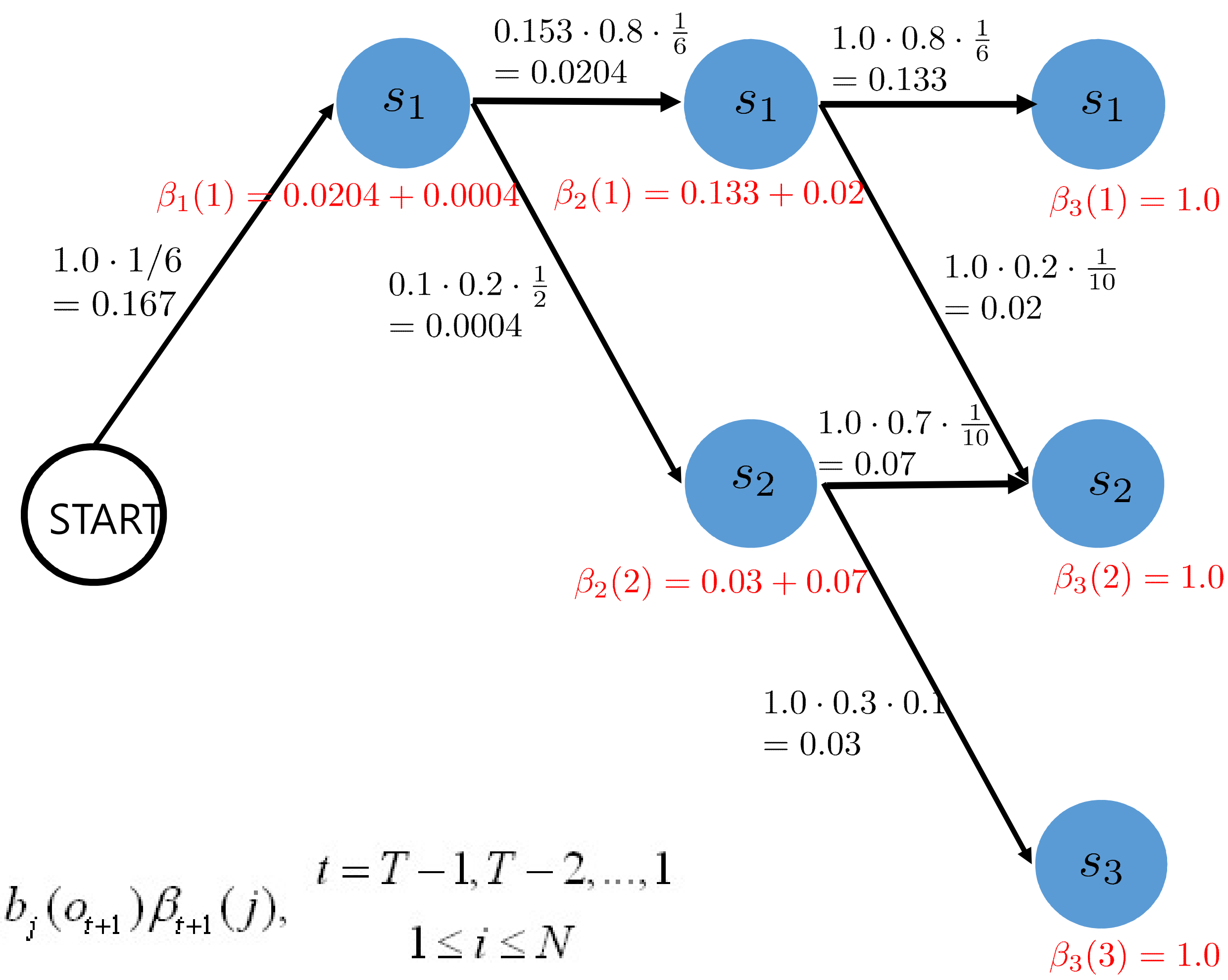
- observation symbol probability distribution

$$B_1 = \begin{bmatrix} b_1(v_1) & b_2(v_1) & b_3(v_1) \\ b_1(v_2) & b_2(v_2) & b_3(v_2) \\ b_1(v_3) & b_2(v_3) & b_3(v_3) \\ b_1(v_4) & b_2(v_4) & b_3(v_4) \\ b_1(v_5) & b_2(v_5) & b_3(v_5) \\ b_1(v_6) & b_2(v_6) & b_3(v_6) \end{bmatrix} = \begin{bmatrix} 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/2 & 1/2 \end{bmatrix}$$

- initial state distribution

$$\boldsymbol{\pi}_1 = [\pi_1, \pi_2, \pi_3] = [1, 0, 0]$$

Sequence: {3, 6, 1}



$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \ldots, 1$$
$$1 \le i \le N$$

The Matching : Viterbi algorithm

to **choose the single best state sequence** which maximizes $P(\mathbf{Q}, O|\lambda)$.

Let us define $\delta_t(i)$ as **the highest probability along a single path**, at time t ,which accounts for the first observation, i.e.

$$\delta_t(i) = \max_{q_1 q_2 \cdots q_{t-1}} P(q_1 q_2 \cdots q_t = s_i, o_1 o_2 \cdots o_t|\lambda)$$

We must keep track of the state sequence which gave the path, at time t to state s_i . We do this in a separate array $\psi_t(i)$.

1. Initialization:

$$\delta_1(i) = \pi_i b_i(o_1) \quad 1 \le i \le N$$

$$\psi_1(i) = 0$$

2. Recursion:

$$\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] b_j(o_t) \quad \begin{array}{c} 2 \le t \le T \\ 1 \le j \le N \end{array}$$

$$\psi_t(i) = \arg\max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] \quad \begin{array}{c} 2 \le t \le T \\ 1 \le j \le N \end{array}$$

3. Termination:

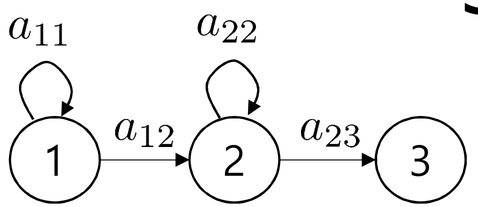
$$P^* = \max_{1 \le i \le N} [\delta_T(i)]$$

$$q_T^* = \arg\max_{1 \le i \le N} [\delta_T(i)]$$

4. Path(state-sequence) backtracking

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \ldots, 1$$

The Matching: Example



$$\lambda_1 = \{\mathbf{A}_1, \mathbf{B}_1, \boldsymbol{\pi}_1\}$$

- state transition probability distribution

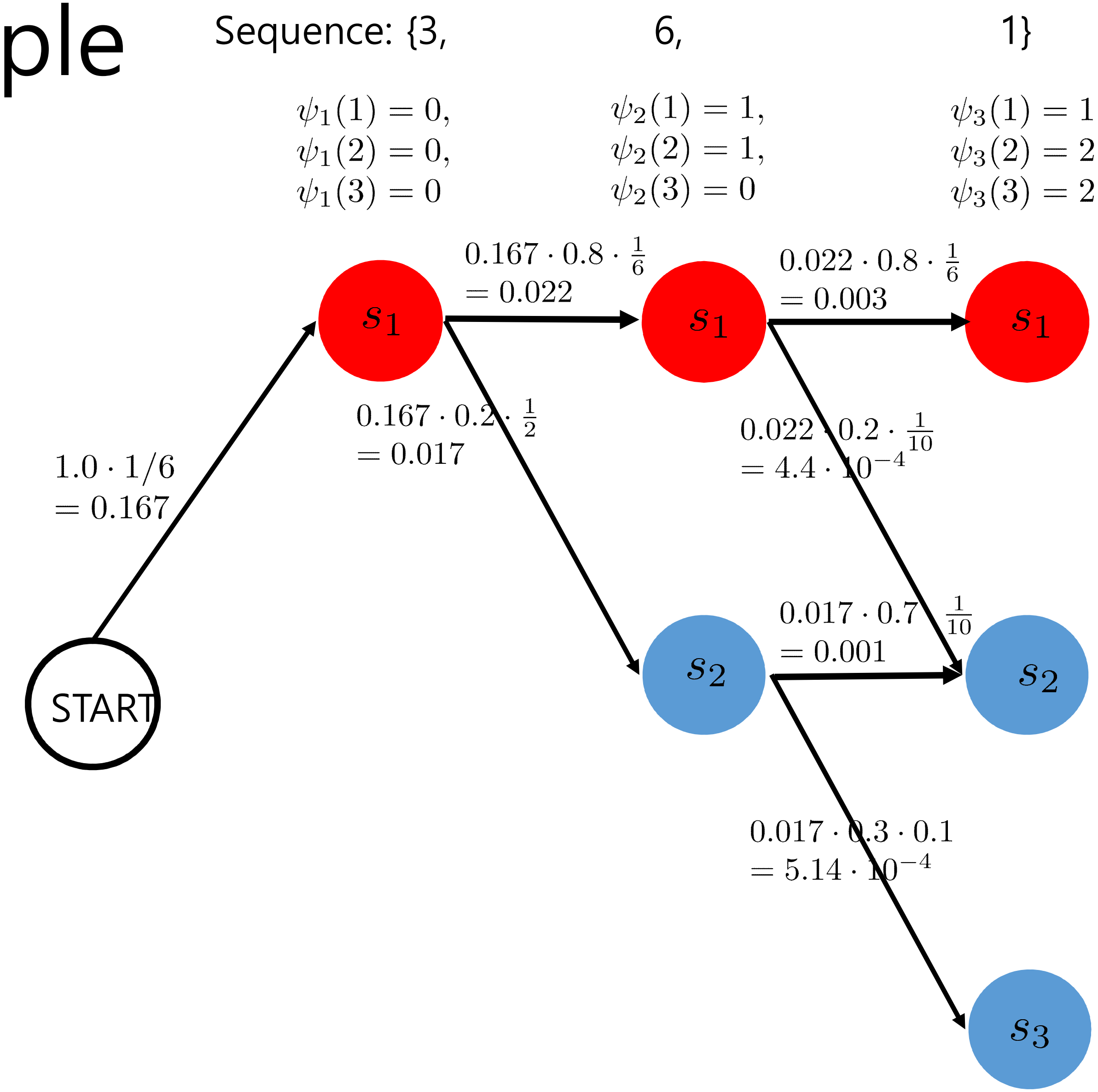
$$A_1 = \begin{bmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & a_{23} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.7 & 0.3 \\ 0 & 0 & 0 \end{bmatrix}$$

- observation symbol probability distribution

$$B_1 = \begin{bmatrix} b_1(v_1) & b_2(v_1) & b_3(v_1) \\ b_1(v_2) & b_2(v_2) & b_3(v_2) \\ b_1(v_3) & b_2(v_3) & b_3(v_3) \\ b_1(v_4) & b_2(v_4) & b_3(v_4) \\ b_1(v_5) & b_2(v_5) & b_3(v_5) \\ b_1(v_6) & b_2(v_6) & b_3(v_6) \end{bmatrix} = \begin{bmatrix} 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/2 & 1/2 \end{bmatrix}$$

- initial state distribution

$$\boldsymbol{\pi}_1 = [\pi_1, \pi_2, \pi_3] = [1, 0, 0]$$



The Training : Baum-Welch Reestimation

- No globally optimum solution is known. However, the Baum-Welch (re)estimation procedures will yield a locally optimal solution iteratively.
- Given the current model estimate λ, the Baum-Welch procedures compute the expected values of model events, then refine the model based on the computed values.
- Let us define $\xi_i(i, j)$ as the probability of being in state s_i at time t and state s_j at time $t+1$ given the model and the observation sequence, i.e.

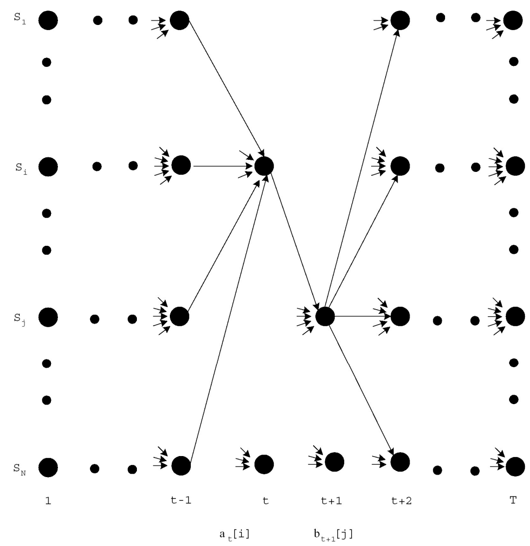
$$\xi_t(i, j) = P(q_t = s_i, q_{t+1} = s_j | O, \lambda) \quad = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)}$$

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i, j)$$

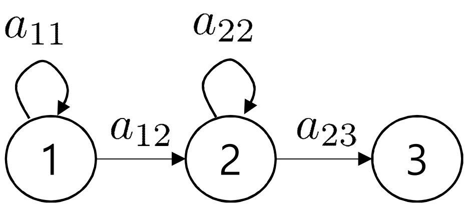
- Summing $\gamma_t(i)$ and $\xi_t(i, j)$ we get

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{ expected number of transitions from } s_i$$

$$\sum_{t=1}^{T-1} \xi_t(i, j) = \text{expected number of transitions from } s_i \text{ to } s_j$$



The Training : Baum-Welch Reestimation



$$\lambda_1 = \{\mathbf{A}_1, \mathbf{B}_1, \boldsymbol{\pi}_1\}$$

$\sum_{t=1}^{T-1} \gamma_t(i) =$ expected number of transitions from s_i

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) = \frac{P(q_t = S_i, O|\lambda)}{P(O,\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_j \alpha_t(j)\beta_t(j)}$$

$$\boxed{\gamma_2(1) = \frac{\alpha_2(1)\beta_2(1)}{\alpha_2(1)\beta_2(1)+\alpha_2(2)\beta_2(2)} = 0.3356}$$

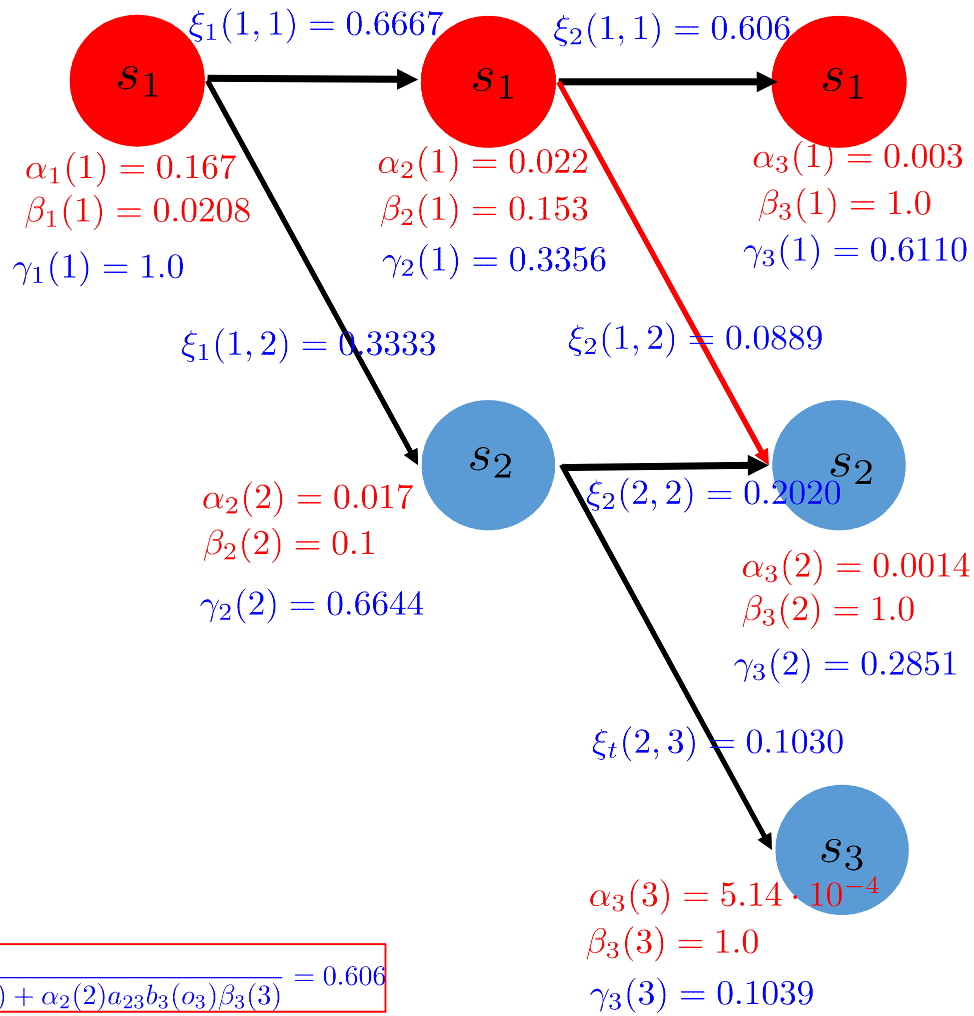
$$\boxed{\gamma_2(2) = \frac{\alpha_2(2)\beta_2(2)}{\alpha_2(1)\beta_2(1)+\alpha_2(2)\beta_2(2)} = 0.6644}$$

$\sum_{t=1}^{T-1} \xi_t(i,j) =$ expected number of transitions from s_i to s_j

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_i \sum_j \alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}$$

$$\boxed{\xi_2(1,1) = \frac{\alpha_2(1)a_{11}b_1(o_3)\beta_3(1)}{\alpha_2(1)a_{11}b_1(o_3)\beta_3(1) + \alpha_2(1)a_{12}b_2(o_3)\beta_3(2) + \alpha_2(2)a_{22}b_2(o_3)\beta_3(2) + \alpha_2(2)a_{23}b_3(o_3)\beta_3(3)} = 0.606}$$

seq:	{3,	6,	1}
state:	{1,	1,	1}



The Training : Baum-Welch Reestimation

$$\bar{\pi} = \text{ expected number of times in state, } s_i \text{ at } t=1$$

$$= \gamma_1(i) \qquad \gamma_1(1)=1.0 \quad \gamma_1(2)=0.0 \quad \gamma_1(3)=0.0$$

$$\bar{a}_{ij} = \frac{\text{expected number of transitions from state } s_i \text{ to } s_j}{\text{expected number of transitions from state } s_i}$$

$$= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\bar{\alpha}_{11} = \frac{\xi_1(1,1)+\xi_2(1,1)}{\gamma_1(1)+\gamma_2(1)} = \frac{0.6667+0.606}{1.0+0.3356} = 0.9529$$

$$\bar{\alpha}_{12} = \frac{\xi_1(1,2)+\xi_2(1,2)}{\gamma_1(1)+\gamma_2(1)} = \frac{0.3333+0.0889}{1.0+0.3356} = 0.3161$$

→ Normalize

$$\bar{\alpha}_{11} = \frac{0.9529}{0.9529+0.3161} = 0.7509$$

$$\bar{\alpha}_{11} = \frac{0.3161}{0.9529+0.3161} = 0.2491$$

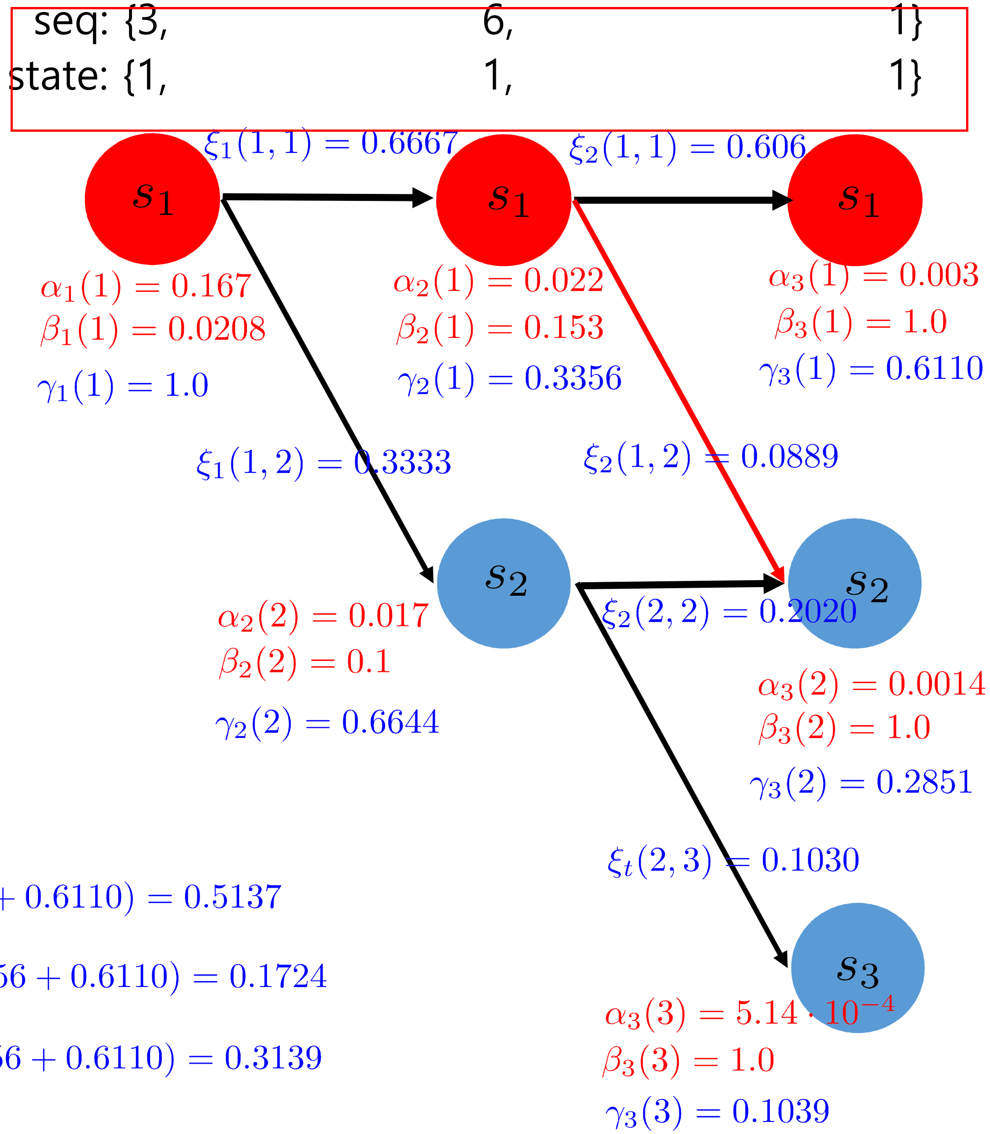
$$\bar{b}_j(k) = \frac{\text{expected number of times in state } s_i \text{ with symbol } v_k}{\text{expected number of times in state } s_j}$$

$$= \frac{\sum_{t=1,\ o_t=v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

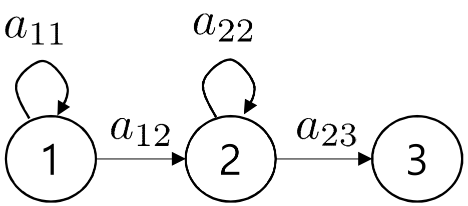
$$b_1(k=3) = \frac{\gamma_1(1)}{\gamma_1(1)+\gamma_2(1)+\gamma_3(1)} = 1.0/(1.0+0.3356+0.6110) = 0.5137$$

$$b_1(k=6) = \frac{\gamma_2(1)}{\gamma_1(1)+\gamma_2(1)+\gamma_3(1)} = 0.3356/(1.0+0.3356+0.6110) = 0.1724$$

$$b_1(k=1) = \frac{\gamma_3(1)}{\gamma_1(1)+\gamma_2(1)+\gamma_3(1)} = 0.6110/(1.0+0.3356+0.6110) = 0.3139$$



The training



$$\lambda_1 = \{\mathbf{A}_1, \mathbf{B}_1, \boldsymbol{\pi}_1\}$$

- state transition probability distribution

$$A_1 = \begin{bmatrix} a_{11} & a_{12} & 0 \\ 0 & a_{22} & a_{23} \\ 0 & 0 & 0 \end{bmatrix} = \boxed{\begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.7 & 0.3 \\ 0 & 0 & 0 \end{bmatrix}}$$

- observation symbol probability distribution

$$B_1 = \begin{bmatrix} b_1(v_1) & b_2(v_1) & b_3(v_1) \\ b_1(v_2) & b_2(v_2) & b_3(v_2) \\ b_1(v_3) & b_2(v_3) & b_3(v_3) \\ b_1(v_4) & b_2(v_4) & b_3(v_4) \\ b_1(v_5) & b_2(v_5) & b_3(v_5) \\ b_1(v_6) & b_2(v_6) & b_3(v_6) \end{bmatrix} = \begin{bmatrix} 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/10 & 1/10 \\ 1/6 & 1/2 & 1/2 \end{bmatrix}$$

- initial state distribution

$$\boldsymbol{\pi}_1 = [\pi_1, \pi_2, \pi_3] = [1, 0, 0]$$

Iter10

$$\hat{A}_1 \quad \begin{matrix} 0.4216 & 0.5784 & 0.0000 \\ 0 & 1.0000 & 0.0000 \\ 0 & 0 & 0 \end{matrix}$$

$$\hat{B}_1 \quad \begin{matrix} 0.1739 & 0.1751 & 0.1807 & 0.1583 & 0.1654 & 0.1466 \\ 0.1351 & 0.1223 & 0.1273 & 0.1345 & 0.1291 & 0.3517 \\ 0.0006 & 0.0005 & 0.0004 & 0.0015 & 0.0000 & 0.9970 \end{matrix}$$

Iter20

$$\hat{A}_1 \quad \begin{matrix} 0.8990 & 0.1010 & 0.0000 \\ 0 & 1.0000 & 0.0000 \\ 0 & 0 & 0 \end{matrix}$$

$$\hat{B}_1 \quad \begin{matrix} 0.1556 & 0.1558 & 0.1558 & 0.1535 & 0.1598 & 0.2194 \\ 0.0847 & 0.0964 & 0.0951 & 0.1052 & 0.0821 & 0.5364 \\ 0.0034 & 0.0411 & 0.0572 & 0.0023 & 0.0089 & 0.8869 \end{matrix}$$

Iter50

$$\hat{A}_1 \quad \boxed{\begin{matrix} 0.8418 & 0.1582 & 0.0000 \\ 0 & 1.0000 & 0.0000 \\ 0 & 0 & 0 \end{matrix}}$$

→ Closes to A_1

$$\hat{B}_1 \quad \begin{matrix} 0.1604 & 0.1653 & 0.1707 & 0.1576 & 0.1712 & 0.1747 \\ 0.1082 & 0.0937 & 0.0981 & 0.1089 & 0.1033 & 0.4878 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{matrix}$$