

Ch17_ 음성인식의 기수 : 은닉 마르코프 모델(HMM)





❖ 특징들의 시간성

- 예) 지진파, 음성, 주식 거래량, 온라인 필기 문자 등
- 이들을 순차^{sequential} 데이터 또는 문맥 의존^{context-dependent} 데이터라 부름

❖ 순차 데이터의 인식

- 시간성의 표현과 정보 추론 방법 필요
- 은닉 마코프 모델은^{HMM} 가장 널리 사용되는 방법

❖ HMM

- 19세기 마코프 모델에 토대를 둠 (7.2 절)
- 1960년대 Baum 등이 은닉 추가하여 HMM으로 확장함 (7.3~7.5절)
- HMM은 많은 분야에서 문제 해결 도구로 활용됨
 - 패턴인식, 컴퓨터비전, 데이터마이닝, 정보검색, 생물 정보학, 신호처리, 데이터베이스
 - 응용 사례) 음성 인식, 온라인 필기 인식, DNA 열 찾기, 제스처 인식, 영어 발음 교정, 음악 인식
- 훌륭한 튜토리얼 논문 [Rabiner 89]

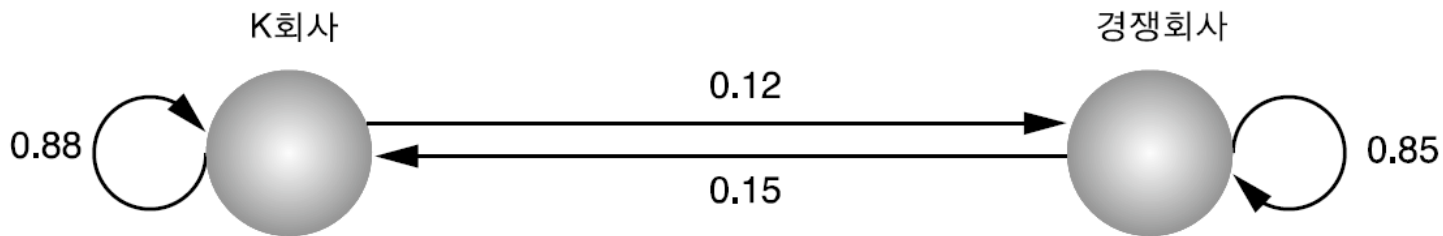


01_ 확률 행렬과 마르코프 연쇄

❖ 확률 행렬과 마르코프 연쇄 (1/3)

- 확률 행렬 : 행렬의 성분이 확률로 이루어진 행렬
- 마르코프 연쇄 : 한 상태에서 다른 상태로 변할 확률이 과거의 자취보다 현재의 상태에만 의존하는 모델

K라는 분유 회사가 있다. 이 회사의 분유가 현재 분유 시장의 25%를 점유하고 있다고 한다. 전년도의 데이터에 의하면 K회사 고객의 88%가 계속 고객으로 남아 있고, 12%는 K회사의 제품에서 경쟁회사의 제품으로 바꾸었다. 경쟁회사 고객의 85%는 여전히 그 제품을 사용하고 있고, 나머지 15%는 경쟁회사의 제품에서 K회사의 제품으로 바꾼 것으로 분석되었다. 이러한 경향이 계속된다고 가정할 경우에 2년 후 그리고 더 오랜 시간이 지난 후 K회사의 시장 점유율은 어떻게 될까?



[그림 17-1] 상태-전이 다이어그램



01_ 확률 행렬과 마르코프 연쇄

❖ 확률 행렬과 마르코프 연쇄 (3/3)

- 상태1 : 고객이 K회사를 선택하여 K회사의 분유를 구매한다.
- 상태2 : 고객이 경쟁회사를 선택하여 그 회사의 분유를 구매한다.

- 천이 행렬 P

$$P = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{bmatrix} 0.88 & 0.12 \\ 0.15 & 0.85 \end{bmatrix} \end{matrix}$$

- 초기치

$$P_{init} = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{bmatrix} 0.25 & 0.75 \end{bmatrix} \end{matrix}$$

- 기간 t 에서의 시스템의 상태

$$s_t = s_{t-1}(P) = s_{t-2}(P)(P) = \dots = s_1(P)^{t-1}$$

- 다음해의 시장 점유율

$$\begin{aligned} s_2 &= s_1 P = [0.25, 0.75] \begin{bmatrix} 0.88 & 0.12 \\ 0.15 & 0.85 \end{bmatrix} \\ &= [(0.25)(0.88) + (0.75)(0.15), (0.25)(0.12) + (0.75)(0.85)] \\ &= [0.3325, 0.6675] \end{aligned}$$



01_ 확률 행렬과 마르코프 연쇄

❖ 확률 행렬과 마르코프 연쇄 (3/3)

- 3년 후의 K회사의 시장 점유율 → 39.27%

$$\begin{aligned} s_3 &= s_2 P = [0.3325, 0.6675] \begin{bmatrix} 0.88 & 0.12 \\ 0.15 & 0.85 \end{bmatrix} \\ &= [0.3927, 0.6073] \end{aligned}$$

- 이 모델에서의 결과는 이전 해의 시장 점유율에만 영향을 받는 것을 알 수 있다.
- 이처럼 한 상태에서 다른 상태로 변할 확률이 현재의 상태에만 의존하는 모델을 “1차 마르코프 연쇄” 라고 한다.



- ❖ 시간성이 없는 데이터
 - 특징들의 선후 관계는 무의미

- ❖ 시간성 있는 데이터 (순차 데이터)
 - 특징들의 선후 관계는 매우 중요함

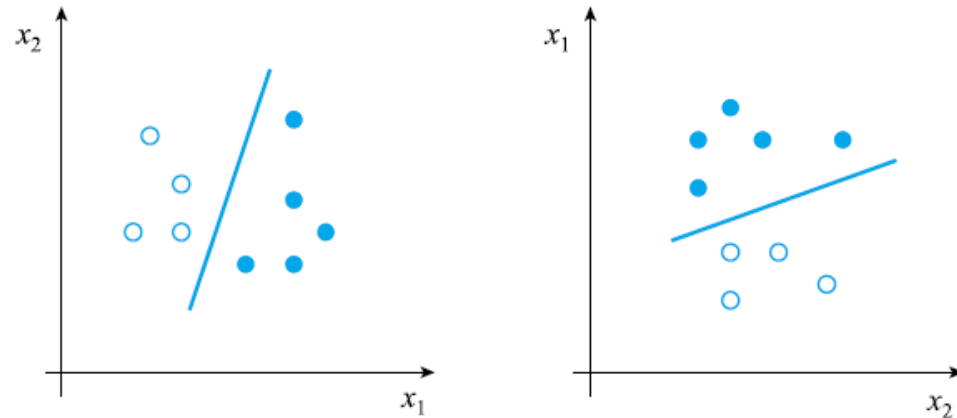


그림 7.1 시간성이 없는 데이터에서 특징의 위치를 바꿈

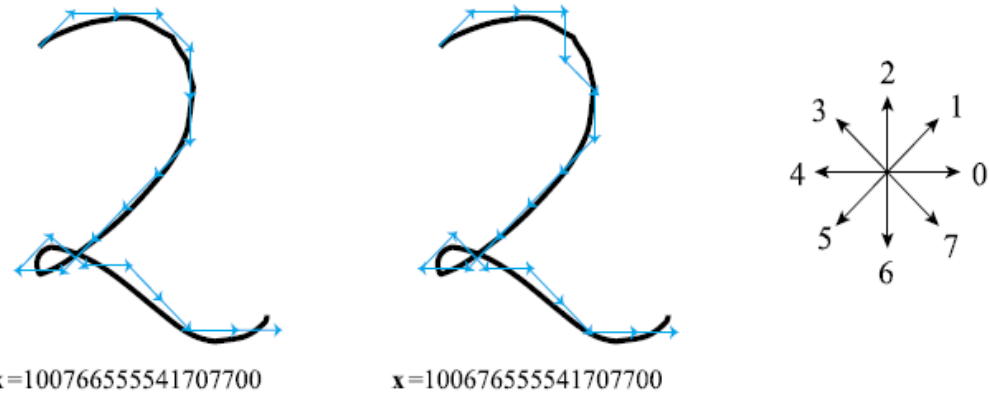


그림 7.2 시간성을 갖는 패턴에서 특징 순서를 바꾸었을 때 나타나는 물리적 왜곡 현상



❖ 순차 데이터

- 가변 길이
- 관측 벡터로 표현

$$\mathbf{O} = (o_1, o_2, \dots, o_{t-1}, o_t, o_{t+1}, \dots, o_T)^T = o_1 o_2 \cdots o_{t-1} o_t o_{t+1} \cdots o_T \quad (7.1)$$

- 관측 o_t 가 가질 수 있는 값의 집합을 알파벳이라 함
 - 알파벳 $V = \{v_1, v_2, \dots, v_m\}$
 - v_i 를 기호라 함

❖ 기호들이 시간에 따라 의존성 가짐

- 예) 그림 7.2의 온라인 숫자
 - $P(o_t=2 | o_{t-1}=6) \approx 0$
 - $P(o_t=5 | o_{t-1}=6) > P(o_t=4 | o_{t-1}=6)$



마코프 모델

❖ 우선 은닉 마코프 모델에서 은닉이 빠진 마코프 모델을 공부하자.

- 러시아 수학자 Andrey Markov가 제안
- 시간 t 에서의 관측은 가장 최근 r 개 관측에만 의존한다는 가정 하의 확률 추론

$$\left. \begin{array}{l}
 r=0\text{이면, 0차 마코프 체인 } P(o_t | o_{t-1}o_{t-2} \cdots o_1) = P(o_t) \\
 r=1\text{이면, 1차 마코프 체인 } P(o_t | o_{t-1}o_{t-2} \cdots o_1) = P(o_t | o_{t-1}) \\
 r=2\text{이면, 2차 마코프 체인 } P(o_t | o_{t-1}o_{t-2} \cdots o_1) = P(o_t | o_{t-1}, o_{t-2})
 \end{array} \right\} (7.2)$$



(a) 0차 마코프체인



(b) 1차 마코프 체인



(c) 2차 마코프 체인

그림 7.3 마코프 체인



❖ 마코프 체인의 합리성은?

- 예) 최근 사흘의 날씨 정보를 가지고 오늘 날씨 추론
 - 그끄저께 해 ($o_{t-3}=\text{해}$), 그저께 해 ($o_{t-2}=\text{해}$), 어제 비 ($o_{t-1}=\text{비}$)
 - 오늘 비올 확률은?
 - 1차 마코프 체인을 사용한다면

$$P(o_t = \text{비} \mid o_{t-1} = \text{비}, o_{t-2} = \text{비}, o_{t-3} = \text{해}) = P(o_t = \text{비} \mid o_{t-1} = \text{비})$$

➤ 즉 1차 마코프 체인에서는 아래 네 개 확률이 같다.

» 정말 같을까?

» 다르다면 오차가 무시할 정도일까?

$$P(o_t = \text{비} \mid o_{t-1} = \text{비}, o_{t-2} = \text{비}, o_{t-3} = \text{비})$$

$$P(o_t = \text{비} \mid o_{t-1} = \text{비}, o_{t-2} = \text{비}, o_{t-3} = \text{해})$$

$$P(o_t = \text{비} \mid o_{t-1} = \text{비}, o_{t-2} = \text{해}, o_{t-3} = \text{비})$$

$$P(o_t = \text{비} \mid o_{t-1} = \text{비}, o_{t-2} = \text{해}, o_{t-3} = \text{해})$$



❖ 마코프 모델

- 1차 마코프 체인 사용
- 2차 이상에서는 추정할 매개 변수가 많아 현실적인 문제 발생
- 알파벳을 구성하는 기호 각각을 **상태**로 간주

❖ 날씨 예

- $V = \{\text{비, 구름, 해}\}$
- 기후 관측에 의해 얻은 날씨 변화 확률

표 7.1 날씨 변화 확률

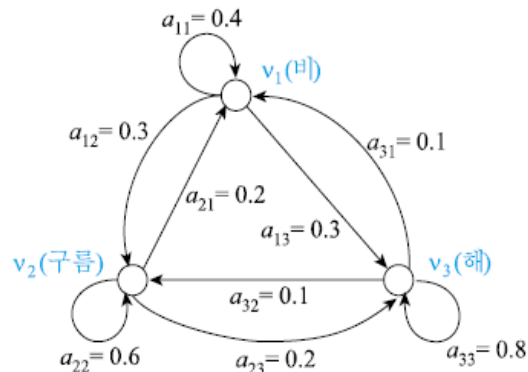
| 오늘 \ 내일 | 비 | 구름 | 해 |
|---------|-----|-----|-----|
| 비 | 0.4 | 0.3 | 0.3 |
| 구름 | 0.2 | 0.6 | 0.2 |
| 해 | 0.1 | 0.1 | 0.8 |



❖ 상태 전이

- 상태 전이 확률 행렬과 상태 전이도

$$\mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{vmatrix}$$



(a) 상태 전이 확률 행렬

(b) 상태 전이도

그림 7.4 마코프 모델을 위한 상태 전이 확률과 상태 전이도

$$\left. \begin{array}{l} a_{ij} = P(o_t = v_j \mid o_{t-1} = v_i) \\ \text{여기서 } a_{ij} \geq 0 \text{ 와 } \sum_{j=1}^m a_{ij} = 1 \text{ 을 만족} \end{array} \right\}$$



❖ 마코프 모델로 무엇을 할 수 있나?

▪ 관측벡터 \mathbf{O} 의 확률 구하기

$$\begin{aligned}
 P(\mathbf{O} | \text{마코프모델}) &= P(\mathbf{O} | \mathbf{A}) \\
 &= P(\mathbf{O}) \\
 &= P(o_1, o_2, \dots, o_T) \\
 &= P(o_1)P(o_2 | o_1)P(o_3 | o_2, o_1) \cdots P(o_t | o_{t-1}, o_{t-2}, \dots, o_1) \cdots P(o_T | o_{T-1}, \dots, o_1) \\
 &= P(o_1)P(o_2 | o_1)P(o_3 | o_2) \cdots P(o_t | o_{t-1}) \cdots P(o_T | o_{T-1}) \\
 &= P(o_1) \prod_{i=1}^{T-1} P(o_{i+1} | o_i)
 \end{aligned} \quad (7.4)$$

▪ 예제 7.1 “오늘 해가 떴는데 내일부터 7일 간의 날씨가 해-해-비-비-해-구름-해일 확률은 얼마인가?”

$$\begin{aligned}
 P(\mathbf{O} | \mathbf{A}) &= P(o_1 = \text{해}, o_2 = \text{해}, o_3 = \text{해}, o_4 = \text{비}, o_5 = \text{비}, o_6 = \text{해}, o_7 = \text{구름}, o_8 = \text{해} | \mathbf{A}) \\
 &= P(\text{해}, \text{해}, \text{해}, \text{비}, \text{비}, \text{해}, \text{구름}, \text{해}) \\
 &= P(\text{해}) P(\text{해} | \text{해}) P(\text{해} | \text{해}) P(\text{비} | \text{해}) P(\text{비} | \text{비}) P(\text{해} | \text{비}) P(\text{구름} | \text{해}) P(\text{해} | \text{구름}) \\
 &= \pi_3 a_{33} a_{33} a_{31} a_{11} a_{13} a_{32} a_{23} \\
 &= 1 * 0.8 * 0.8 * 0.1 * 0.4 * 0.3 * 0.1 * 0.2 \\
 &= 1.536 * 10^{-4}
 \end{aligned}$$



❖ 또 다른 예, 온라인 필기 숫자 인식

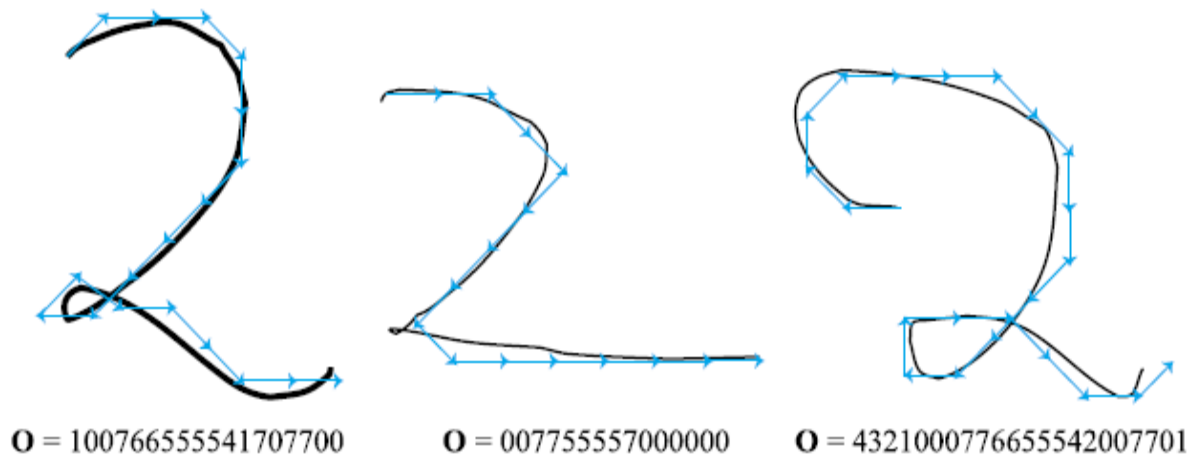


그림 7.5 온라인 필기 숫자 인식에서 부류 2의 훈련 샘플들



❖ 또 다른 예, 온라인 필기 숫자 인식

- 상태 전이 확률 행렬을 구하면,

표 7.2 상태 전이의 발생 회수

| $t-1 \backslash t$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|--------------------|----|---|---|---|---|---|---|---|
| 0 | 11 | 1 | 0 | 0 | 0 | 0 | 0 | 5 |
| 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 2 | 8 | 0 | 1 |
| 6 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 0 |
| 7 | 4 | 0 | 0 | 0 | 0 | 1 | 2 | 4 |

$$A = \begin{pmatrix} 11/17 & 1/17 & 0 & 0 & 0 & 0 & 0 & 5/17 \\ 2/3 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2/11 & 8/11 & 0 & 1/11 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 4/11 & 0 & 0 & 0 & 0 & 1/11 & 2/11 & 4/11 \end{pmatrix}$$

- 초기 확률 행렬을 구하면,

$$\pi = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7, \pi_8) = \left(\frac{1}{3}, \frac{1}{3}, 0, 0, \frac{1}{3}, 0, 0, 0\right)$$



❖ 또 다른 예, 온라인 필기 숫자 인식

- 이렇게 만든 마코프 모델로 그림 7.6의 샘플이 발생할 확률을 구하면,

“체인 코드 1 → 0 → 7 → 6 → 5 → 5 → 7 → 0 → 0으로 표현되는 샘플이 발생할 확률은 얼마인가?”

$$\begin{aligned}
 P(\mathbf{O}|\mathbf{A}) &= P(o_1=1, o_2=0, o_3=7, o_4=6, o_5=5, o_6=5, o_7=7, o_8=0, o_9=0|\mathbf{A}) \\
 &= P(1) P(0|1) P(7|0) P(6|7) P(5|6) P(5|5) P(7|5) P(0|7) P(0|0) \\
 &= \pi_2 a_{21} a_{18} a_{87} a_{76} a_{66} a_{68} a_{81} a_{11} \\
 &= (1/3) * (2/3) * (5/17) * (2/11) * (1/2) * (8/11) * (1/11) * (4/11) * (11/17) \\
 &= 0.9243 * 10^{-4}
 \end{aligned}$$

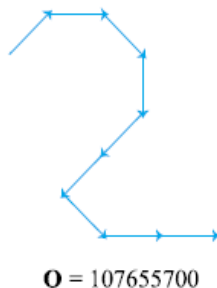


그림 7.6 이 샘플이 발생할 확률은 0.00009243



❖ 마코프 모델과 0차, 2차 마코프 체인과의 비교 (온라인 필기 숫자 예)

▪ 0차 마코프 체인

표 7.3 0차 마코프 체인을 위한 확률 표

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-----|---------------|-------------|-------------|-------------|-------------|---------------|-------------|---------------|
| t | 19 (19/55) | 4 (4/55) | 2 (2/55) | 1 (1/55) | 3 (3/55) | 11 (11/55) | 4 (4/55) | 11 (11/55) |

$$\begin{aligned}
 P(\mathbf{O}|\mathbf{A}_0) &= P(o_1 = 1, o_2 = 0, o_3 = 7, o_4 = 6, o_5 = 5, o_6 = 5, o_7 = 7, o_8 = 0, o_9 = 0 | \mathbf{A}_1) \\
 &= P(1)P(0)P(7)P(6)P(5)P(5)P(7)P(0)P(0) \\
 &= (4/55) * (19/55) * (11/55) * (4/55) * (11/55) * (11/55) * (11/55) * (19/55) \\
 &\quad * (19/55) \\
 &= 0.3489 * 10^{-6}
 \end{aligned}$$



❖ 마코프 모델과 0차, 2차 마코프 체인과의 비교 (온라인 필기 숫자 예)

- 2차 마코프 체인
 - 추정할 매개 변수가 64*8개로 증가

표 7.4 2차 마코프 체인을 위한 전이 확률 표

| $(t-2, t-1) \backslash t$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---------------------------|------------|------------|------------|------------|------------|------------|------------|------------|
| 00 | 5 (5/9) | 0 (0/9) | 0 (0/9) | 0 (0/9) | 0 (0/9) | 0 (0/9) | 0 (0/9) | 4 (4/9) |
| 01 | ... | | | | | | | |
| ... | | | | | | | | |
| <i>ij</i> | | | | | | | | |
| ... | | | | | | | | |
| 77 | | | | | | | | ... |

❖ 꼭 이해하고 기억해야 할 것! (MM과 HMM의 근본적인 차이점)

- 마코프 모델에서는 상태를 나타내는 노드에 관측 (비, 해, 구름) 그 자체가 들어 있다. 즉 상태를 볼 수 있다.
- HMM에서는 상태를 볼 수 없다. 즉 상태가 은닉된다.



Andrey Andreyevich Markov

(1856년 6월 14일 ~ 1922년 7월 20일) 러시아

Markov는 러시아의 수학자이다. 그를 유명하게 만든 이론은 Markov 사슬인데, 이 이론은 시간성 데이터를 처리하는데 현재 가장 널리 쓰이는 은닉 마코프 모델의 이론적 토대가 되었다. 그는 Markov 사슬 이론의 응용으로 A.S. Pushkin의 시 'Eugeny Oregin'에 나타나는 자음과 모음의 분포를 분석하는데 적용하였다. Markov는 1917년에 일어난 러시아 혁명이라는 격변기에서 정치적 완고함을 보이기도 하였다. 일례로 1908년 학생 소요가 한창일 때 그가 교수로 근무하던 Saint Petersburg 대학의 학생들 동향을 관찰하라는 정부 명령이 떨어졌다. 그는 교수가 '정부의 요원'이 될 수 없다고 거부하였고 그것을 계기로 학교를 그만두게 되었고 나중에 복직되었다. 그의 삶과 수학에의 공헌에 대한 상세한 내용은 [Basharin04]를 참고하기 바란다.



[Basharin04] Gely P. Basharin, Amy N. Langville, and Valeriy A. Naumov, "The life and work of A. A. Markov," *Linear Algebra and Its Applications*, Vol.386, pp.3-26, 2004.



❖ 마르코프 가정, 마르코프 모델, 마르코프 과정 (1/3)

- 어느 지역에 매일 변하는 날씨에 대한 마르코프 모델.
- 단, 이 지역의 날씨 하루를 단위로 변하며, 관측 가능한 날씨는 맑음(☀),비(☁),흐림(☁) 가운데 날씨 예측은 과거 날씨의 이력을 바탕으로 내일 날씨가 어떻게 될 것인가를 결정
- 날씨 예측에 관한 통계적인 모델을 설정해 보자. 1일전 (q_{n-1}), 2일전 (q_{n-2}), 3일전 (q_{n-3}), 등의 과거의 날씨에 따라 오늘 (q_n), 날씨가 어떠했다는 통계가 있을 경우,

$$P(q_n | q_{n-1}, q_{n-2}, \dots, q_1)$$

- 즉, n 째 날의 날씨는 $q_n \in \{\text{맑음(☀)}, \text{비(☁)}, \text{흐림(☁)}\}$ 이고, 그 날씨가 될 확률은 알려져 있는 과거($q_{n-2}, q_{n-1} \dots$) 의 날씨에 의존
- 확률을 사용하면 과거 날씨 이력을 바탕으로 내일, 모레의 날씨도 확률적으로 예측을 할 수 있을 것이다. 예를 들어, 만약 과거 3일간의 날씨가 {맑음(☀),비(☁),흐림(☁)} 의 순서로 관측되었다면, 내일 날씨가 비가 올 확률은 다음과 같은 식으로 표현된다.

$$P(q_4 = \text{☁} | q_3 = \text{☁}, q_2 = \text{☀}, q_1 = \text{☀})$$



❖ 마르코프 가정, 마르코프 모델, 마르코프 과정 (2/3)

- n 이 클 수록 우리가 수집해야 하는 관측 데이터도 많아져야 한다는 것
- 가변 $n = 6$ 이라고 가정하면, 우리가 수집해야 하는 과거 날씨 통계 데이터가 무려 $3^{(6-1)} = 243$ 개나 필요하게 된다. 그러므로 다음과 같은 “마르코프 가정” 이라고 하는 좀더 간단한 가정이 필요하다.

- 그래서, 어떤 열 q_1, q_2, \dots, q_n 에 대하여 1차 마르코프 가정:

$$P(q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \dots) = P(q_t = S_j \mid q_{t-1} = S_i)$$

- 2차 마르코프 가정: $P(q_1, q_2, \dots, q_n) = \prod_{i=1}^n P(q_i \mid q_{i-1})$

- 마르코프 가정을 사용하여 어떤 열 $\{q_1, q_2, \dots, q_n\}$ 이 관측될 확률은 과거와 현재의 관측 결과의 결합 확률(joint probability)로 표현

$$\begin{aligned} &P(q_1, q_2, \dots, q_{T-1}, q_T) \\ &= P(q_1)P(q_2 \mid q_1)P(q_3 \mid q_1, q_2), \dots, P(q_{T-1} \mid q_1, q_2, \dots, q_{T-2})P(q_T \mid q_1, q_2, \dots, q_{T-1}) \\ &= P(q_1)P(q_2 \mid q_1)P(q_3 \mid q_2), \dots, P(q_{T-1} \mid q_{T-2})P(q_T \mid q_{T-1}) \end{aligned}$$



02_ 마르코프 가정, 마르코프 모델, 마르코프 과정

❖ 마르코프 가정, 마르코프 모델, 마르코프 과정 (3/3)

- 우변의 두 상태 간의 천이 확률이 시간에 관하여 독립이라고 가정하면, 상태 천이 확률은 다음과 같이 표현할 수 있다.

$$\left\{ \begin{array}{l} a_{ij} \geq 0, \forall_{i,j} \\ \sum_{j=1}^N a_{ij} = 1, \forall_i \end{array} \right\} \quad \ast \quad a_{ij} = P(q_t = S_j \mid q_{t-1} = S_i)$$

- 1차 마르코프 모델에서 이들 확률 값들을 상태 천이 다이어그램으로 표현 할 수 있다. 날씨 영역은 3개의 상태($S = \{ \text{☀}, \text{☁}, \text{☂} \}$)를 가지고 매일 위의 천이 확률 표와 같은 확률값 ($P(q_n | q_{n-1})$)에 따라 가능한 다른 상태로의 천이가 됨.

[표 17-1] 날씨 상태 천이 확률표와 상태 천이 행렬

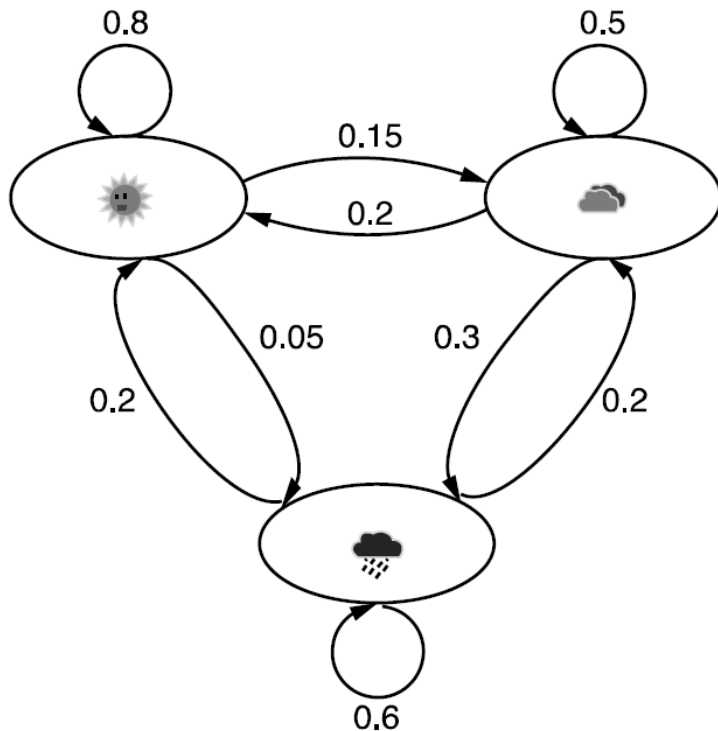
| 오늘의 날씨 | 내일의 날씨 | | |
|--------|--------|------|-------|
| | 맑음(☀) | 비(☂) | 흐림(☁) |
| 맑음(☀) | 0.8 | 0.05 | 0.15 |
| 비(☂) | 0.2 | 0.6 | 0.2 |
| 흐림(☁) | 0.2 | 0.3 | 0.5 |

$$\begin{pmatrix} 0.8 & 0.05 & 0.15 \\ 0.2 & 0.6 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$



❖ 마르코프 가정, 마르코프 모델, 마르코프 과정 (3/3)

- 천이 확률값들은 1차 마르코프 모델의 상태 천이 다이어그램으로 표현할 수 있다. 날씨영역은 세 가지 상태($S = \{☀, ☁, ☔\}$)를 나타낸다. 그리고 날씨는 [표 17-1]의 천이 확률표와 같이 확률값 $P(q_n|q_{n-1})$ 에 따라 다른 상태로의 천이가 이루어진다.



[그림 17-2] 날씨 상태 천이 다이어그램



❖ 마르코프 가정, 마르코프 모델, 마르코프 과정 (3/3)

오늘 날씨(q_1)가 (☀)일 경우에 내일 날씨(q_2)가 (☀)이 되고 모레 날씨(q_3)가 (☁)가 될 확률은 얼마일까?

$$\begin{aligned} &P(q_2 = \text{☀}, q_3 = \text{☁} | q_1 = \text{☀}) \\ &= P(q_3 = \text{☁} | q_2 = \text{☀}, q_1 = \text{☀}) \cdot P(q_2 = \text{☀} | q_1 = \text{☀}) \\ &= P(q_3 = \text{☁} | q_2 = \text{☀}) \cdot P(q_2 = \text{☀} | q_1 = \text{☀}) \leftarrow \text{마르코프 가정} \quad (17.12) \\ &= 0.05 \times 0.8 \\ &= 0.04 \end{aligned}$$



❖ 마르코프 가정, 마르코프 모델, 마르코프 과정 (3/3)

[그림 17-2]의 상태 천이 다이어그램에서 이동하면서 해당되는 확률을 곱해도 같은 결과가 된다.

$$\begin{aligned} P(q_1 = \text{☀}, q_2 = \text{☀}, q_3 = \text{☁}) \\ &= P(q_1 = \text{☀})P(q_2 = \text{☀} | q_1 = \text{☀})P(q_3 = \text{☁} | q_1 = \text{☀}, q_2 = \text{☀}) \\ &= P(q_1 = \text{☀})P(q_2 = \text{☀} | q_1 = \text{☀})P(q_3 = \text{☁} | q_1 = \text{☀}) \\ &= 1.0 \times 0.8 \times 0.05 \\ &= 0.04 \end{aligned} \tag{17.13}$$



❖ 마르코프 가정, 마르코프 모델, 마르코프 과정 (3/3)

그렇다면 어제, 오늘의 날씨가 각각 $q_1 = \text{☔}$, $q_2 = \text{☁}$ 일 때, 내일 날씨가 $q_3 = \text{☀}$ 이 될 확률은 얼마일까?

$$\begin{aligned} P(q_3 = \text{☀} | q_2 = \text{☁}, q_1 = \text{☔}) \\ &= P(q_3 = \text{☀} | q_2 = \text{☁}) \leftarrow \text{마르코프 가정} \\ &= 0.2 \end{aligned} \tag{17.14}$$

마르코프 모델은 상태 천이의 형태에 따라 날씨 HMM 모델과 같은 에르고딕(ergodic) 모델과 음성인식에 많이 적용되는 좌우(left-right) 모델이 있다.



❖ 마르코프 모델 (1/2)

- 어떤 주어진 시점에서 N개의 가능한 상태 $S=\{S_1, S_2, \dots, S_N\}$ 에 있을 수 있는 시스템
- 일정한 간격의 지속 시간을 가지고 새로운 상태로 천이가 일어나게 되는 시스템
- 상태간의 천이는 확률적으로 표현되어 질 수 있는 시스템
- 일반적으로, 시스템이 어떠한 상태 ($q_t = S_i$)에 있을 확률은 그 시스템의 시간적인 이력을 완전히 표현하는 어떤 함수로 표현 가능
- 이러한 함수를 간략히 하기 위해서 시간 t 에서의 시스템의 상태는 오직 시간 $t-1$ 에서 상태에만 따른다고 가정 하면(마르코프 가정) 다음의 수식이 가능.

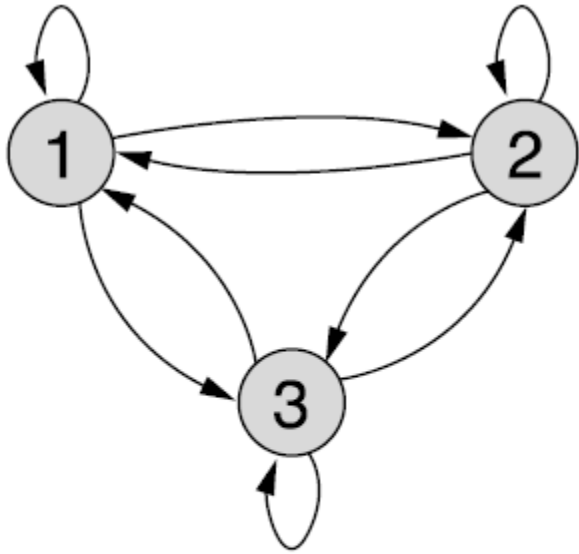
$$P(q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \dots) = P(q_t = S_j | q_{t-1} = S_i)$$

- 이것을 “1차 마르코프-과정(Markov Process)”이라고 한다. 고차 마르코프 과정은 더 큰 지연 ($t-2, t-3, \dots$)에 따른 표현 방법이다.

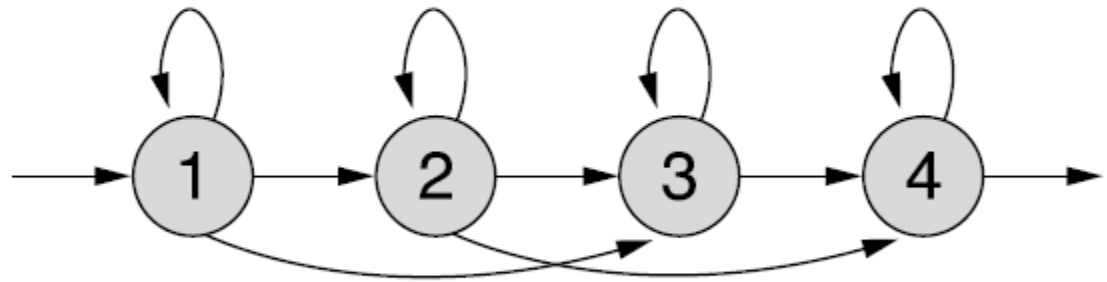


❖ 마르코프 모델 (2/2)

- 마르코프 모델은 상태 천이의 형태에 따라 날씨 HMM모델과 같은 에르고딕 (ergodic)모델과 음성인식에 많은 적용되는 좌우(left-right)모델이 있다.



(a) 에르고딕(ergodic) 모델



(b) 좌우(left-right) 모델

[그림 17-3] 마르코프 모델의 종류



❖ 날씨 HMM 모델 (1/4)

- 마르코프 모델에서 설명한 날씨 모델을 은닉 마르코프 모델로 바꾸어 날씨를 숨겨 보자. 그런데 날씨를 숨긴다는 가정이 어떤 상황이 될까?
- 여러분이 어느 외딴 집에 갇혀 있다고 가정해보자. 날씨를 아는 방법은 매일 밥을 가져다 주는 감시자가 우산을 가지고 있는지(☂) 안 가지고(☀) 있는지를 확인하는 것이 오늘 날씨를 알 수 있는 유일한 실마리가 된다고 하자.
- 방에 갇혀 있으므로 날씨를 직접 관측 할 수 없는 것이다. 아래 표와 같은 우산 확률을 가정 할 수 있을 것이다. 날씨가 맑으면 감시자가 우산을 가지고 있을 확률은 0.1 비가 오면 그 확률이 0.8, 흐린 날씨이면 그 확률이 0.3이 된다는 의미이다.

[표 17-2] 날씨에 따라 우산을 가지고 있을 확률

| 날씨 | 우산 확률 |
|-------|-------|
| 맑음(☀) | 0.1 |
| 비(☂) | 0.8 |
| 흐림(☁) | 0.3 |



❖ 날씨 HMM 모델 (2/4)

날씨 마르코프 과정에서는 날씨의 관측이 가능하므로, 어떤 날씨 열 $\{q_1, q_2, \dots, q_n\}$ 이 관측될 확률은 (17.15)와 같이 결합 확률(joint probability)로 바로 표현할 수 있었다.

$$P(q_1, q_2, \dots, q_n) = \prod_{i=1}^n P(q_i | q_{i-1}) \tag{17.15}$$

그러나 이제 실제 날씨는 직접 관측이 불가능하며 은닉(hidden)되어 있다. 그러므로 어떤 날씨 $q_n \in \{\text{맑음}(\text{☀}), \text{비}(\text{☁}), \text{흐림}(\text{☁})\}$ 이 관측될 확률은 감시자가 i 번째 날에 우산을 가져 왔다면 $o_i = \text{☂}$ 로, 우산을 가져오지 않았다면 $o_i = \text{☂}$ 가 되어 우산 열 o_i 의 관측을 통해서만 추정할 수 있을 뿐이다. 이것은 조건부 확률 $P(q_i | o_i)$ 에 해당하며 베이즈의 정리에 따라 다음과 같이 표현된다.

$$P(q_i | o_i) = \frac{P(o_i | q_i)P(q_i)}{P(o_i)} \tag{17.16}$$



❖ 날씨 HMM 모델 (3/4)

만약 확률 과정에서 날씨열($Q = \{q_1, q_2, \dots, q_n\}$)과 우산열($O = \{o_1, o_2, \dots, o_n\}$)이 주어진다
면, 관측되는 우산열에 따른 날씨열이 일어날 조건부 확률은 (17.17)과 같다.

$$P(q_1, \dots, q_n | o_1, \dots, o_n) = \frac{P(o_1, \dots, o_n | q_1, \dots, q_n)P(q_1, \dots, q_n)}{P(o_1, \dots, o_n)} \quad (17.17)$$

(17.17)에서 날씨열의 확률($P(q_1, \dots, q_n)$)과 우산열의 확률($P(o_1, \dots, o_n)$)이 주어진다
면, $P(o_1, \dots, o_n | q_1, \dots, q_n)$ 는 모든 i 에 대하여 독립이므로 $\prod_{i=1}^n P(o_i | q_i)$ 가 성립하게 된다.

감시자가 우산을 가지고 있는지(☂), 안 가지고 있는지(☀)의 우산열에 대한 관측 확률에서 바
깡 날씨가 (맑음(☀), 비(☂), 흐림(☁)) 중의 하나가 될 날씨열에 대한 확률을 구하기 위해서



❖ 날씨 HMM 모델 (4/4)

날씨에 독립인 우산열이 발생할 사전 확률 $P(o_1, \dots, o_n)$ 은 동일하므로 제거될 수 있다. 그러므로 실제 확률에 비례하는 우도(likelihood) 확률(L)은 (17.18)과 같이 표현된다.

$$\begin{aligned} P(q_1, \dots, q_n | o_1, \dots, o_n) &\propto \\ L(q_1, \dots, q_n | o_1, \dots, o_n) &= P(o_1, \dots, o_n | q_1, \dots, q_n)P(q_1, \dots, q_n) \end{aligned} \quad (17.18)$$

만약 1차 마르코프 가정일 경우에 (17.18)은 (17.19)와 같이 간략화된다.

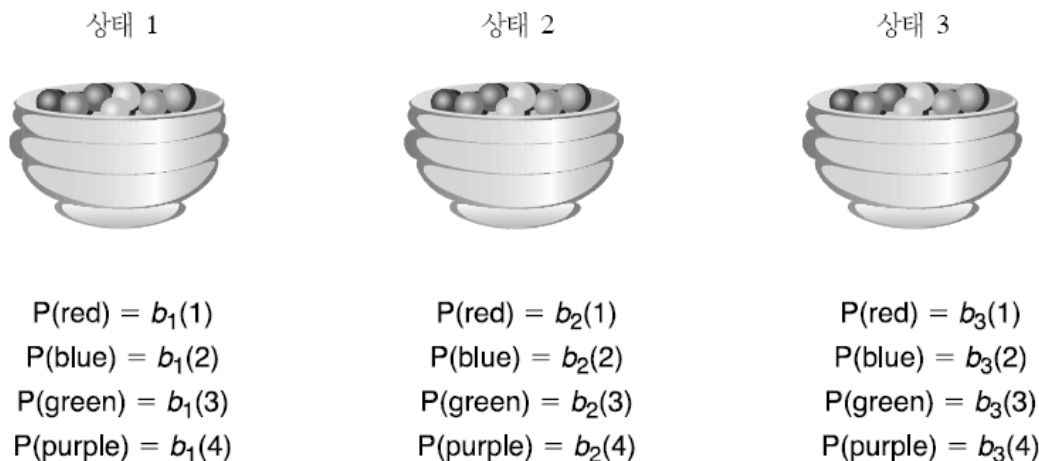
$$P(q_1, \dots, q_n | o_1, \dots, o_n) \propto L(q_1, \dots, q_n | o_1, \dots, o_n) = \prod_{i=1}^n P(o_i | q_i) \prod_{i=1}^n P(q_i | q_{i-1}) \quad (17.19)$$



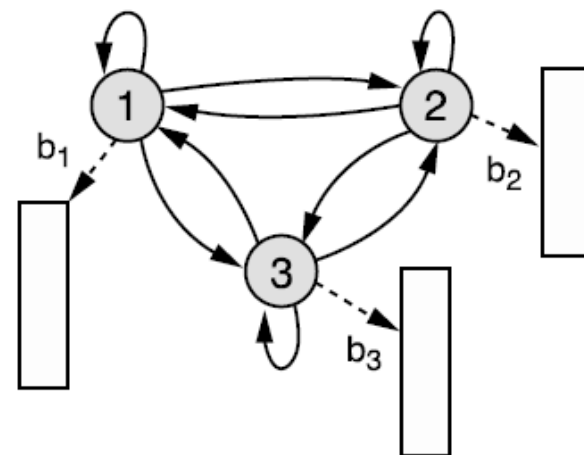
03_ 은닉 마르코프 모델

❖ 커튼 뒤에 숨어진 항아리 속의 공 HMM 모델(1/2)

어떤 방에 커튼이 드리워져 있고 커튼 뒤에는 어떤 사람과 네 가지 색깔(red, blue, green, purple)의 공이 들어 있는 3개의 항아리가 있다. 커튼 뒤의 사람이 어떤 확률 과정에 따라서 항아리를 선택하고 항아리 속에서 공을 꺼낸다. 그리고 여러분에게 공을 보여주고 항아리 속으로 그 공을 도로 집어넣는 과정을 반복한다.



(a) 항아리 속 공 HMM 모델의 출력 확률



(b) 항아리 속 공 HMM 모델의 상태 천이도



❖ 커튼 뒤에 숨어진 항아리 속의 공 HMM 모델(2/2)

- 이 과정에서 은닉되어 있는 것은 무엇일까?
- 이 과정에서 관측 가능한 상태는 무엇일까?
- 이 모델의 경우, HMM은 다음과 같은 구성 성분들로 모델링 된다.

① N : 모델의 상태 수(항아리의 수): $S = \{S_1, S_2, \dots, S_N\}$

② M : 이산적인 관측 심벌 개수(항아리 속에 있는 공의 색깔): $O = \{o_1, o_2, \dots, o_M\}$

③ 상태 천이 확률 행렬 $A = a_{ij}$

※ $a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$, $1 \leq i, j < N$: 상태 i 에서 상태 j 로 천이 확률

④ 관측 심벌 확률 분포 $B = b_j(k)$,

※ $b_j(k) = P(O_t = o_k | q_t = S_j)$, $1 \leq k < M$: 상태 j 에서 k 가 관측될 확률

⑤ 초기 상태 분포 $\pi = \pi_i$,

※ $\pi_i = P(q_1 = S_i)$, $1 \leq i < N$



❖ 커튼 뒤에 숨어진 항아리 속의 공 HMM 모델(2/2)

결국 2개의 스칼라양(N, M)과 3개의 확률 파라미터로 모델링된다.

$$\lambda = (A, B, \pi) \quad (17.23)$$

한편 HMM은 발생 모델과 인식 모델로 이용할 수 있다.

발생 모델(generative model)은 HMM에서 관측열을 발생하는 모델로, (17.23)의 모든 파라미터들이 주어진다면(모델링되어 있다면) 이 모델에서 관측열 $O = \{o_1, o_2, o_3, \dots\}$ 을 다음과 같은 과정으로 생성할 수 있다.

단계 1: 초기 상태 분포 π 로 초기 상태 s_1 을 선택한다.

단계 2: $t=1$ 로 설정한다.

단계 3: 관측 분포 $b_j(k)$ 에 따르는 관측 O_t 을 생성한다.

단계 4: 그 상태에서 상태 천이 확률(a_{ij})에 따라 새로운 상태 s_{t+1} 로 옮긴다.

단계 5: $t=t+1$ 로 설정하고 $t \geq T$ 때까지 단계 3~5의 과정을 반복한다.



❖ 확률 평가(Probability Evaluation) 문제

관측열 $O = \{o_1, o_2, o_3, o_4, \dots\}$ 과 여러 개의 HMM 모델 $\lambda = (A, B, \pi)$ 이 주어져 있을 때, 이 관측열은 어떠한 모델에서 발생할 확률이 가장 큰가? 즉, 관측열에 대한 우도 $P(O|\lambda)$ 를 어떻게 효과적으로 계산할 것인가?

앞의 날씨 모델의 마지막 예제에서 이 문제에 관한 아주 간단한 예를 들었다. 날씨 모델에서는 관측값이 하나일 경우였다. 이제, 이를 확장하여 주어진 모델에서 관측열이 출력될 확률이 얼마인지를 효과적으로 계산할 수 있어야 한다. 그 확률값에 따라 최적의 모델을 다수의 모델로부터 선택(인식)할 수 있기 때문이다. 이를 ‘확률 평가(probability evaluation)의 문제’라고 하는데, 이 문제는 전향(forward)과 후향(backward) 알고리즘에 의하여 해결된다.



❖ 최적 상태열(Optimal State Sequence)를 찾는 문제

관측열 $O = \{o_1, o_2, o_3, o_4, \dots\}$ 과 모델 $\lambda = (A, B, \pi)$ 이 주어져 있을 때, 관측열을 가장 잘 설명하는 최적 상태열(optimal state sequence) $Q = \{q_1, q_2, q_3, q_4, \dots\}$ 을 어떻게 찾을 것인가?

이는 관측열을 생성한 정확한 상태열을 알아내는 것은 어렵겠지만, 가장 최적의 숨겨져 있는 상태열을 어떻게 찾아낼 것인지를 문제다. 이 문제는 비터비(Viterbi) 알고리즘을 이용해 해결할 수 있다.



❖ 파라미터 추정(Parameter Estimation)의 문제

우도 $P(O|\lambda)$ 를 최대화하는 모델 $\lambda=(A, B, \pi)$ 의 각 파라미터(Parameter)를 어떻게 추정(Estimation)하여 구할 것인가?

어떤 모델에서 발생한 관측열들의 집합이 주어질 경우, 이 관측열을 가장 잘 설명하는 모델의 파라미터들을 어떻게 최적화하여 구할 것인지 하는 문제이다. 모델에 대해서 관측열의 출력이 최대가 되도록 모델의 파라미터들을 조절하는 문제이다. 이러한 문제를 ‘학습의 문제’라고 하는데 대부분의 HMM 응용 분야에서 성능을 결정하는 중요한 사항이다. 이 문제는 바움-웰치(Baum-Welch) 재추정 알고리즘으로 해결된다.



❖ 확률 평가 문제와 해법(1/2)

관측열 $O = \{o_1, o_2, o_3, o_4, \dots\}$ 과 여러 개의 HMM 모델 $\lambda = (A, B, \pi)$ 이 주어져 있을 때, 이 관측열이 어떠한 모델에서 발생할 확률이 가장 큰가? 즉, 관측열에 대한 우도 $P(O|\lambda)$ 를 어떻게 효과적으로 계산할 것인가?

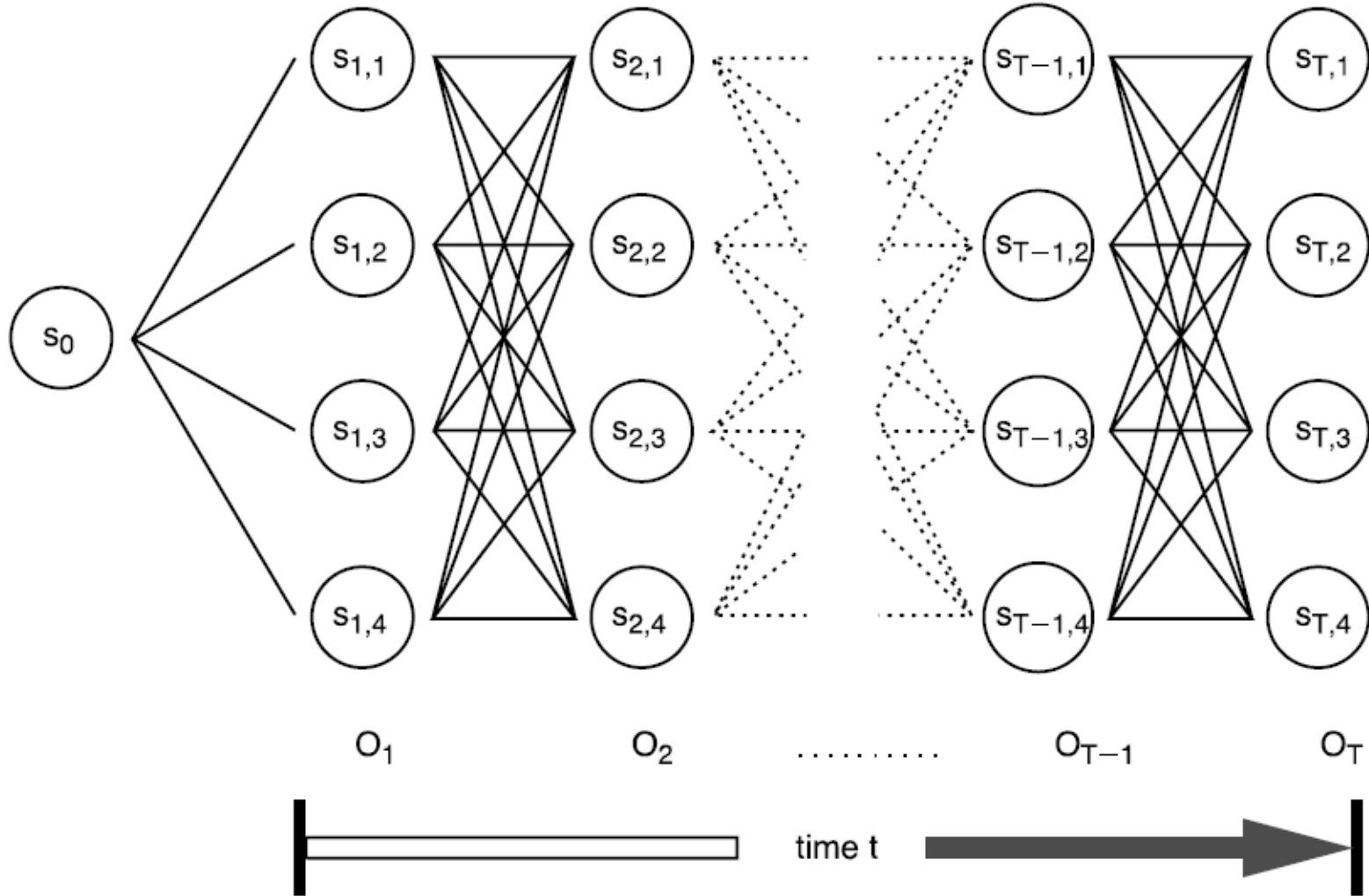
→ 전향(Forward)과 후향(Backward) 알고리즘

HMM으로 모델링된 각 모델에서 어떠한 관측열이 발생할 확률을 계산하기 위해서는 길이가 T 인 관측열의 모든 가능한 상태열로 구성된 [그림 17-5]와 같은 트렐리스(trellis) 상에서 각 확률을 계산하면 된다. 이때 모든 상태열의 가능한 가지 수는 N^T 가지가 된다.



05_ 확률 평가 문제와 해법

❖ 확률 평가 문제와 해법(1/2)



[그림 17-5] 관측열과 상태 트렐리스



❖ 확률 평가 문제와 해법(2/2)

파라미터들(λ)로 구성되어 HMM에서 발생하는 상태열($Q = q_1, q_2, \dots, q_N$)의 발생 확률은 한 상태에서 다음 상태까지 천이 확률의 곱과 같다.

$$P(Q | \lambda) = \pi_{q_1} \prod_{t=1}^{N-1} a_{q_t, q_{t+1}} = \pi_{q_1} \cdot a_{q_1, q_2} \cdot a_{q_2, q_3} \cdots a_{q_{N-1}, q_N} \quad (17.24)$$

그리고 주어진 상태열을 따라 관측되는 관측열의 확률 계산은 다음과 같다.

파라미터들(λ)로 구성되어 HMM에서 발생하여 관측되는 관측열이 $O = \{o_1, o_2, \dots, o_N\}$ 이고, 상태열이 $Q = \{q_1, q_2, \dots, q_N\}$ 이면, 트렐리스에서 상태열의 경로 Q 를 따라 발생하는 사건 관측열(O)의 발생 확률은 가능한 경로를 따라 계산된 출력 확률의 곱으로 다음과 같이 계산된다.

$$P(O | Q, \lambda) = \prod_{t=1}^N P(o_t | q_t, \lambda) = b_{q_1}(o_1) \cdot b_{q_2}(o_2) \cdots b_{q_N}(o_N) \quad (17.25)$$



❖ 확률 평가 문제와 해법(2/2)

그리고 관측열(O)과 어떤 경로상에서 상태열(Q)의 결합 확률은 O 와 Q 가 동시에 일어날 확률 $P(O, Q | \lambda)$ 로 표현되며, 앞에서 정의된 두 식의 곱으로 재구성된다.

$$\begin{aligned} P(O, Q | \lambda) &= P(O | Q, \lambda) \cdot P(Q | \lambda) \\ &= \pi_{q_1} \cdot b_{q_1}(o_1) \cdot a_{q_1, q_2} b_{q_2}(o_2) \cdots a_{q_{N-1}, q_N} b_{q_N}(o_N) \quad \text{※ Bayes 정리 (17.26)} \end{aligned}$$

또한 HMM에 따른 관측열($O = \{o_1, o_2, \dots, o_N\}$)의 확률은 파라미터들(λ)로 구성된 HMM이 허용하는 모든 가능한 상태열에서 열의 결합우도의 합으로 확장된다.

$$P(O | \lambda) = \sum_{all Q} P(O, Q | \lambda) \quad (17.27)$$



❖ 전향(forward) 과정 알고리즘

- 전향 변수 $\alpha_t(i)$ 를 정의

$$\alpha_t(i) = P(O, q_t = S_i | \lambda) = P(o_1, o_2, \dots, o_t, q_t = S_i | \lambda)$$

- 단계 1 : 초기화

$$\alpha_t(i) = \pi_i b_i(o_1) \quad * \quad 1 \leq i \leq N$$

- 2 단계 : 유도

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \quad * \quad \begin{cases} 1 \leq t \leq T-1 \\ 1 \leq j \leq N \end{cases}$$

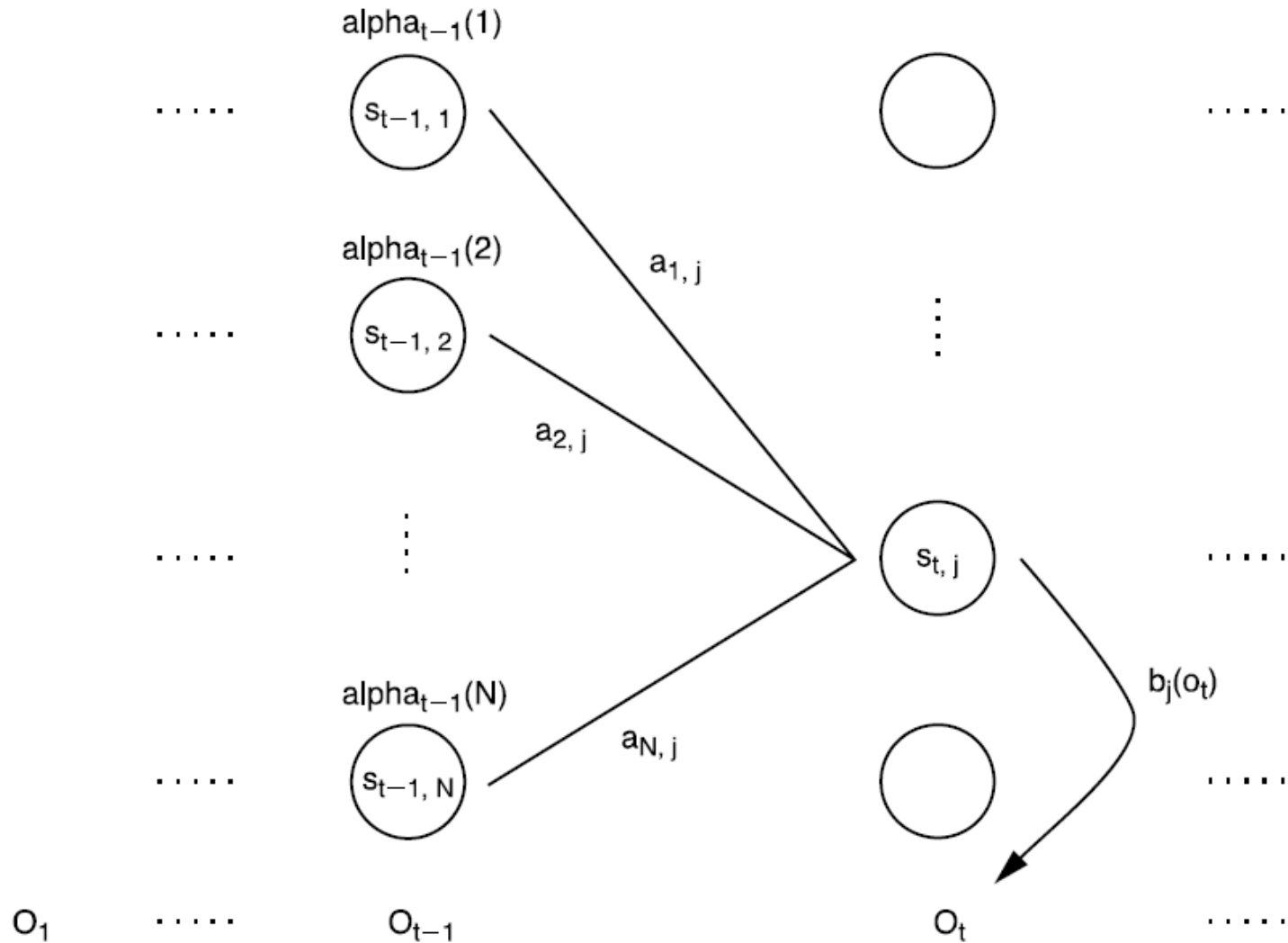
- 3 단계 : 종료

$$P(O | \lambda) = \sum_{i=1}^N \alpha_T(i)$$



05_ 확률 평가 문제와 해법

❖ 전향(forward) 과정 알고리즘



[그림 17-6] 전향 과정



❖ 후향(backward) 과정 알고리즘

- 전향 변수 $\beta_{t+1}(i)$ 를 정의

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | q_t = S_i, \lambda)$$

- 단계 1 : 초기화

$$\beta_T(i) = 1 \quad * \quad 1 \leq i \leq N$$

- 2 단계 : 유도

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \quad * \quad \begin{cases} t = T-1, T-2, \dots, 1 \\ 1 \leq i \leq N \end{cases}$$

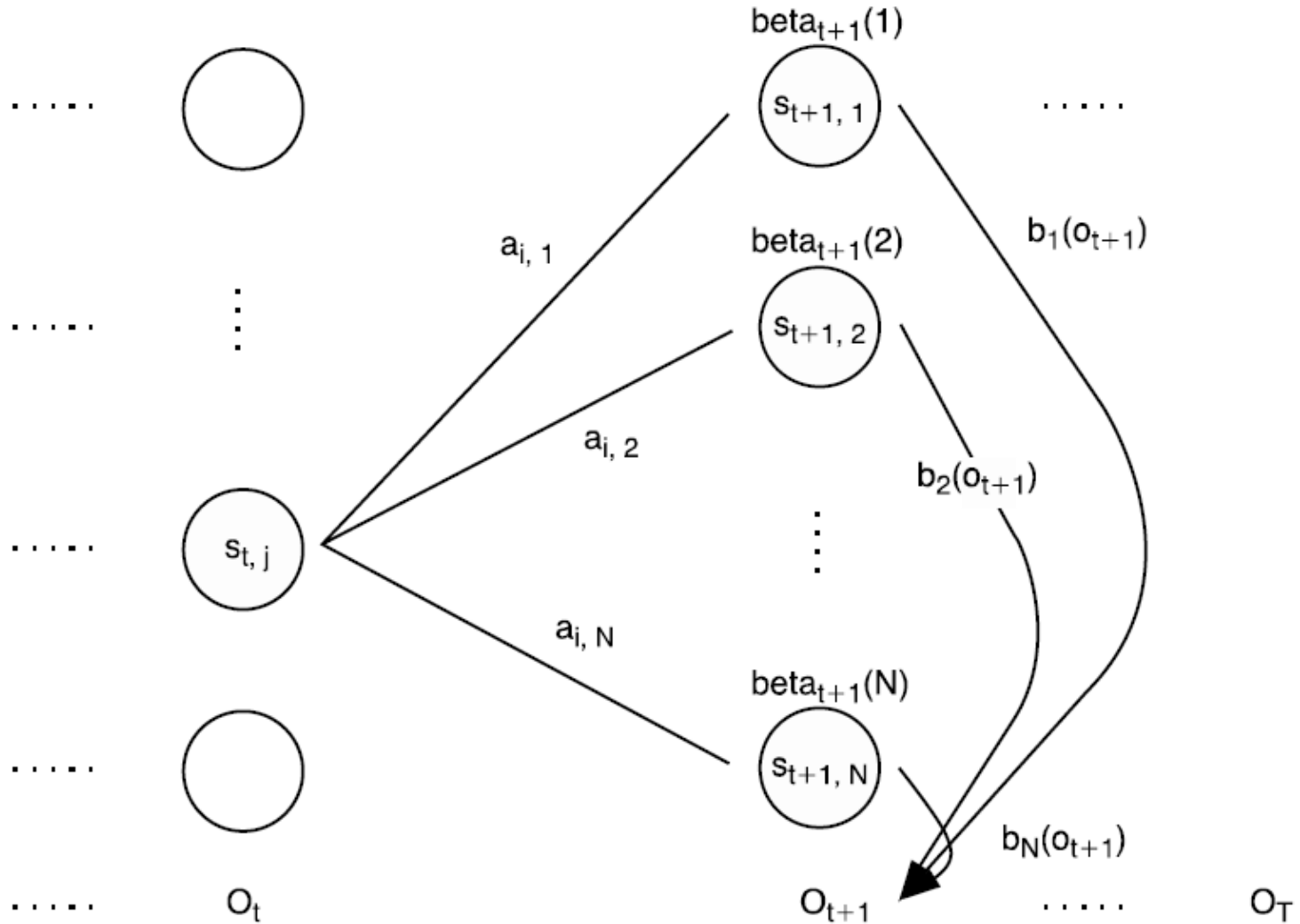
- 3 단계 : 종료

$$P(O | \lambda) = \sum_{i=1}^N \pi_i b_i(o_1) \beta_1(i)$$



05_ 확률 평가 문제와 해법

❖ 후향(backward) 과정 알고리즘



[그림 17-7] 후향 과정



05_ 확률 평가 문제와 해법

❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(1/6)

초기 확률 분포(π), 상태 천이 행렬(A)과 각 상태에서의 출력 확률(B)이 다음과 같은 동전 던지기 3상태 HMM 모델을 가정한다.

$$\pi = [1/3 \quad 1/3 \quad 1/3], \quad A = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{matrix} & \begin{matrix} H & T \end{matrix} \\ \begin{matrix} 1 \\ 1/2 \\ 1/3 \end{matrix} & \begin{bmatrix} 0 \\ 1/2 \\ 2/3 \end{bmatrix} \end{matrix}$$

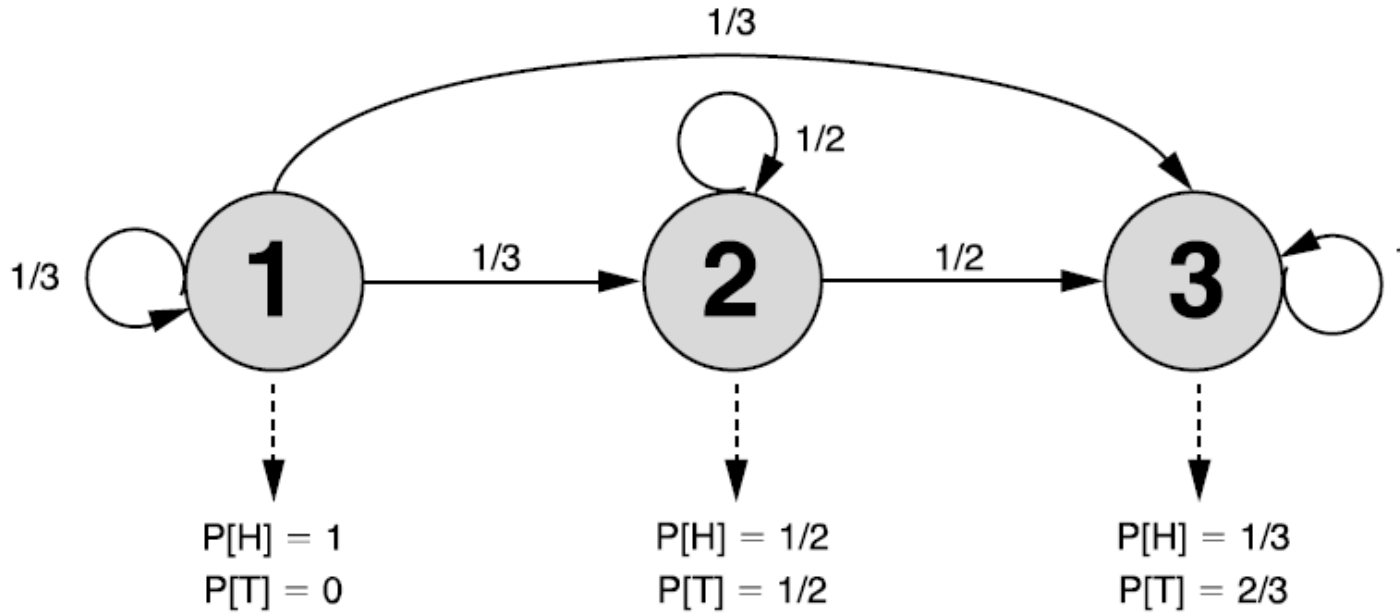
이 모델에서 관측열이 $O = \{T, H, T\}$ 와 같이 될 확률값은 얼마가 될까?



05_ 확률 평가 문제와 해법

❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(1/6)

- 동전 던지기 3상태 HMM 모델의 상태 천이도



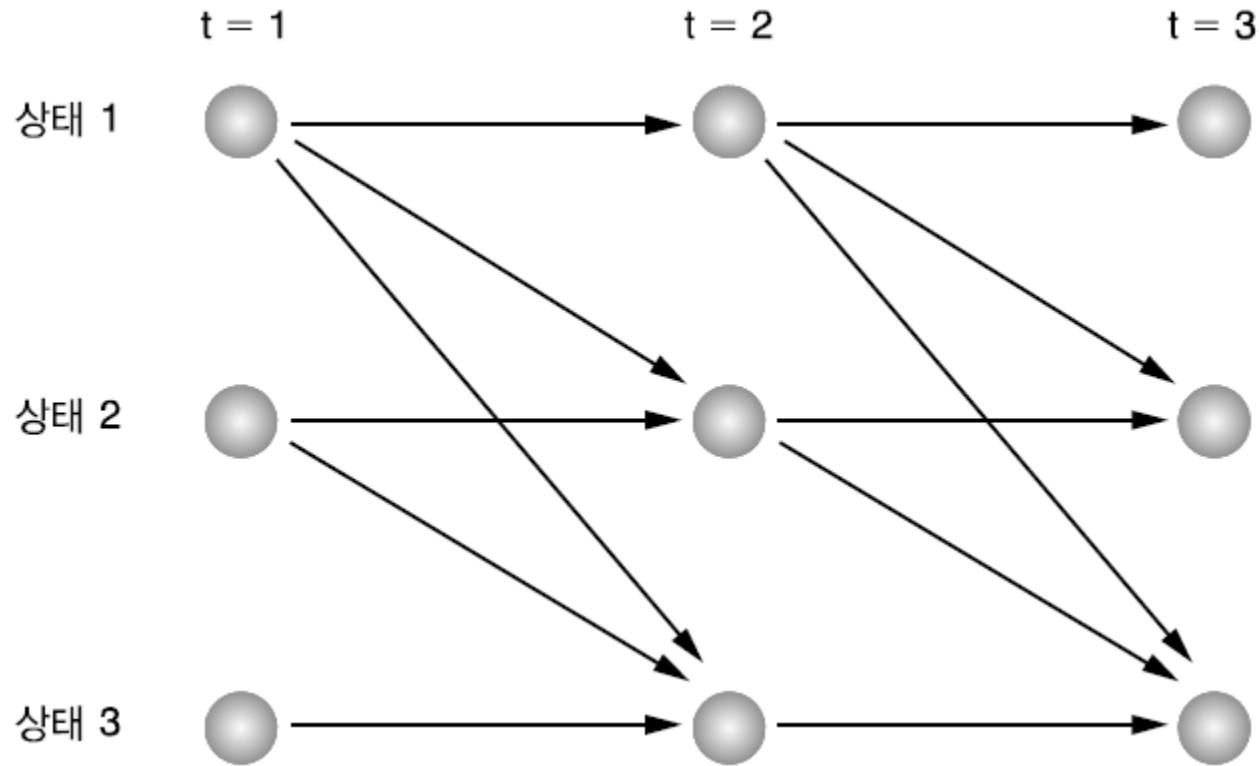
[그림 17-8] 상태 천이도



05_ 확률 평가 문제와 해법

❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(2/6)

- 시간에 따른 상태 트렐리스(trellis)도를 그려보면 다음과 같다.



[그림 17-9] 트렐리스도

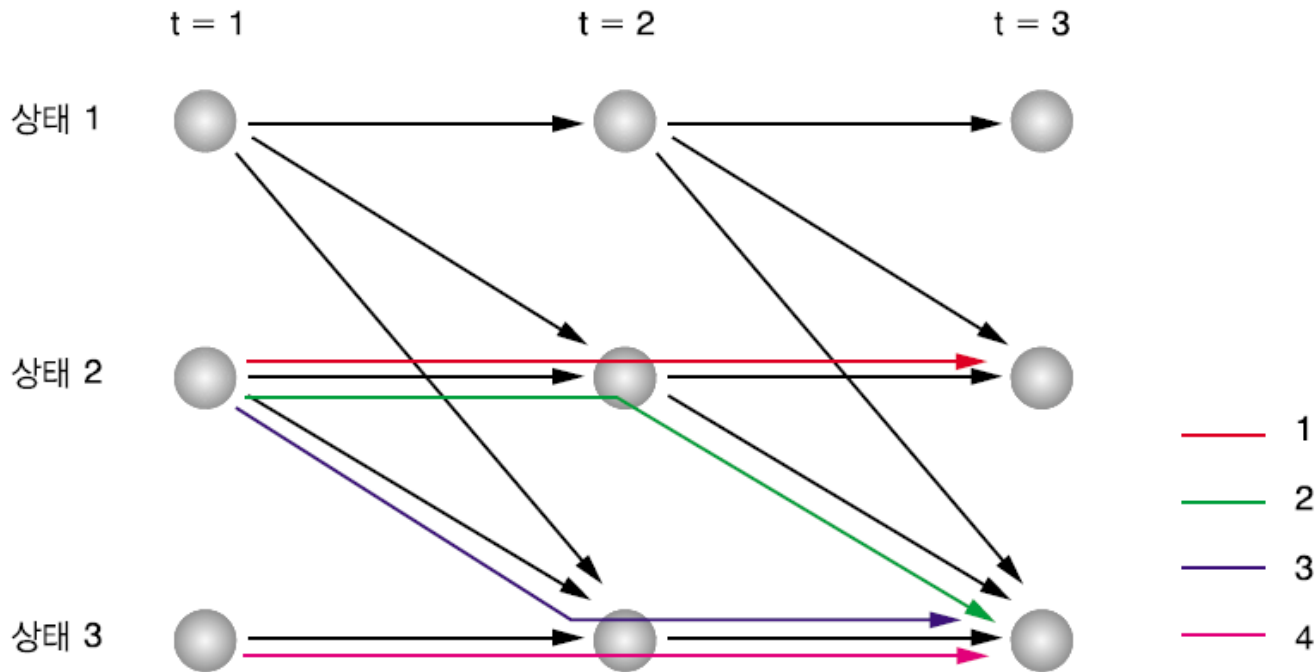


05_ 확률 평가 문제와 해법

❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(2/6)

- 전향, 후향 알고리즘을 이용하지 않고, 모든 가능한 상태열에 대해 확률을 계산하는 경우

- [그림 17-8]의 상태 천이도와 [그림 17-9]의 트렐리스도에서 q_1 에서 q_3 까지 천이하면서 출력심벌을 T-H-T로 발생 가능한 상태는 총 10가지지만 상태 1에서 시작하는 경우의 가능한 경로는 6가지다. 그런데 이 경우 T가 발생할 확률은 $P(T) = 0$ 이므로, [그림 17-10]과 같이 네 경우만 남게 된다.



[그림 17-10] 트렐리스 상에서 가능한 모든 경로



❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(2/6)

- 전향, 후향 알고리즘을 이용하지 않고, 모든 가능한 상태열에 대해 확률을 계산하는 경우
- [그림 17-8]의 상태 천이도와 [그림 17-9]의 트렐리스도에서 q_1 에서 q_3 까지 천이하면서 출력심벌을 T-H-T로 발생 가능한 상태는 총 10가지지만 상태 1에서 시작하는 경우의 가능한 경로는 6가지다. 그런데 이 경우 T가 발생할 확률은 $P(T) = 0$ 이므로, [그림 17-10]과 같이 네 경우만 남게 된다.



05_ 확률 평가 문제와 해법

❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(2/6)

$$P_1(T, H, T) = \pi_1 b_1(o_1 = T) a_{22} b_2(o_2 = H) a_{22} b_3(o_3 = T) = \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = 0.0104$$

$$P_1(THT) = \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{96} = 0.0104$$

$$P_2(THT) = \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{72} = 0.0139$$

$$P_3(THT) = \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot 1 \cdot \frac{2}{3} = \frac{1}{54} = 0.0185$$

$$P_4(THT) = \frac{1}{3} \cdot \frac{2}{3} \cdot 1 \cdot \frac{1}{3} \cdot 1 \cdot \frac{2}{3} = \frac{4}{81} = 0.0494$$

$$P(O) = P_1(T, H, T) + P_2(T, H, T) + P_3(T, H, T) + P_4(T, H, T) = 0.0922$$



05_ 확률 평가 문제와 해법

❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(3/6)

- 전향 알고리즘을 이용하는 경우
- 1단계 : 초기화

$$\alpha_1(i) = \pi_i b_i(o_1) \quad * 1 \leq i \leq 3$$

| | |
|-------------|--|
| $t = 0$ 일 때 | $\alpha_1(1) = \pi_1 b_1(o_1 = T) = 1/3 \cdot 0 = 0$ $\alpha_1(2) = \pi_2 b_2(o_1 = T) = 1/3 \cdot 1/2 = 1/6$ $\alpha_1(3) = \pi_3 b_3(o_1 = T) = 1/3 \cdot 2/3 = 2/9$ |
|-------------|--|



05_ 확률 평가 문제와 해법

❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(3/6)

- 전향 알고리즘을 이용하는 경우
- 2단계 : 유도

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^3 \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \quad * \begin{cases} 1 \leq t \leq 2 \\ 1 \leq j \leq 3 \end{cases}$$

| | | |
|-----|-----|---|
| t=1 | j=1 | $\alpha_2(1) = [\alpha_1(1) \cdot a_{11} + \alpha_1(2) \cdot a_{21} + \alpha_1(3) \cdot a_{31}] \cdot b_1(o_2 = H)$ $= 0$ |
| | j=2 | $\alpha_2(2) = [\alpha_1(1) \cdot a_{12} + \alpha_1(2) \cdot a_{22} + \alpha_1(3) \cdot a_{32}] \cdot b_2(o_2 = H)$ $= \left[\frac{1}{6} \cdot \frac{1}{2} \right] \cdot \frac{1}{2} = \frac{1}{24} = 0.0417$ |
| | j=3 | $\alpha_2(3) = [\alpha_1(1) \cdot a_{13} + \alpha_1(2) \cdot a_{23} + \alpha_1(3) \cdot a_{33}] \cdot b_3(o_2 = H)$ $= \left[\frac{1}{6} \cdot \frac{1}{2} + \frac{2}{9} \cdot 1 \right] \cdot \frac{1}{3} = \left[\frac{1}{12} + \frac{2}{9} \right] \cdot \frac{1}{3} = 0.1019$ |



05_ 확률 평가 문제와 해법

❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(4/6)

- 전향 알고리즘을 이용하는 경우
- 2단계 : 유도

| | | |
|-------|-------|---|
| $t=2$ | $j=1$ | $\alpha_3(1) = [\alpha_2(1) \cdot a_{11} + \alpha_2(2) \cdot \alpha_{21} + \alpha_2(3) \cdot a_{31}] \cdot b_1(o_3 = T) = 0$ |
| | $j=2$ | $\alpha_3(2) = [\alpha_2(1) \cdot a_{12} + \alpha_2(2) \cdot \alpha_{22} + \alpha_2(3) \cdot a_{32}] \cdot b_2(o_3 = T)$ $= \left[0.0417 \cdot \frac{1}{2} \right] \cdot \frac{1}{2} = 0.0104$ |
| | $j=3$ | $\alpha_3(3) = [\alpha_2(1) \cdot a_{13} + \alpha_2(2) \cdot \alpha_{23} + \alpha_2(3) \cdot a_{33}] \cdot b_3(o_3 = T)$ $= \left[0.0417 \cdot \frac{1}{2} + 0.1019 \cdot 1 \right] \cdot \frac{2}{3} = 0.0818$ |

- 3단계 : 종료

$$P(O | \lambda) = \sum_{i=1}^3 \alpha_3(i)$$

$$P(O | \lambda) = \alpha_3(1) + \alpha_3(2) + \alpha_3(3) = 0.0922$$



05_ 확률 평가 문제와 해법

❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(5/6)

- 후향 알고리즘을 이용하는 경우
- 1단계 : 초기화

$$\beta_T(i) = 1 \quad * \quad 1 \leq i \leq 3$$

| | |
|-----------|--|
| $t=3$ 일 때 | $\beta_3(1) = 1$ $\beta_3(2) = 1$ $\beta_3(3) = 1$ |
|-----------|--|

- 2단계 : 유도

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \quad * \quad \begin{cases} t = 2, 1 \\ 1 \leq i \leq 3 \end{cases}$$



05_ 확률 평가 문제와 해법

❖ 전향, 후향 알고리즘을 통한 확률 평가 비교(5/6)

▪ 2단계 : 유도

| | | |
|-----|-----|--|
| t=2 | j=1 | $\beta_2(1) = \sum_{j=1}^N a_{1j} b_j(o_3) \beta_3(j) = [a_{11} b_1(o_3 = T) \beta_3(1) + a_{12} b_2(o_3 = T) \beta_3(2) + a_{13} b_3(o_3 = T) \beta_3(3)]$ $= \left[\frac{1}{3} \cdot \frac{1}{2} \cdot 1 + \frac{1}{3} \cdot \frac{2}{3} \cdot 1 \right] = 0.3889$ |
| | j=2 | $\beta_2(2) = \sum_{j=1}^N a_{2j} b_j(o_3) \beta_3(j) = [a_{21} b_1(o_3 = T) \beta_3(1) + a_{22} b_2(o_3 = T) \beta_3(2) + a_{23} b_3(o_3 = T) \beta_3(3)]$ $= \left[\frac{1}{2} \cdot \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{2}{3} \cdot 1 \right] = 0.5833$ |
| | j=3 | $\beta_2(3) = \sum_{j=1}^N a_{3j} b_j(o_3) \beta_3(j) = [a_{31} b_1(o_3 = T) \beta_3(1) + a_{32} b_2(o_3 = T) \beta_3(2) + a_{33} b_3(o_3 = T) \beta_3(3)]$ $= 1 \cdot \frac{2}{3} \cdot 1 = 0.6667$ |
| t=1 | j=1 | $\beta_1(1) = \sum_{j=1}^N a_{1j} b_j(o_2) \beta_2(j) = [a_{11} b_1(o_2 = H) \beta_2(1) + a_{12} b_2(o_2 = H) \beta_2(2) + a_{13} b_3(o_2 = H) \beta_2(3)]$ $= \left[\frac{1}{3} \cdot 1 \cdot 0.3889 + \frac{1}{3} \cdot \frac{1}{2} \cdot 0.5833 + \frac{1}{3} \cdot \frac{1}{3} \cdot 0.6667 \right] = 0.3009$ |
| | j=2 | $\beta_1(2) = \sum_{j=1}^N a_{2j} b_j(o_2) \beta_2(j) = [a_{21} b_1(o_2 = H) \beta_2(1) + a_{22} b_2(o_2 = H) \beta_2(2) + a_{23} b_3(o_2 = H) \beta_2(3)]$ $= \left[\frac{1}{2} \cdot \frac{1}{2} \cdot 0.5833 + \frac{1}{2} \cdot \frac{1}{3} \cdot 0.6667 \right] = 0.2569$ |
| | j=3 | $\beta_1(3) = \sum_{j=1}^N a_{3j} b_j(o_2) \beta_2(j) = [a_{31} b_1(o_2 = H) \beta_2(1) + a_{32} b_2(o_2 = H) \beta_2(2) + a_{33} b_3(o_2 = H) \beta_2(3)]$ $= \left[1 \cdot \frac{1}{3} \cdot 0.6667 \right] = 0.2222$ |



❖ 최적 상태열 문제와 해법(1/6)

관측열 $O = \{o_1, o_2, o_3, o_4, \dots\}$ 과 모델 $\lambda = (A, B, \pi)$ 이 주어져 있을 때, 관측열을 가장 잘 설명하는 최적 상태열 $Q = \{q_1, q_2, q_3, q_4, \dots\}$ 을 어떻게 찾을 것인가?

→ 비터비 알고리즘

- 최적 상태열을 탐색하기 위해서는 최적화 기준(criteria) 설정이 가능
- 각 시간 t 에서 개별적으로 가장 그럴듯한 상태 q_t 를 찾는 방법
- 하나의 최적 상태 열의 경로를 찾는 방법 (예: 사후확률 $P(q|O, \lambda)$ 을 최대화하는 경로)
- 전향-후향 변수 $\gamma_t(i)$ 를 정의

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) = \frac{P(O, q_t = S_i | \lambda)}{P(O | \lambda)} = \frac{P(O, q_t = S_i | \lambda)}{\sum_{i=1}^N P(O, q_t = S_i | \lambda)}$$



❖ 최적 상태열 문제와 해법(2/6)

- $\gamma_t(i)$ 는 $\alpha_t(i)$ 와 $\beta_t(i)$ 의 곱과 같으므로,

$$\gamma_t(i) = \frac{P(O, q_t = S_i | \lambda)}{\sum_{i=1}^N P(O, q_t = S_i | \lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_t(i)\beta_t(i)}$$

- 각 시간에서는 최적 상태(q_t^*)는 각 시간에서 $\gamma_t(i)$ 가 최대가 되는 상태

$$q_t^* = \arg \max [\gamma_t(i)] \quad \text{※ } \forall t = 1 \dots T$$



❖ 최적 상태열 문제와 해법(3/6)

관측열 $O = \{o_1, o_2, o_3, o_4, \dots, o_T\}$ 가 주어져 있을 때, 이러한 관측열을 발생하는 단일한 최적 관측열 $Q = \{q_1, q_2, q_3, q_4, \dots, q_T\}$ 을 찾기 위해서 또 다른 변수 하나를 정의한다.

$$\delta_t(i) = \max_{q_1 q_2 \dots q_{t-1} q_t} P[q_1 q_2 \dots q_{t-1} q_t = S_i, o_1 o_2 o_3 \dots o_t | \lambda] \quad (17.49)$$

이 변수는 $\delta_t(i)$ 가 시간 t 에서 첫 번째 t 개의 관측과 상태 S_i 에서 끝나는 단일 패스에서 가장 확률이 큰 최상의 스코어를 의미한다.

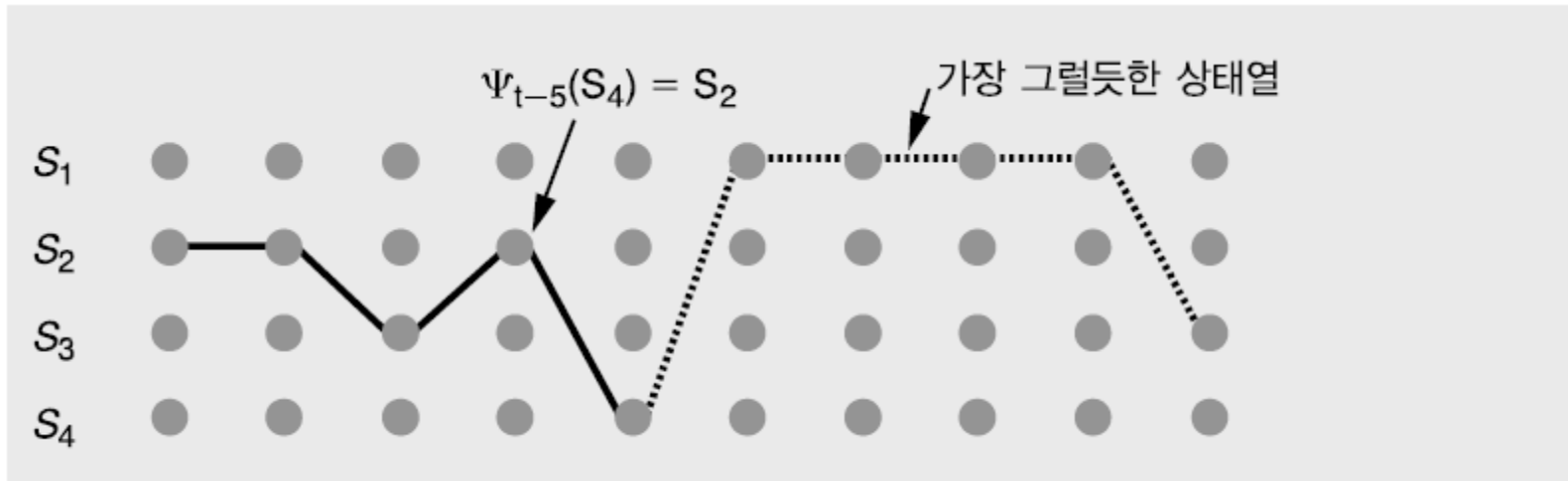
점화식으로 $\delta_{t+1}(j)$ 는 다음과 같이 표현되고 계산된다.

$$\delta_{t+1}(j) = \max_i [\delta_t(i) a_{ij}] b_j(o_{t+1}) \quad (17.50)$$



06_ 최적 상태열 문제와 해법

❖ 최적 상태열 문제와 해법(3/6)



[그림 17-11] 최적 상태열 찾기

이제, 각 시간 t 에서 j 에 대하여 $\delta_t(i)$ 를 최대화하는 상태의 트랙을 배열 $\Psi_t(j)$ 에 저장하고 역추적(백트래킹)을 통하여 최적 상태열을 탐색한다.



❖ 최적 상태열 문제와 해법(4/6) - 비터비 알고리즘

▪ 1단계 : 초기화

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

$$\psi_1(i) = 0 \quad * \text{이전 상태 없음}$$

▪ 2단계 : 반복

$$\left. \begin{aligned} \delta_t(j) &= \max_{1 \leq i \leq N} \left[\delta_{t-1}(i) a_{ij} \right] b_j(o_t) \\ \Psi_t(j) &= \arg \max_{1 \leq i \leq N} \left[\delta_{t-1}(i) a_{ij} \right] \end{aligned} \right\} * \begin{aligned} &2 \leq t \leq T \\ &1 \leq j \leq N \end{aligned}$$

▪ 3단계 : 종료

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

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

▪ 4단계 : 최적 상태열 역추적(백트래킹)

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



06_ 최적 상태열 문제와 해법

❖ 최적 상태열 문제와 해법(5/6) -비터비 알고리즘에 의한 최적 상태열 탐색

- 1단계 : 초기화 $\delta_1(i) = \pi_i b_i(o_1) \quad \ast 1 \leq i \leq N$
 $\Psi_1(i) = 0 \quad \ast$ 이전 상태 없음

$$\delta_1(1) = \pi_1 b_1(o_1 = T) = \frac{1}{3} \cdot 0 = 0$$

$$\delta_1(2) = \pi_2 b_2(o_1 = T) = \frac{1}{3} \cdot \frac{1}{2} = 0.1667$$

$$\delta_1(3) = \pi_3 b_3(o_1 = T) = \frac{1}{3} \cdot \frac{2}{3} = 0.2222$$

- 2단계 : 반복 $\left. \begin{array}{l} \delta_t(j) = \max_{1 \leq i \leq N} \left[\delta_{t-1}(i) a_{ij} \right] b_j(o_t) \\ \Psi_t(j) = \arg \max_{1 \leq i \leq N} \left[\delta_{t-1}(i) a_{ij} \right] \end{array} \right\} \ast \begin{array}{l} 2 \leq t \leq T \\ 1 \leq j \leq N \end{array}$



06_ 최적 상태열 문제와 해법

▪ 2단계 : 반복

| | | | |
|-----|-----|--|--------------------|
| t=2 | j=1 | $\delta_2(1) = \max [\delta_1(1) \cdot a_{11}, \delta_1(2) \cdot a_{12}, \delta_1(3) \cdot a_{13}] \cdot b_1(o_2 = H)$ $= \max \left[0 \cdot \frac{1}{3}, 0.1667 \cdot \frac{1}{3}, 0.2222 \cdot \frac{1}{3} \right] \cdot 1$ $= \max [0, 0.0556, 0.0741] \cdot 1 = 0.0741$ | $\Psi_2(1) = 3$ |
| | j=2 | $\delta_2(2) = \max [\delta_1(1) \cdot a_{21}, \delta_1(2) \cdot a_{22}, \delta_1(3) \cdot a_{23}] \cdot b_2(o_2 = H)$ $= \max \left[0 \cdot 0, 0.1667 \cdot \frac{1}{2}, 0.2222 \cdot \frac{1}{2} \right] \cdot \frac{1}{2}$ $= \max [0, 0.0833, 0.1111] \cdot \frac{1}{2} = 0.0556$ | $\Psi_2(2) = 3$ |
| | j=3 | $\delta_2(3) = \max [\delta_1(1) \cdot a_{31}, \delta_1(2) \cdot a_{32}, \delta_1(3) \cdot a_{33}] \cdot b_3(o_2 = H)$ $= \max [0 \cdot 0, 0.1667 \cdot 0, 0.2222 \cdot 1] \cdot \frac{1}{3}$ $= \max [0, 0, 0.2222] \cdot \frac{1}{3} = 0.0741$ | $\Psi_2(3) = 3$ |
| t=3 | j=1 | $\delta_3(1) = \max [\delta_2(1) \cdot a_{11}, \delta_2(2) \cdot a_{12}, \delta_2(3) \cdot a_{13}] \cdot b_1(o_3 = T)$ $= \max \left[0.1111 \cdot \frac{1}{3}, 0.0556 \cdot \frac{1}{3}, 0.0741 \cdot \frac{1}{3} \right] \cdot 0$ $= 0$ | $\Psi_3(1) = 1, 3$ |
| | j=2 | $\delta_3(2) = \max [\delta_2(1) \cdot a_{21}, \delta_2(2) \cdot a_{22}, \delta_2(3) \cdot a_{23}] \cdot b_2(o_3 = T)$ $= \max \left[0.0741 \cdot 0, 0.0556 \cdot \frac{1}{2}, 0.0741 \cdot \frac{1}{2} \right] \cdot \frac{1}{2}$ $= 0.0185$ | $\Psi_3(2) = 3$ |
| | j=3 | $\delta_3(3) = \max [\delta_2(1) \cdot a_{31}, \delta_2(2) \cdot a_{32}, \delta_2(3) \cdot a_{33}] \cdot b_3(o_3 = T)$ $= \max [0.0741 \cdot 0, 0.0556 \cdot 0, 0.0741 \cdot 1] \cdot \frac{2}{3}$ $= 0.0494$ | $\Psi_3(3) = 3$ |



❖ 최적 상태열 문제와 해법(6/6) -비터비 알고리즘에 의한 최적 상태열 탐색

- 3단계 : 종료

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

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

- 4단계 : 최적 상태열 역추적(백트래킹)

$$q_t^* = \Psi_{t+1}(q_{t+1}^*) \quad \text{※ } t = T-1, T-2, \dots, 1$$



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

우도 $P(O|\lambda)$ 를 최대화하는 모델 $\lambda=(A, B, \pi)$ 의 각 파라미터를 어떻게 추정하여 구할 것인가?

→ Baum-Welch 재추정 알고리즘

출력열이 주어질 때, 특정 시간, 특정 상태에 있을 확률은 전향-후향 알고리즘으로 다음과 같이 상태 q_i 에서 끝나는 관측값 $O(o_1, \dots, o_t)$ 의 확률(전향 확률)과 상태 S_i 에서 시작하는 관측값 $O=(o_{t+1}, \dots, o_T)$ 의 확률(후향 확률)의 곱으로 표현할 수 있다.

$$P(q_t = S_i | O) = P(o_1, \dots, o_t, q_t = S_i, o_{t+1}, \dots, o_T) = \alpha_t(i) \cdot \beta_t(i) \quad (17.62)$$

그리고 관측열이 주어질 경우에 시간 t 에서 상태 S_i 에 존재할 확률은 조건부 확률로 다음과 같이 표현된다.

$$P(q_t = S_i | O) = \frac{\alpha_t(i) \cdot \beta_t(i)}{P(O | \lambda)} \quad (17.63)$$



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

분자는 전향 확률과 후향 확률의 곱으로 계산될 수 있지만, 분모는 같은 결과가 나오는 여러 가지 다른 방법으로 계산될 수 있다. 예를 들어, 최종 시간에서 모든 상태의 전향 확률을 합해도 되고($\sum_i \alpha_T(i)$) 혹은 시간 0에서 후향 확률을 모든 상태에서 합해도 된다($\sum_i \beta_0(i)$). 또는 특정 시간(t)의 모든 상태에서 $\alpha_t(i)\beta_t(i)$ 를 합해도 같은 값을 얻을 수 있다. 이 중에서 전향 확률을 사용한다면 (17.63)은 다음과 같이 표현된다.

$$\gamma_t(i) = P(q_t = S_i | O) = \frac{\alpha_t(i) \cdot \beta_t(i)}{\sum_i \alpha_T(i)} \quad (17.64)$$

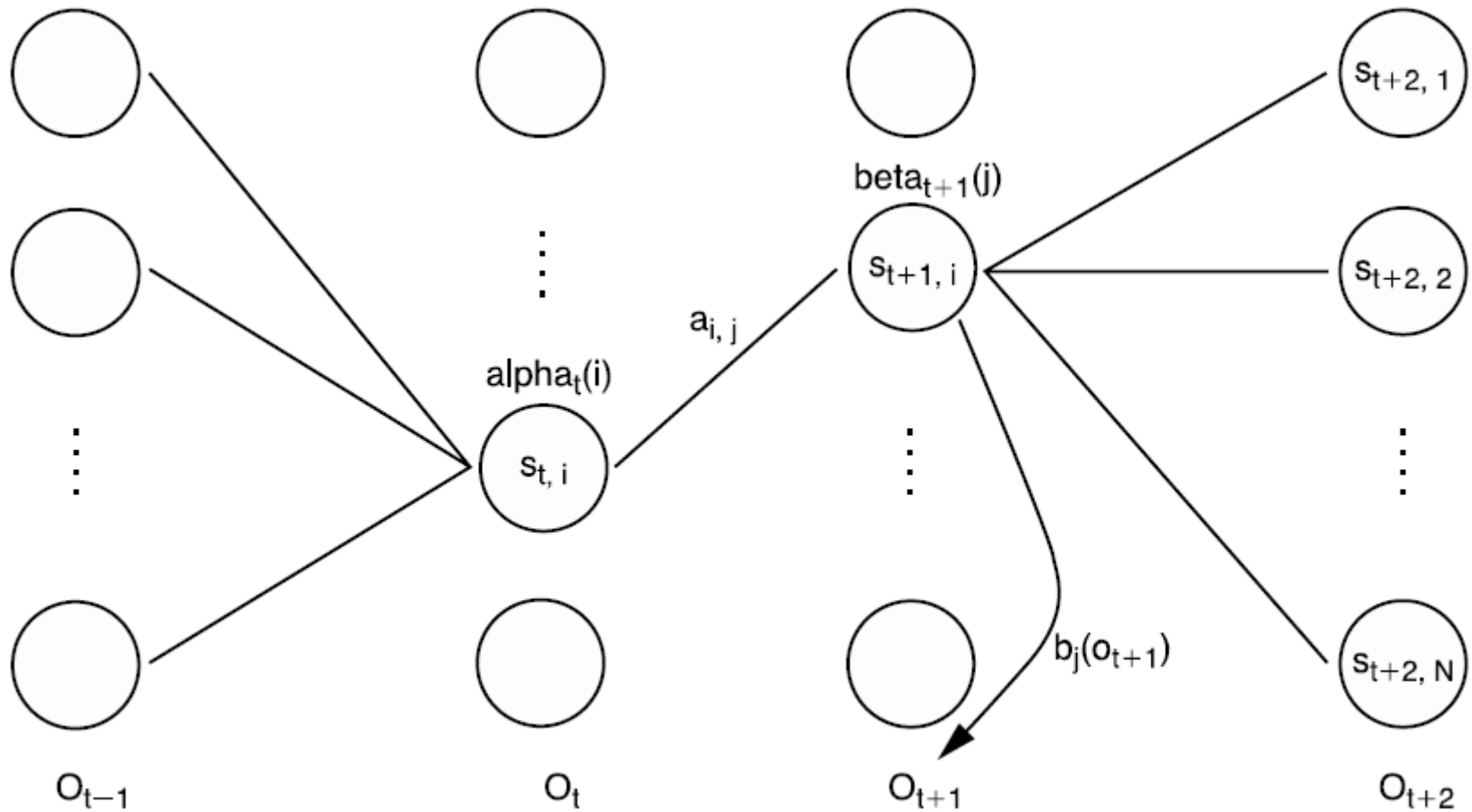
이제 천이 확률의 추정을 위해서 위의 수식을 약간 변형한다. 관측열 $O = \{o_1, o_2, \dots, o_T\}$ 가 주어질 때, S_i 에서 S_j 로 천이할 확률(a_{ij})은 다음과 같이 표현된다. 여기서 $\xi_t(i, j)$ 라는 변수 하나를 정의한다. 이 변수는 시간 t 에서는 상태 S_i 에 머물고, 시간 $t+1$ 에서는 상태 S_j 에 있을 확률을 나타낸다.

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \quad (17.65)$$



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘



[그림 17-12] 전후향 알고리즘



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

전향 변수 $\alpha_t(i)$ 와 후향 변수 $\beta_t(i)$ 그리고 조건부 확률 정의에서 (17.65)를 다시 쓰면,

$$\xi_t(i, j) = \frac{P(q_t = S_i, q_{t+1} = S_j, O | \lambda)}{P(O | \lambda)}$$

$$\frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{P(O | \lambda)} \tag{17.66}$$

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

비터비 알고리즘에서 변수 $\gamma_t(i)$ 는 모델 λ 와 관측열 O 가 주어져 있을 때, 시간 t 에서 상태 S_i 에 있을 확률로 정의하였다. $\xi_t(i, j)$ 를 $\gamma_t(i)$ 와 관련하여 다시 쓰면 다음과 같다.

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \tag{17.67}$$



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

시간상에서 $\gamma_t(i)$ 의 합은 상태 S_i 에 올 확률로 해석할 수 있다. S_i 에서 시작하는 천이 수는 다음과 같다.

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{“상태 } S_i \text{에서 시작하는 천이 확률(기대값)”} \quad (17.68)$$

유사하게, 시간 $t=1$ 부터 $t=T-1$ 까지 $\xi_t(i, j)$ 의 합은 상태 S_i 에서 상태 S_j 까지 천이 확률로 해석할 수 있다.

$$\sum_{t=1}^{T-1} \xi_t(i, j) = \text{“상태 } S_i \text{에서 } S_j \text{로의 천이 확률(기대값)”} \quad (17.69)$$

추정 파라미터 변수는 다음과 같이 정의된다.

$$\bar{\pi}_i = \gamma_1(i) = \text{“시간 } t=1 \text{에서 상태 } i \text{에 있을 확률”}$$



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

상태 S_i 에서 상태 S_j 로 천이하는 수의 기대값을 상태 S_i 에서 천이될 수 있는 수의 기대값으로 나눈 값은 다음과 같다.

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{17.70}$$

상태 S_j 에 존재하면서 관측 심볼 v_k 가 발생할 횟수의 기대값을 상태 S_j 에 존재하는 횟수의 기대값으로 나눈 값은 다음과 같다.

$$\bar{b}_j(k) = \frac{\sum_{t=1, o_t=v_k}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \tag{17.71}$$



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

- 단계 1: 초기화 $\lambda = (A, B, \pi)$
- 단계 2: $\lambda = \lambda'$ 가 될 때까지 반복(EM 알고리즘)
 - 최대화(Maximization) 단계

① 전향 알고리즘으로 전향 변수 계산

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = S_i | \lambda) \quad \ast \text{ for } \forall t, i \quad (17.72)$$

② 후향 알고리즘으로 후향 변수 계산

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | q_t = S_i, \lambda) \quad \ast \text{ for } \forall t, i \quad (17.73)$$

③ $1 \leq t \leq T$ 에 대하여 다음의 전향, 후향 변수를 계산한다.

$$\gamma_t(i, j) = P[o, q_{t-1} = S_i, q_t = S_j] = \alpha_{t-1} \alpha_{ij} b_{ij}(o_t) \beta_t(j) \quad (17.74)$$



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

- 추정(Expectation) 단계
 - 초기 확률값을 추정한다.
 - 상태 S_i 에서 상태 S_j 천이 확률을 추정한다.

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (17.75)$$

- 상태 S_i 에서 상태 S_j 로 천이할 확률과 출력 k 를 추정한다.

$$\bar{b}_j(k) = \frac{\sum_{t=1, o_t=v_k}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (17.76)$$

- $\lambda' = (\pi', A', B')$ 라고 둔다.



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

+ 전향-후향 변수 $\gamma_t(i)$ 유도 과정

$$\begin{aligned}
& P\left[O, q_{t-1} = S_i, q_t = S_j\right] \\
&= P\left[o_1, \dots, o_t, q_{t-1} = S_i, q_t = S_j, o_{t+1}, \dots, o_T\right] \\
&= P\left[o_{t+1}, \dots, o_T \mid o_1, \dots, o_t, q_t = S_i, q_t = S_j\right] P\left[o_1, \dots, o_t, q_{t-1} = S_i, q_t = S_j\right] \\
&= P\left[o_{t+1}, \dots, o_T \mid q_t = S_i, q_t = S_j\right] P\left[o_1, \dots, o_t, q_{t-1} = S_i, q_t = S_j\right] \\
&= P\left[o_{t+1}, \dots, o_T \mid q_t = S_j\right] P\left[o_1, \dots, o_t, q_{t-1} = S_i, q_t = S_j\right] \\
&= P\left[o_{t+1}, \dots, o_T \mid q_t = S_j\right] P\left[o_t \mid o_1, \dots, o_{t-1}, q_{t-1} = S_i, q_t = S_j\right] P\left[o_1, \dots, o_{t-1}, q_{t-1} = S_j, q_t = S_j\right] \\
&= P\left[o_{t+1}, \dots, o_T \mid q_t = S_j\right] P\left[o_t \mid q_{t-1} = S_j, q_t = S_j\right] P\left[o_1, \dots, o_{t-1}, q_{t-1} = S_j, q_t = S_j\right] \\
&= P\left[o_{t+1}, \dots, o_T \mid q_t = S_j\right] P\left[o_t \mid q_{t-1} = S_j, q_t = S_j\right] \\
&\quad P\left[q_t = S_j \mid o_1, \dots, o_{t-1}, q_{t-1} = S_i, q_t = S_i\right] P\left[o_1, \dots, o_{t-1}, q_{t-1} = S_j\right] \\
&= P\left[o_{t+1}, \dots, o_T \mid q_t = S_j\right] P\left[o_t \mid q_{t-1} = S_i, q_t = S_j\right] \\
&\quad P\left[q_t = S_j \mid q_{t-1} = S_i, q_t = S_i\right] P\left[o_1, \dots, o_{t-1}, q_{t-1} = S_i\right] \\
&= \beta_t(j) b_{ij}(o_t) a_{ij} \alpha_{t-1}(i)
\end{aligned}$$



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

■ 활용 예

① 시간 t 에서 상태 S_i 에 존재할 확률

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)}$$

| | | |
|-------|-------|---|
| $t=1$ | $i=1$ | $\gamma_1(1) = \frac{\alpha_1(1)\beta_1(1)}{P(O \lambda)} = \frac{0 \cdot 0.3009}{0.0922} = 0$ |
| | $i=2$ | $\gamma_1(2) = \frac{\alpha_1(2)\beta_1(2)}{P(O \lambda)} = \frac{(1/6) \cdot 0.2569}{0.0922} = 0.4621$ |
| | $i=3$ | $\gamma_1(3) = \frac{\alpha_1(3)\beta_1(3)}{P(O \lambda)} = \frac{(2/9) \cdot 0.2222}{0.0922} = 0.5350$ |



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

| | | |
|-------|-------|--|
| $t=2$ | $i=1$ | $\gamma_2(1) = \frac{\alpha_2(1)\beta_2(1)}{P(O \lambda)} = \frac{0 \cdot 0.3889}{0.0922} = 0$ |
| | $i=2$ | $\gamma_2(2) = \frac{\alpha_2(2)\beta_2(2)}{P(O \lambda)} = \frac{0.0417 \cdot 0.5833}{0.0922} = 0.2638$ |
| | $i=3$ | $\gamma_2(3) = \frac{\alpha_2(3)\beta_2(3)}{P(O \lambda)} = \frac{0.1019 \cdot 0.6667}{0.0922} = 0.7368$ |

| | | |
|-------|-------|---|
| $t=3$ | $i=1$ | $\gamma_3(1) = \frac{\alpha_3(1)\beta_3(1)}{P(O \lambda)} = \frac{0 \cdot 1}{0.0922} = 0$ |
| | $i=2$ | $\gamma_3(2) = \frac{\alpha_3(2)\beta_3(2)}{P(O \lambda)} = \frac{0.0104 \cdot 1}{0.0922} = 0.1127$ |
| | $i=3$ | $\gamma_3(3) = \frac{\alpha_3(3)\beta_3(3)}{P(O \lambda)} = \frac{0.0818 \cdot 1}{0.0922} = 0.8872$ |

② 시간 t 에서 상태 q_i 에 시간 $t+1$ 에서 상태 q_{t+1} 에 존재할 확률

$$\xi_t(i, j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)}$$



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

| | | | |
|-----|-----|-----|--|
| t=1 | i=1 | j=1 | $\xi_{51}(1,1) = \frac{\alpha_1(1)a_{11}b_1(O_2)\beta_2(1)}{P(O \lambda)}$ $= \frac{0 \cdot 0.333 \cdot 1 \cdot 0.3889}{0.0922} = 0$ |
| | | j=2 | $\xi_{51}(1,2) = \frac{\alpha_1(1)a_{12}b_2(O_2)\beta_2(2)}{P(O \lambda)}$ $= \frac{0 \cdot 0.333 \cdot 0.5 \cdot 0.5833}{0.0922} = 0$ |
| | | j=3 | $\xi_{51}(1,3) = \frac{\alpha_1(1)a_{13}b_3(O_2)\beta_2(3)}{P(O \lambda)}$ $= \frac{0 \cdot 0.333 \cdot 0.333 \cdot 0.6667}{0.0922} = 0$ |

| | | | |
|-----|-----|-----|--|
| t=1 | i=2 | j=1 | $\xi_{51}(2,1) = \frac{\alpha_1(2)a_{21}b_1(O_2)\beta_2(1)}{P(O \lambda)}$ $= \frac{0.1666 \cdot 0.1 \cdot 0.3889}{0.0922} = 0$ |
| | | j=2 | $\xi_{51}(2,2) = \frac{\alpha_1(2)a_{22}b_2(O_2)\beta_2(2)}{P(O \lambda)}$ $= \frac{0.1666 \cdot 0.5 \cdot 0.5 \cdot 0.5833}{0.0922} = 0.2634$ |
| | | j=3 | $\xi_{51}(2,3) = \frac{\alpha_1(2)a_{23}b_3(O_2)\beta_2(3)}{P(O \lambda)}$ $= \frac{0.1666 \cdot 0.5 \cdot 0.333 \cdot 0.6667}{0.0922} = 0.2005$ |



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

| | | | |
|-----|-----|-----|--|
| t=1 | i=3 | j=1 | $\xi_1(3,1) = \frac{\alpha_1(3)a_{31}b_1(O_2)\beta_2(1)}{P(O \lambda)}$ $= \frac{0.2222 \cdot 0 \cdot 0.3889 \cdot 1}{0.0922} = 0$ |
| | | j=2 | $\xi_1(3,2) = \frac{\alpha_1(3)a_{32}b_2(O_2)\beta_2(2)}{P(O \lambda)}$ $= \frac{0.2222 \cdot 0 \cdot 0.5 \cdot 0.5833}{0.0922} = 0$ |
| | | j=3 | $\xi_1(3,3) = \frac{\alpha_1(3)a_{33}b_3(O_2)\beta_2(3)}{P(O \lambda)}$ $= \frac{0.2222 \cdot 1 \cdot 0.3333 \cdot 0.6667}{0.0922} = 0.5355$ |
| t=2 | i=1 | j=1 | $\xi_2(1,1) = \frac{\alpha_2(1)a_{11}b_1(O_3)\beta_3(1)}{P(O \lambda)}$ |
| | | j=2 | $\xi_2(1,2) = \frac{\alpha_2(1)a_{12}b_2(O_3)\beta_3(2)}{P(O \lambda)}$ |
| | | j=3 | $\xi_2(1,3) = \frac{\alpha_2(1)a_{13}b_3(O_3)\beta_3(3)}{P(O \lambda)}$ |



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

| | | | |
|-------|-------|-------|---|
| $t=2$ | $i=2$ | $j=1$ | $\xi_2(2,1) = \frac{\alpha_2(2)a_{21}b_1(O_3)\beta_3(1)}{P(O \lambda)}$ |
| | | $j=2$ | $\xi_2(2,2) = \frac{\alpha_2(2)a_{22}b_2(O_3)\beta_3(2)}{P(O \lambda)}$ |
| | | $j=3$ | $\xi_2(2,3) = \frac{\alpha_2(2)a_{23}b_3(O_3)\beta_3(3)}{P(O \lambda)}$ |

| | | | |
|-------|-------|-------|---|
| $t=2$ | $i=3$ | $j=1$ | $\xi_2(3,1) = \frac{\alpha_2(3)a_{31}b_1(O_3)\beta_3(1)}{P(O \lambda)}$ |
| | | $j=2$ | $\xi_2(3,2) = \frac{\alpha_2(3)a_{32}b_2(O_3)\beta_3(2)}{P(O \lambda)}$ |
| | | $j=3$ | $\xi_2(3,3) = \frac{\alpha_2(3)a_{33}b_3(O_3)\beta_3(3)}{P(O \lambda)}$ |



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

③ 시간 t 에서 최대 경로 확률

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_t} P(q_1 q_2 \dots q_t = i, O_1 O_2 \dots O_t | \lambda)$$

$$\bar{\pi}_i = \gamma_1(i) \quad \times 1 \leq i \leq N$$

$$i=1 \quad \bar{\pi}_1 = \gamma_1(1) = 0$$

$$i=2 \quad \bar{\pi}_2 = \gamma_1(2) = 0.4621$$

$$i=3 \quad \bar{\pi}_3 = \gamma_1(3) = 0.5350$$

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

| | | |
|-------|-------|---|
| $i=1$ | $j=1$ | $a_{11} = \frac{\sum_{t=1}^{T-1} \xi_t(1,1)}{\sum_{t=1}^{T-1} \gamma_t(1)} = \frac{\sum_{t=1}^{T-1} \xi_t(1,1)}{[\gamma_1(1) + \gamma_2(1)]}$ |
| | $j=2$ | $a_{12} = \frac{\sum_{t=1}^{T-1} \xi_t(1,2)}{\sum_{t=1}^{T-1} \gamma_t(1)} = \frac{\sum_{t=1}^{T-1} \xi_t(1,2)}{[\gamma_1(1) + \gamma_2(1)]}$ |
| | $j=3$ | $a_{13} = \frac{\sum_{t=1}^{T-1} \xi_t(1,3)}{\sum_{t=1}^{T-1} \gamma_t(1)} = \frac{\sum_{t=1}^{T-1} \xi_t(1,3)}{[\gamma_1(1) + \gamma_2(1)]}$ |



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

| | | |
|-------|-------|---|
| $i=2$ | $j=1$ | $a_{21} = \frac{\sum_{t=1}^2 \xi_t(2,1)}{\sum_{t=1}^2 \gamma_t(2)} = \frac{\sum_{t=1}^2 \xi_t(2,1)}{[\gamma_1(2) + \gamma_2(2)]}$ |
| | $j=2$ | $a_{22} = \frac{\sum_{t=1}^2 \xi_t(2,2)}{\sum_{t=1}^2 \gamma_t(2)} = \frac{\sum_{t=1}^2 \xi_t(2,2)}{[\gamma_1(2) + \gamma_2(2)]}$ |
| | $j=3$ | $a_{23} = \frac{\sum_{t=1}^2 \xi_t(2,3)}{\sum_{t=1}^2 \gamma_t(2)} = \frac{\sum_{t=1}^2 \xi_t(2,3)}{[\gamma_1(2) + \gamma_2(2)]}$ |

| | | |
|-------|-------|---|
| $i=3$ | $j=1$ | $a_{31} = \frac{\sum_{t=1}^2 \xi_t(3,1)}{\sum_{t=1}^2 \gamma_t(3)} = \frac{\sum_{t=1}^2 \xi_t(3,1)}{[\gamma_1(3) + \gamma_2(3)]}$ |
| | $j=2$ | $a_{32} = \frac{\sum_{t=1}^2 \xi_t(3,2)}{\sum_{t=1}^2 \gamma_t(3)} = \frac{\sum_{t=1}^2 \xi_t(3,2)}{[\gamma_1(3) + \gamma_2(3)]}$ |
| | $j=3$ | $a_{33} = \frac{\sum_{t=1}^2 \xi_t(3,3)}{\sum_{t=1}^2 \gamma_t(3)} = \frac{\sum_{t=1}^2 \xi_t(3,3)}{[\gamma_1(3) + \gamma_2(3)]}$ |



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

※ $t=1: v_1(T) = o_1$ $t=2: v_2(H) = o_2$ $t=3: v_3(T) = o_3$

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



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

| | | |
|-----|-----|---|
| j=1 | k=1 | $\bar{b}_1(1) = \frac{\sum_{t=1, o_t=u_1}^3 \gamma_t(1)}{\sum_{t=1}^3 \gamma_t(1)} = \frac{\gamma_1(1) + \gamma_3(1)}{\gamma_1(1) + \gamma_2(1) + \gamma_3(1)}$ |
| | k=2 | $\bar{b}_1(2) = \frac{\sum_{t=1, o_t=u_2}^3 \gamma_t(1)}{\sum_{t=1}^3 \gamma_t(1)} = \frac{\gamma_2(1)}{\gamma_1(1) + \gamma_2(1) + \gamma_3(1)}$ |
| | k=3 | $\bar{b}_1(3) = \frac{\sum_{t=1, o_t=u_3}^3 \gamma_t(1)}{\sum_{t=1}^3 \gamma_t(1)} = \frac{\gamma_1(1) + \gamma_3(1)}{\gamma_1(1) + \gamma_2(1) + \gamma_3(1)}$ |

| | | |
|-----|-----|---|
| j=2 | k=1 | $\bar{b}_2(1) = \frac{\sum_{t=1, o_t=u_1}^3 \gamma_t(2)}{\sum_{t=1}^3 \gamma_t(2)} = \frac{\gamma_1(2) + \gamma_3(2)}{\gamma_1(2) + \gamma_2(2) + \gamma_3(2)}$ |
| | k=2 | $\bar{b}_2(2) = \frac{\sum_{t=1, o_t=u_2}^3 \gamma_t(2)}{\sum_{t=1}^3 \gamma_t(2)} = \frac{\gamma_2(2)}{\gamma_1(2) + \gamma_2(2) + \gamma_3(2)}$ |
| | k=3 | $\bar{b}_2(3) = \frac{\sum_{t=1, o_t=u_3}^3 \gamma_t(2)}{\sum_{t=1}^3 \gamma_t(2)} = \frac{\gamma_1(2) + \gamma_3(2)}{\gamma_1(2) + \gamma_2(2) + \gamma_3(2)}$ |



07_ 파라미터 추정의 문제와 해법

❖ Baum-Welch (Baum-Welch) 재추정 알고리즘

| | | |
|-------|-------|---|
| $j=3$ | $k=1$ | $\bar{b}_3(1) = \frac{\sum_{t=1, o_t=v_1}^3 \gamma_t(3)}{\sum_{t=1}^3 \gamma_t(3)} = \frac{\gamma_1(3) + \gamma_3(3)}{\gamma_1(3) + \gamma_2(3) + \gamma_3(3)}$ |
| | $k=2$ | $\bar{b}_3(2) = \frac{\sum_{t=1, o_t=v_2}^3 \gamma_t(3)}{\sum_{t=1}^3 \gamma_t(3)} = \frac{\gamma_2(3)}{\gamma_1(3) + \gamma_2(3) + \gamma_3(3)}$ |
| | $k=3$ | $\bar{b}_3(3) = \frac{\sum_{t=1, o_t=v_3}^3 \gamma_t(3)}{\sum_{t=1}^3 \gamma_t(3)} = \frac{\gamma_1(3) + \gamma_3(3)}{\gamma_1(3) + \gamma_2(3) + \gamma_3(3)}$ |