SOON CHUN HYANG UNIVERSITY

Hidden Markov Models

SCH Univ.
Dept. of AI and Bigdata
Lee Howoo

Senseable AI Lab

개요

순차 데이터

$$X = (X_1, \cdots X_i, X_{i+1}, \cdots X_d)$$

$$X = (X_1, \cdots X_t, X_{t+1}, \cdots X_T)$$

순차 데이터

$$X = (X_1, \cdots X_i, X_{i+1}, \cdots X_d)$$

$$X = (X_{\color{red}1}, \cdots X_{\color{red}t}, X_{\color{red}{t+1}}, \cdots X_{\color{red}T})$$

→ Sequential Data (순차 데이터)

순차 데이터 예시

- 시간 특성이 있는 데이터
- Sequence data (Discrete)

Equipment Sequential data in manufacturing processes

→ **Time**

	공정1	공정2	공정3	공정4	공정5	공정6	공정7	공정8	Class
1	설비1	설비2	설비1	설비1	설비2	설비1	설비4	설비2	정상
2	설비2	설비4	설비2	설비4	설비2	설비4	설비2	설비4	정상
3	설비1	설비3	설비1	설비1	설비3	설비1	설비3	설비1	정상
…	…	…	…	…	…	…	…	…	…
2998	설비3	설비3	설비3	설비3	설비3	설비3	설비3	설비3	불량
2999	설비4	설비1	설비2	설비4	설비6	설비5	설비5	설비1	불량
3000	설비5	설비5	설비1	설비3	설비1	설비5	설비1	설비3	불량

순차 데이터 예시

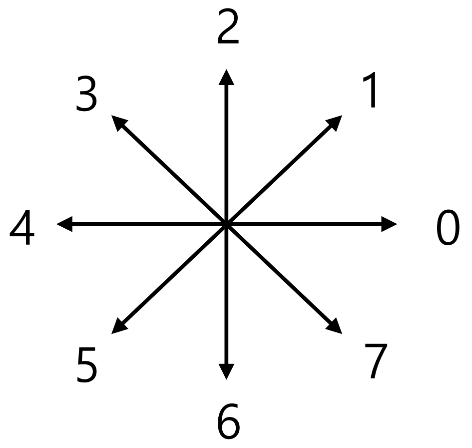
순차 데이터:$X = (X_1, \cdots X_t, X_{t+1}, \cdots X_T)$

순차 데이터 인식 예시

숫자 이미지

2

상태(state) 정의

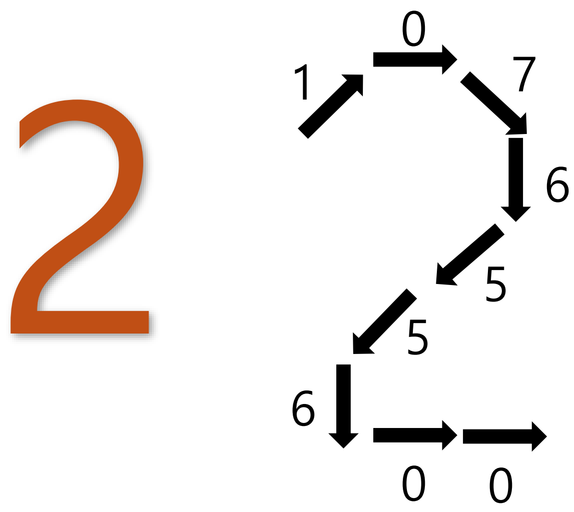


순차 데이터 예시

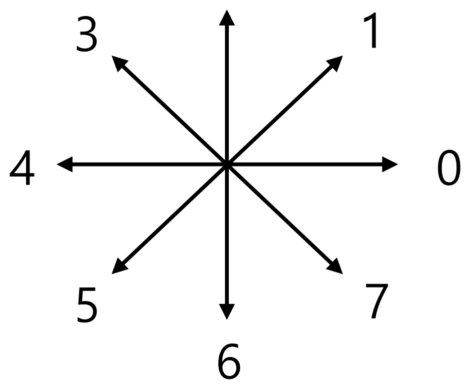
순차 데이터:$X = (X_1, \cdots X_t, X_{t+1}, \cdots X_T)$

순차 데이터 인식 예시

숫자 이미지



상태(state) 정의

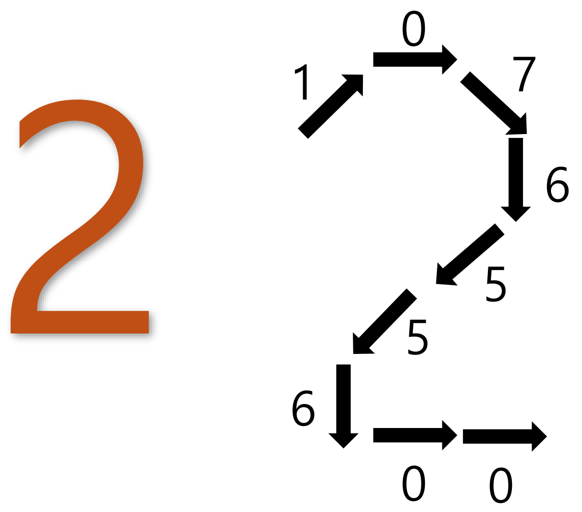


순차 데이터 예시

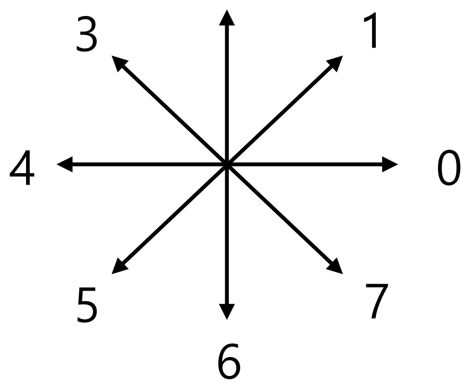
순차 데이터: $X = (X_{1}, \cdots X_{t}, X_{t+1}, \cdots X_{T})$

순차 데이터 인식 예시

숫자 이미지



상태(state) 정의



$$X = (1, 0, 7, 6, 5, 5, 6, 0, 0)$$

Hidden Markov Model 개념

- Hidden Markov Model (HMM) 이란?
    - 순차데이터를 확률적 (Stochastic)으로 모델링하는 생성 모델 (Generative Model)

Hidden Markov Model

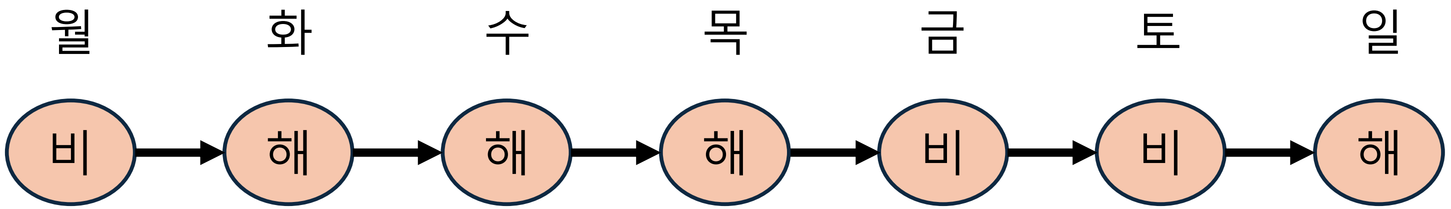
Hidden → 숨겨져 있는 (관측이 되지 않는)

Markov Model → 마코프 모델

Markov Model

- Markov Model 이란?
  - State로 이루어진 Sequence를 상태 전이 확률 행렬로 표현하는 것
  - 상태 전이 확률 행렬 예시

Sequence



State: {비, 해}

State transition probability matrix :
(상태 전이 확률 행렬)

To

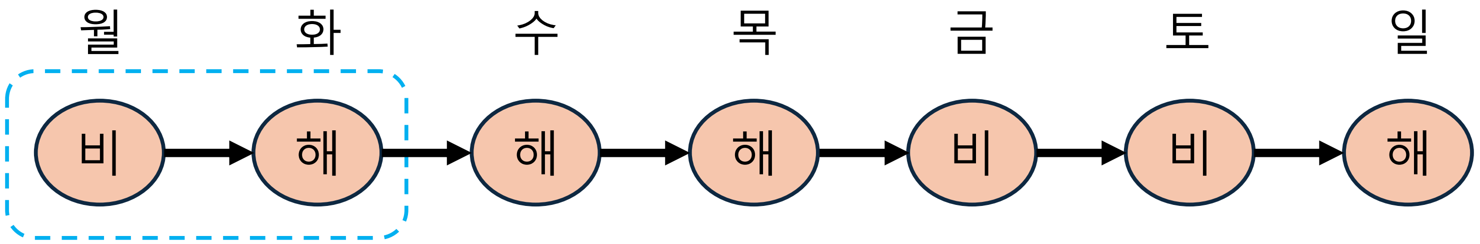
From

$$\begin{array}{c|cc} & 비 & 해 \\ \hline 비 & 0 & 0 \\ 해 & 0 & 0 \end{array} \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Markov Model

- Markov Model 이란?
  - State로 이루어진 Sequence를 상태 전이 확률 행렬로 표현하는 것
  - 상태 전이 확률 행렬 예시

Sequence



State: {비, 해}

State transition probability matrix :
(상태 전이 확률 행렬)

To

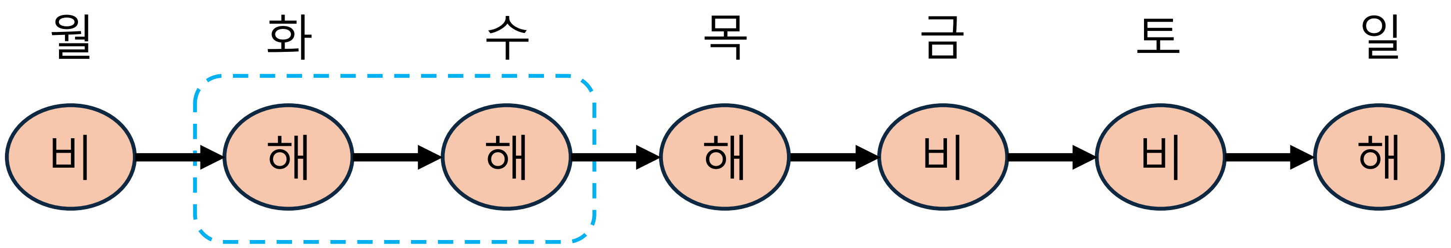
From

$$\begin{array}{c c} & \begin{array}{cc} \text{비} & \text{해} \end{array} \\ \begin{array}{c} \text{비} \\ \text{해} \end{array} & \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \end{array}$$

Markov Model

- Markov Model 이란?
  - State로 이루어진 Sequence를 상태 전이 확률 행렬로 표현하는 것
  - 상태 전이 확률 행렬 예시

Sequence



State: {비, 해}

State transition probability matrix :
(상태 전이 확률 행렬)

To

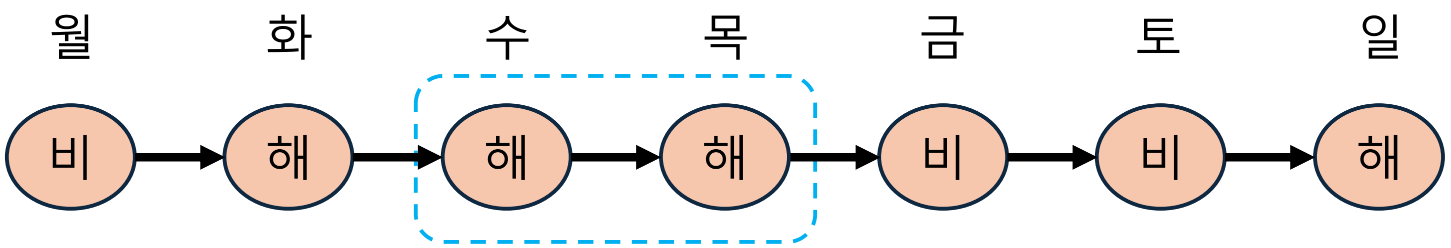
From

$$
\begin{array}{cc} & \begin{array}{cc} \text{비} & \text{해} \end{array} \\ \begin{array}{c} \text{비} \\ \text{해} \end{array} & \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \end{array}
$$

Markov Model

- Markov Model 이란?
  - State로 이루어진 Sequence를 상태 전이 확률 행렬로 표현하는 것
  - 상태 전이 확률 행렬 예시

Sequence



State: {비, 해}

State transition probability matrix :
(상태 전이 확률 행렬)

To

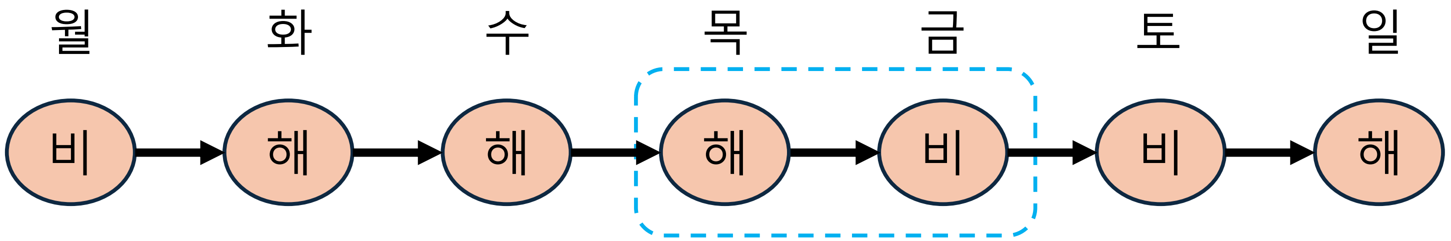
From

$$
\begin{array}{cc} & \begin{array}{cc} 비 & 해 \end{array} \\ \begin{array}{c} 비 \\ 해 \end{array} & \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix} \end{array}
$$

Markov Model

- Markov Model 이란?
  - State로 이루어진 Sequence를 상태 전이 확률 행렬로 표현하는 것
  - 상태 전이 확률 행렬 예시

Sequence



State: {비, 해}

State transition probability matrix :
(상태 전이 확률 행렬)

To

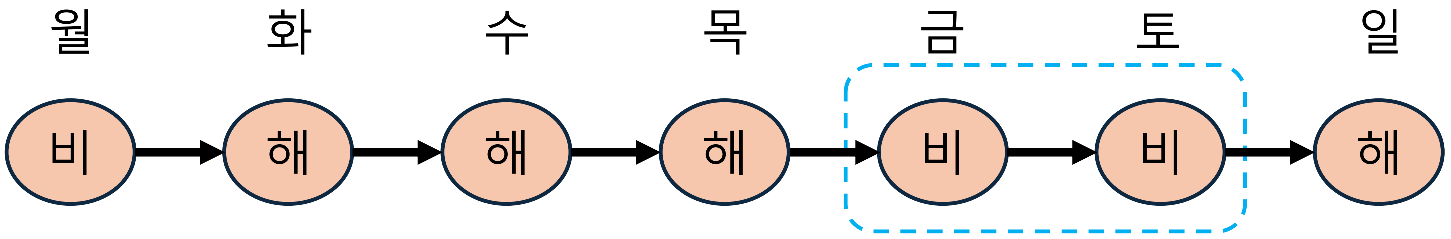
From

$$\begin{array}{c c} & \begin{array}{cc} 비 & 해 \end{array} \\ \begin{array}{c} 비 \\ 해 \end{array} & \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} \end{array}$$

Markov Model

- Markov Model 이란?
  - State로 이루어진 Sequence를 상태 전이 확률 행렬로 표현하는 것
  - 상태 전이 확률 행렬 예시

Sequence



State: {비, 해}

State transition probability matrix :
(상태 전이 확률 행렬)

To

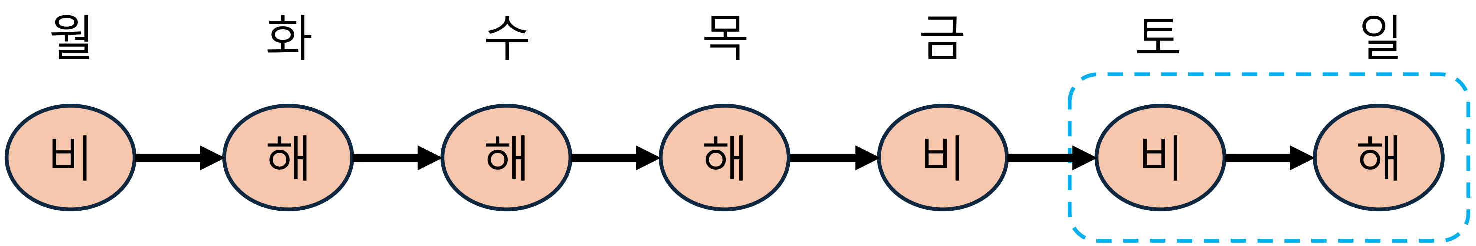
From

$$\begin{array}{c c} & \begin{array}{cc} 비 & 해 \end{array} \\ \begin{array}{c} 비 \\ 해 \end{array} & \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \end{array}$$

Markov Model

- Markov Model 이란?
  - State로 이루어진 Sequence를 상태 전이 확률 행렬로 표현하는 것
  - 상태 전이 확률 행렬 예시

Sequence



State: {비, 해}

State transition probability matrix :
(상태 전이 확률 행렬)

To

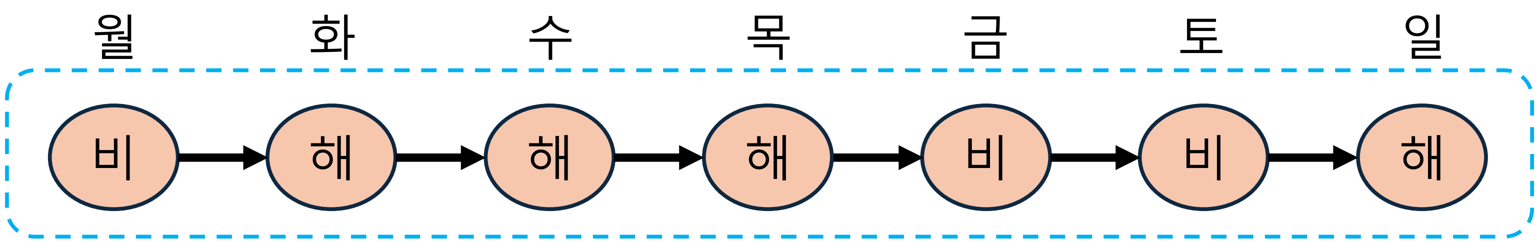
From

$$\begin{array}{c|cc} & 비 & 해 \\ \hline 비 & 1 & 2 \\ 해 & 1 & 2 \end{array} \quad \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$$

Markov Model

- Markov Model 이란?
  - State로 이루어진 Sequence를 상태 전이 확률 행렬로 표현하는 것
  - 상태 전이 확률 행렬 예시

Sequence



State: {비, 해}

State transition probability matrix :
(상태 전이 확률 행렬)

To

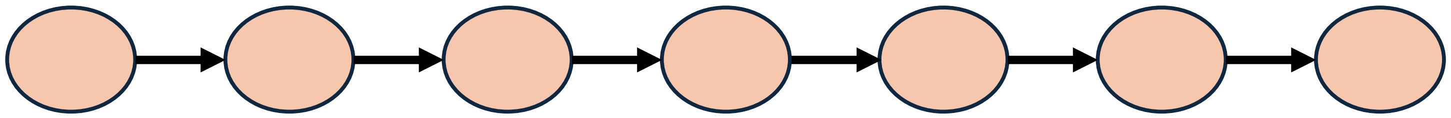
From

$$
\begin{array}{c|cc} & 비 & 해 \\ \hline 비 & 1/3 & 2/3 \\ 해 & 1/3 & 2/3 \end{array}
\quad
\begin{bmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{bmatrix}
$$

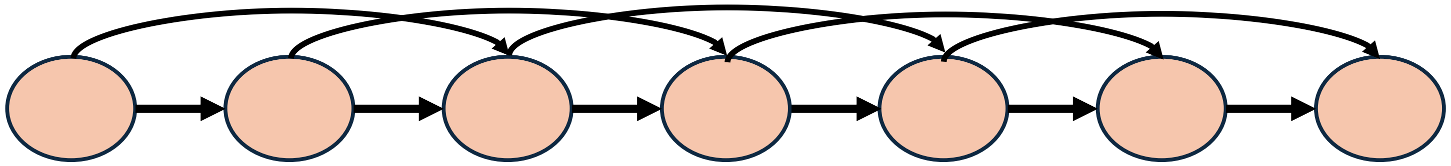
Markov Model

- Markov Model 이란?
  - Markov 가정: 시간 t에서 관측은 가장 최근 r개의 관측에만 의존한다는 가정
  - 한 상태에서 다른 상태로의 전이는 이전 상태의 긴 이력을 필요치 않다는 가정
    - If r = 1, $p(s_t|s_{t-1}, s_{t-2} \cdots s_1) = p(s_t|s_{t-1})$
    - If r = 2, $p(s_t|s_{t-1}, s_{t-2} \cdots s_1) = p(s_t|s_{t-1}, s_{t-2})$
    - …

First order Markov model



Second order Markov model



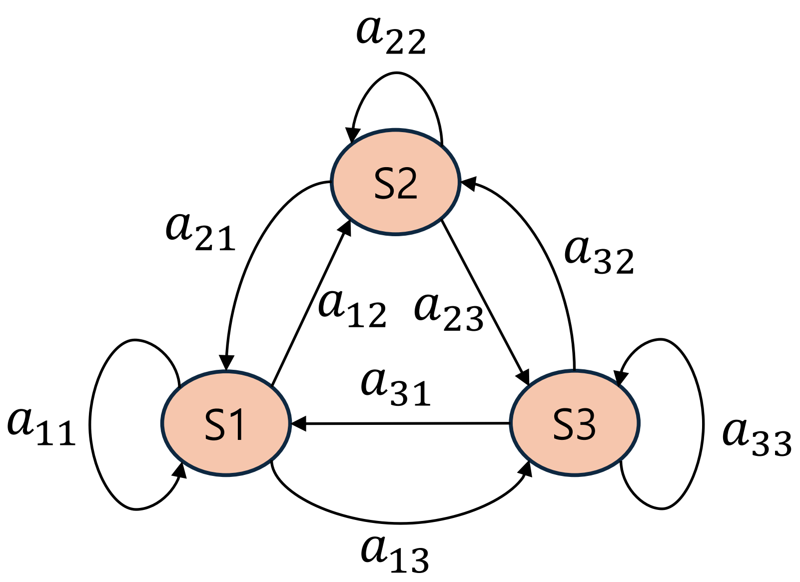
Markov Model

- Parameters of the Markov model
  - 상태 전이 확률 행렬 A(a_{ij})

Matrix Form

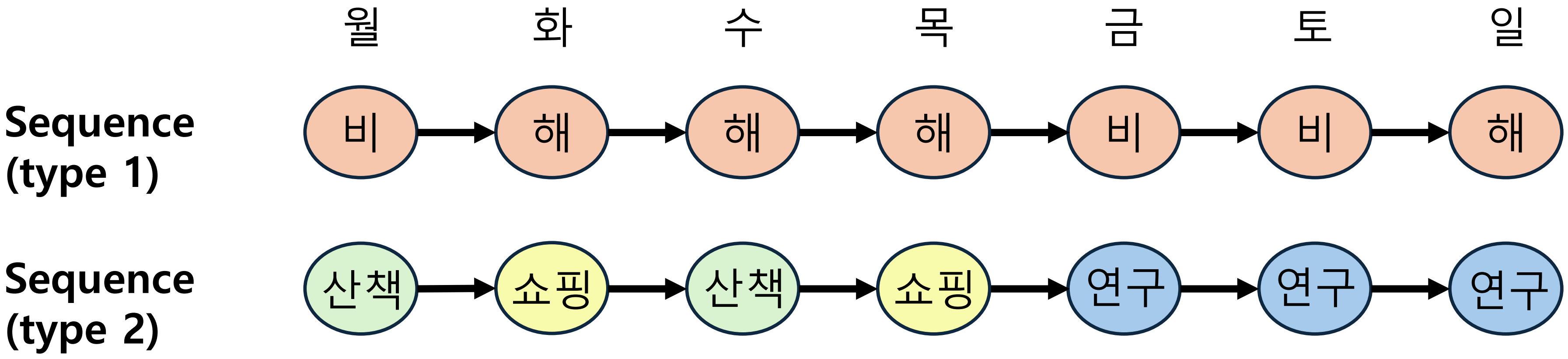
$$T-1 \quad \begin{array}{c} \\ S1 \\ S2 \\ S3 \end{array} \overset{T}{\begin{array}{ccc} S1 & S2 & S3 \\ \end{array}} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} =$$

Graph Form



Hidden Markov Model 개념

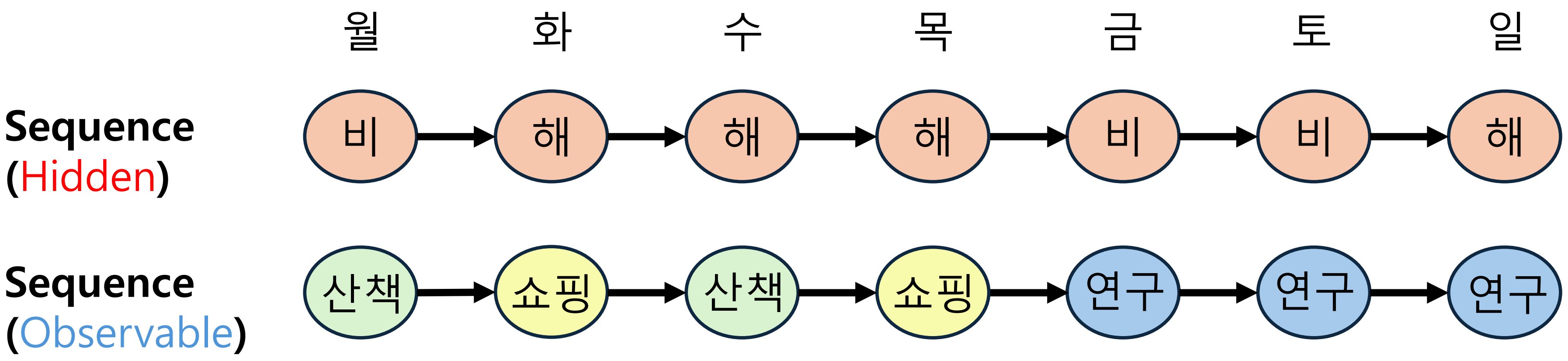
- Hidden Markov Model 이란?
  - Type 1 state sequence : $(s_1, s_2, s_3, \cdots, s_{T-1}, s_T)$
  - Type 2 state sequence : $(o_1, o_2, o_3, \cdots, o_{T-1}, o_T)$



같은 시간에 발생한 두 종류의 state sequence 각각의 특성과 그들의 관계를 이용해 모델링

Hidden Markov Model 개념

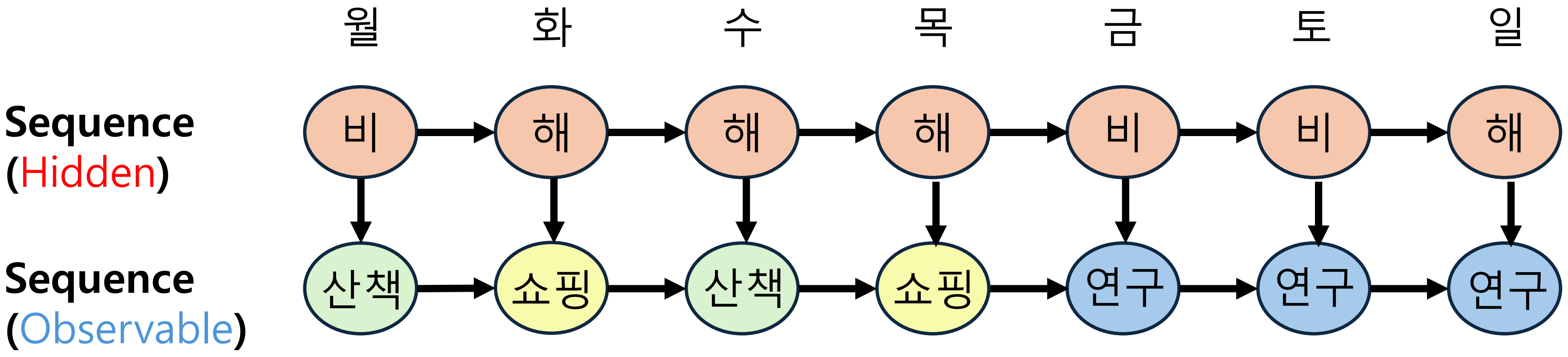
- Hidden Markov Model 이란?
  - Hidden state sequence : $(s_1, s_2, s_3, \cdots, s_{T-1}, s_T)$
  - Observable state sequence : $(o_1, o_2, o_3, \cdots, o_{T-1}, o_T)$



Type 1 Sequence는 숨겨져 있고 **(**Hidden**)**, Type 2 Sequence는 관측이 가능 (Observable)

Hidden Markov Model 개념

- Hidden Markov Model 이란?
  - Hidden state sequence : $(s_1, s_2, s_3, \cdots, s_{T-1}, s_T)$
  - Observable state sequence : $(o_1, o_2, o_3, \cdots, o_{T-1}, o_T)$

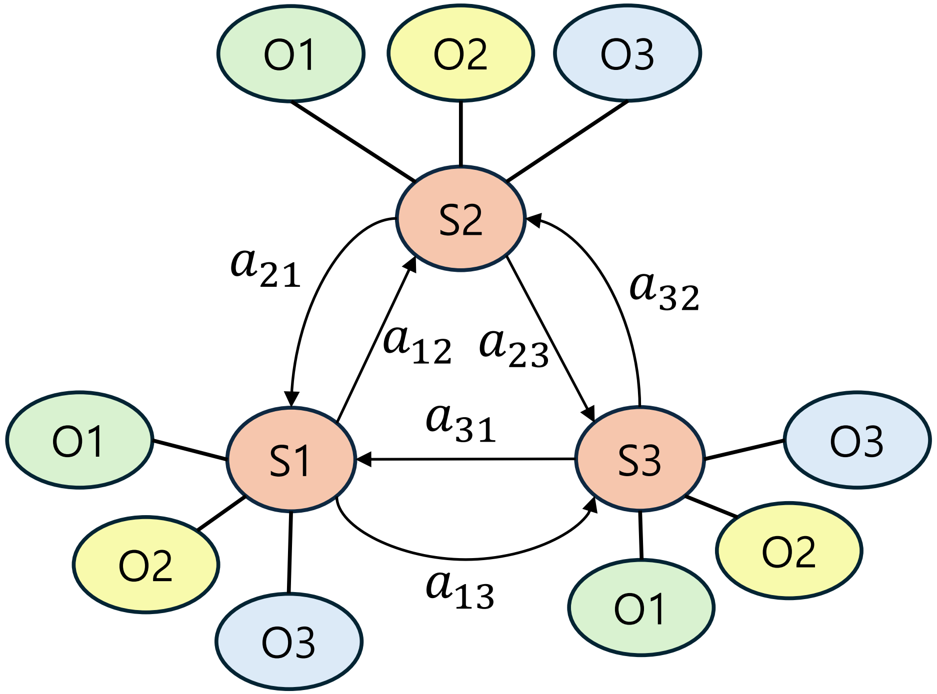


- Hidden sequence가 Markov assumption을 따름 -> 순차적 특성을 반영
- Observable sequence는 순차적 특성을 반영하는 Hidden sequence에 종속

Hidden Markov Model 개념

- Hidden Markov Model

Hidden Markov model (Graph Form)

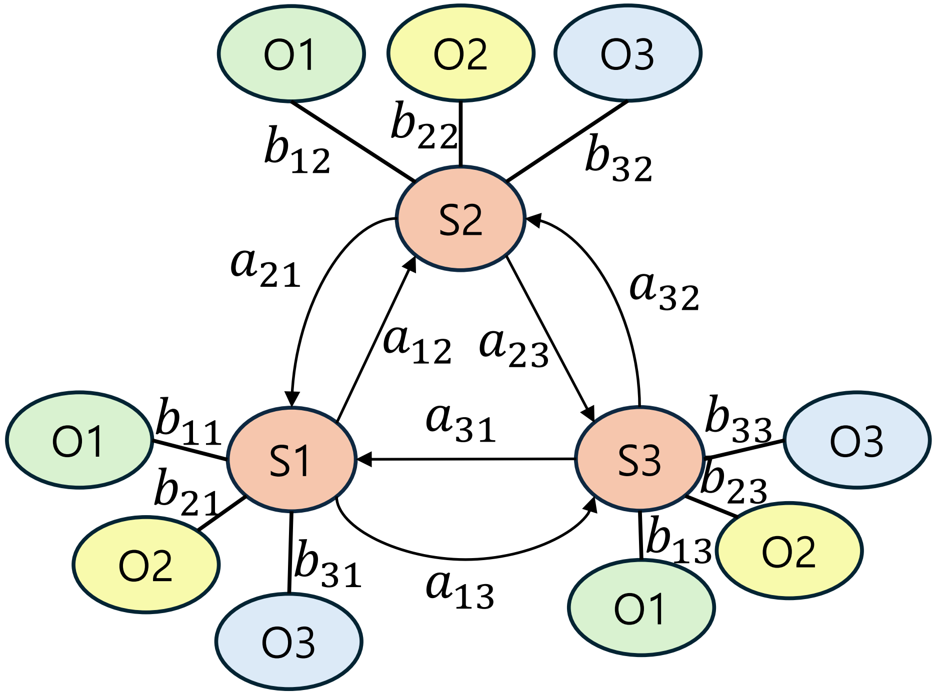


Hidden Markov Model 예제

- Hidden Markov Model 예제 (Observable vs Hidden)
  - 특정인의 행동 (쇼핑, 연구, 산책)에 따라 날씨 추측
  - 빈대떡 소비량에 따른 날씨 추측
  - 중국음식 배달량에 따른 날씨 추측
  - 공 색깔 정보에 따라 그 공이 담겨있는 상자의 종류 추측
  - DNA 염기서열 (ACGT …)에서 어느 부분이 유전자 (Gene)인지 추측
  - 주어진 단어의 품사 추측

Hidden Markov Model – Parameters

- Parameters of a hidden Markov model
  - $A(a_{ij})$: State transition probability matrix (상태전이확률 행렬)
  - $\mathrm{B}(b_{jk})$: Emission probability matrix (방출확률 행렬)

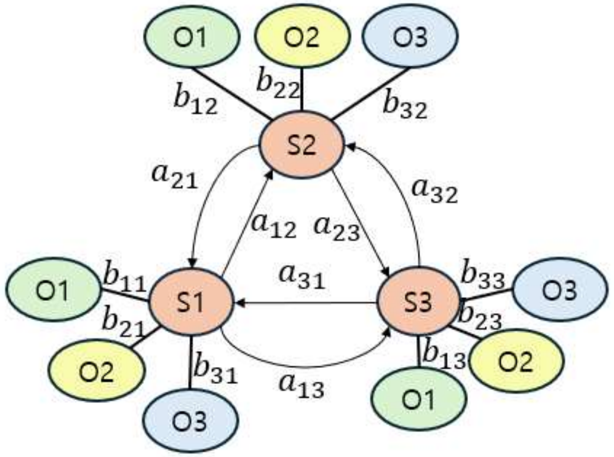


$$A = \begin{array}{c} \\ S1 \\ S2 \\ S3 \end{array} \begin{array}{c} \begin{array}{ccc} S1 & S2 & S3 \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \end{array}$$

$$B = \begin{array}{c} \\ O1 \\ O2 \\ O3 \end{array} \begin{array}{c} \begin{array}{ccc} S1 & S2 & S3 \end{array} \\ \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \end{array}$$

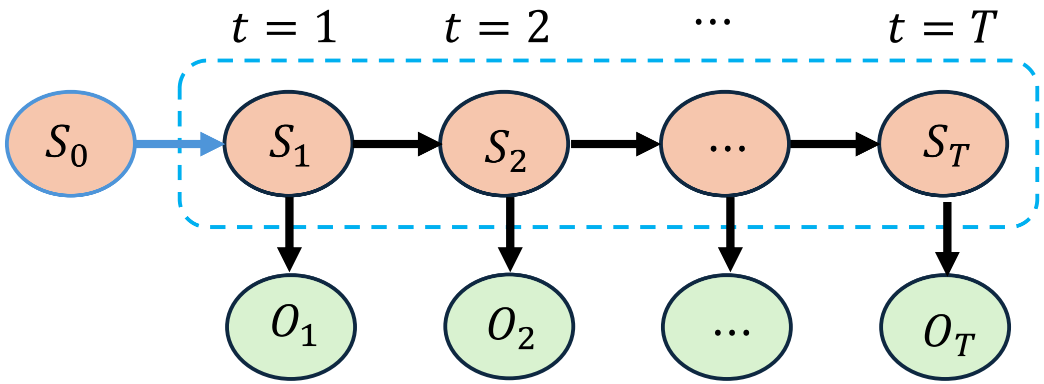
Hidden Markov Model – Parameters

- Parameters of a hidden Markov model
  - $A(a_{ij})$: State transition probability matrix
  - $\mathrm{B}(b_{jk})$: Emission probability matrix
  - $\pi(\pi_i)$: Initial state probability matrix



Hidden states

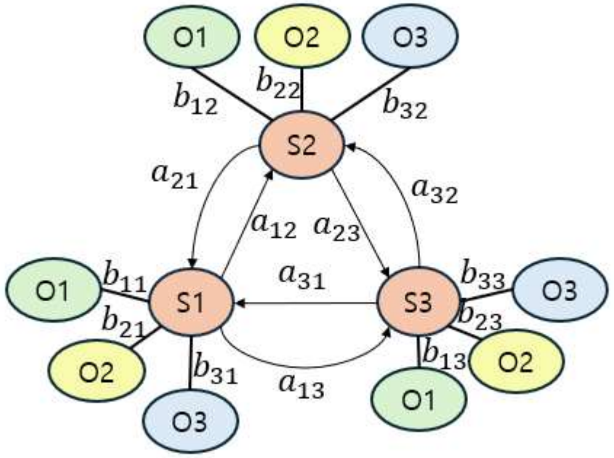
Observations

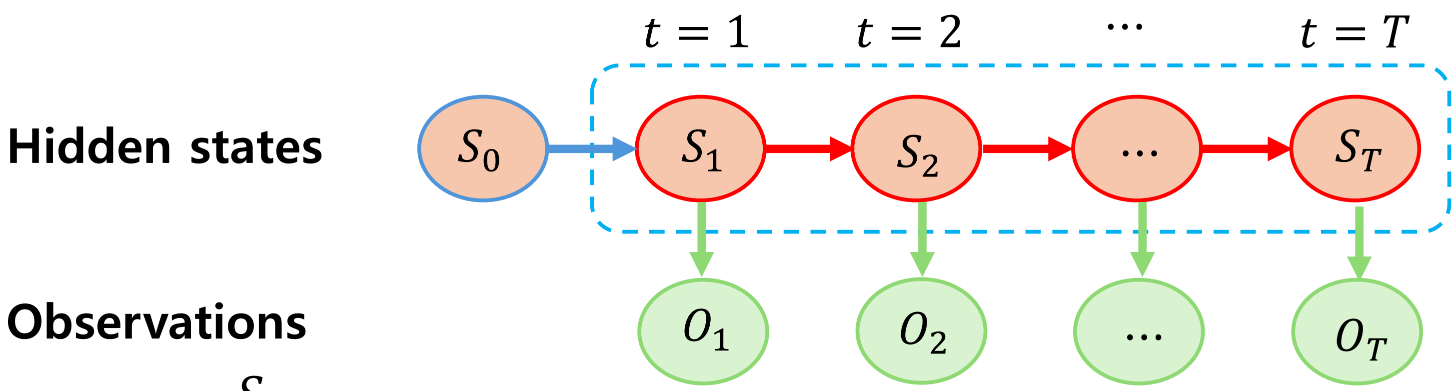


$$\pi = \begin{matrix} S1 \\ S2 \\ S3 \end{matrix} \overset{S_{t=0}}{\begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix}}$$

Hidden Markov Model – Parameters

- Parameters of a hidden Markov model (λ)
  - $A(a_{ij})$: State transition probability matrix
  - $B(b_{jk})$: Emission probability matrix
  - $\pi(\pi_i)$: Initial state probability matrix

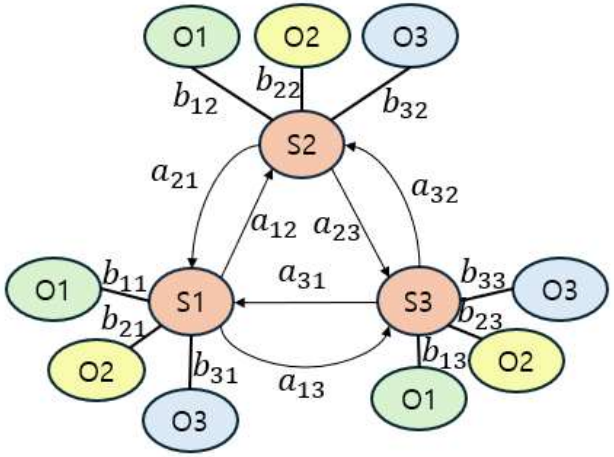




$$\pi = \begin{array}{c} \\ S1 \\ S2 \\ S3 \end{array} \begin{array}{c} S_{t=0} \\ \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} \end{array} \quad A = \begin{array}{c} \\ S1 \\ S2 \\ S3 \end{array} \begin{array}{c} \quad S1 \quad\quad S2 \quad\quad S3 \\ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \end{array} \quad B = \begin{array}{c} \\ O1 \\ O2 \\ O3 \end{array} \begin{array}{c} \quad S1 \quad\quad S2 \quad\quad S3 \\ \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \end{array}$$

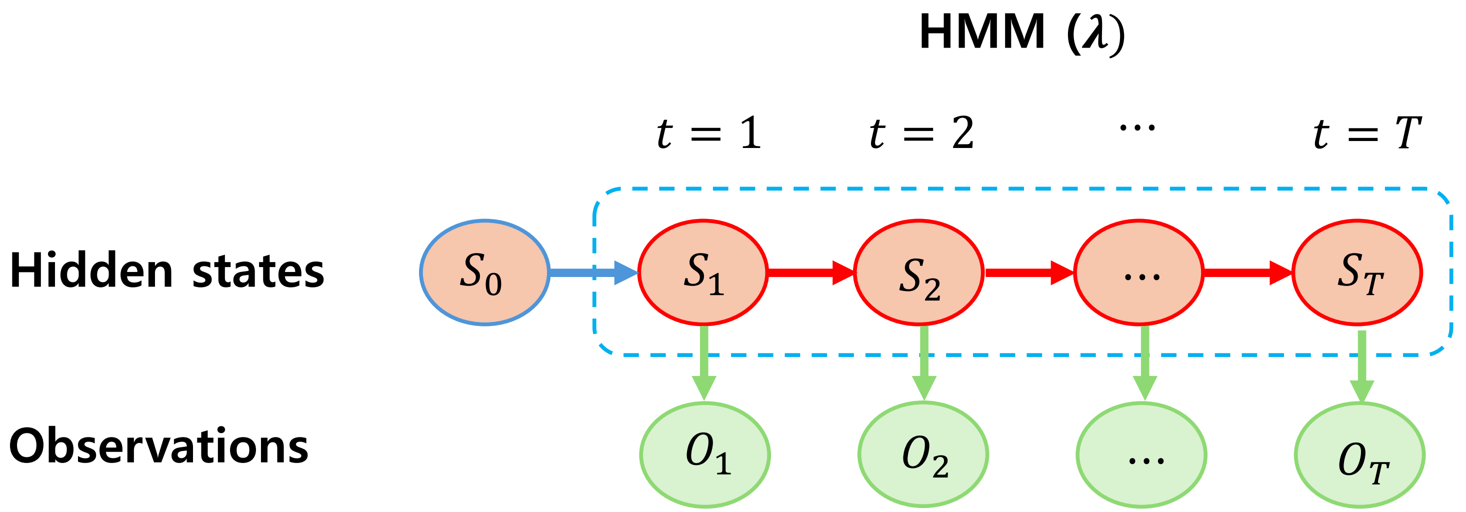
Hidden Markov Model – Parameters

- Parameters of a hidden Markov model : $\lambda = \{A, B, \pi\}$
  - State transition probability matrix (A) : $A = |a_{ij}|$
    - A -> HMM이 작동하는 도중 다음 상태를 결정
    - $a_{ij} = p(q_{t+1} = s_j | q_t = s_i), 1 \le i, j \le n$
    - $\sum_{j=1}^{n} a_{ij} = 1$
  - Emission probability matrix (B) : $B = |b_j(v_k)|$
    - B는 HMM이 어느 상태에 도달하였을 때, 그 상태에서 관측될 확률 결정
    - $b_j(v_k)$: 은닉 상태 b_j에서 관측치 v_k가 도출될 확률
    - $b_j(v_k) = P(o_t = v_k | q_t = s_j),\ 1 \le j \le n,\ 1 \le k \le m$
    - $\sum_{j=1}^{n} b_j(v_k) = 1$
  - Initial state probability matrix (π) : $\pi = |\pi_i|$
    - π -> HMM 을 가동시킬 때 어느 상태에서 시작할지 결정
    - π_i -> s_i 에서 시작할 확률
    - $\sum_{j=1}^{n} \pi_i = 1$



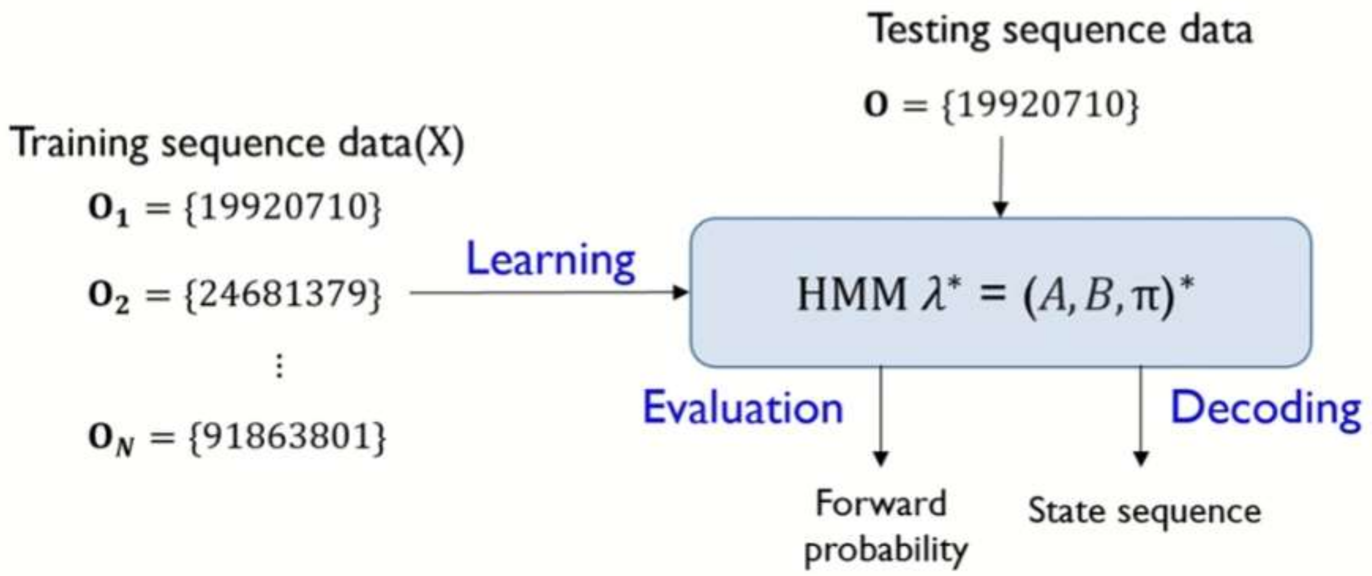
Main Problems of Hidden Markov Models

- Hidden Markov Model $\lambda = [A, B, \pi]$
- Three problems of hidden Markov model
  - Given HMM (λ^*) and **O**, find the probability of **O** -> **<u>Evaluation</u>** problem
  - Given HMM (λ^*) and **O,** find the optimal hidden state sequence **(S)** -> **<u>Decoding</u>** problem
  - Given $X = \{O_{1,}, \cdots, O_N\}$, find the HMM ($\lambda^*$) -> **<u>Learning</u>** problem



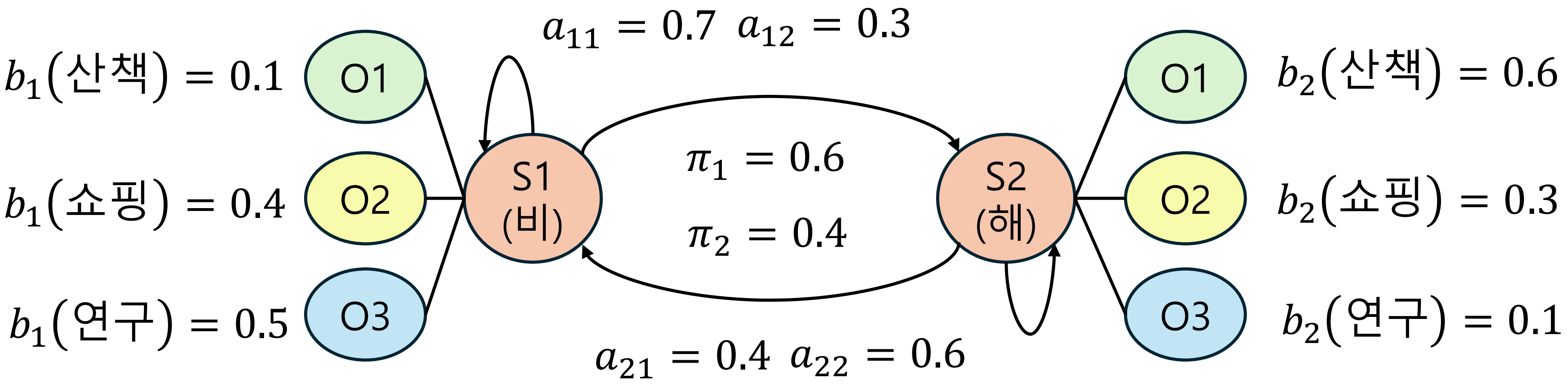
Main Problems of Hidden Markov Models

- Hidden Markov Model $\lambda = [A, B, \pi]$
- Three problems of hidden Markov model
  - Given HMM (λ^*) and **O**, find the probability of **O** -> **Evaluation** problem
  - Given HMM (λ^*) and **O,** find the optimal hidden state sequence **(S)** -> **Decoding** problem
  - Gicen $\mathrm{X} = \{O_{1,}, \cdots, O_N\}$, find the HMM ($\lambda^*$) -> **Learning** problem



Hidden Markov Models - Evaluation

- Evaluation problem
  - Problem : Given HMM (λ^*) and **O**, find the probability of **O**
  - Solution : Forward algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑할 확률은?



$\boldsymbol{O} = (o_1 = \text{산책}, o_2 = \text{산책}, o_3 = \text{연구}, o_4 = \text{쇼핑})$

$\boldsymbol{Probability(O)} = (o_1 = \text{산책}, o_2 = \text{산책}, o_3 = \text{연구}, o_4 = \text{쇼핑}) = ?$

Hidden Markov Models - Evaluation

$$\boldsymbol{Probability}\ (o_1 = \text{산책}, o_2 = \text{산책}, o_3 = \text{연구}, o_4 = \text{쇼핑})$$

$$\boldsymbol{Probability}\left(o_1 = \text{산책}, o_2 = \text{산책}, o_3 = \text{연구}, o_4 = \text{쇼핑}\ \middle|\ \text{비}, \text{비}, \text{비}, \text{비}\right)$$

Hidden Markov Models - Evaluation

1	비	비	비	비
2	비	비	비	해
3	비	비	해	비
4	비	해	비	비
5	해	비	비	비
6	해	해	비	비
7	비	해	해	비
8	비	비	해	해
9	해	비	비	해
10	해	비	해	비
11	비	해	비	해
12	비	해	해	비
13	비	해	해	해
14	해	비	해	해
15	해	해	비	해
16	해	해	해	해

$\boldsymbol{Probability}\ (o_1 = \text{산책}, o_2 = \text{산책},\ o_3 = \text{연구},\ o_4 = \text{쇼핑})$

$\boldsymbol{Probability}\ \left(o_1 = \text{산책}, o_2 = \text{산책},\ o_3 = \text{연구},\ o_4 = \text{쇼핑} \mid \text{비}, \text{비}, \text{비}, \text{비}\right)$

총 경우의 수 = $2^4 = 16$개

Hidden Markov Models - Evaluation

1	비	비	비	비
2	비	비	비	해
3	비	비	해	비
4	비	해	비	비
5	해	비	비	비
6	해	해	비	비
7	비	해	해	비
8	비	비	해	해
9	해	비	비	해
10	해	비	해	비
11	비	해	비	해
12	비	해	해	비
13	비	해	해	해
14	해	비	해	해
15	해	해	비	해
16	해	해	해	해

$\boldsymbol{Probability}\ (o_1 = \text{산책}, o_2 = \text{산책},\ o_3 = \text{연구},\ o_4 = \text{쇼핑})$

$\boldsymbol{Probability}\ \left(o_1 = \text{산책}, o_2 = \text{산책},\ o_3 = \text{연구},\ o_4 = \text{쇼핑}\ \middle|\ \text{비},\text{비},\text{비},\text{비}\right)$

총 경우의 수 = $2^4 = 16$개

총 상태 개수 = N

Sequence 길이 = T

총 경우의 수 = N^T

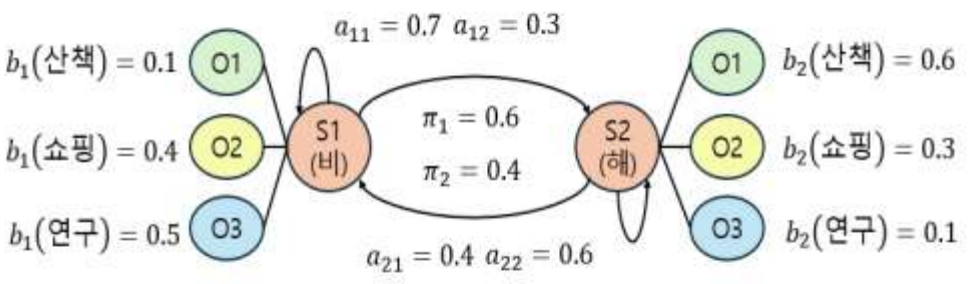
총 상태 개수 = 3

Sequence 길이 = 20

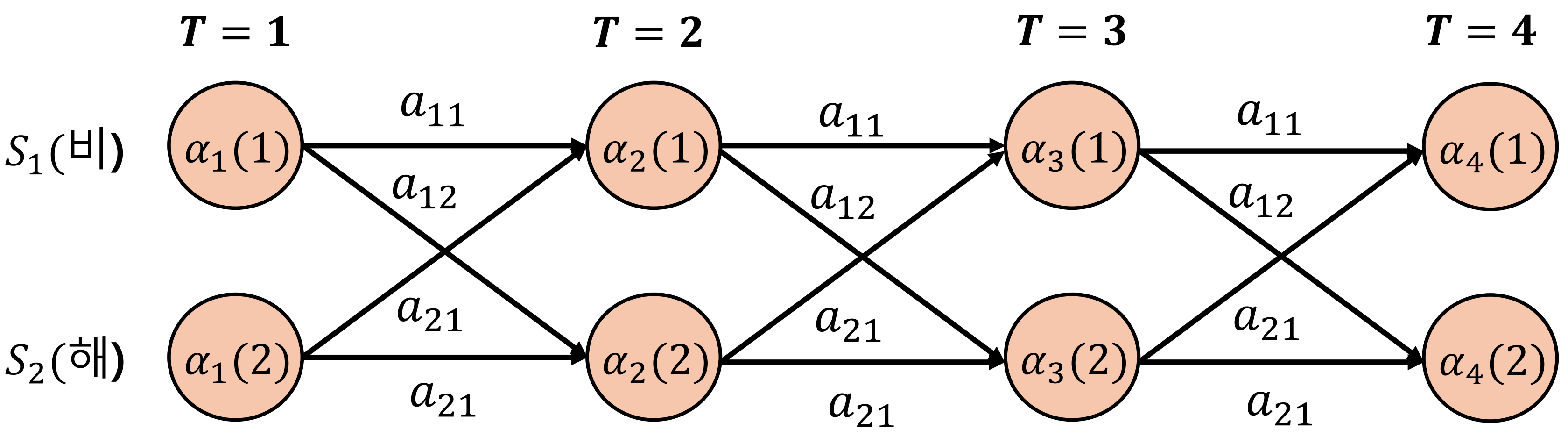
총 경우의 수 = $3^{20} = 3{,}486{,}784{,}401$개 (약 35억)

Hidden Markov Models - Evaluation

- Evaluation problem
  - Problem : Given HMM (λ^*) and **O**, find the probability of **O**
  - Solution : Forward algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑할 확률은?



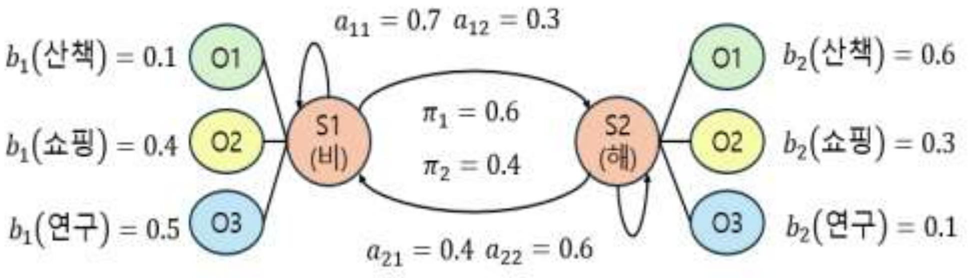
states



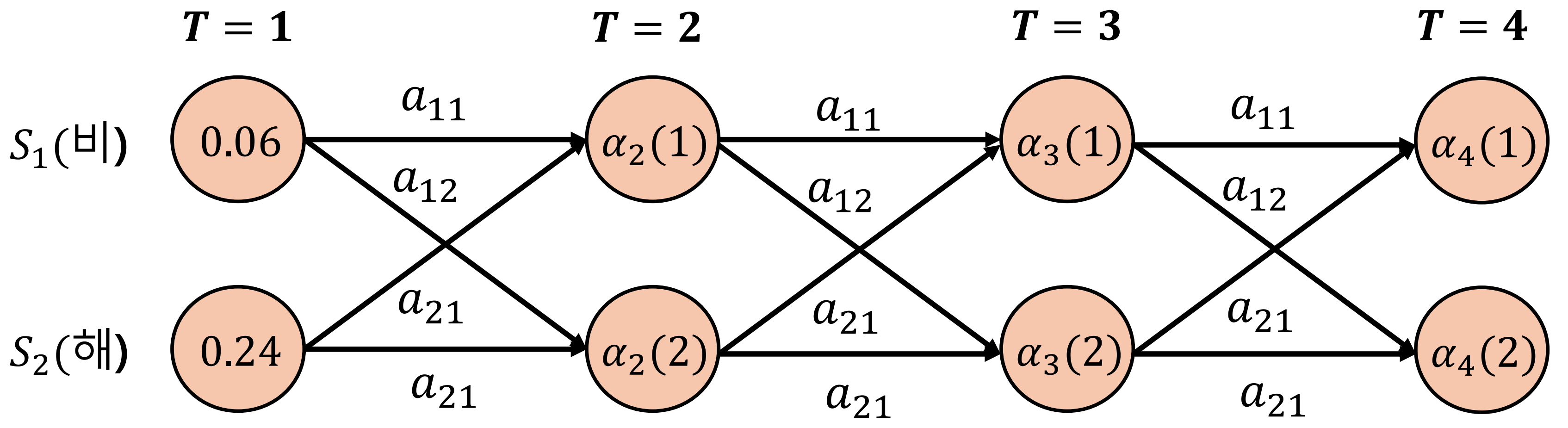
Observations $o_1 =$ 산책 $o_2 =$ 산책 $o_3 =$ 연구 $o_4 =$ 쇼핑

$$\alpha_1(1) = \pi_1 \cdot b_1(\text{산책}) = 0.6 \cdot 0.1 = 0.06$$
$$\alpha_1(2) = \pi_2 \cdot b_2(\text{산책}) = 0.4 \cdot 0.6 = 0.24$$

Hidden Markov Models - Evaluation



- Evaluation problem
  - Problem : Given HMM (λ^*) and **O**, find the probability of **O**
  - Solution : Forward algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑할 확률은?



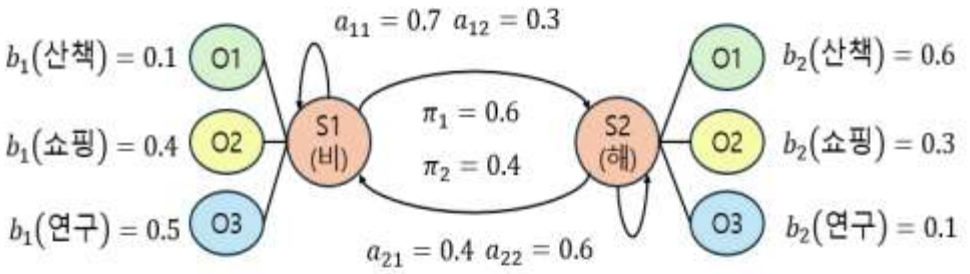
states

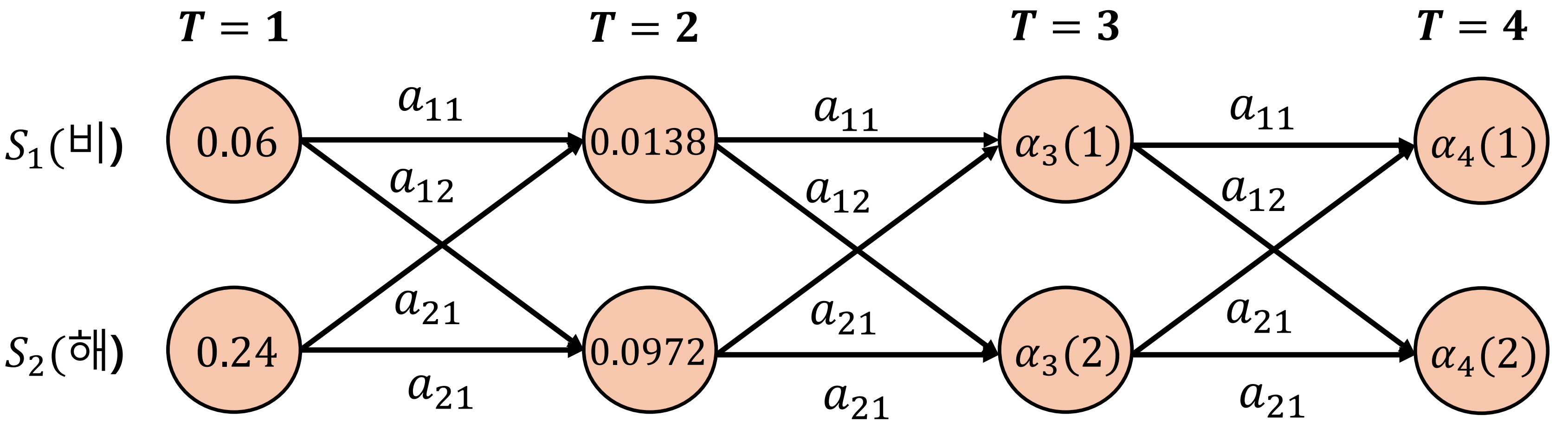
Observations $o_1 =$ 산책 $o_2 =$ 산책 $o_3 =$ 연구 $o_4 =$ 쇼핑

$$\alpha_2(1) = (\alpha_1(1) \cdot a_{11} + \alpha_1(2) \cdot a_{21}) \cdot b_1\left(\text{산책}\right) = (0.06 \cdot 0.7 + 0.24 \cdot 0.4) \cdot 0.1 = 0.0138$$
$$\alpha_2(2) = (\alpha_1(1) \cdot a_{12} + \alpha_1(2) \cdot a_{22}) \cdot b_2\left(\text{산책}\right) = (0.06 \cdot 0.3 + 0.24 \cdot 0.6) \cdot 0.6 = 0.0972$$

Hidden Markov Models - Evaluation

- Evaluation problem
  - Problem : Given HMM (λ^*) and **O**, find the probability of **O**
  - Solution : Forward algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑할 확률은?



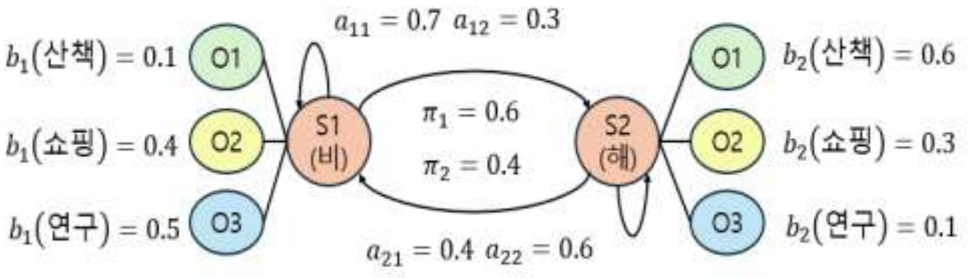


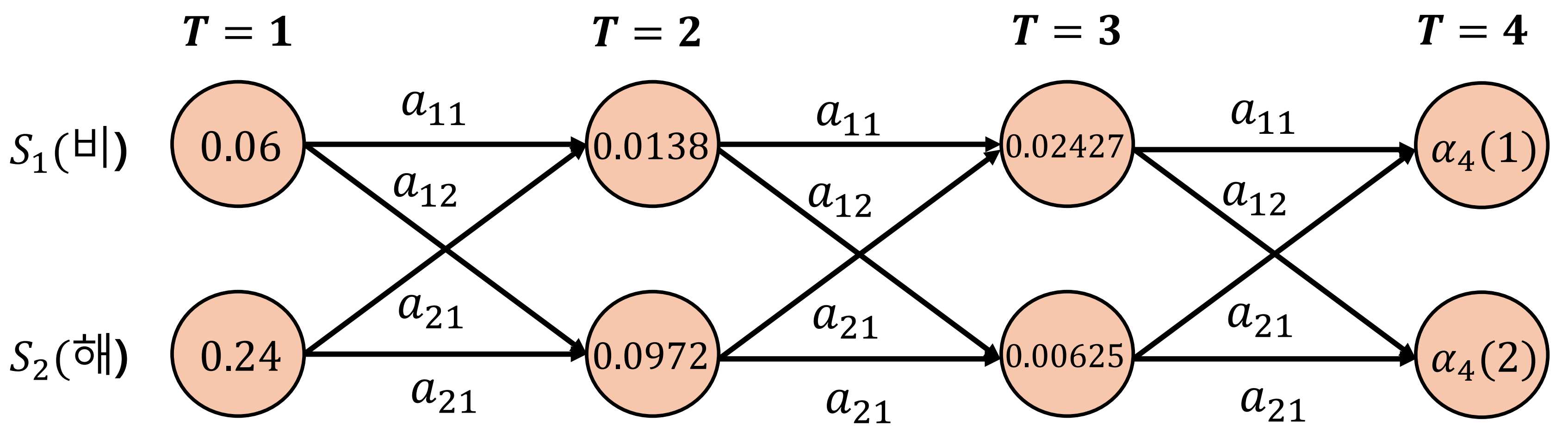
Observations o_1 = 산책 o_2 = 산책 o_3 = 연구 o_4 = 쇼핑

$$\alpha_3(1) = (\alpha_2(1) \cdot a_{11} + \alpha_2(2) \cdot a_{21}) \cdot b_1(\text{연구}) = (0.0138 \cdot 0.7 + 0.0972 \cdot 0.4) \cdot 0.5 = 0.02427$$
$$\alpha_3(2) = (\alpha_2(1) \cdot a_{12} + \alpha_2(2) \cdot a_{22}) \cdot b_2(\text{연구}) = (0.0138 \cdot 0.3 + 0.0972 \cdot 0.6) \cdot 0.1 = 0.00625$$

Hidden Markov Models - Evaluation

- Evaluation problem
  - Problem : Given HMM (λ^*) and **O**, find the probability of **O**
  - Solution : Forward algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑할 확률은?



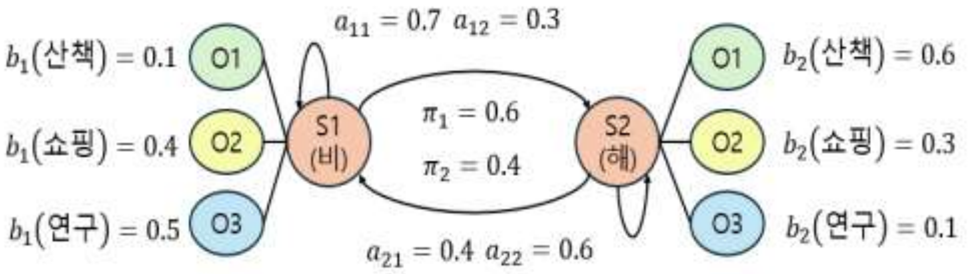


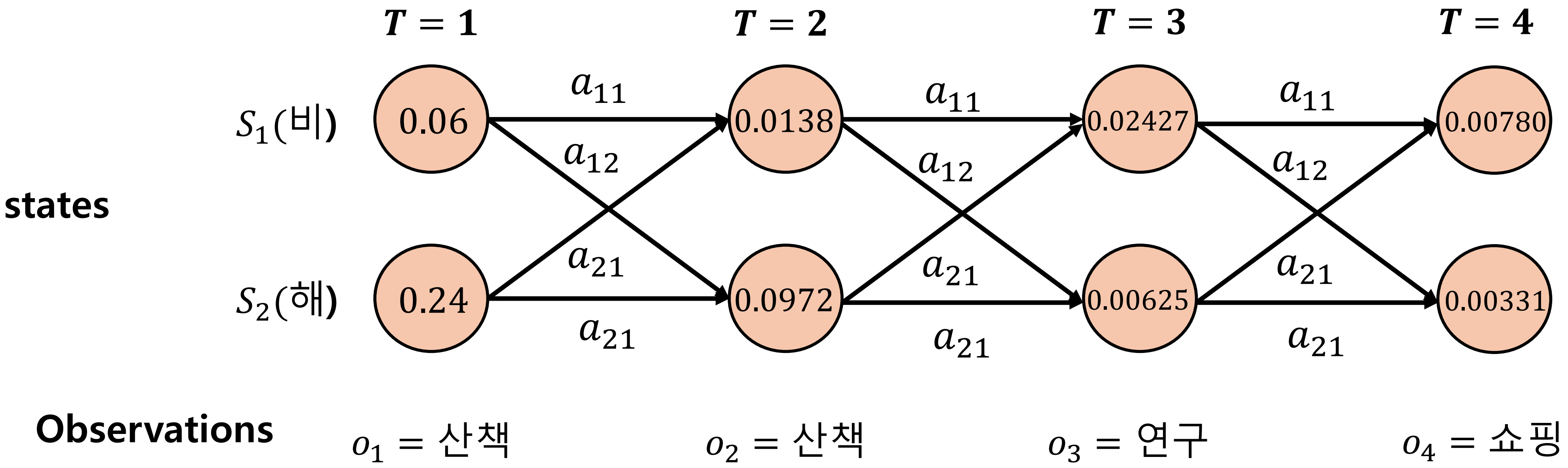
Observations o_1 = 산책 o_2 = 산책 o_3 = 연구 o_4 = 쇼핑

$$\alpha_4(1) = (\alpha_3(1) \cdot a_{11} + \alpha_3(2) \cdot a_{21}) \cdot b_1\left(\text{쇼핑}\right) = (0.002427 \cdot 0.7 + 0.00625 \cdot 0.4) \cdot 0.4 = 0.00780$$
$$\alpha_4(2) = (\alpha_3(1) \cdot a_{12} + \alpha_3(2) \cdot a_{22}) \cdot b_2\left(\text{쇼핑}\right) = (0.002427 \cdot 0.3 + 0.00625 \cdot 0.6) \cdot 0.4 = 0.00331$$

Hidden Markov Models - Evaluation

- Evaluation problem
  - Problem : Given HMM (λ^*) and **O**, find the probability of **O**
  - Solution : Forward algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑할 확률은?

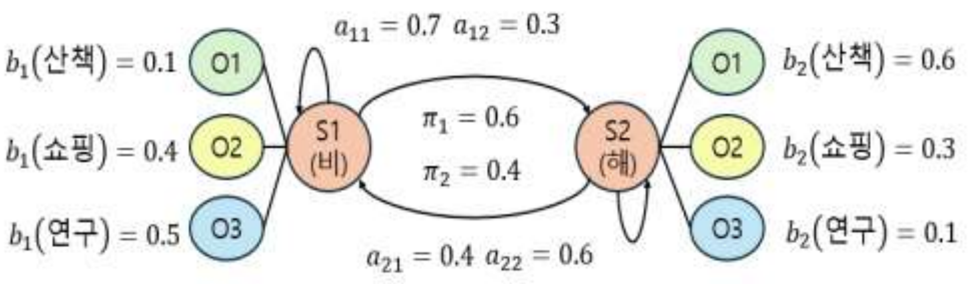


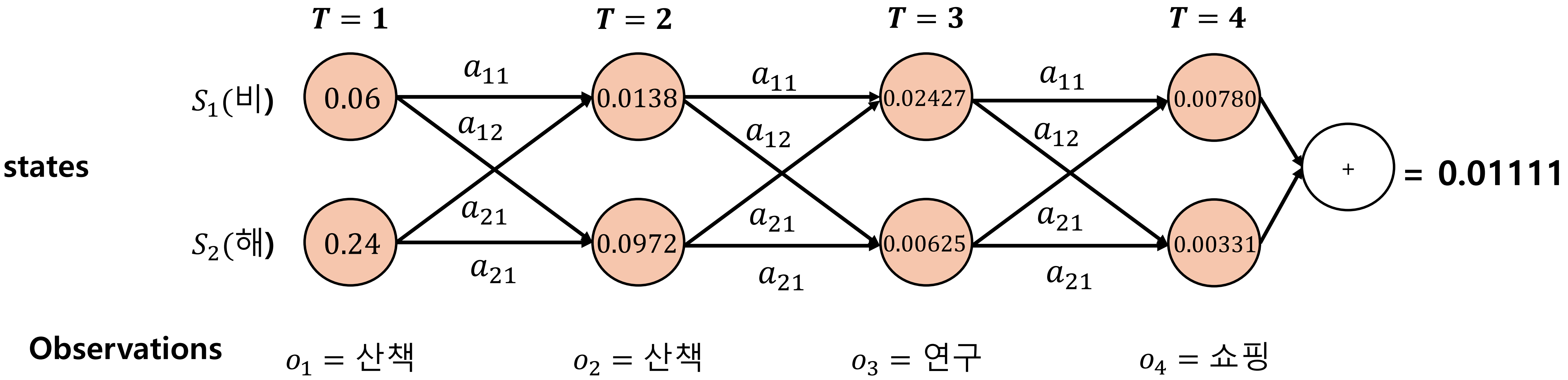


$$\alpha_4(1) = (\alpha_3(1) \cdot a_{11} + \alpha_3(2) \cdot a_{21}) \cdot b_1\,(쇼핑) = (0.002427 \cdot 0.7 + 0.00625 \cdot 0.4) \cdot 0.4 = 0.00780$$
$$\alpha_4(2) = (\alpha_3(1) \cdot a_{12} + \alpha_3(2) \cdot a_{22}) \cdot b_2\,(쇼핑) = (0.002427 \cdot 0.3 + 0.00625 \cdot 0.6) \cdot 0.4 = 0.00331$$

Hidden Markov Models - Evaluation

- Evaluation problem
  - Problem : Given HMM (λ^*) and **O**, find the probability of **O**
  - Solution : Forward algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑할 확률은?





$$\alpha_4(1) = (\alpha_3(1) \cdot a_{11} + \alpha_3(2) \cdot a_{21}) \cdot b_1\left(\text{쇼핑}\right) = (0.002427 \cdot 0.7 + 0.00625 \cdot 0.4) \cdot 0.4 = 0.00780$$
$$\alpha_4(2) = (\alpha_3(1) \cdot a_{12} + \alpha_3(2) \cdot a_{22}) \cdot b_2\left(\text{쇼핑}\right) = (0.002427 \cdot 0.3 + 0.00625 \cdot 0.6) \cdot 0.4 = 0.00331$$

Hidden Markov Models - Evaluation

- Forward Algorithm for Evaluation Problem
  - Problem : Given HMM (λ^*) and **O**, find the probability of **O**
  - Solution : Forward algorithm

Forward probability (전방 확률, $\alpha_t(i)$)

Forward probability = $p(O|\lambda) = \sum_{j=1}^{n} \alpha_T(j)$

$$\alpha_1(i) = \pi_i b_i(o_1), 1 \leq i \leq n$$

$$\alpha_t(i) = \left[\sum_{j=1}^{n} \alpha_{t-1}(i) a_{ji}\right] \cdot b_i(o_t), 2 \leq i \leq T, 1 \leq i \leq n$$

-> Forward probability 는 주어진 Sequence O가 HMM에 속할 확률

-> HMM 1 (정박사), HMM 2 (조박사)가 있을 때, 어느 HMM에 속할 확률이 높을지?

-> Sequence classification 문제에 활용 가능!

Hidden Markov Models - Evaluation

Forward probability

앞으로(시간 순으로) 확률 계산 (Forward algorithm)

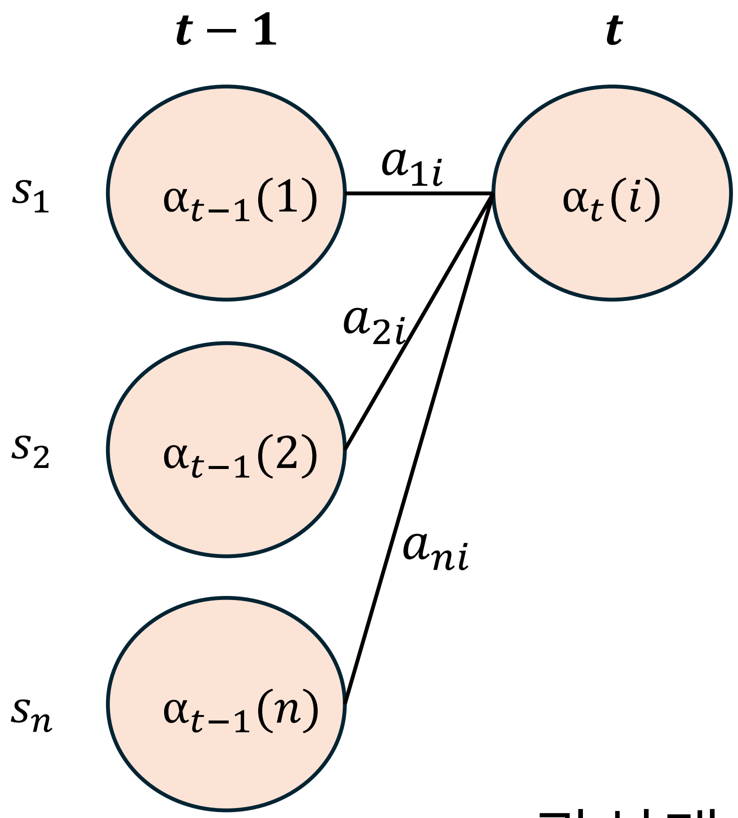
↕

Backward probability

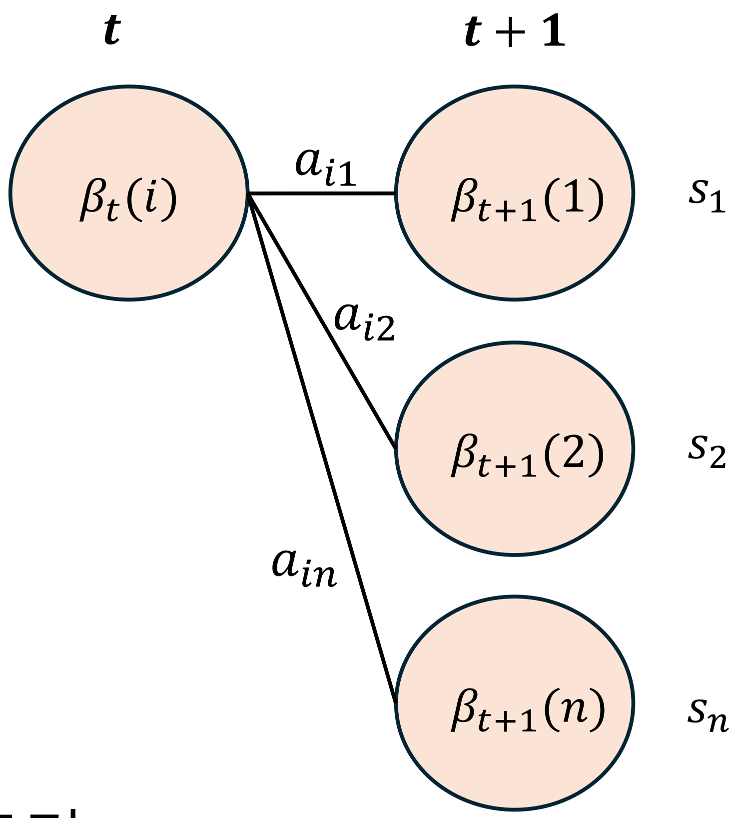
뒤로(시간의 역순으로) 확률 계산 (Backward algorithm)

Hidden Markov Models - Evaluation

Forward probability



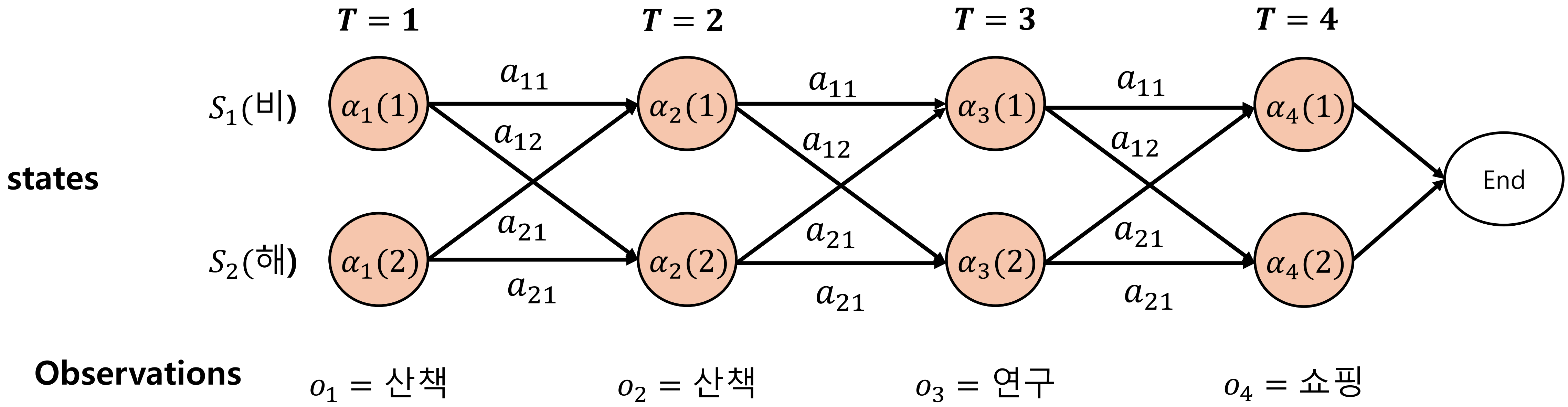
Backward probability



관심대상 (t시점) state를 i로 표기

Hidden Markov Models - Evaluation

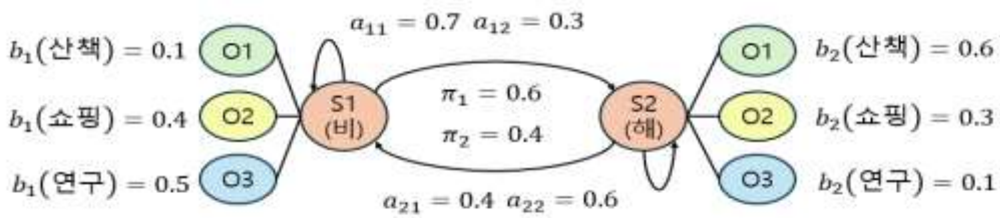
- An example of **Backward** algorithm
  - Problem : Given HMM (λ^*) and **O**, find the **backward** probability of **O**
  - Solution : **Backward** algorithm

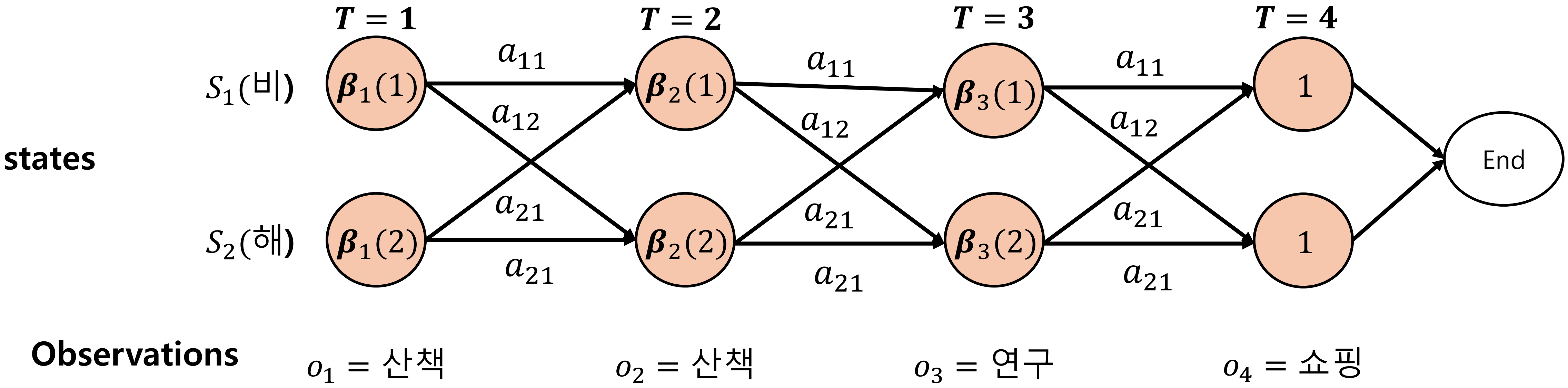


α -> β $\quad$ $\beta_{t=T}(i) = 1$

Hidden Markov Models - Evaluation

- An example of **Backward** algorithm
  - Problem : Given HMM (λ^*) and **O**, find the **backward** probability of **O**
  - Solution : **Backward** algorithm

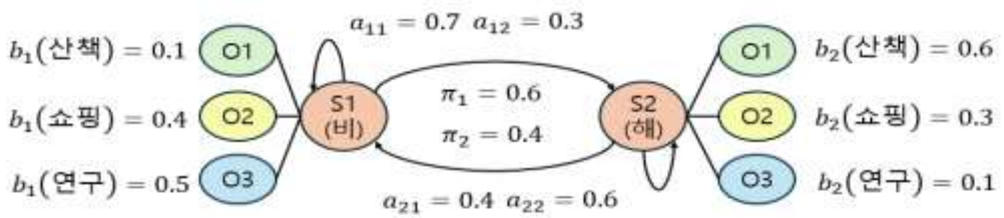


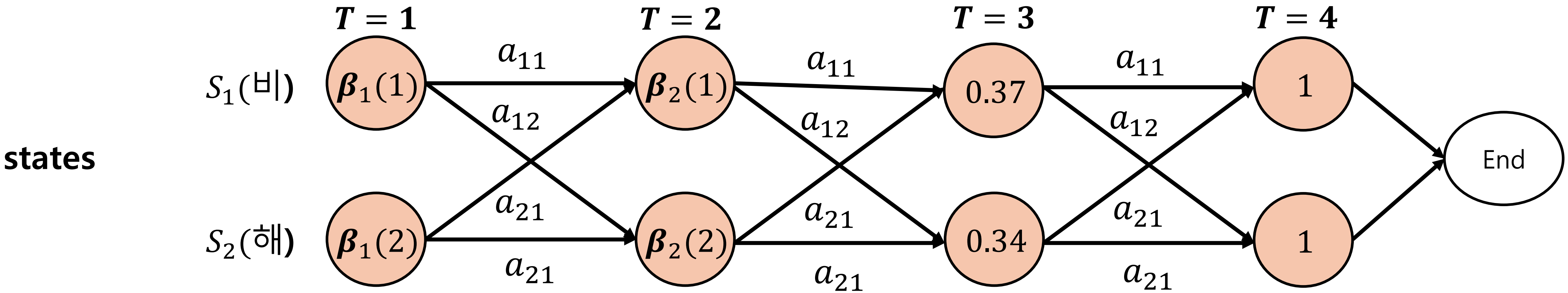


$$\beta_3(1) = (\beta_4(1)a_{11} \cdot b_1(\text{쇼핑}) + \beta_4(2) \cdot a_{12} \cdot b_2(\text{쇼핑})) = (1 \cdot 0.7 \cdot 0.4 + 1 \cdot 0.3 \cdot 0.3) = 0.37$$
$$\beta_3(2) = (\beta_4(1)a_{21} \cdot b_1(\text{쇼핑}) + \beta_4(2) \cdot a_{22} \cdot b_2(\text{쇼핑})) = (1 \cdot 0.4 \cdot 0.4 + 1 \cdot 0.6 \cdot 0.3) = 0.34$$

Hidden Markov Models - Evaluation

- An example of **Backward** algorithm
  - Problem : Given HMM (λ^*) and **O**, find the **backward** probability of **O**
  - Solution : **Backward** algorithm



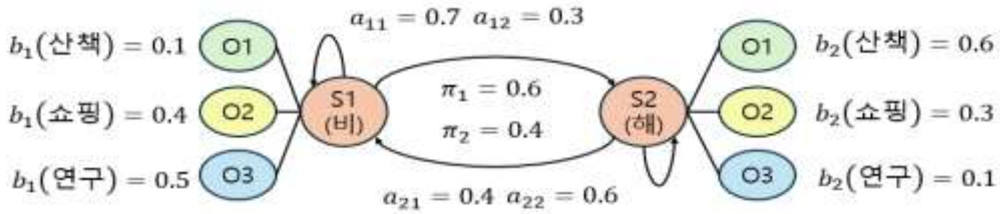


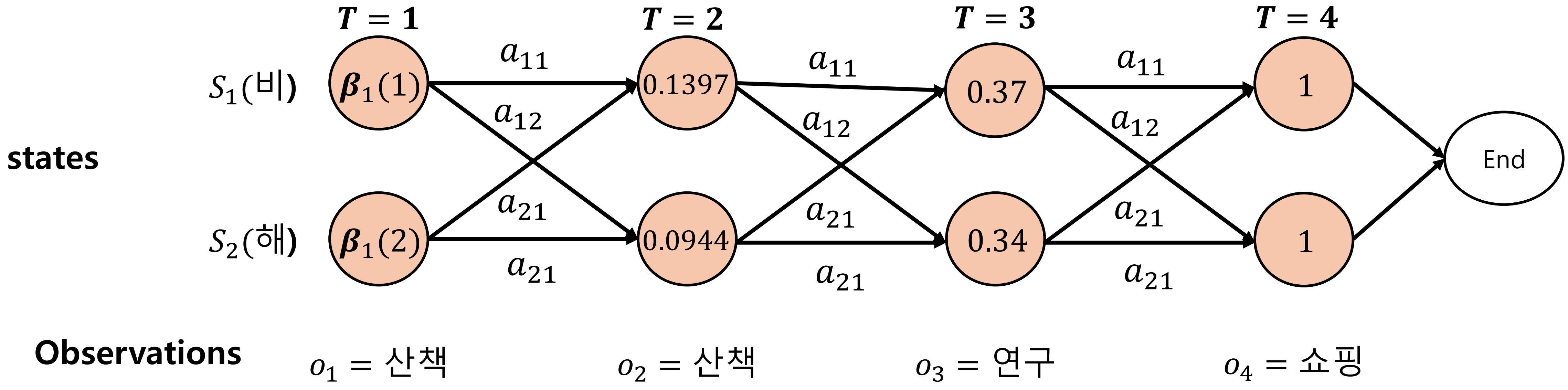
Observations $o_1 =$ 산책 $o_2 =$ 산책 $o_3 =$ 연구 $o_4 =$ 쇼핑

$$\beta_2(1) = (\beta_3(1)a_{11} \cdot b_1(연구) + \beta_3(2) \cdot a_{12} \cdot b_2(연구)) = (0.37 \cdot 0.7 \cdot 0.5 + 0.34 \cdot 0.3 \cdot 0.1) = 0.1397$$
$$\beta_2(2) = (\beta_3(1)a_{21} \cdot b_1(연구) + \beta_3(2) \cdot a_{22} \cdot b_2(연구)) = (0.37 \cdot 0.4 \cdot 0.5 + 0.34 \cdot 0.6 \cdot 0.1) = 0.0944$$

Hidden Markov Models - Evaluation

- An example of **Backward** algorithm
  - Problem : Given HMM (λ^*) and **O**, find the **backward** probability of **O**
  - Solution : **Backward** algorithm

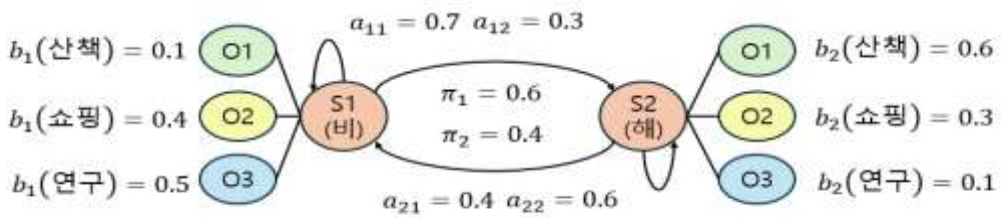


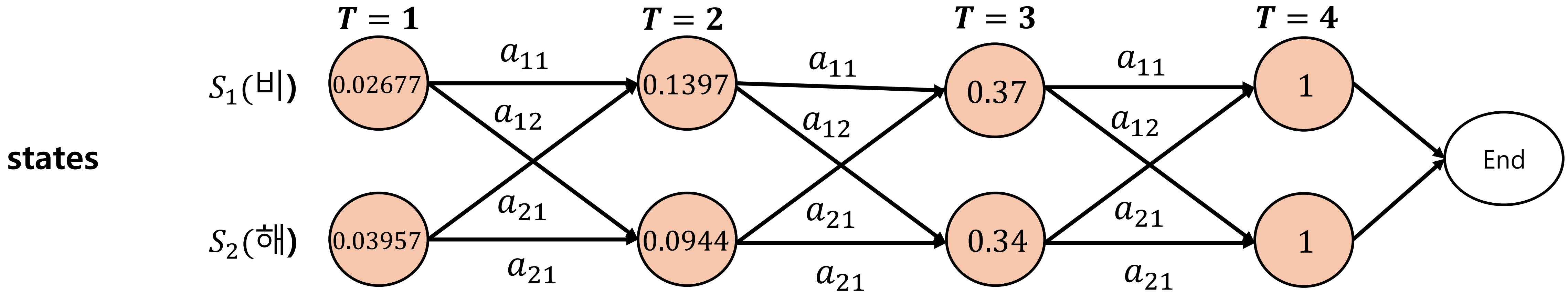


$$\beta_1(1) = (\beta_2(1)a_{11} \cdot b_1(산책) + \beta_2(2) \cdot a_{12} \cdot b_2(산책)) = (0.1397 \cdot 0.7 \cdot 0.1 + 0.0944 \cdot 0.3 \cdot 0.6) = 0.02677$$
$$\beta_1(2) = (\beta_2(1)a_{21} \cdot b_1(산책) + \beta_2(2) \cdot a_{22} \cdot b_2(산책)) = (0.1397 \cdot 0.4 \cdot 0.1 + 0.0944 \cdot 0.6 \cdot 0.6) = 0.03957$$

Hidden Markov Models - Evaluation

- An example of **Backward** algorithm
  - Problem : Given HMM (λ^*) and **O**, find the **backward** probability of **O**
  - Solution : **Backward** algorithm





Observations $o_1 =$ 산책 $o_2 =$ 산책 $o_3 =$ 연구 $o_4 =$ 쇼핑

$\boldsymbol{Probability(O)} = (o_1 = 산책, o_2 = 산책, o_3 = 연구, o_4 = 쇼핑)$
$= (\beta_1(1) \cdot \pi_1 \cdot b_1(산책) + \beta_1(2) \cdot \pi_2 \cdot b_3(산책)) = (0.02677 \cdot 0.6 \cdot 0.1 + 0.03957 \cdot 0.4 \cdot 0.6) = 0.01111$

Hidden Markov Models - Evaluation

- Forward probability로부터 계산

$\boldsymbol{Probability(O)} = (o_1$ = 산책, o_2 = 산책, o_3 = 연구, o_4 = 쇼핑$) = 0.0780 + 0{,}00331 = 0.01111$

- Forward probability로부터 계산

$\boldsymbol{Probability(O)} = (o_1$ = 산책, o_2 = 산책, o_3 = 연구, o_4 = 쇼핑$)$
$= (\beta_1(1) \cdot \pi_1 \cdot b_1($산책$) + \beta_1(2) \cdot \pi_2 \cdot b_3($산책$)) = (0.02677 \cdot 0.6 \cdot 0.1 + 0.03957 \cdot 0.4 \cdot 0.6) = 0.01111$

Probability_Forward (O) = Probability_Backward (O)

Hidden Markov Models - Evaluation

Forward probability

$$\alpha_t(i) = \left[\sum_{j=1}^{n} \alpha_{t-1}(i) a_{ji}\right] \cdot b_i(o_t), 2 \le i \le T, 1 \le i \le n$$

$\alpha_{t=1}(i) = \pi_i b_i(o_{t=1}) \cdots$ 초기시점

Backward probability

$$\beta_t(i) = \left[\sum_{j=1}^{n} \beta_{t+1}(i) \cdot a_{ij} \cdot b_j(o_{t+1})\right], 1 \le i \le T-1, 1 \le i \le n$$

$\beta_{t=T}(i) = 1 \cdots$ 마지막시점

Hidden Markov Models - Evaluation

Forward probability

$$\alpha_t(i) = \left[\sum_{j=1}^{n} \alpha_{t-1}(i) a_{ji}\right] \cdot b_i(o_t), 2 \le i \le T, 1 \le i \le n$$

- **Backward probability -> HMM Training에 활용 (Learning)**

$\alpha_{t=1}(i) = \pi_i b_i(o_{t=1})$ ⋯ 첫시점

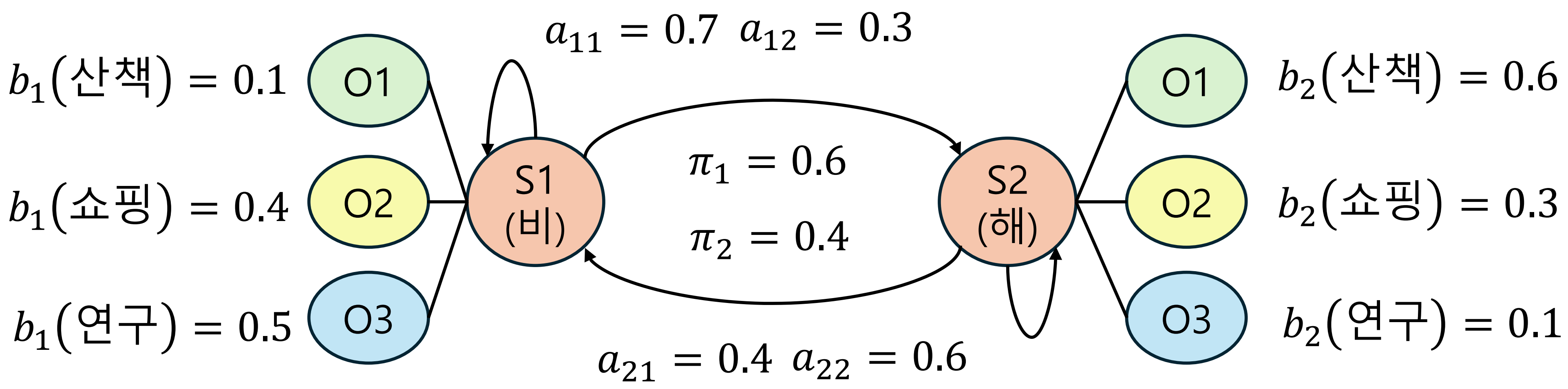
Backward probability

$$\beta_t(i) = \left[\sum_{j=1}^{n} \beta_{t+1}(i) \cdot a_{ij} \cdot b_j(o_{t+1})\right], 1 \le i \le T-1, 1 \le i \le n$$

$\beta_{t=T}(i) = 1$ ⋯ 마지막시점

Hidden Markov Models – Decoding

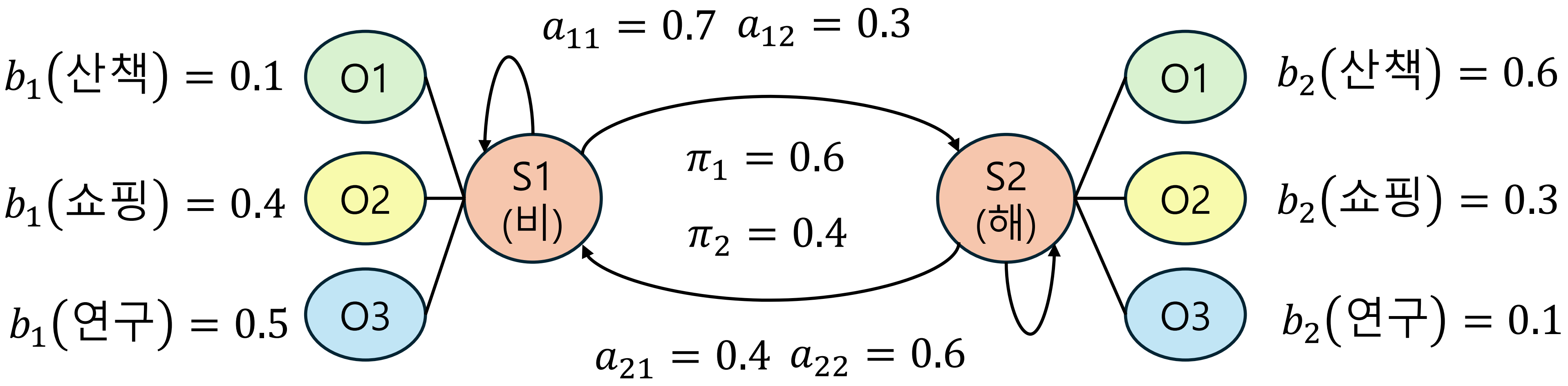
- Decoding problem (디코딩) -> HMM의 핵심
  - Problem : Given HMM (λ^*) and **O**, find the optimal **S** (가장 그럴싸한 은닉상태의 시퀀스 결정)
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?



$\boldsymbol{O} = (o_1 = \text{산책}, o_2 = \text{산책}, o_3 = \text{연구}, o_4 = \text{쇼핑})$

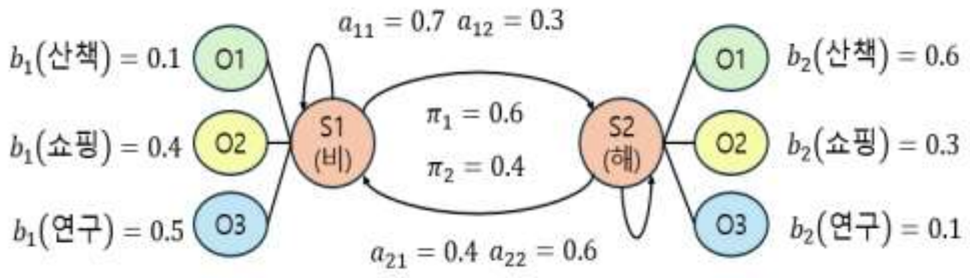
Hidden Markov Models – Decoding

- Decoding problem (디코딩) -> HMM의 핵심
    - Problem : Given HMM (λ^*) and **O**, find the optimal **S** (가장 그럴싸한 은닉상태의 시퀀스 결정)
    - Solution : viterbi algorithm
    - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?



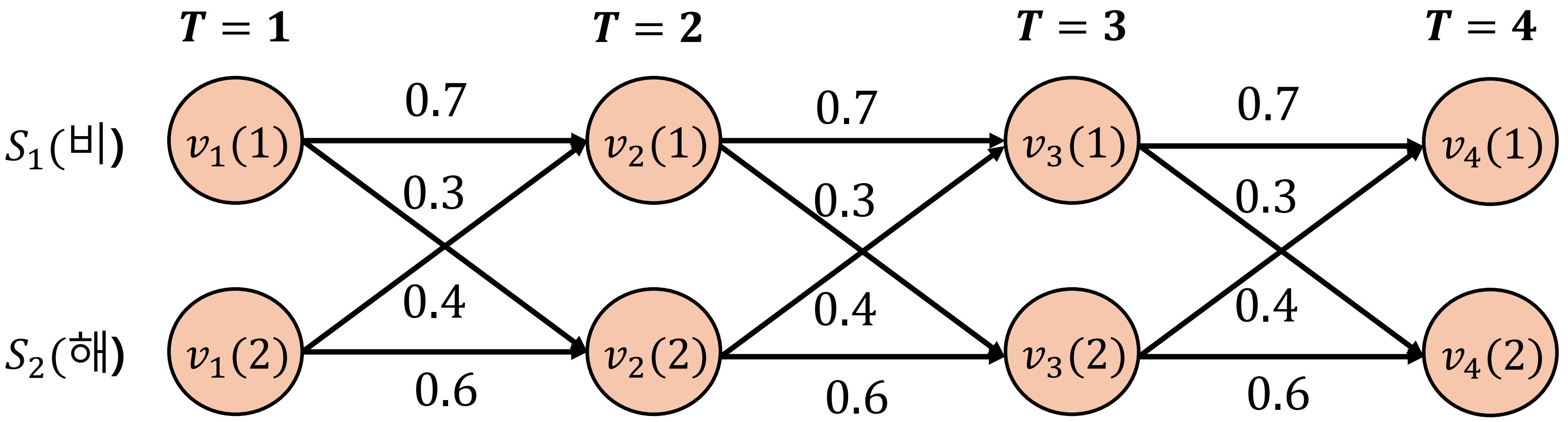
$\boldsymbol{O} = (o_1 = 산책, o_2 = 산책, o_3 = 연구, o_4 = 쇼핑)$

Hidden Markov Models – Decoding



- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?

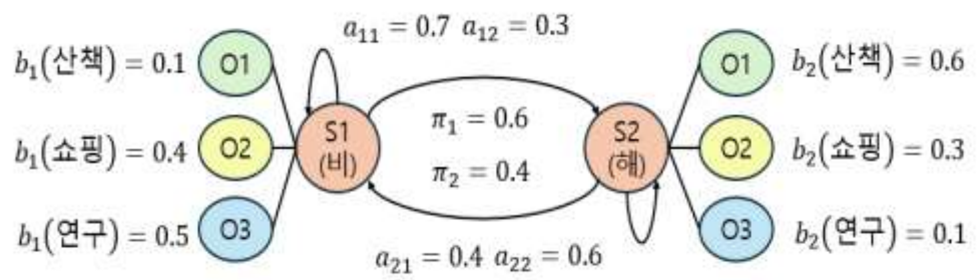
states



Observations $o_1 =$ 산책 $o_2 =$ 산책 $o_3 =$ 연구 $o_4 =$ 쇼핑

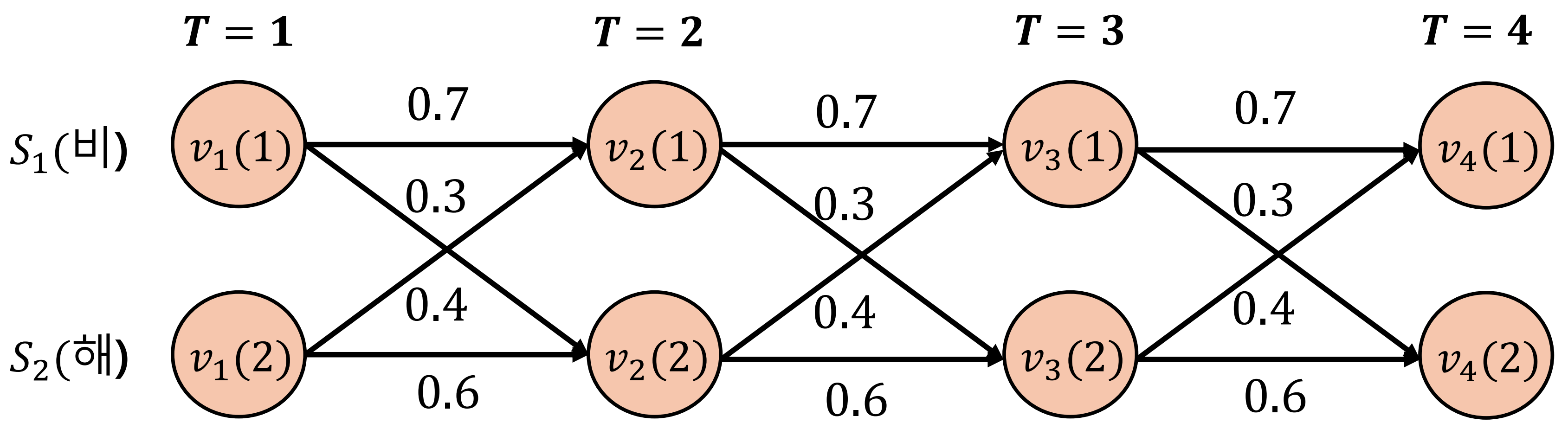
$$v_1(1) = \pi_1 \cdot b_1(\text{산책}) = 0.6 \cdot 0.1 = 0.06$$
$$v_1(2) = \pi_2 \cdot b_2(\text{산책}) = 0.4 \cdot 0.6 = 0.24$$

Hidden Markov Models – Decoding



- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?

states



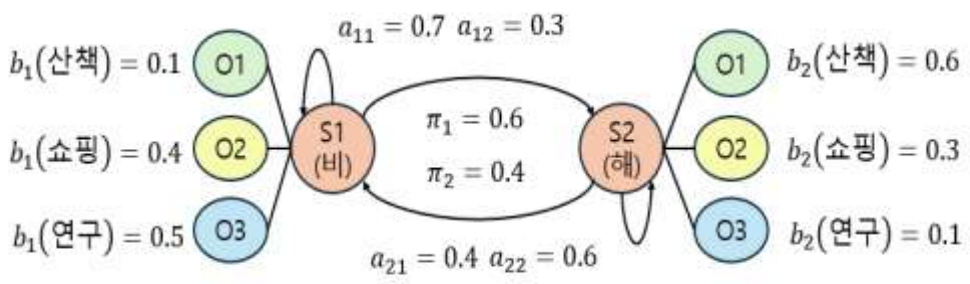
Observations o_1 = 산책 o_2 = 산책 o_3 = 연구 o_4 = 쇼핑

$v_1(1) = \pi_1 \cdot b_1(\text{산책}) = 0.6 \cdot 0.1 = 0.06$

$v_1(2) = \pi_2 \cdot b_2(\text{산책}) = 0.4 \cdot 0.6 = 0.24$

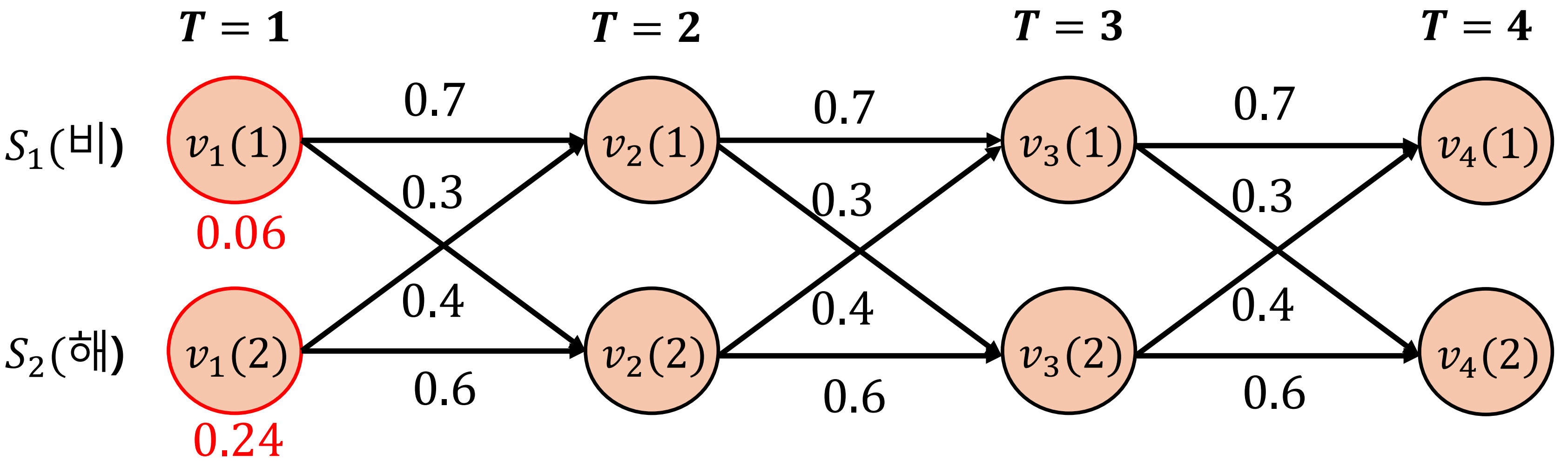
Vierbi 확률 $v_t(i)$: t번째 시점의 i 은닉상태의 확률

Hidden Markov Models – Decoding



- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?

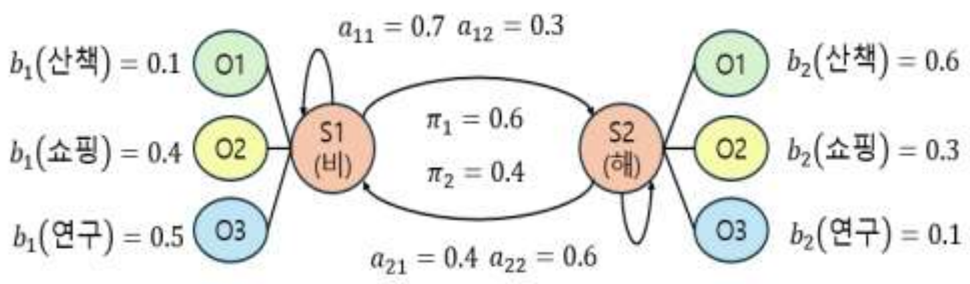
states



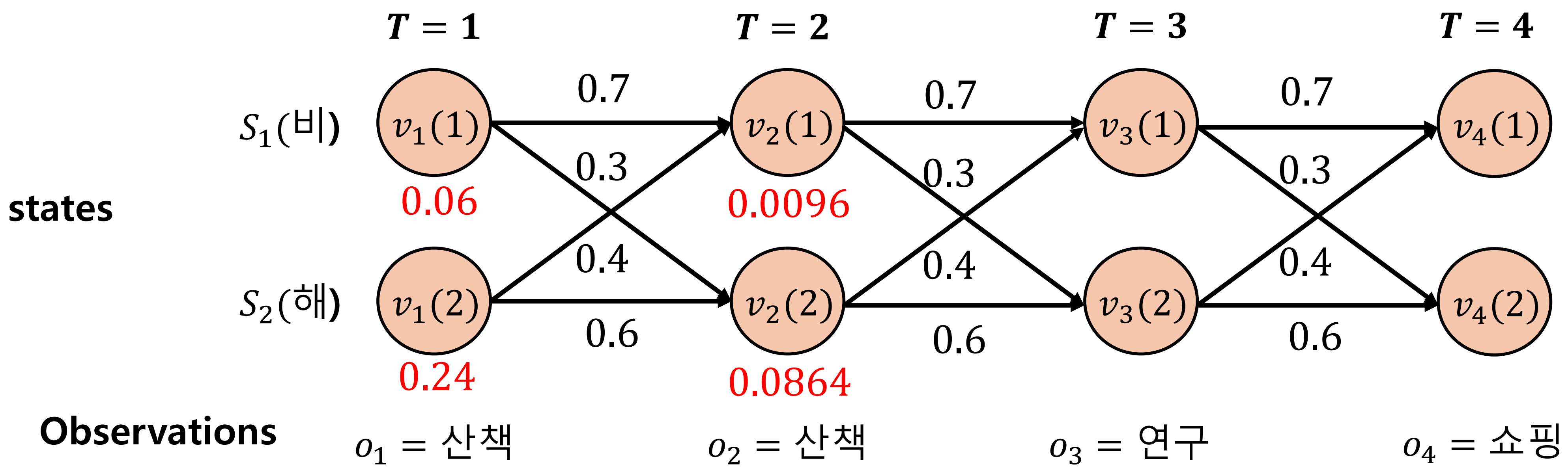
Observations $o_1 =$ 산책 $o_2 =$ 산책 $o_3 =$ 연구 $o_4 =$ 쇼핑

$$v_1(1) = \pi_1 \cdot b_1(산책) = 0.6 \cdot 0.1 = 0.06$$
$$v_1(2) = \pi_2 \cdot b_2(산책) = 0.4 \cdot 0.6 = 0.24$$

Hidden Markov Models – Decoding

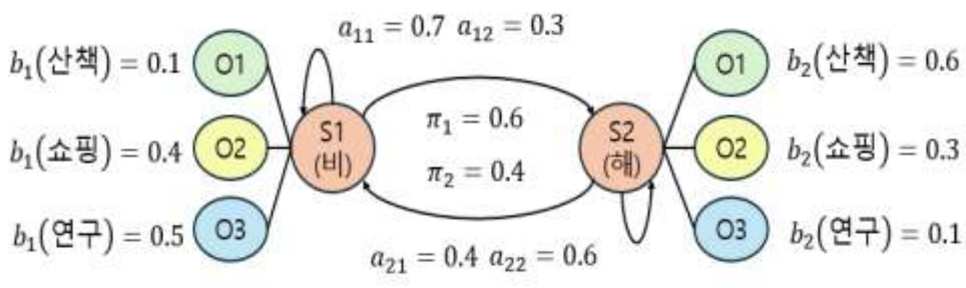


- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?

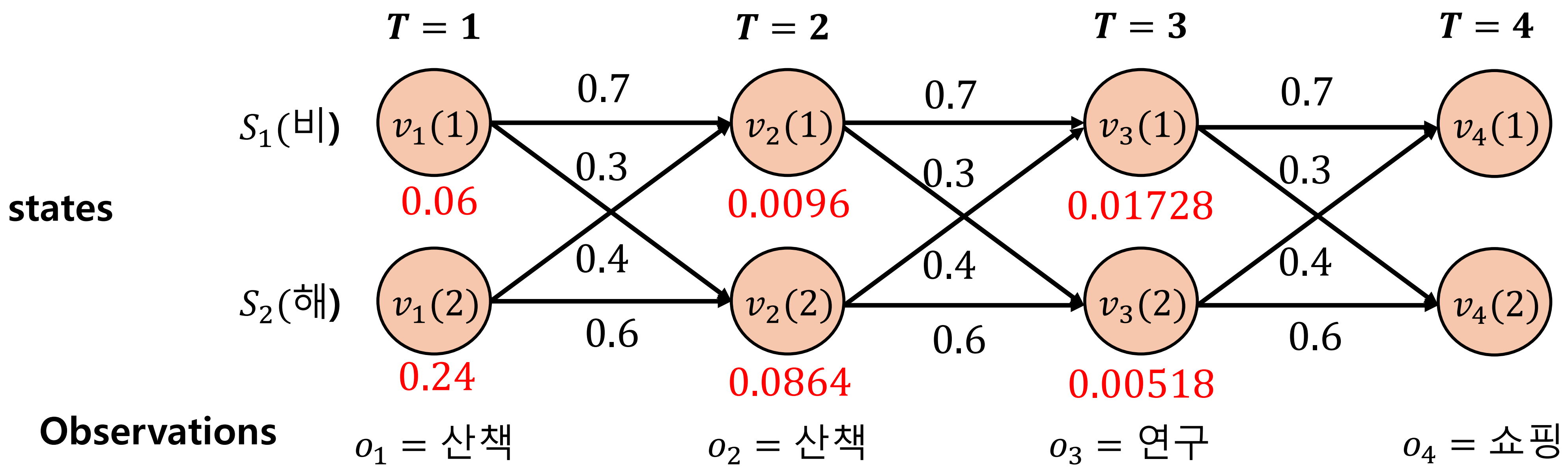


$$v_2(1) = \max(v_1(1) \cdot a_{11}, v_1(2) \cdot a_{21}) \cdot b_1(\text{산책}) = \max(0.042, 0.096) \cdot 0.1 = 0.0096, \tau_2(1) = 2$$
$$v_2(2) = \max(v_1(1) \cdot a_{12}, v_1(2) \cdot a_{22}) \cdot b_2(\text{산책}) = \max(0.018, 0.144) \cdot 0.6 = 0.0864, \tau_2(2) = 2$$

Hidden Markov Models – Decoding



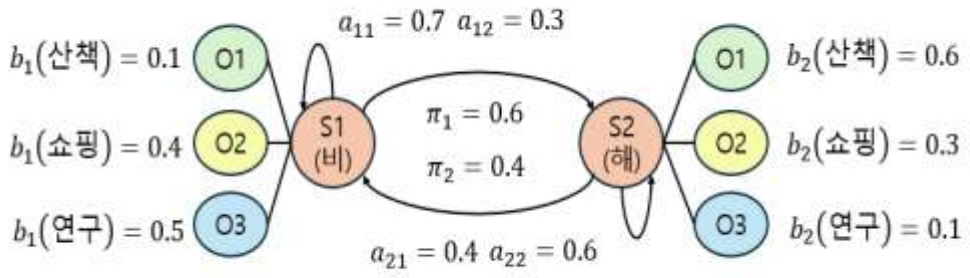
- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?



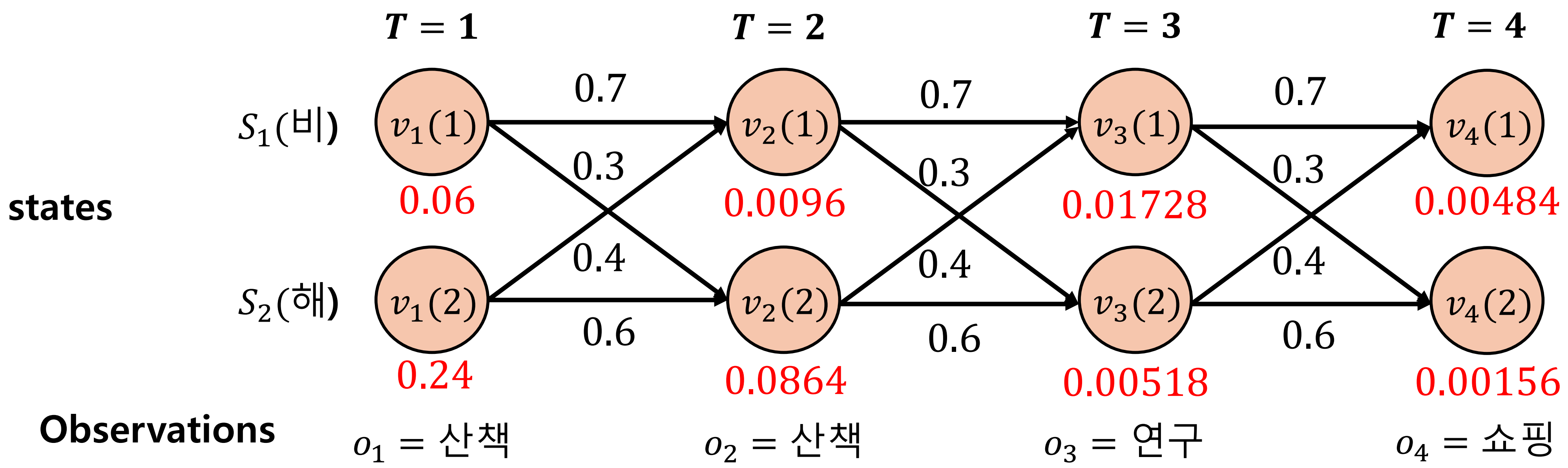
$$v_3(1) = \max(v_2(1) \cdot a_{11}, v_2(2) \cdot a_{21}) \cdot b_1(\text{연구}) = \max(0.00672, 0.03456) \cdot 0.5 = 0.01728, \tau_3(1) = 2$$

$$v_3(2) = \max(v_2(1) \cdot a_{12}, v_2(2) \cdot a_{22}) \cdot b_2(\text{연구}) = \max(0.00288, 0.05184) \cdot 0.1 = 0.00518, \tau_3(2) = 2$$

Hidden Markov Models – Decoding

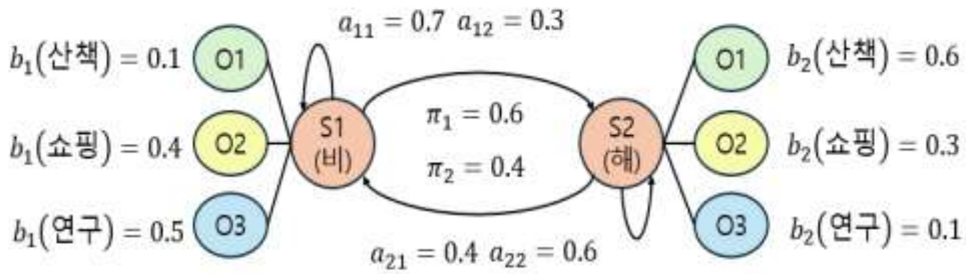


- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?

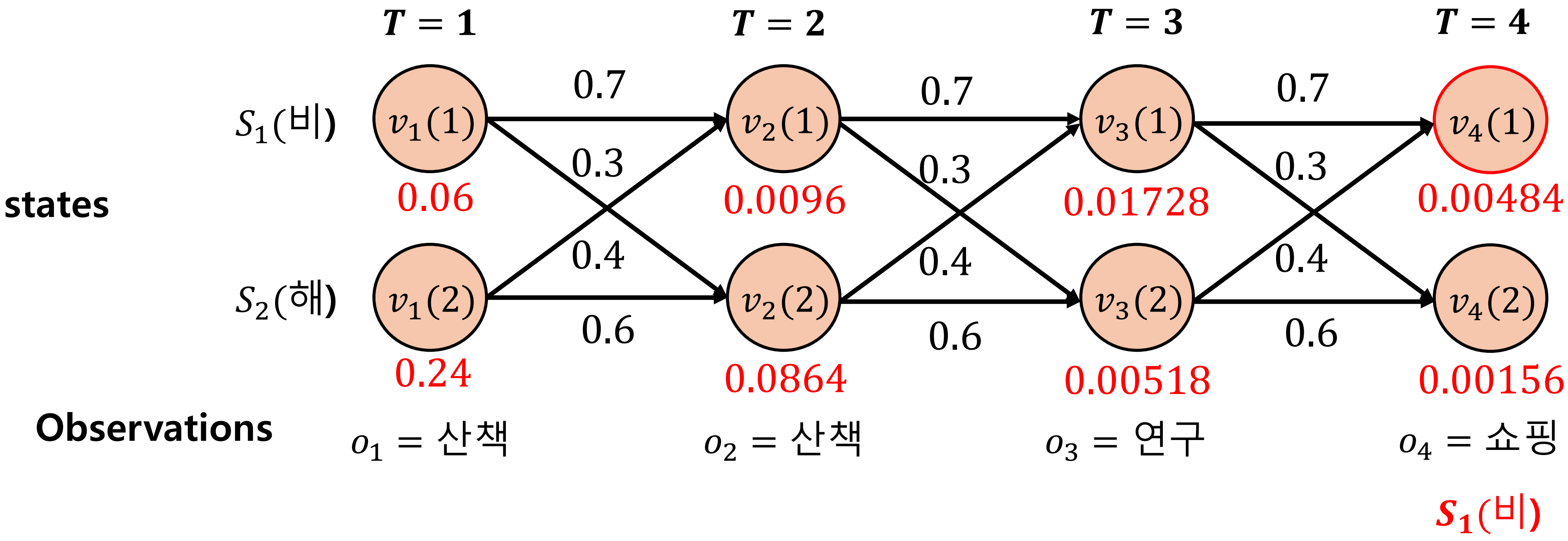


$$v_4(1) = \max(v_3(1) \cdot a_{11}, v_3(2) \cdot a_{21}) \cdot b_1(\text{쇼핑}) = \max(0.012096, 0.00207) \cdot 0.4 = 0.00484, \tau_4(1) = 1$$
$$v_4(2) = \max(v_3(1) \cdot a_{12}, v_3(2) \cdot a_{22}) \cdot b_2(\text{쇼핑}) = \max(0.005184, 0.003108) \cdot 0.3 = 0.00156, \tau_4(2) = 1$$

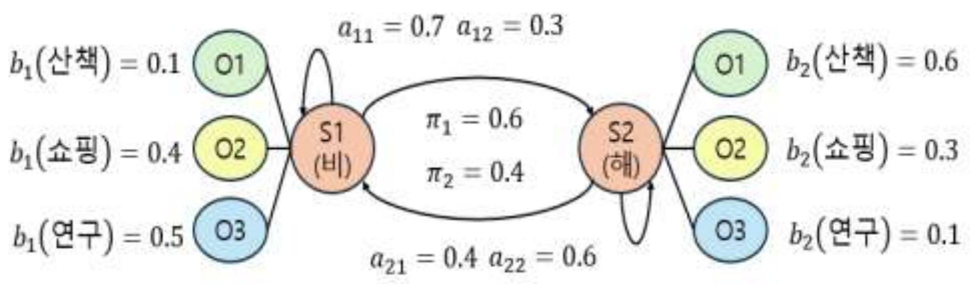
Hidden Markov Models – Decoding



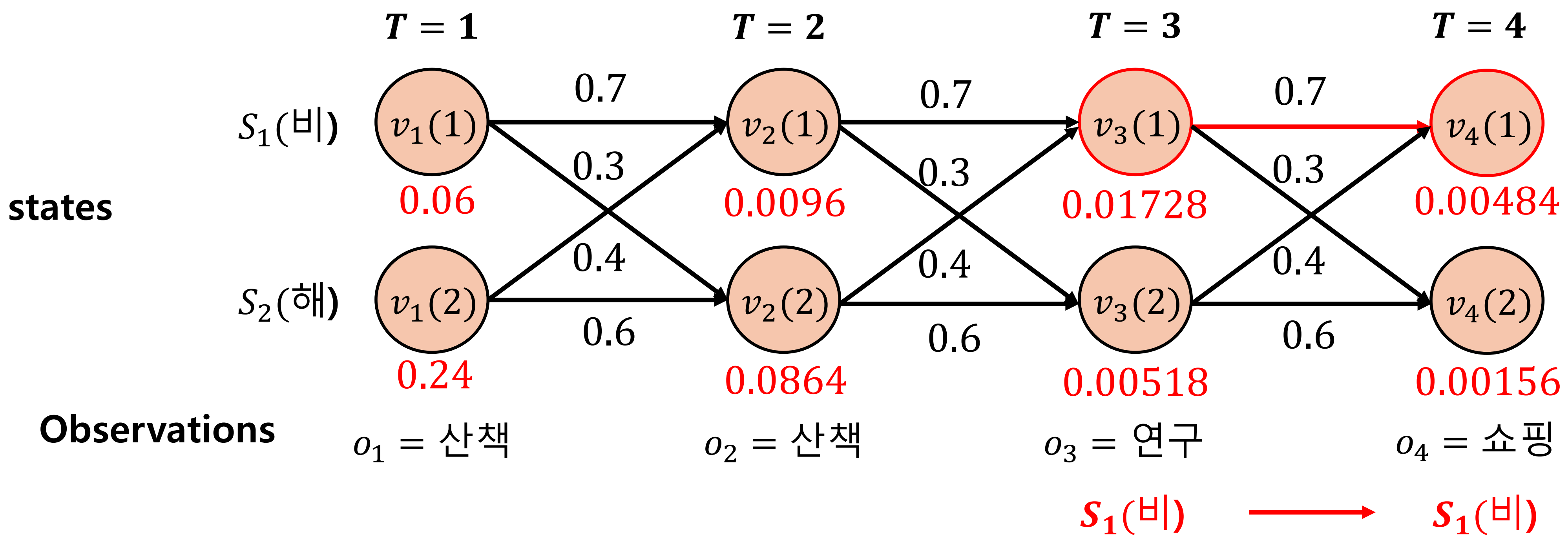
- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?



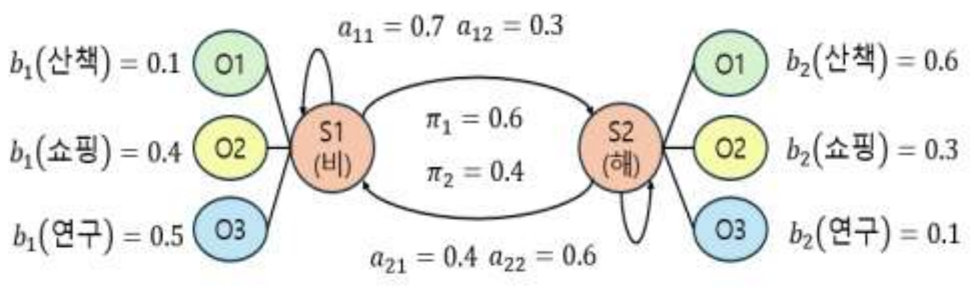
Hidden Markov Models – Decoding



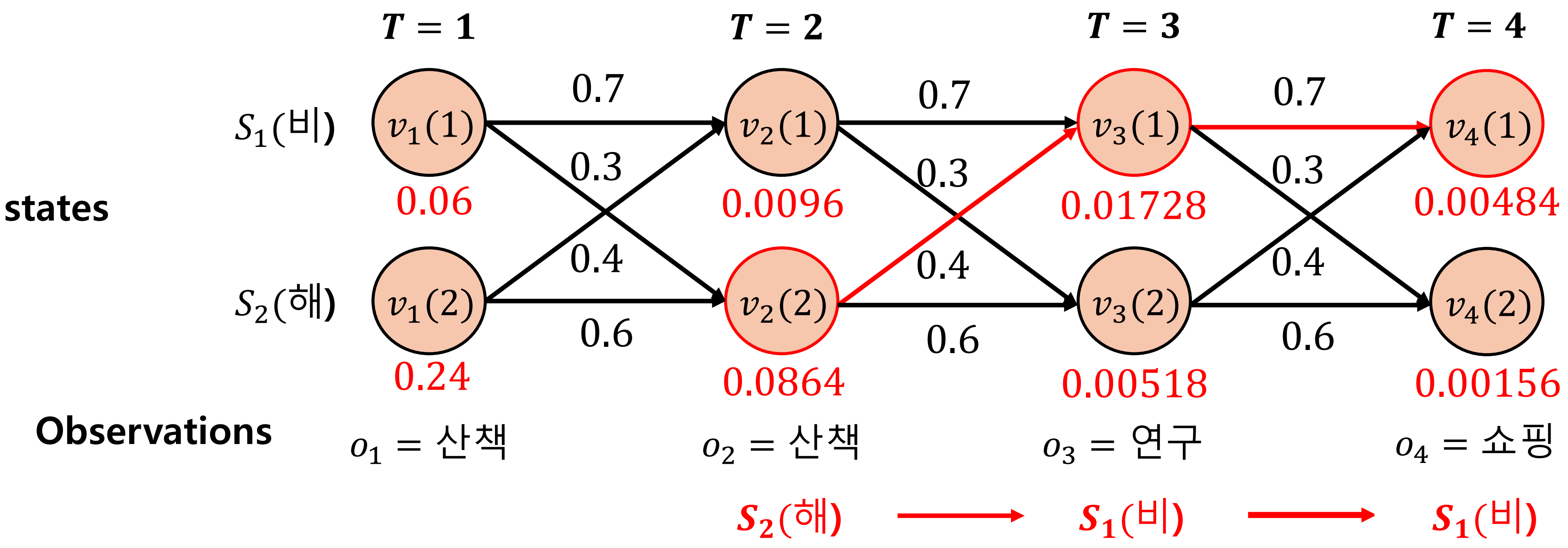
- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?



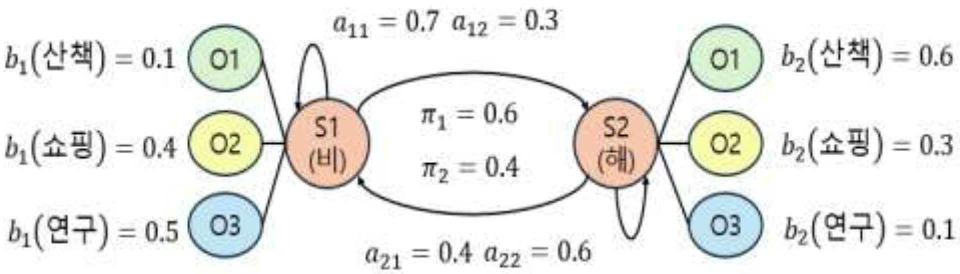
Hidden Markov Models – Decoding



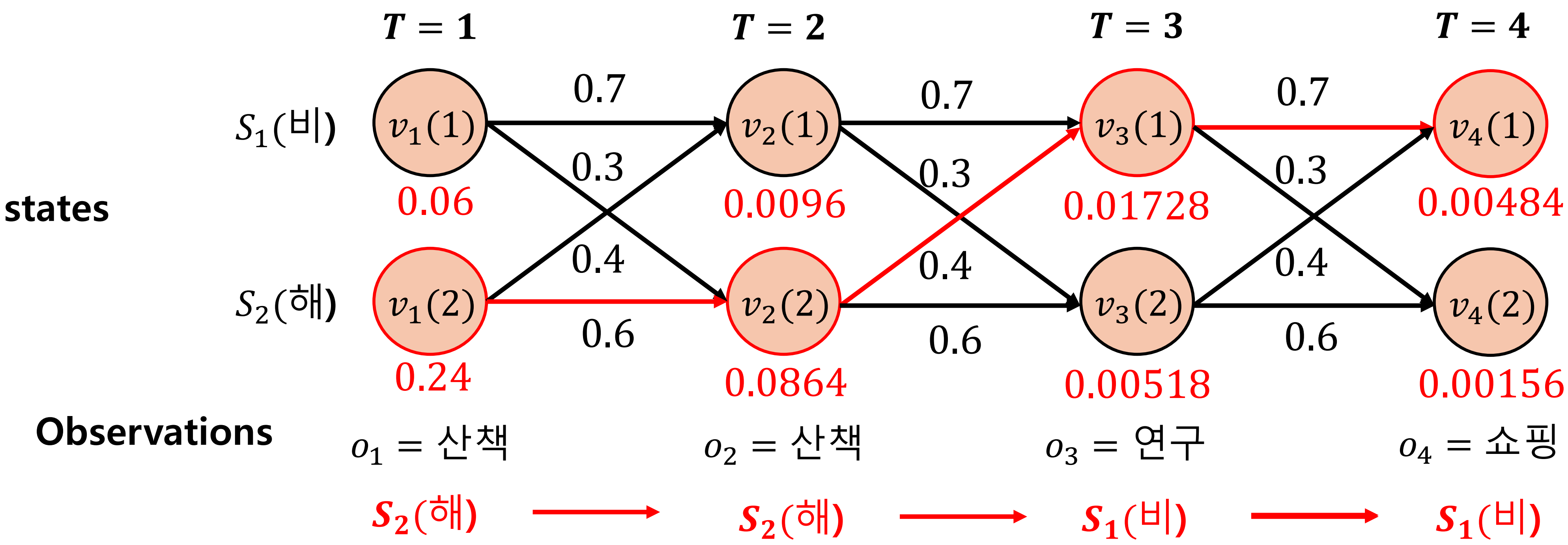
- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?



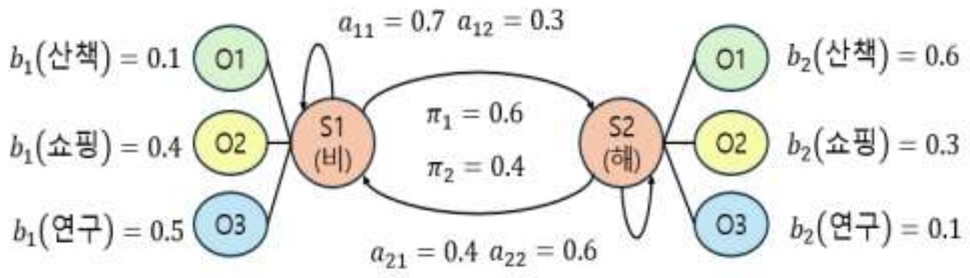
Hidden Markov Models – Decoding



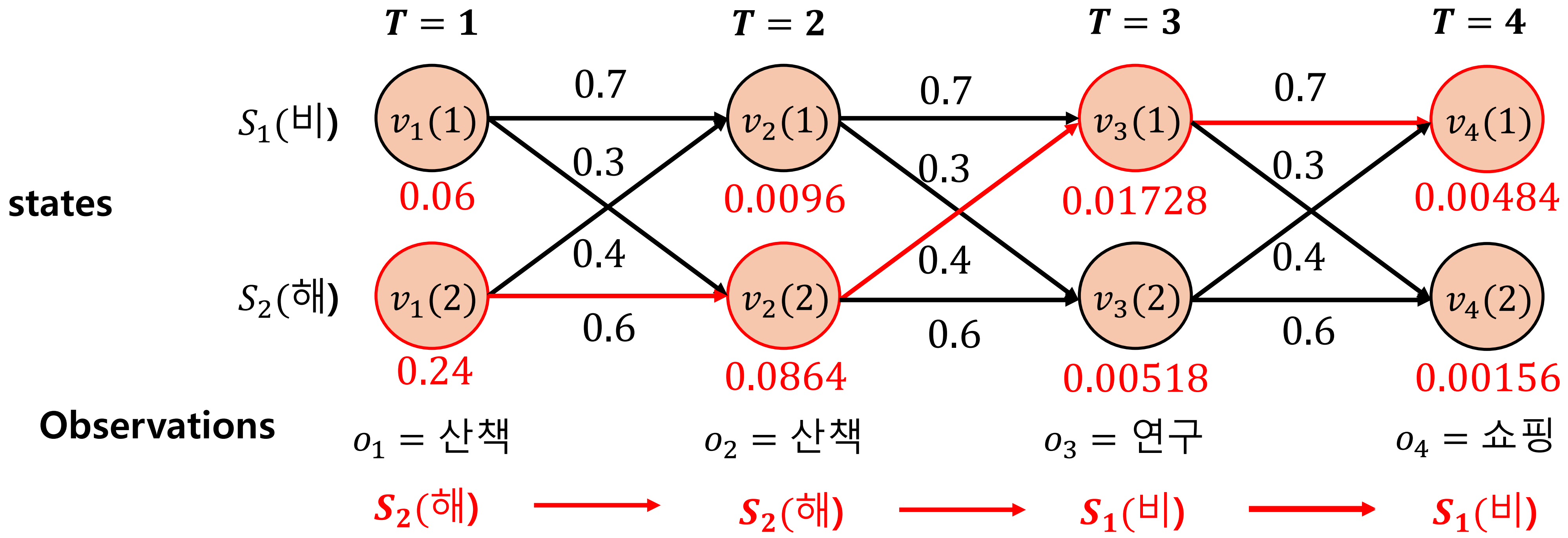
- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?



Hidden Markov Models – Decoding



- Decoding problem: Given HMM (λ^*) and **O**, find the optimal **S**
  - Solution : viterbi algorithm
  - Example : 정 박사가 오늘 산책, 내일 산책, 모레 연구, 글피 쇼핑을 했다면, 각 날들의 날씨는?



$$\hat{Q}_t = (\hat{q}_1 = s_2, \hat{q}_2 = s_2, \hat{q}_3 = s_1, \hat{q}_4 = s_1)$$

Hidden Markov Models – Decoding

- Viterbi Algorithm for Decoding Problem

Viterbi algorithm

$$v_t(i) \quad = \max_{q_1,q_2,\cdots q_{t-1}} p(o_1, o_2, \cdots o_t, q_1, \mathrm{q}_2, \cdots q_{t-1}, q_t = s_i|\lambda)$$

$$= \left[\max_{1\le j\le n} v_{t-1}(i) a_{ji}\right] \cdot b_i(o_t)$$

$$v_1(i) = \pi_i b_i(o_1)$$

$$v_t(i) = \left[\max_{1\le j\le n} v_{t-1}(i) a_{ji}\right] \cdot b_i(o_t), 2 \le t \le T, 1 \le i \le n$$

$$\tau_t(i) = \operatorname*{argmax}_{1\le j\le n} \left[v_{t-1}(i) a_{ji}\right], 2 \le t \le T, 1 \le i \le n$$

$$\hat{q}_T = \operatorname*{argmax}_{1\le j\le n} v_T(j)$$

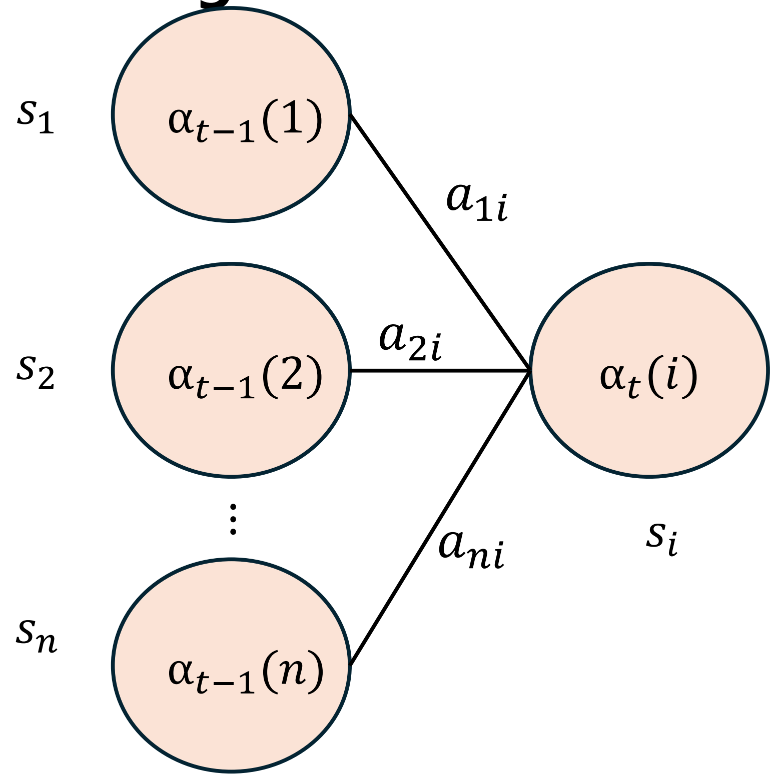
$$\hat{q}_t = \tau_{t+1}(\hat{q}_{t+1}), t = T-1, T-2, \cdots, 1$$

$$\hat{Q}_t = (\hat{q}_1, \hat{q}_2, \cdots, \hat{q}_t)$$

Hidden Markov Models – Decoding

- Summary of Evaluation and Decoding

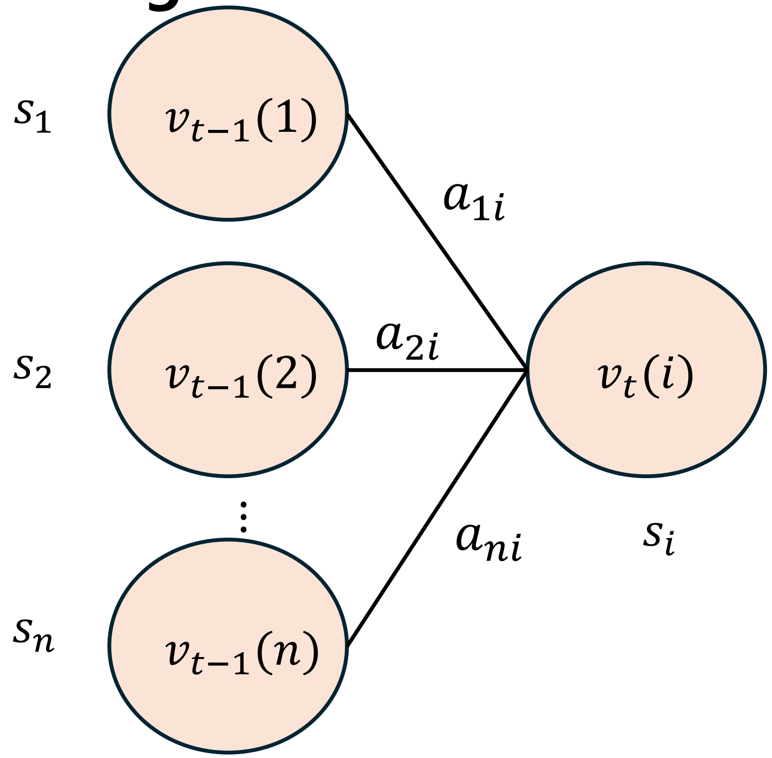
Forward algorithm for evaluation



$$\alpha_t(i) = \left[\sum_{j=1}^{n} \alpha_{t-1}(i) a_{ji}\right] \cdot b_i(o_t)$$

가능한 모든 경우의 확률의 합

Viterbi algorithm for decoding



$$v_t(i) = \left[\max_{1 \le j \le n} v_{t-1}(i) a_{ji}\right] \cdot b_i(o_t)$$

가능한 모든 경우의 확률의 최댓값

Hidden Markov Models – Parameter Learning

- Learing problem (학습)
  - 관측 벡터 O의 확률을 최대로 하는 HMM (λ) (parameter) 을 찾아야 함
  - 여러 개 관측시퀀스를 줄테니 최적의 HMM parameter를 찾아야 함

$$\text{HMM } (\lambda^*) = \underset{\lambda}{argmax}\ P(O|\lambda)$$

HMM parameters : $\{A, B, \pi\}$

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

Hidden Markov Models – Parameter Learning

- Learing problem (학습)
  - HMM (λ) 설정했을 때, 관측치가 발생할 확률을 최대화 하는 λ^*를 찾아야 함

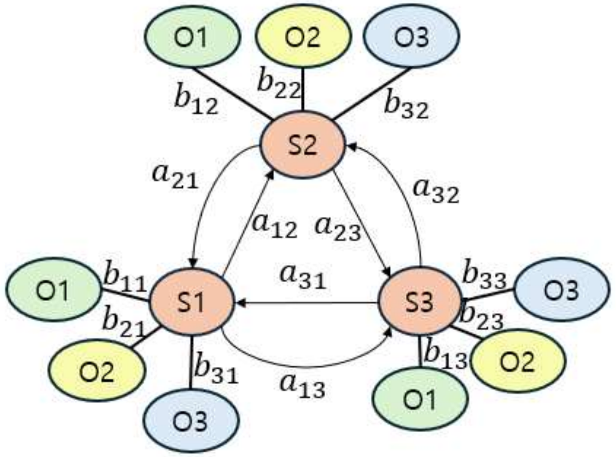
$$\text{HMM}\ (\lambda^*) = \underset{\lambda}{argmax}\ P(O|\lambda)$$

⇩

- Maximum likelihood method (최대우도법)
  - 데이터를 기반으로 확률변수의 파라미터를 구하는 방법
  - 어떤 모수가 주어졌을 때, 원하는 값들이 나올 우도(확률)를 최대로 만드는 파라미터를 선택

HMM parameters : $\{A, B, \pi\}$

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$



Hidden Markov Models – Parameter Learning

- Prodedure of HMM Learning
  - Input : HMM (λ) architecture
  - Output : HMM (λ^*) = $\{A^*, B^*, \pi^*\}$
  - Algorithm

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

1. HMM 초기화
2. 적절한 방법으로 P(O|HMM(λ^{new})) > P(O|HMM(λ))를 찾음
3. 만족스러우면 $\hat{\lambda}$ = HMM(λ^{new})으로 설정하고 멈춤 혹은 2번 반복

Hidden Markov Models – Parameter Learning

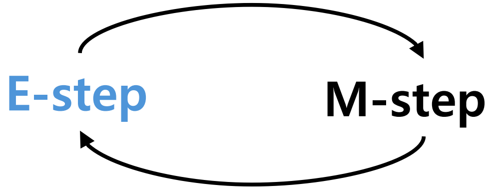
- Learning problem (학습)
  - Problem : Given X = $\{O_1, \cdots, O_N\}$, find the HMM (λ^*)
  - Solution : Baum-Welch algorithm (or forward-backward algorithm)

Baum-Welch Algorithm

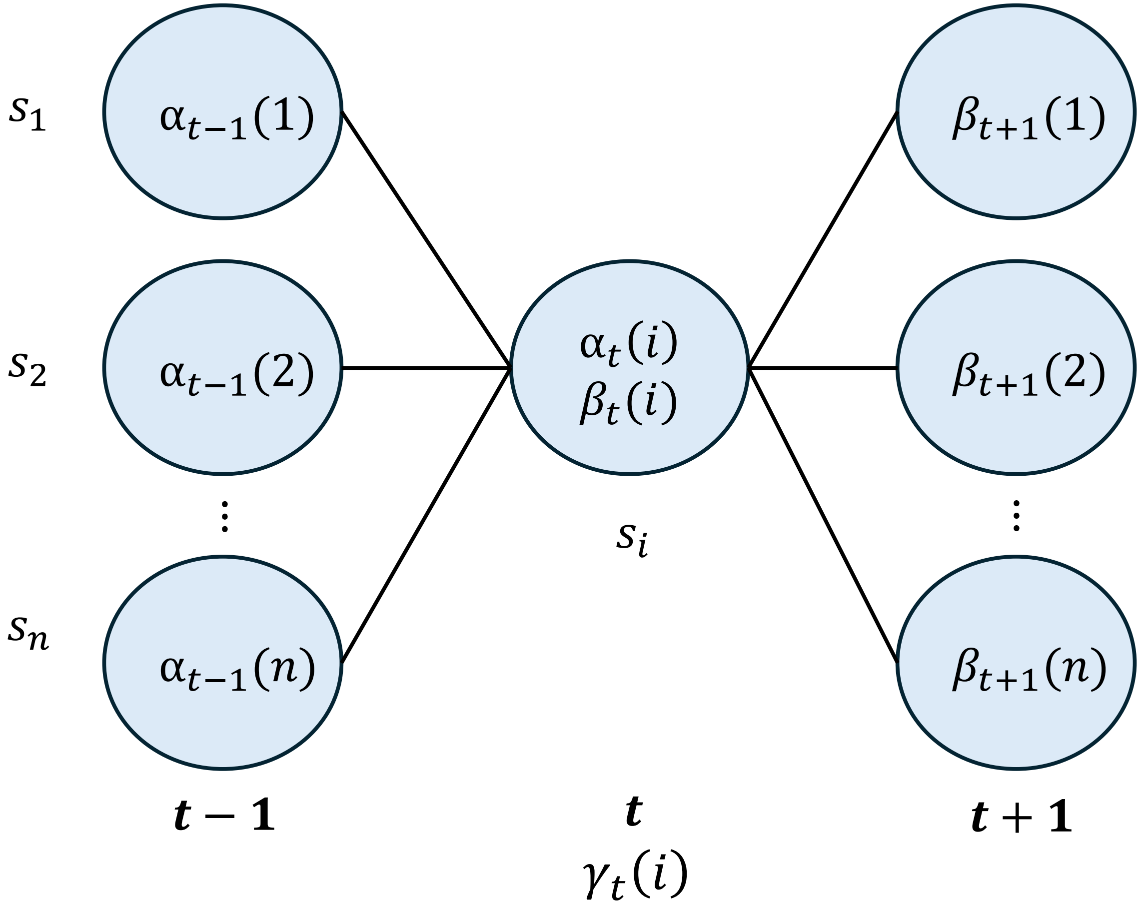
E-step
(hidden state estimation)

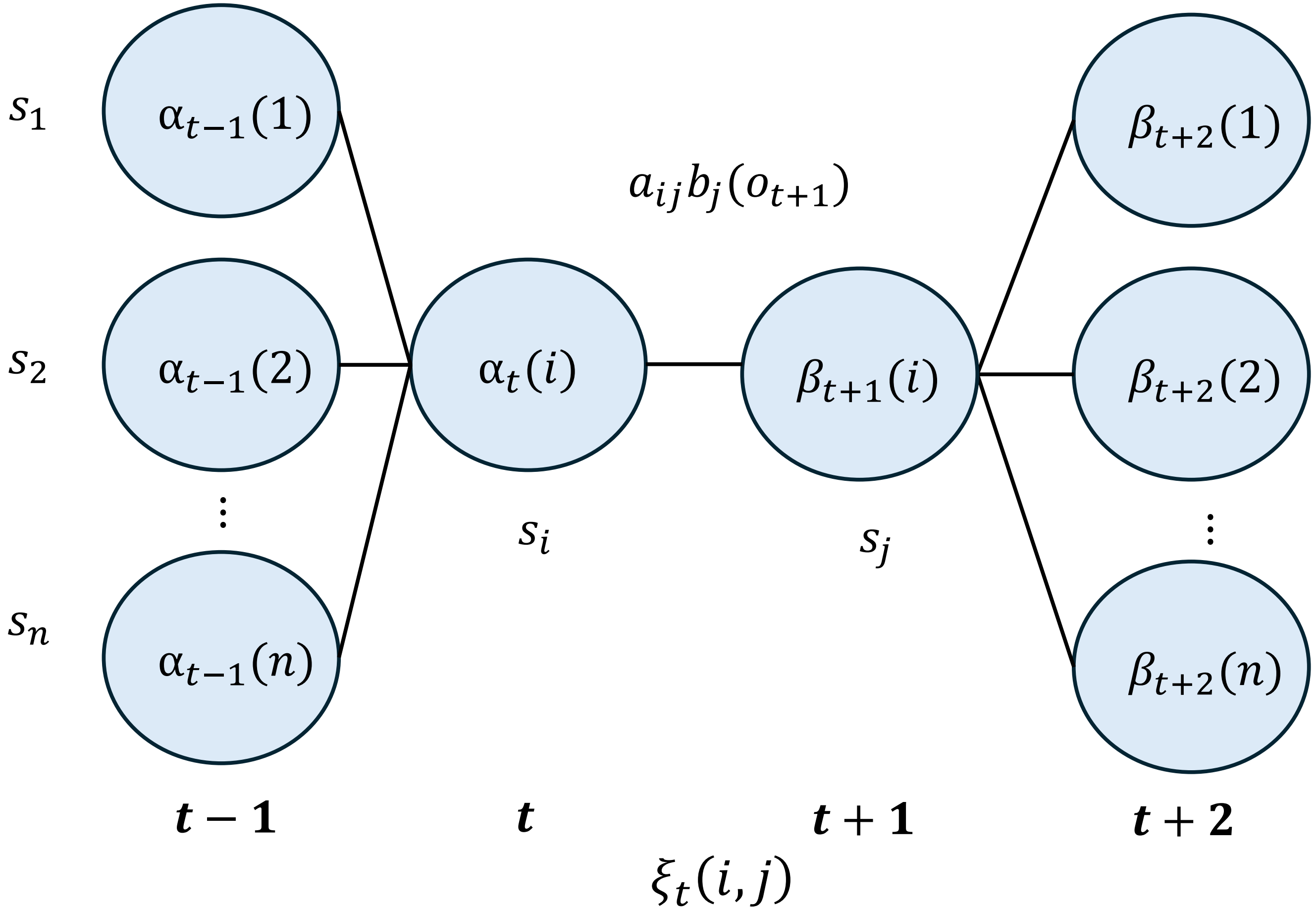
M-step
(HMM (λ) update)

Hidden Markov Models – Parameter Learning



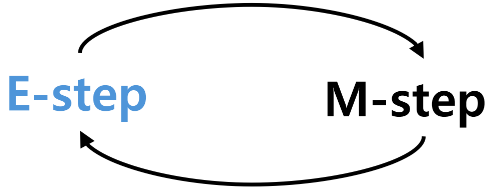
- Learning problem (학습)
    - Baum-Welch algorithm이 사용하는 $\gamma_t(i)$와 $\xi_t(i,j)$의 역할



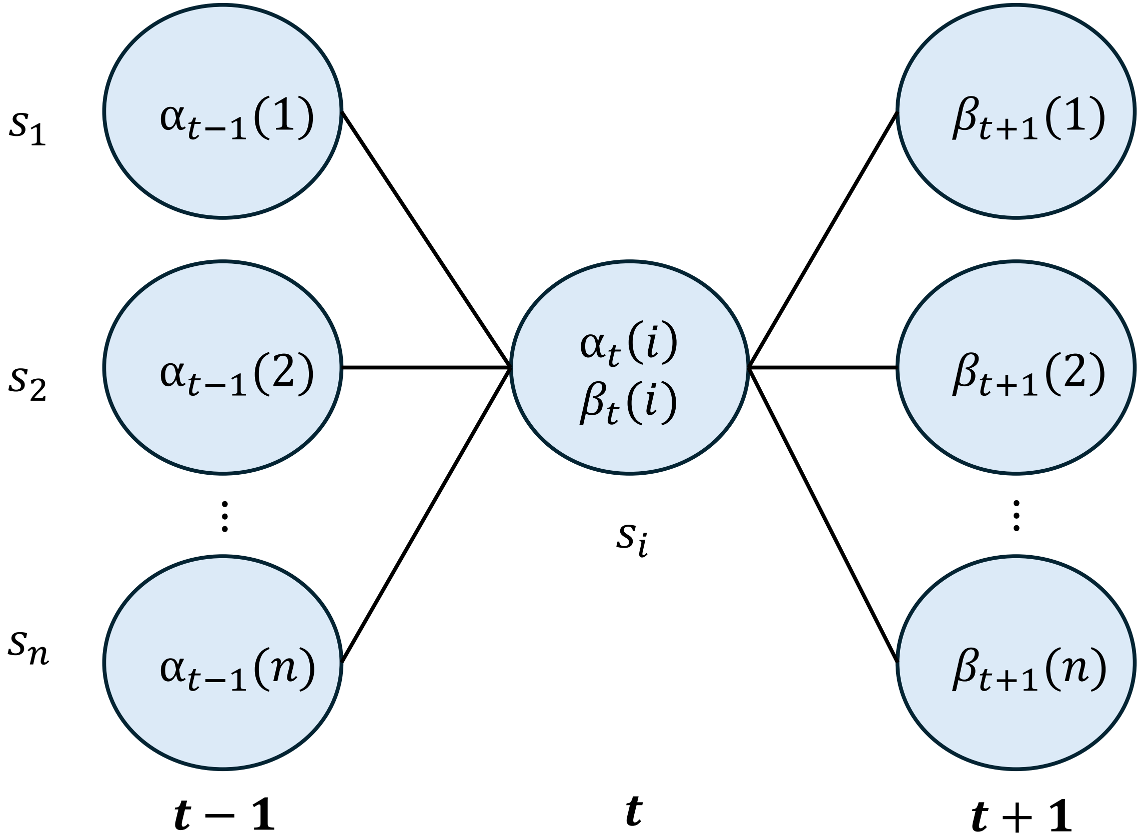


α: forward probability , β: backward probability

Hidden Markov Models – Parameter Learning



- Learning problem (학습)
  - $\gamma_t(i)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i일 확률
  - $\xi_t(i,j)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i, t+1시점 상태는 s_j일 확률

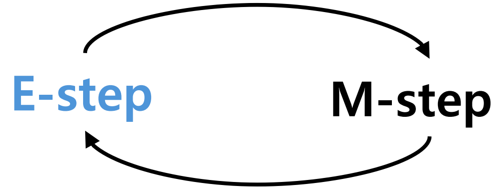


$$\alpha_t(i) = p(O_1, O_2, \cdots, O_t, q_t = s_i|\lambda) = \left[\sum_{j=1}^{n} \alpha_{t-1}(i) a_{ji}\right] \cdot b_i(o_t)$$

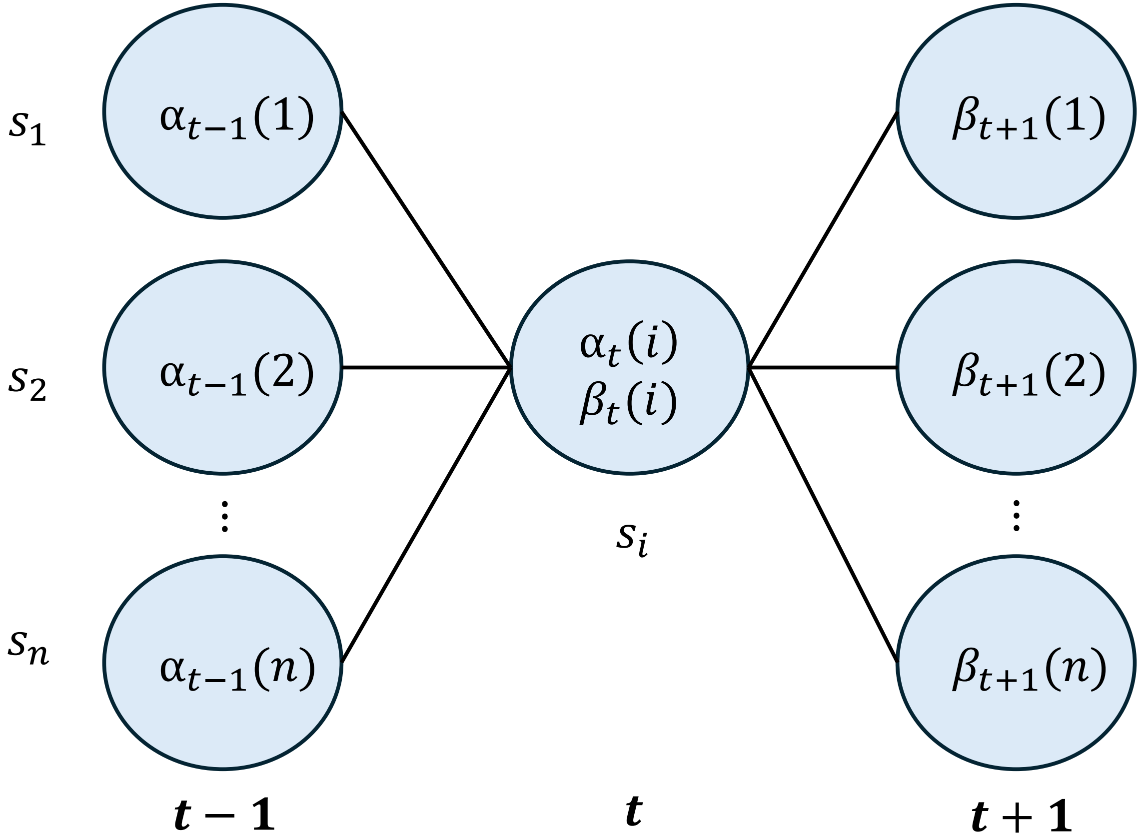
$$\beta_t(i) = p(O_{t+1}, O_{t+2}, \cdots, O_T|q_t = s_i, \lambda) = \left[\sum_{j=1}^{n} a_{ij} b_j(o_t) \beta_t(j)\right]$$

$\alpha_t(i)\beta_t(i)$: t번째 시점에 상태 i를 지나는 모든 경로에 해당하는 확률의 합

Hidden Markov Models – Parameter Learning



- Learning problem (학습)
  - $\gamma_t(i)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i일 확률
  - $\xi_t(i,j)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i, t+1시점 상태는 s_j일 확률

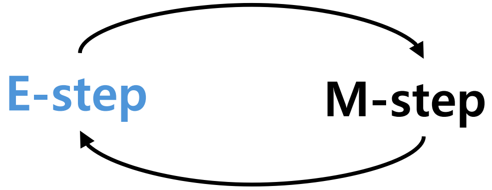


$$\gamma_t(i) = p(q_t = s_i | O, \lambda)$$

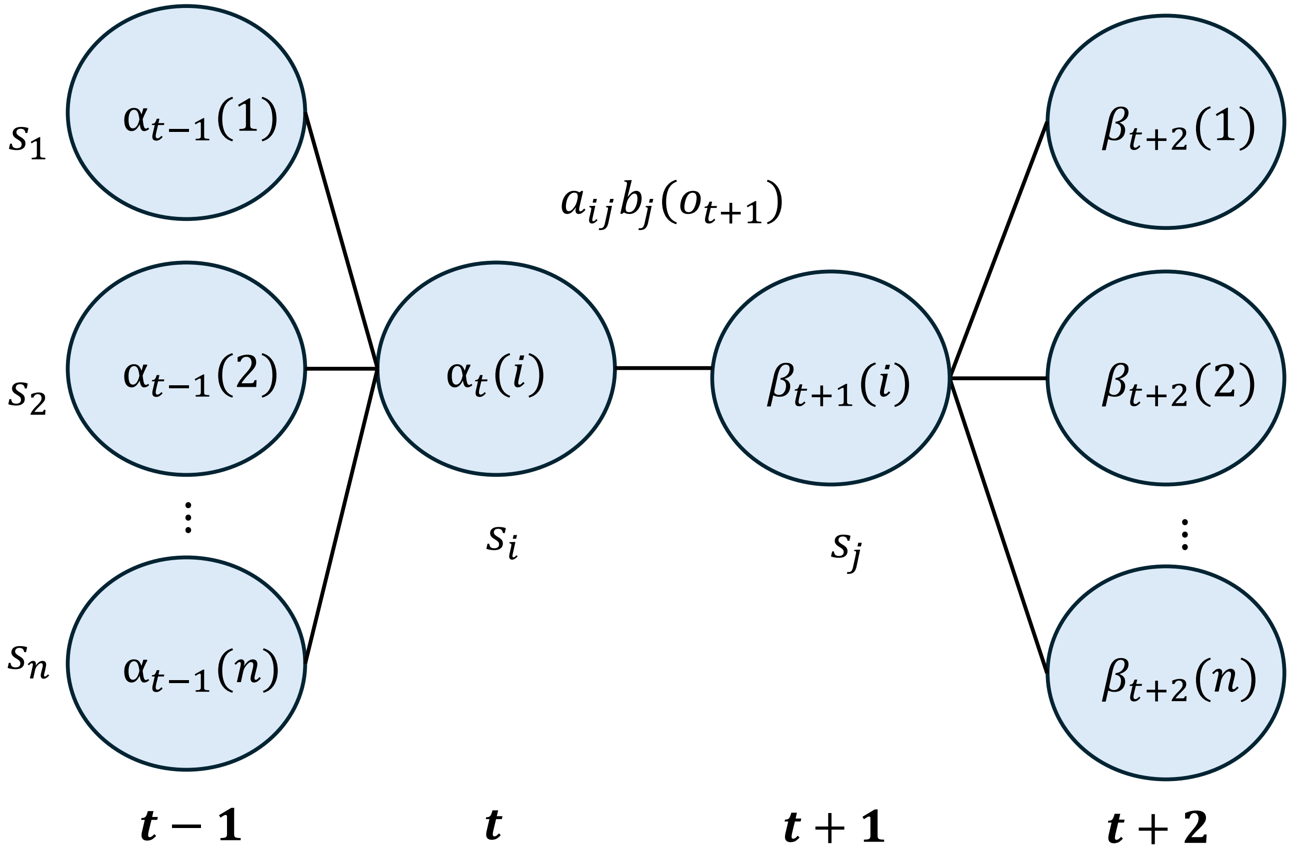
$$= \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{n} \alpha_t(j)\beta_t(j)}, 1 \le t \le T, 1 \le i \le n$$

$$= \frac{s_i \text{일 확률}}{s_{1\sim n} \text{일 확률 모두 더한 값}}, 1 \le t \le T, 1 \le i \le n$$

Hidden Markov Models – Parameter Learning



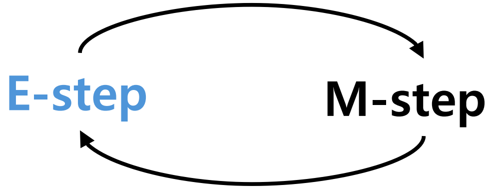
- Learning problem (학습)
  - $\gamma_t(i)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i일 확률
  - $\xi_t(i,j)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i, t+1시점 상태는 s_j일 확률



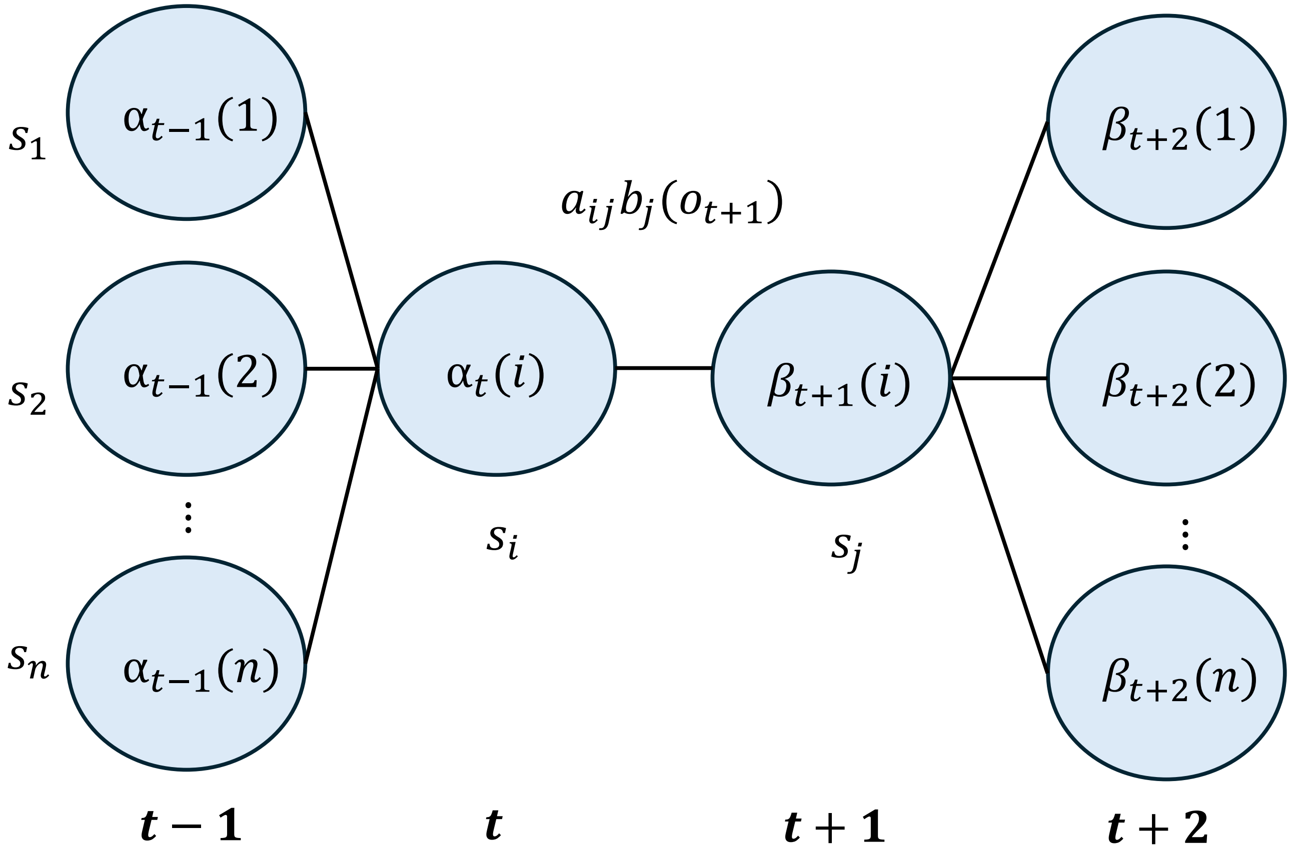
$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_t(i)\,a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}$$

$$1 \le t \le T-1, 1 \le i,j \le n$$

Hidden Markov Models – Parameter Learning



- Learning problem (학습)
  - $\gamma_t(i)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i일 확률
  - $\xi_t(i,j)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i, t+1시점 상태는 s_j일 확률

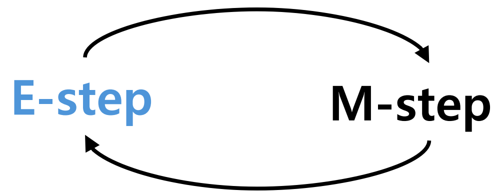


i상태와 j상태 연결

$$\xi_t(i,j) = \frac{\alpha_t(i)\,a_{ij}b_j(o_{t+1})\,\beta_{t+1}(j)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_t(i)\,a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}$$

$$1 \le t \le T-1, 1 \le i,j \le n$$

Hidden Markov Models – Parameter Learning



- Learning problem (학습)
  - $\gamma_t(i)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i일 확률
  - $\xi_t(i,j)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i, t+1시점 상태는 s_j일 확률

$$\gamma_t(i) = p(q_t = s_i | O, \lambda)$$

$$= \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{n} \alpha_t(j)\beta_t(j)}, 1 \le t \le T, 1 \le i \le n$$

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_t(i)\, a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}$$

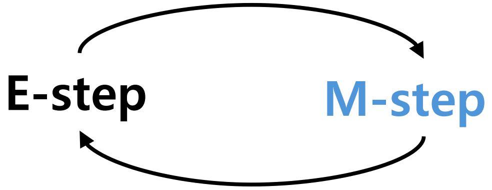
$$1 \le t \le T-1, 1 \le i,j \le n$$

E-step은 α, β를 계산하여 $\gamma_t(i)$, $\xi_t(i,j)$를 구하는 것

M-step은 $\gamma_t(i)$, $\xi_t(i,j)$를 이용하여 HMM (λ) 개선 -> HMM(λ^{new})

개선? P(O|HMM(λ^{new})) > P(O|HMM(λ)) -> 즉, HMM(λ^{new})의 evaluation시 높은 확률

Hidden Markov Models – Parameter Learning



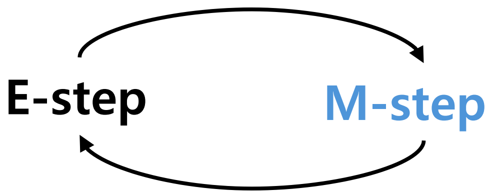
- Learning problem (학습)
  - Baum-Welch algorithm 이 사용하는 hidden state $\gamma_t(i)$와 $\xi_t(i,j)$의 역할
  - $\gamma_t(i)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_t일 확률
  - $\xi_t(i,j)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i, t+1시점 상태는 s_j일 확률

$$\pi_i^{new} = t\text{가 1일때 } s_i\text{에 있을 확률} \Rightarrow \gamma_{t=1}(i), \quad 1 \le i \le n$$

Expected number of times in state s_i at time t = 1

Hidden Markov Models – Parameter Learning

- Learning problem (학습)
  - Baum-Welch algorithm 이 사용하는 hidden state $\gamma_t(i)$와 $\xi_t(i,j)$의 역할
  - $\gamma_t(i)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_t일 확률
  - $\xi_t(i,j)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i, t+1시점 상태는 s_j일 확률

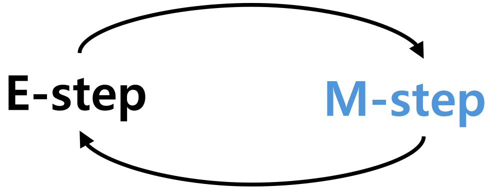


$$a_{ij}{}^{new} = \frac{s_i\text{에서}s_j\text{로 전이할 기대값}}{s_i\text{에서 전이할 기대값}}$$

$$= \frac{(t = 1\text{일 때, } s_i\text{에서}s_j\text{로 전이할 확률}) + \cdots + (t = T-1\text{일 때,} s_i\text{에서}s_j\text{로 전이할 확률})}{(t = 1\text{일 때, } s_i\text{에서 전이할 확률}) + \cdots + (t = T-1\text{일 때,} s_i\text{에서 전이할 확률})}$$

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

Hidden Markov Models – Parameter Learning



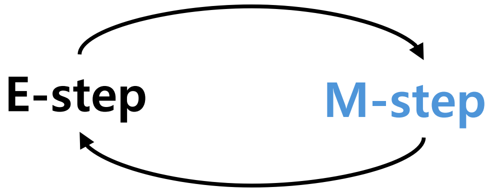
- Learning problem (학습)
  - Baum-Welch algorithm 이 사용하는 hidden state $\gamma_t(i)$와 $\xi_t(i,j)$의 역할
  - $\gamma_t(i)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_t일 확률
  - $\xi_t(i,j)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i, t+1시점 상태는 s_j일 확률

$$b_i(v_k)^{new} = \frac{s_i\text{에서 } v_k\text{를 관측할 확률 }(i\text{상태하에서 관측치가 } v_k\text{일 확률})}{s_i\text{에 있을 확률}(i\text{상태가 나타날 확률})}$$

$$= \frac{o_t = v_t\text{인 모든 } t\text{에 대해 } s_i\text{에 있을 확률의 합}}{(t=1\text{일 때, } s_i\text{에 있을 확률}) + \cdots + (t = T-1\text{일 때, } s_i\text{에 있을 확률})}$$

$$= \frac{\sum_{t=1, st. o_t = v_t}^{T} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)}, 1 \le i \le n, 1 \le k \le m$$

Hidden Markov Models – Parameter Learning



- Learning problem (학습)
  - Baum-Welch algorithm 이 사용하는 hidden state $\gamma_t(i)$와 $\xi_t(i,j)$의 역할
  - $\gamma_t(i)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_t일 확률
  - $\xi_t(i,j)$: HMM (λ), O 주어졌을 때, t시점 상태가 s_i, t+1시점 상태는 s_j일 확률

$$\pi_i{}^{new} = \gamma_t(i), \quad 1 \le i \le n$$

$$a_{ij}{}^{new} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, 1 \le i,j \le n$$

$$b_i(v_k)^{new} = \frac{\sum_{t=1, st. o_t = v_t}^{T} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)}, 1 \le i \le n, 1 \le k \le m$$

Hidden Markov Models – Parameter Learning

- Prodedure of HMM Learning
  - Input : HMM (λ) architecture
  - Output : HMM (λ^*) $= \{A^*, B^*, \pi^*\}$
  - Algorithm

    1. HMM 초기화

    2. 적절한 방법으로 P(O|HMM(λ^{new})) > P(O|HMM(λ))를 찾음

    3. 만족스러우면 $\hat{\lambda}$ = HMM(λ^{new})으로 설정하고 멈춤 혹은 2번 반복

      - 만족이란?

        a. 특정 threshold(r) 값 이하인 경우 만족

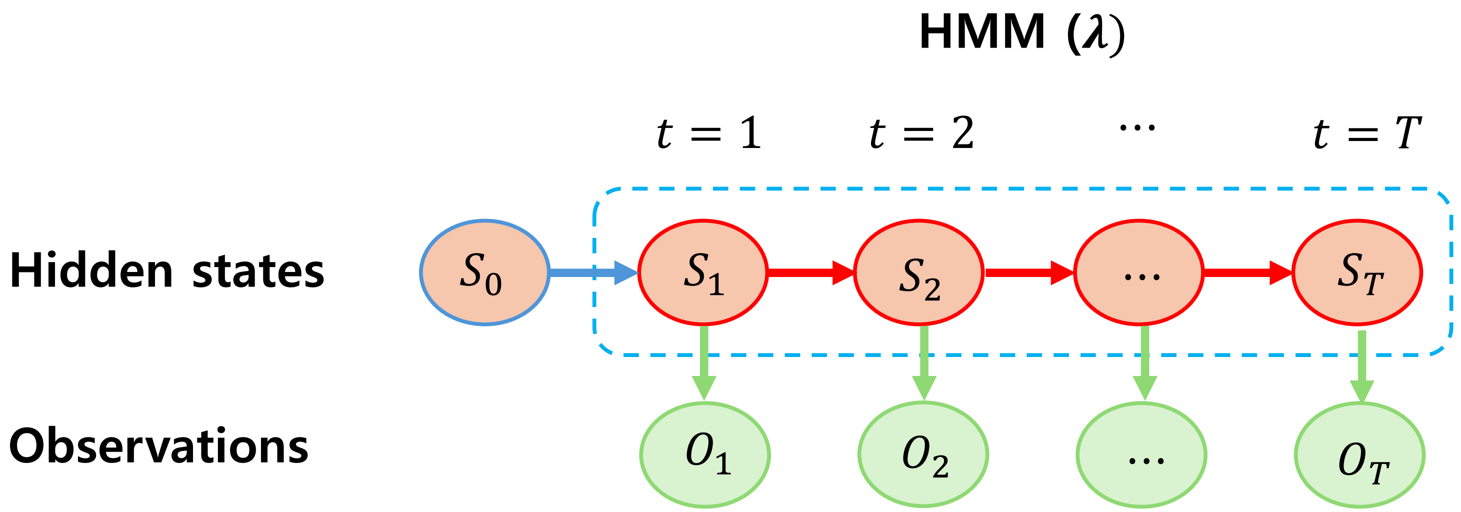
        If P(O|HMM(λ^{new})) $-$ P(O|HMM(λ)) is less than r, Stop

        b. 특정 threshold(r) 값 이상인 경우 만족

        If P(O|HMM(λ^{new})) is greater than r, Stop

Hidden Markov Models – Parameter Learning

- Summary of Hidden Markov Model $\lambda = [A, B, \pi]$
- Three problems of hidden Markov model
  - Given HMM (λ^*) and **O**, find the probability of **O** -> <u>**Evaluation**</u> problem
  - Given HMM (λ^*) and **O,** find the optimal **S** -> <u>**Decoding**</u> problem
  - Given $X = \{O_{1,}, \cdots, O_N\}$, find the HMM ($\lambda^*$) -> <u>**Learning**</u> problem



Thank you