



계층적 은닉 마르코프 모델: 분석 및 응용

샤이 파인

컴퓨터 과학 연구소, 히브리 대학교, 예루살렘 91904, 이스라엘

fshai@cs.huji.ac.il

요람 싱어

AT&T 연구소, 180 Park Avenue, Florham Park, NJ 07932

singer@research.att.com

나프탈리 티시비

컴퓨터 과학 연구소, 히브리 대학교, 예루살렘 91904, 이스라엘

tishby@cs.huji.ac.il

편집자:데이비드 하우슬러

추상적인. 우리는 Hierarchical Hidden Markov Models(HHMM)이라고 하는 널리 사용되는 숨겨진 Markov 모델의 재귀적 계층적 일반화를 소개, 분석 및 시연합니다. 우리의 모델은 특히 언어, 손글씨 및 음성에서 많은 자연적 순서로 나타나는 복잡한 다중 규모 구조에 의해 동기 부여되었습니다. 우리는 그러한 구조의 모델링에 대한 체계적인 자율적 접근 방식을 추구합니다. 표준 Baum-Welch(정방향-역방향) 알고리즘을 확장하여 레이블이 지정되지 않은 데이터에서 모델 매개변수를 추정하는 효율적인 절차를 도출합니다. 그런 다음 관찰 시퀀스의 자동 계층적 구문 분석을 위해 훈련된 모델을 사용합니다. 모델의 두 가지 응용 프로그램과 해당 매개 변수 추정 절차를 설명합니다. 첫 번째 애플리케이션에서는 자연스러운 영어 텍스트의 계층적 모델을 구성하는 방법을 보여줍니다. 이러한 모델에서 계층 구조의 다른 수준은 텍스트의 다른 길이 척도에 있는 구조에 해당합니다. 두 번째 애플리케이션에서는 HHMM을 사용하여 필기체 필기의 문자 조합을 나타내는 반복된 획을 자동으로 식별하는 방법을 보여줍니다.

키워드: 통계 모델, 시간 패턴 인식, 숨겨진 변수 모델, 필기체 필기

1. 소개

히든 마르코프 모델(HMM)은 음성 및 필기 인식(Rabiner & Juang, 1986, Nag et al., 1985) 및 전산 분자 생물학(Krogh et al., 1993, Baldi 등, 1994). 은닉 마르코프 모델은 자연어 모델링에도 사용됩니다(예를 들어(Jelinek, 1985) 참조). 대부분의 이러한 응용 프로그램에서 모델의 토폴로지는 미리 결정되고 모델 매개변수는 이 맥락에서 Baum-Welch(또는 정방향-역방향) 알고리즘으로 알려진 EM 절차(Dempster et al., 1977)에 의해 추정됩니다(Baum & 페트리, 1966). 일부 최근 작업은 모델 구조의 추론도 탐구했습니다(Stolcke & Omohundro, 1994). 그러나 위의 대부분의 응용 프로그램에서 길이 척도의 다양성과 시퀀스의 재귀적 특성으로 인해 어려움이 있습니다. 이러한 어려움 중 일부는 확률적 문맥 자유 문법(SCFG)을 사용하여 극복할 수 있습니다. 일반적으로 SCFG에 의해 유도된 관찰된 시퀀스의 가능성은 모델 매개변수의 작은 변화에 따라 크게 달라지기 때문에 확률적 문법의 매개변수는 추정하기 어렵습니다. 또한 inside-outside 알고리즘(Lari & Young, 1990)이라고 하는 SCFG의 매개변수 추정을 위한 일반적인 알고리즘은 관찰된 시퀀스의 길이에 3차 시간 복잡도가 있습니다. 일반적으로 SCFG에 의해 유도된 관찰된 시퀀스의 가능성은 모델 매개변수의 작은 변화에 따라 크게 달라지기 때문에 확률적 문법의 매개변수는 추정하기 어렵습니다. 또한 inside-outside 알고리즘(Lari & Young, 1990)이라고 하는 SCFG의 매개변수 추정을 위한 일반적인 알고리즘은 관찰된 시퀀스의 길이에 3차 시간 복잡도가 있습니다. 일반적으로 SCFG에 의해 유도된 관찰된 시퀀스의 가능성은 모델 매개변수의 작은 변화에 따라 크게 달라지기 때문에 확률적 문법의 매개변수는 추정하기 어렵습니다. 또한 inside-outside 알고리즘(Lari & Young, 1990)이라고 하는 SCFG의 매개변수 추정을 위한 일반적인 알고리즘은 관찰된 시퀀스의 길이에 3차 시간 복잡도가 있습니다.

이 논문에서 우리는 자연 시퀀스의 자기유사 및 계층적 구조, 즉 계층적 숨겨진 마르코프 모델을 활용하는 SCFG에 대한 더 간단한 대안을 제시합니다. 우리의 주된 동기는 음성, 손글씨 또는 텍스트와 같은 자연어에 존재하는 다양한 확률적 수준과 길이 척도를 더 잘 모델링할 수 있도록 하는 것입니다. 이러한 모델의 또 다른 중요한 속성은 계층의 상위 수준을 통해 관찰 시퀀스에서 장기간에 걸쳐 상관 관찰을 추론하는 기능입니다. 내부-외부 알고리즘에서 영감을 받은 추정 체계를 통해 모델 매개변수를 효율적으로 추정하는 방법을 보여줍니다. 우리가 제안하는 모델의 구조는 상당히 일반적이며 하위 모델의 활성화 횟수를 임의로 허용합니다. 이 추정 절차는 전체 계산 시간이 관찰 길이의 2차만 되도록 효율적으로 근사화할 수 있습니다. 따라서 실행 시간을 합리적으로 유지하면서 모델에서 오랜 시간 상관 관계를 캡처할 수 있습니다. 우리는 자연스러운 영어 텍스트의 다중 해상도 구조를 학습하여 모델의 적용 가능성과 추정 절차를 시연합니다. 결과 모델은 문장 부호, 빈번한 문자 조합 및 구문 어미와 같은 다양한 시간 척도의 "시간 전 문가" 형성을 보여줍니다. 우리는 또한 필기체 필기체의 문자 조합을 나타내는 반복된 획의 자율 학습을 위해 계층적 숨겨진 Markov 모델의 학습 알고리즘을 사용합니다.

이 논문은 다음과 같이 구성되어 있습니다. 섹션 2에서는 계층적 숨겨진 Markov 모델을 소개하고 설명합니다. 3절에서는 계층적 은닉 마르코프 모델의 매개변수에 대한 추정 절차를 도출합니다. 섹션 4에서는 모델과 추정 체계를 활용하는 두 가지 응용 프로그램을 설명하고 시연합니다. 마지막으로 섹션 5에서는 관련 작업에 대해 논의하고 모델의 몇 가지 가능한 일반화를 설명하고 결론을 내립니다. 프레젠테이션을 간단하게 유지하기 위해 대부분의 기술적 세부 사항은 기술 부록으로 미루었습니다. 논문에 사용된 기호와 변수에 대한 요약은 부록 C에 나와 있습니다.

2. 모델 설명

계층적 은닉 마르코프 모델(HHMM)은 구조화된 단단계 확률 프로세스입니다. HHMM은 각각의 숨겨진 상태를 자체적으로 "자율적인" 확률 모델로 만들어 표준 HMM을 일반화합니다. 즉, 각 상태도 HHMM입니다. 따라서 HHMM의 상태는 단일 기호가 아닌 시퀀스를 방출합니다. HHMM은 상태의 하위 상태 중 하나를 재귀적으로 활성화하여 시퀀스를 생성합니다. 이 하위 상태는 또한 하위 상태로 구성될 수 있으므로 하위 상태 중 하나 등을 활성화합니다. 이 재귀 활성화 프로세스는 생산 상태라고 하는 특수 상태에 도달하면 종료됩니다. 프로덕션 상태는 일반적인 HMM 상태 출력 메커니즘을 통해 실제로 출력 기호를 내보내는 유일한 상태입니다. 생산 상태에서 방출되는 출력 기호는 출력 기호 집합에 대한 확률 분포에 따라 선택됩니다. 관찰 가능한 기호를 직접 방출하지 않는 숨겨진 상태를 내부 상태라고 합니다. 내부 상태에 의한 하위 상태의 활성화를 수직 전이라고 합니다. 수직 전환(하위 수준 상태로의 추가 수직 전환을 포함할 수 있음)이 완료되면 제어가 재귀 활성화 체인을 시작한 상태로 돌아갑니다. 그런 다음 동일한 수준 내에서 수평 전환이라고 하는 상태 전환이 수행됩니다. 그만큼 수직 전환(하위 수준 상태로의 추가 수직 전환을 포함할 수 있음)이 완료되면 제어가 재귀 활성화 체인을 시작한 상태로 돌아갑니다. 그런 다음 동일한 수준 내에서 수평 전환이라고 하는 상태 전환이 수행됩니다. 그만큼

일련의 상태 및 수직 전환은 루트 상태가 계층 구조의 맨 위에 있는 노드이고 잎이 생성 상태인 트리 구조를 유도합니다. 표기법을 단순화하기 위해 전체 기본 트리 구조를 가진 HHMM으로 분석을 제한합니다. 즉, 모든 리프는 루트 상태에서 동일한 거리에 있습니다. 일반적인 구조를 가진 HHMM의 분석은 여기에 제시된 분석을 간단하게 일반화한 것입니다. 본 논문에서 기술된 실험은 일반 토폴로지로 수행되었다.

우리는 모든 HHMM이 표준 단일 레벨 HMM으로 표현될 수 있다는 점에 주목하고 싶습니다. HMM의 상태는 완전히 연결된 구조를 가진 해당 HHMM의 생산 상태입니다. 즉, 상태 중 하나에서 다른 상태로 이동할 확률이 0이 아닌 경우입니다. 그러나 동등한 HMM에는 섹션 4에서 설명한 애플리케이션에서 활용하는 다중 레벨 구조가 없습니다.

이제 HHMM에 대한 공식적인 설명을 제공합니다. \mathbf{r} Σt 는 유한 알파벳입니다. 우리는 Σ 에 의해 가능한 모든 문자열 집합 $\text{ov}\Sigma$ 어. 관찰 순서는 유한한 문자열입니다.

Σ로 표시 $영형 = 영형_1 영형_2 \dots 영형_T$, HHMM의 상태는 b로 표시됩니다. $큐와이 \langle C \rangle$ 내 $\langle C \rangle \in \{1, \dots, C\}$ 어디 L 상태 인덱스이고 계층 인덱스입니다. 루트의 계층 인덱스는 1생산 상태 C

이다. 내부 상태는 동일한 번호를 가질 필요가 없습니다.

하위 상태. 따라서 우리는 내부의 하위 상태 수를 나타냅니다.

문맥에서 명확할 때마다 상태 인덱스를 생략하고 상태를 표시합니다. C/t 수준
 ~에 의해 \mathcal{C}_t 모델 구조(토폴로지) 외에도 HHMM은 내부 상태와 출력 분포 벡터 간의 상태 전이 확률이 특징입니다.

생산 상태. 즉, 각 내부 s에 대해 $\sum_{i=1}^D x_{i,s}$ 가 $D-1$ 이하인 상태가 있습니다

b로 표시된 전이 확률 행렬 f 와 $i \in \mathcal{I}$ 에 대하여 $f_{ij} = P(X_{t+1} = j | X_t = i)$ 이다.

하위 상태 $\circ \pi_{C|D}$. 마찬가지로, $\Pi_{\pi_{C|D}} = \{\pi_{\pi_{C|D}}(\pi_{C|D+1})\} = \{ \Pi(\pi_{C|D+1}/q_C) \}$ 하위 상태에 대한 초기 분포 벡터입니다. $\pi_{C|D}$ 형 cf. 통계가 발생할 확률입니다. $\pi_{C|D}$ 자형 C 처음에 상태를 활성화합니다 $\pi_{C|D+1}$. 만약에 $\pi_{C|D+1}$ 차례로 내부 상태입니다. $\Pi/\Omega_N(C(\pi_{C|D+1}))$ 수직 전환을 할 확률로 해석될 수도 있습니다. $\pi_{C|D} + 1$, 테1부모로부터

상태 \mathbf{y}_t . 각 생산 통계 \mathbf{y}_t 이 자형 \mathbf{d}_t 출력 확률 벡터에 의해서만 매개변수화됩니다.

$B_{\text{큐디}} = \{B_{\text{큐디}}(\text{케이})\}$, 어디 $B_{\text{큐디}}(\text{케이}) = \Pi(\sigma_{\text{케이}}/q_{\text{디}})$ 생산 상태가 될 확률입니다. 큐디기호를 출력합니다 $\sigma_{\text{케이}} \in \Sigma$. 매개변수의 전체 세트는 다음과 같이 표시됩니다.

$$\lambda = \{\lambda_{\vec{c}}\}_{\vec{c} \in \{1, \dots, D\}^D} \in \{ \{ \vdash_{\vec{c}} \}_{\vec{c} \in \{1, \dots, D-1\}}, \{ \prod_{\vec{c}} \}_{\vec{c} \in \{1, \dots, D-1\}}, \{ B_{\vec{c}} \} \}.$$

임의의 토폴로지와 매개변수가 있는 HHMM의 그림이 그림 1에 나와 있습니다.

요약하면 문자열은 루트 상태에서 시작하여 무작위로 루트의 하위 상태 중 하나를 선택하여 생성됩니다. 유형 1, 마찬가지로 각 내부 통계에 대해 큰

입력, 하나의 오_ㄱ의 하위 상태는 다음에 따라 무작위로 선택됩니다. _ㄱ최초 확률 벡터 _ㄱ 하위 상태 중 하나를 재귀적으로 활성화하는 선택된 하위 상태로 작업이 진행됩니다. 이러한 재귀 작업은 프로덕션이 완료될 때까지 수행됩니다. _ㄱ

세인트 아이츠

상태 출력 확률 벡터에 따라 단일 기호가 생성되는 지점에 도달합니다. β/β_{c} . 그런 다음 제어가 활성 상태로 돌아갑니다. κ/κ_{c} 이후. 재귀 문자열 생성이 완료되면 재귀를 시작한 내부 상태는 레벨의 상태 전이 행렬에 따라 같은 레벨의 다음 상태를 선택하고 새로 선택된 상태는 새로운 재귀 문자열 생성 프로세스를 시작합니다. 각 레벨(상단 제외)에는

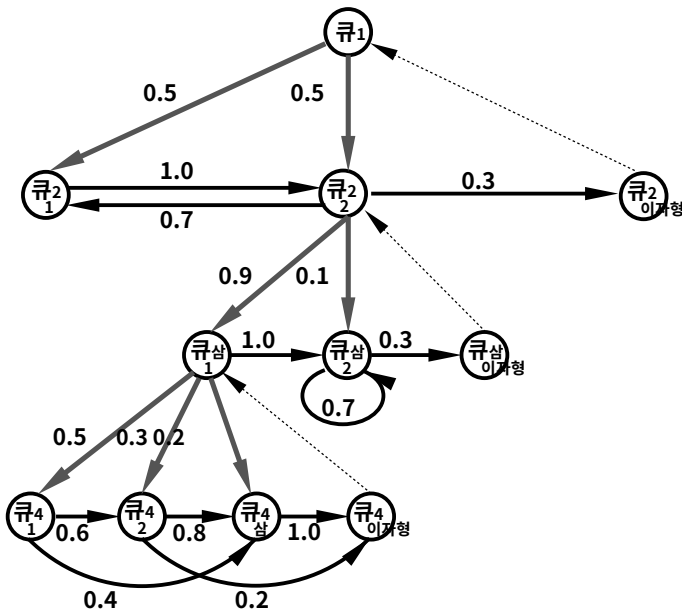


그림 1.4단계 HHMM의 그림. 회색 및 검은색 가장자리는 각각 수직 및 수평 전환을 나타냅니다. 파선의 가는 모서리는 각 레벨의 끝 상태에서 레벨의 상위 상태로의 복귀(강제)를 나타냅니다. 단순화를 위해 생산 상태는 그림에서 생략되었습니다.

최종 상태, 표시 큐4_이퍼형 확률 상태를 종료하는 실제 수단입니다.

활성화 프로세스. 최종 상태에 도달하면 제어가 전체 계층 구조의 상위 상태로 돌아갑니다. 최상위 수준 (루트 상태)은 최종 상태가 없는 유일한 수준입니다. 따라서 모든 재귀 활성화의 제어가 루트 상태로 돌아왔을 때 관찰 시퀀스의 생성이 완료됩니다. 그런 다음 루트 상태는 새로운 확률적 문자열 생성을 시작할 수 있습니다. 우리는 루트 상태에서 유한한 수의 단계로 모든 상태에 도달할 수 있다고 가정합니다. 즉, 모델이 강하게 연결되어 있습니다.

3. 추론과 학습

HMM의 경우와 마찬가지로 HHMM을 사용하는 응용 프로그램에서는 일반적으로 세 가지 자연적인 문제가 발생합니다.

시퀀스의 우도 계산: HHMM과 그 매개변수가 주어졌을 때 λ 등 =

$\{\lambda_{\text{큐}i}\}$, 확률을 찾으십시오 $\pi(\omega/\lambda)$ 시퀀스의 λ 열자 λ 열자

가장 가능성 있는 상태 시퀀스 찾기: HHMM이 주어지면 매개변수 λ =

$\{\lambda_{\text{큐}i}\}$, 관찰 시퀀스 ω , 관찰 시퀀스를 생성할 가능성이 가장 높은 단일 상태 활성화 시퀀스를 찾습니다.

모델의 매개변수 추정: HHMM의 구조와 하나 또는

더 많은 관찰 시퀀스 $\{이자형\text{영형}_{Ej}\}$, 가장 가능성이 높은 매개변수 찾기 λ 이자형 \sim 의 t 모델, $\lambda \models$ 인수 최대 $\lambda \Pi(\{영형_{Ej}\}/\lambda)$.

HHMM에 대한 위의 문제에 대한 솔루션은 계층 구조 및 다중 스케일 속성으로 인해 HMM보다 더 복잡합니다. 예를 들어, 관찰 시퀀스가 주어진 가장 가능성 있는 상태 시퀀스는 도달할 가능성이 가장 높은 상태의 간단한 인덱스 시퀀스가 아니라 상태 활성화의 다중 해상도 구조입니다. 이제 가장 간단한 것부터 시작하여 이러한 문제에 대한 해결책을 제시합니다. 우리는 다음과 같은 용어를 사용할 것입니다. 77

τ_{tim} 에서 작업을 시작했습니다. E_{if} the (pos-
상당히 비어 있음) 하위 시퀀스 $\text{영형}_{E_{if}} \times \dots \times \text{영형}_{E_{if}-1}$ 이전에 생성된 큐_{i} 자형 $\tau_{부모}$ 에 의해 활성화되었습니다.
상태와 기호 E_{if} 도달한 생산 상태 중 하나에 의해 생성되었습니다. $\text{큐}_{E_{if}}$ $\tau_{로}$. 마디면
유사하게 우리는 sta 라고 말합니다. $\text{큐}_{\tau_{E_{if}}}$ τ_{tim} 에서 작업을 완료했습니다. E_{if} 에이프 $\text{영형}_{E_{if}}$ 마지막 상징이었다
도달한 모든 생산 상태에 의해 생성된 $\text{큐}_{\text{정말로}}$ $\tau_{오, 엠}$ 컨트롤이 반환되었습니다. $\text{큐}_{\text{영형}_{E_{if}}}$ $\tau_{로}$
~에서 $\text{큐}_{\text{정말로}}$ $\tau_{오, 엠}$.

3.1. 시퀀스의 우도 계산

HHMM의 각 내부 상태는 하위 상태를 사용하여 관찰의 하위 문자열을 생성할 수 있는 자울 모델로 볼 수 있으므로 효율적인 우도 평가 절차는 재귀적이어야 합니다. 각각 쿼에스따
티 가능성을 계산

하위 문자열 생성 ω g 로 표시 $\Pi(\omega/\lambda, \overline{\kappa}_d)$. 이 순간을 가정
 확률은 루트 st를 제외하고 제공됩니다. $\overline{\kappa} \vdash \text{t.eLet } L = (L_1, L_2, \dots, L_{|\mathcal{L}|})$ 관찰 시퀀스 생성 중에 방문
 한 두 번째 수준의 상태 인덱스 *영형*
 전자 = $\text{영형}_1, \alpha_2, \dots, \alpha_{E_f}$ E_f 길 이 E_f 마지막으로 입력된 상태에 유의하십시오.
 두 번째 수준은 $\overline{\kappa}_2$ 끝, 따라서 $\overline{\kappa}_{2_{\text{불}}} = \overline{\kappa}_2$ 끝. 허락하다 $\tau_{\text{제}i}$ 의 시간적 위치일 것
 sta에 의해 생성된 첫 번째 기호 $\overline{\kappa}_{\text{제}i}$ 이 인덱스의 전체 목록을 로 표시하고 에 의해 활성화되도록 합
 $\tau = (n, n_2, \dots, \tau_{|\mathcal{L}|})$. 부터 $\overline{\kappa}_2$ L_1 니다. $\overline{\kappa}_1$ 처음 단계에서 $\overline{\kappa}_{d2}$ 끝은
 활성화된 두 번째 수준의 마지막 상태, 우리는 $\tau_{\text{시간}1a}$ = 위의 정보가 주 α_{E_f1} 그리고 $\tau_{\text{일}} = E_f$. 그만큼
 진 전체 시퀀스의 우도는,

[illegible]

무조건 가능성을 계산하려면 가능한 모든 전환 시간을 합산해야 합니다. τ 및 주 지수 L /에스. 그러한 조합이 기하급수적으로 많기 때문에 분명히 이것은 실현 가능하지 않습니다. 다행스럽게도 HHMM의 구조 덕분에 동적 프로그래밍을 사용하여 Baum-Welch 알고리즘의 일반화된 버전을 고안할 수 있습니다. 일반화된 순방향 확률 α , (\cdot) 로 정의됩니다.

$\alpha(E_i, E_i + K_{E_i}, \mathbb{C}P^{d_i-1}) = \Pi(\text{영형}_{E_i} \cdots \text{영형}_{E_i + K_{E_i}}, q_{d_i} \text{ 파이에 마쳤다 } E_i + K_{E_i} / \mathbb{C}P^{d_i-1})$ 에 시작됨.

그런, $\alpha(E_i, E_i)$ 케이, 큐디나, 큐_{i-1} 부분 관찰 시퀀스가
영형₁...영형_i 케이Sta에 의해 생성되었습니다. 큐_{디-1} 그리고 그 큐_디 마지막으로 활성화된 상태 b 큐_디와 i-1

o 세대 동안 $\text{영형}_{E_i} \cdots \text{영형}_{E_{t+K}} \text{케이}$. 각 subst의 작동에 유의하십시오. $\overline{\text{큐}}_{t+1}$. 이자형¹하다 반드시 tim에서 끝나지는 않음 $E_i + \text{케이}$ 그리고 그 $\text{영형}_{E_i} \cdots \text{영형}_{E_{t+K}} \text{케이}$ 다음에 의해 생성된 더 긴 시퀀스의 접두사가 될 수 있습니다. $\overline{\text{큐}}_{t+1}$. 시퀀스가 발생할 확률을 계산하려면 $\text{영형}_{N_{E_i}}$ 씨 · 이자형 · $\text{영형}_{E_{t+K}} \text{케이}$ 에 의해 생성되었습니다 $\overline{\text{큐}}_{t+1}$, 우리는 가능한 모든 상태를 합산해야 합니다. L 에서 veenl 덩 $\overline{\text{큐}}_{t+1}$

끝,

$$\prod_{i=1}^L (\text{영형}_{E_i} \cdots \text{영형}_{E_{t+K}} \text{케이} | q_{t+1}) = \frac{1}{q_{t+1}^L} \alpha(E_i, E_{t+K} \text{케이}, \overline{\text{큐}}_{t+1}, \overline{\text{큐}}_{t+1}) \text{ } t \text{ } \overline{\text{큐}}_{t+1} \text{ 끝.}$$

마지막으로, 전체 관찰 시퀀스의 우도는 가능한 모든 시작 상태를 합산하여 얻습니다. $\overline{\text{큐}}_0$ 마워

it)e,

$$\prod_{i=1}^L (\omega / \lambda) = \frac{1}{L!} \alpha(1, E_i, q_{t+1}, \overline{\text{큐}}_1).$$

일반화의 정의 α

레벨의 상태에 대한 d 변수

$L \text{엘} - 1, \alpha(E_i, E_{t+K} \text{케이} | L) \text{HMM의}$

$\text{케이}, \overline{\text{큐}}_t, \overline{\text{큐}}_{t+1}$, 정방향 변수의 정의와 동일합니다. α 르

$E_{t+K} \text{케이} (L) \text{HMM의}$

이 수준으로만 구성되고 출력 확률 벡터는 다음과 같이 정의됩니다.

생산 상태 $\overline{\text{큐}}_{t+1}$ L 의 평가 α 변수는 재귀 상향식으로 수행됩니다.

그런 식으로 일 α 내부 s의 하위 상태에 대해 계산된 e 값 $\overline{\text{큐}}$ 를 결정하는 데 사용되는 $\text{tatre} \alpha$ 의 값 $\overline{\text{큐}}$.

요약하면 각 내부 직원에 대해 $\overline{\text{큐}}_{t+1}$ ewe는 그것을 계산해야 합니다 α 다음을 기반으로 각 하위 시퀀스의 지 귀적 분해를 사용하여 관찰 시퀀스의 각 가능한 하위 시퀀스에 대한 svalue α 의 값 $\overline{\text{큐}}$ 의 하위 상태. 따라서 평가의 시간복잡도는 αHMM i의 모든 상태에 대한 값 영형

초 $(NT_{\text{삼}})$, 어디 N 총 상태 수이고

$E/\text{관찰}$ 시퀀스의 길이입니다. 유사한 방식으로 일반화된 역방향 변수 β 정의되고,

$$\beta(E_i, E_{t+K} \text{케이}, \overline{\text{큐}}_{t+1}, \overline{\text{큐}}_{t+1}) = \prod (\text{영형}_{E_i} \cdots \text{영형}_{E_{t+K}} \text{케이} | q_{t+1} \text{ } L \text{에서 시작, } q_{t+1} \text{파이에 마쳤다 } E_{t+K} \text{케이}).$$

계산에 대한 자세한 설명 α

그리고 β 부록 A에 나와 있습니다.

3.2. 가장 가능성 있는 상태 시퀀스 찾기

가장 가능성 있는 상태 시퀀스는 상태의 다중 척도 목록입니다. $\overline{\text{큐}}_{\text{shtatde}}$ 가 문자열을 생성했습니다.

$\text{영형}_{L_1} \cdots \text{영형}_{\text{제이}}$ 이면 부모 상태에서 str를 생성했습니다. $\text{영형}_{\text{안에}}$ $\text{케이} \text{영} \cdots \text{영형}_{\text{엘}}$, 그런 $\text{케이} \leq L_1$

그리고 $\text{제이} \leq \text{엘}$. 따라서 문자열 $\text{영형}_{L_1} \cdots \text{영형}_{\text{제이}}$ sta의 하위 상태로 세분됩니다. $\overline{\text{큐}}_{\text{teto}}$ 겹치지 않는 하위 시퀀스. 이 목록은 th를 유도하는 데 사용되는 동일한 추론 라인에 따라 효율적으로 계산할 수 있습니다. α 합산을 최대화로 대체하는 evariables. HMM에 대해 가장 가능성 있는 상태 시퀀스를 찾는 프로세스는 Viterbi 알고리즘(Viterbi, 1967)으로 알려져 있으므로 HHMM에 대한 수정된 알고리즘을 일반화된 Viterbi 알고리즘이라고 합니다.

의 정의와 유사하다. α 우리가 정의하는 변수 $\delta(E_i, E_{t+K} \text{케이}, \overline{\text{큐}}_{t+1}, \overline{\text{큐}}_{t+1})$ ~이 되기 위해 가장 가능성 있는(계층적) 상태 서열 유전자의 우도 $\text{영형}_{L_1} \text{에티}$ · 안에 $\text{영형}_{\text{영형}_{E_{t+K}} \text{케이}}$ 주어진 저것 $\overline{\text{큐}}_{t+1}$ 팀에 입력되었습니다 $E/\text{전자}$, 하위 상태 $\overline{\text{큐}}_{t+1}$ 마지막으로 활성화된 상태였습니다. $\overline{\text{큐}}_{\text{비}} \text{cy}$, 그리고 컨트롤이 반환된 $\overline{\text{큐}}_{t+1}$ 시간에 $E_{t+K} \text{케이}$. 실제 계층적 구문 분석에 관심이 있기 때문에

상태로의 시퀀스에서 우리는 또한 두 가지 추가 변수를 유지합니다. ψ (전자 E, 타 + 케이, 큐디 나, 큐디)
활성화될 가능성이 가장 높은 상태의 인덱스입니다. \bar{c}_{t-1} 와 i_1 활성화하기 전에 \bar{c}_{t-1} 그리고 $E_t = \tau(E_t, E_{t-1}, \bar{c}_{t-1}, \bar{c}_{t-1})$ ($E_t \leq E_{t-1} \leq E_{t-1} + K_{t-1}$) 때 이다 \bar{c}_{t-1} 나 에 의해 활성화되었습니다 \bar{c}_{t-1} . 이것들을 감안할 때 두 개의 변수 가장 가능성 있는 계층적 상태 시퀀스는 목록을 스캔하여 얻습니다. ψ 그리고 τ 루트 상태에서 프로덕션 상태로. 너비 우선 검색이 스캐닝에 사용되는 경우 상태는 레벨 색인별로 위에서 아래로 나열됩니다. 깊이 우선 검색이 사용되면 활성화 시간별로 상태가 나열됩니다. 단순히 합산을 최대화로 대체했기 때문에 일반화된 Viterbi 알고리즘의 시간 복잡도는 일반화된 Baum-Welch의 시간과 동일합니다.
영형

나 N라이 E/삼). 의사 코드

이 알고리즘에 대한 설명은 부록 B에 나와 있습니다.
우리는 또한 다음에서 가장 가능성 있는 상태 시퀀스에 대한 근사치를 찾는 휴리스틱을 고안했습니다. 영형(N_{T_2}) 시각. 이 휴리스틱은 서로 다른 상태에 의해 유도된 시퀀스에 대한 분포가 서로 실질적으로 다르다고 가정합니다. 따라서 가장 가능성 있는 상태 시퀀스를 찾는 데 있어 수평 전이의 영향은 무시할 수 있습니다. 따라서 전향 확률을 무시하는 근사 검색을 수행합니다. 즉, 우리는 각 sta를 취급합니다 \bar{c}_{t-1} 이웃 국가의 영향을 무시하는 자율적 모델로서의 HHMM o \bar{c}_{t-1} 수준에서 \bar{c}_{t-1} . 따라서 각 내부 노드에서 최대 작업을 하나만 수행하여 전체 실행 시간을 줄입니다.

영형이 자행 E/2). 하지만

이 근사에 대한 이론적 정당성은 없으며, 실험에서 근사 검색에 의해 발견된 가장 가능성 있는 상태 시퀀스가 정확한 일반화된 Viterbi 알고리즘에 의해 발견된 상태 시퀀스와 매우 유사하다는 것을 발견했습니다 (섹션 4에 설명된 실험 참조).

3.3. HHMM의 매개변수 추정

HHMM에 대한 최대 우도 매개변수 추정 절차는 재귀적으로 관찰을 생성하는 확률적 수직 전환도 고려해야 하기 때문에 Baum-Welch 알고리즘의 일반화입니다. 따라서 경로 var 외에도 α iables 및 β 앞으로 및 '뒤로' 전환에 해당하는 '아래로' 및 '위로' 전환에 해당하는 추가 경로 변수를 추가합니다. 기대 단계에서 사용되는 변수는 다음과 같습니다.

$\xi(E_t, \bar{c}_{t-1}, \bar{c}_{t-1}, \bar{c}_{t-1})$ 수평 전환을 수행할 확률입니다. \bar{c}_{t-1} 미디엄 \bar{c}_{t-1} 에게 \bar{c}_{t-1} 이, 두 하위 상태 o \bar{c}_{t-1} 에프 \bar{c}_{t-1} , 시간에 E_t 생산 후 o 영형에프 E_t 그리고 생산 전에 영형 E_{t+1} , $\xi(E_t, \bar{c}_{t-1}, \bar{c}_{t-1}, \bar{c}_{t-1}) = \Pi(\text{영형}_1 \cdots \text{영형}_{E_t}, q_{\bar{c}_{t-1}} \rightarrow \bar{c}_{t-1}, \text{영형}_{E_t+1} \cdots \text{영형}_{E_t} | \lambda)$.

기반으로 ξ 우리는 두 개의 보조 변수를 정의합니다 γ 에스 안 에 그리고 γ 밖으로 재추정 단계를 단순화합니다.

$\gamma_{\text{안}}(E_t, \bar{c}_{t-1}, \bar{c}_{t-1})$ st로의 수평 전이를 수행할 확률 \bar{c}_{t-1} 나 테비포 영형 E_t 생성되었다 $\gamma_{\text{안}}$ 다음을 사용하여 계산됩니다. ξ 모든 하위 상태를 합산하여 \bar{c}_{t-1} 영형 \bar{c}_{t-1} 수평 전환을 수행할 수 있는 \bar{c}_{t-1} 에게 \bar{c}_{t-1} , $\gamma_{\text{안}}(E_t, \bar{c}_{t-1}, \bar{c}_{t-1}) = \frac{1}{q_{\bar{c}_{t-1}}} \sum_{\bar{c}_{t-1}} \xi(E_t - 1, q_{\bar{c}_{t-1}}, q_{\bar{c}_{t-1}}, \bar{c}_{t-1})$. \bar{c}_{t-1}

$\gamma_{\text{밖으로}}(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})$ 통계를 떠날 확률입니다. $\bar{\kappa}_{d-1}$ 자형 $d-1$ 수평 전환을 수행하여 동일한 lev의 모든 주에 d 세대 이후에 $\gamma_{\text{안에서}}$ 에 E_l 유사 $\gamma_{\text{안에서}}$, $\gamma_{\text{밖으로}}$ 다음을 사용하여 계산됩니다. 모든 하위 상태를 합산하여 $\bar{\kappa}_{d-1}$ 에 E_l 도달할 수 있는 ~에서 $\bar{\kappa}_{d-1}$ 단일 수평 전환으로

$$\gamma_{\text{밖으로}}(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) = \frac{1}{q} \sum_{\kappa_{d-1}=1}^{q_{d-1}} \xi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}, q_{d-1} \kappa_{d-1}) .$$

수직 이동의 확률을 추정하는 데 사용되는 경로 변수 χ 에

$\chi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})$ 확률은 stat $\bar{\kappa}_{d-1}$ 자형 $d-1$ 팀에 입력되었습니다. E_l 전에 $\gamma_{\text{안에서}}$ 생성되었습니다. 처음에는 sta를 활성화하기로 선택했습니다. $\bar{\kappa}_{d-1}$,

$$\chi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) = \frac{1}{q} (\gamma_{\text{안에서}} \cdots \gamma_{\text{안에서}}(E_{l-1}, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) \downarrow, \gamma_{\text{안에서}} \cdots \gamma_{\text{안에서}}(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) .$$

위의 경로 변수와 주어진 현재 매개변수 세트를 기반으로 다음 기대치가 계산됩니다.

$\sum_{E_l=1}^{E_{l-1}} \xi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})$: 예상되는 수평 전환 수 fr $\bar{\kappa}_{d-1}$ 자형 $d-1$ 하위 상태 o $\bar{\kappa}_{d-1}$ 에 E_l 도달 다

$\sum_{E_l=1}^{E_{l-1}} \gamma_{\text{안에서}}(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) = \sum_{\kappa_{d-1}=1}^{q_{d-1}} \sum_{E_l=1}^{E_{l-1}} \xi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}, q_{d-1} \kappa_{d-1})$: 예상 호-stat로의 izontal 전환 $\bar{\kappa}_{d-1}$ 자형 $d-1$ lev의 인접 하위 상태 중 하나에서 d 자형. 엘

$\sum_{E_l=1}^{E_{l-1}} \gamma_{\text{밖으로}}(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) = \sum_{\kappa_{d-1}=1}^{q_{d-1}} \sum_{E_l=1}^{E_{l-1}} \xi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}, q_{d-1} \kappa_{d-1})$: 예상 호-stat에서 izontal 전환 $\bar{\kappa}_{d-1}$ 자형 $d-1$ lev의 인접 하위 상태 중 하나에 d 자형. 엘

$\sum_{E_l=1}^{E_{l-1}} \chi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})$: 예상되는 수직 전환 수. $\bar{\kappa}_{d-1}$ 자형 $d-1$ 에 E_l 도달 다

$\sum_{L=1}^{q_{d-1}} \sum_{E_l=1}^{E_{l-1}} \chi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})$: 예상되는 수직 전환 수 $\bar{\kappa}_{d-1}$ 자형 $d-1$ 에 E_l 도달 다. 엘

$\sum_{E_l=1}^{E_{l-1}} \chi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) + \sum_{E_l=2}^{E_{l-1}} \gamma_{\text{안에서}}(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) = \sum_{E_l=1}^{E_{l-1}} \gamma_{\text{밖으로}}(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})$: 예상 생산 단계로의 수직 전환 수 $\bar{\kappa}_{d-1}$ 자형 $d-1$ 에 E_l 도달 다.

완전한 파생 o $\bar{\kappa}_{d-1}$ 에 E_l , $\gamma_{\text{안에서}}$, $\gamma_{\text{밖으로}}$, 그리고 χ 부록 A에 나와 있습니다. 현재 매개변수에서 위의 기대치를 계산한 후 다음과 같이 새로운 매개변수 세트를 재추정합니다.

$$\pi_{\bar{\kappa}_{d-1}}(\bar{\kappa}_{d-1}) = \chi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}), \quad (1)$$

$$\pi_{\bar{\kappa}_{d-1}}(\bar{\kappa}_{d-1}) = \frac{\sum_{E_l=1}^{E_{l-1}} \chi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})}{\sum_{\kappa_{d-1}=1}^{q_{d-1}} \sum_{E_l=1}^{E_{l-1}} \chi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})} \quad (2 < d < D), \quad (2)$$

$$\pi_{\bar{\kappa}_{d-1}}(\bar{\kappa}_{d-1}) = \frac{\sum_{E_l=1}^{E_{l-1}} \xi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})}{\sum_{\kappa_{d-1}=1}^{q_{d-1}} \sum_{E_l=1}^{E_{l-1}} \xi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})} = \sum_{E_l=1}^{E_{l-1}} \xi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) , \quad (3)$$

$$\pi_{\bar{\kappa}_{d-1}}(\bar{\kappa}_{d-1}) = \frac{\sum_{E_l=1}^{E_{l-1}} \chi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) + \sum_{E_l=2}^{E_{l-1}} \gamma_{\text{안에서}}(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})}{\sum_{E_l=1}^{E_{l-1}} \chi(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1}) + \sum_{E_l=2}^{E_{l-1}} \gamma_{\text{안에서}}(E_l, \bar{\kappa}_{d-1}, \bar{\kappa}_{d-1}^{-1})} . \quad (4)$$

좋은 매개변수 집합을 찾기 위해 다음을 계산하는 기대 단계를 반복합니다. τ, χ , 보조 경로 변수, 그리고 Equ를 사용합니다. (1)-(4) 매개변수의 새로운 추정치를 찾습니다. 지루하긴 하지만 이 반복 절차의 위 단계가 EM 알고리즘의 기대 및 최대화 단계에 해당하는지 확인하는 것은 매우 간단합니다. 따라서 이 절차는 우도 함수의 정지 지점(일반적으로 로컬 최대값)으로 수렴하는 것을 보장합니다.

4. 신칭

이 섹션에서는 다음과 같은 복잡한 시퀀스 모델링 작업에 대한 HHMM 사용 및 매개변수 추정에 대한 두 가지 예를 논의하고 제공합니다.

4.1. 영어 텍스트의 다단계 구조

자연 텍스트와 같은 복잡한 시퀀스의 확률적 분석의 주요 목표 중 하나는 시퀀스에서 멀리 떨어져 나타나는 이벤트 간의 상관 관계를 캡처하는 모델을 설계하는 것입니다. Observable Markov 모델은 이러한 작업에 널리 사용되었습니다(예를 들어(Ron et al., 1996) 및 참조 문헌 참조). 그러나 이러한 모델의 상태는 관찰 가능한 하위 시퀀스를 기반으로 구성되기 때문에 암묵적인 장기 통계 상관 관계를 캡처할 수 없습니다. 여기서 우리는 HHMM에 기반한 대안적 접근법을 제안하고 이 접근법이 위의 어려움을 부분적으로 극복할 수 있다는 몇 가지 실험적 증거를 제공합니다.

우리는 두 개의 HHMM을 구축하고 고전 영어 이야기로 구성된 자연 텍스트에 대해 훈련했습니다. 관찰 알파벳에는 소문자와 대문자, 공백 및 문장 부호가 포함됩니다. 훈련을 위해 근사치를 사용했습니다.5테 00라이0평균 길이의 문장 의50문자. 정확히 동일한 텍스트에 대해 두 개의 HHMM을 교육했습니다. HHMM은 다음과 같았습니다.

- 두 개의 레벨로 구성된 얇은 HHMM. 이 HHMM h의 루트 상태4광고 하위 상태. 두 번째 레벨의 각 상태7모든 프로젝트션 상태인 lshuabdstates. 따라서 모든 생산 상태는 동일한 수준이었습니다. 이 HHMM의 구조는 그림 3의 상단에 나와 있습니다.
- 세 가지 레벨로 구성된 불균형 HHMM. 이 HHMM은 각 내부 상태에서 가변적인 수의 하위 상태를 가졌습니다. 이 HHMM의 구조는 모든 수준에서 생산 상태를 가졌기 때문에 균형이 맞지 않았습니다. 두 번째 HHMM의 그림은 그림 2와 그림 3의 하단 부분에 나와 있습니다.

위의 HHMM에 일반화된 Baum-Welch 매개변수 추정 절차를 적용했습니다. 우리는 훈련 후 첫 번째 HHMM의 하위 상태에 의해 유도된 문자열에 대한 분포가 동일한 데이터에 대해 훈련된 표준 HMM에 의해 유도된 분포와 매우 유사하다는 것을 발견했습니다. 대조적으로, 두 번째 HHMM에 의해 유도된 분포는 실질적으로 다르며 몇 가지 흥미로운 현상을 드러냈습니다. 첫째, 두 번째 HHMM에 의해 유도된 분포는 하위 상태에 따라 크게 다릅니다. 문자열 세트는

중첩이 거의 없도록 각 상태에서 생성될 가능성이 가장 높습니다. 둘째, 상태의 다중 규모 동작을 관찰했습니다. 구체적으로, 깊은 상태에서 생성될 가능성이 가장 높은 문자열은 대략 음성 단위, 즉 다음과 같은 문자열에 해당한다는 것을 발견했습니다. 나에스ng, th, wh, 그리고 우. 계층 구조를 올라가면 두 번째 및 세 번째 수준의 상태는 다음과 같이 자주 사용되는 단어 및 구문인 문자열을 생성합니다. 아니다, 것이다 그리고 어디. 마지막으로 계층 구조의 최상위에서 루트 상태는 문장 규모에 해당하는 분포를 유도했습니다. 예를 들어 루트 상태(따라서 전체 HHMM)에서 생성된 문자열은 문장 부호로 끝날 가능성이 높습니다. 우리는 또한 불균형 HHMM의 훈련이 끝날 때 수평 전환 확률이 매우 높다는 것을 발견했습니다. 이는 동일한 수준에서 상태 간의 강력한 Markov 종속성을 반영합니다. 따라서 각 상태에 의해 유도된 분포는 소수의 문자열에 매우 집중되어 있을 뿐만 아니라 심층 HHMM의 재귀적 활성화에 의해 생성될 수 있는 문자열 집합도 자연 텍스트에 자주 나타나는 구문 구조에 대해 강하게 편향되어 있습니다. 다음 응용 프로그램에서 설명하는 것처럼 훈련된 HHMM 또는 그 하위 모델은 이제 텍스트 분류와 같은 더 복잡한 작업에서 빌딩 블록으로 사용될 수 있습니다. 이러한 고도로 편향된 분포는 그림 2와 3에 설명되어 있습니다. 그림 2에서는 훈련 시작과 끝에서 수평 전환 확률을 제공합니다. 그림 3에는 두 HHMM의 각 하위 상태(하단의 깊은 불균형 HHMM과 상단의 얇은 균형 HHMM)에서 생성된 가장 가능성 있는 문자열이 나열되어 있습니다. 풍부한 모델이 전체 단어와 문장 조각을 포함하는 훨씬 더 다양한 문자열을 개발했다는 것이 그림에서 분명합니다. 그림 3에는 두 HHMM의 각 하위 상태(하단의 깊은 불균형 HHMM과 상단의 얇은 균형 HHMM)에서 생성된 가장 가능성 있는 문자열이 나열되어 있습니다. 풍부한 모델이 전체 단어와 문장 조각을 포함하는 훨씬 더 다양한 문자열을 개발했다는 것이 그림에서 분명합니다. 그림 3에는 두 HHMM의 각 하위 상태(하단의 깊은 불균형 HHMM과 상단의 얇은 균형 HHMM)에서 생성된 가장 가능성 있는 문자열이 나열되어 있습니다. 풍부한 모델이 전체 단어와 문장 조각을 포함하는 훨씬 더 다양한 문자열을 개발했다는 것이 그림에서 분명합니다.

4.2. 필기체 필기의 비지도 학습

(Singer & Tishby, 1994)에서는 필기의 진동 모델을 기반으로 필기체 필기에 대한 동적 부호화 기법을 제안하고 분석하였다. 이 체계는 필기체 필기를 효율적으로 인코딩하는 이산 기호 세트를 통해 연속 펜 궤적에서 문자열로의 역 매핑을 수행합니다. 이러한 기호를 모터 제어 명령이라고 합니다. 모터 제어 명령은 생성 모델을 사용하여 펜 궤적으로 다시 변환될 수 있으며 동적 인코딩 체계에 의해 제거되는 노이즈 없이 필기를 재구성할 수 있습니다. 각 가능한 제어 명령은 for의 데카르트 곱으로 구성됩니다. **엑스**

미디엄 \times 와이어디 **엑스**,와이 $\in \{0,1,2,삼,4,5\}$, 따라서 알파벳 36개의 서로 다른 기호로 구성되어 있습니다. 이러한 기호는 양자화된 수평 및 수직 진폭 변조와 해당 위상 지연을 나타냅니다.

다른 로마 문자는 위의 기호에 대해 다른 시퀀스에 매핑됩니다. 더욱이, 서로 다른 필기체가 있고 인간의 운동 시스템에 소음이 존재하기 때문에 동일한 필기체 글자가 여러 가지로 쓰여질 수 있습니다. 이로 인해 동일한 필기체로 작성된 단어를 나타내는 다른 인코딩이 생성됩니다. 필기체 스크립트를 분석하고 인식하는 시스템에서 바람직한 첫 번째 단계는 완전한 필기체 오행 경로에 해당하는 시퀀스의 분포를 근사화하는 확률 모델을 구축하는 것입니다. 우리는 운동 제어 명령을 관찰 알파벳으로 사용하고 훈련 세트의 다른 필기체 단어에 해당하는 HHMM을 구축했습니다. 예를 들어, 우리는 60워의 에미디엄디유지하 다5개 수준이 있는 HHMM의 매개변수를 추정합니다. 이 HHMM은 불균형한 구조를 가지고 있었고 모든 수준에서 생산 상태를 가졌습니다. HHMM의 토폴로지 설계에서 추가 지식을 고려했습니다.

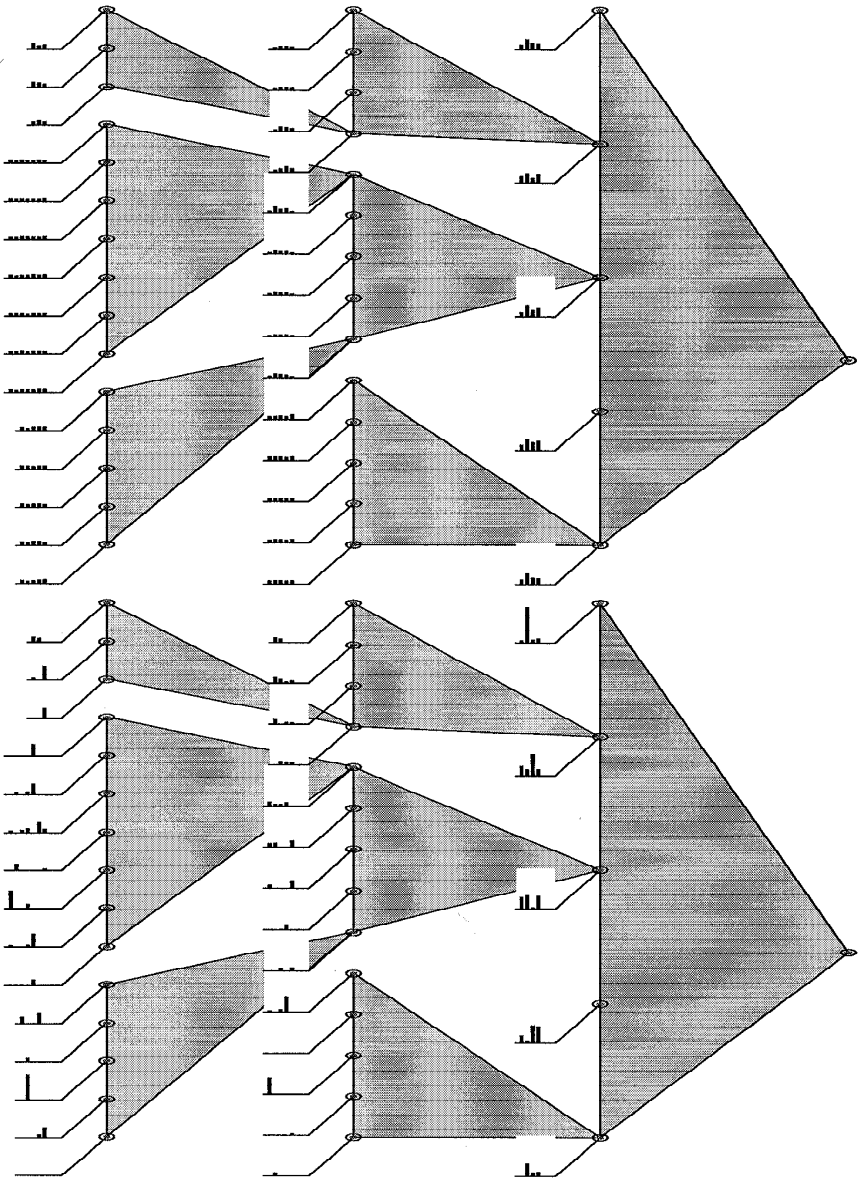


그림 2. 불균형 깊이의 HHMM에 대한 교육 시작(상단) 및 종료 시 전환 분포. 삼각형은 영어 텍스트로 훈련되었습니다. 초기 분포는 거의 균일하지만 최종 분포는 다른 수준의 다른 주 주변에서 급격하게 정점에 이릅니다.

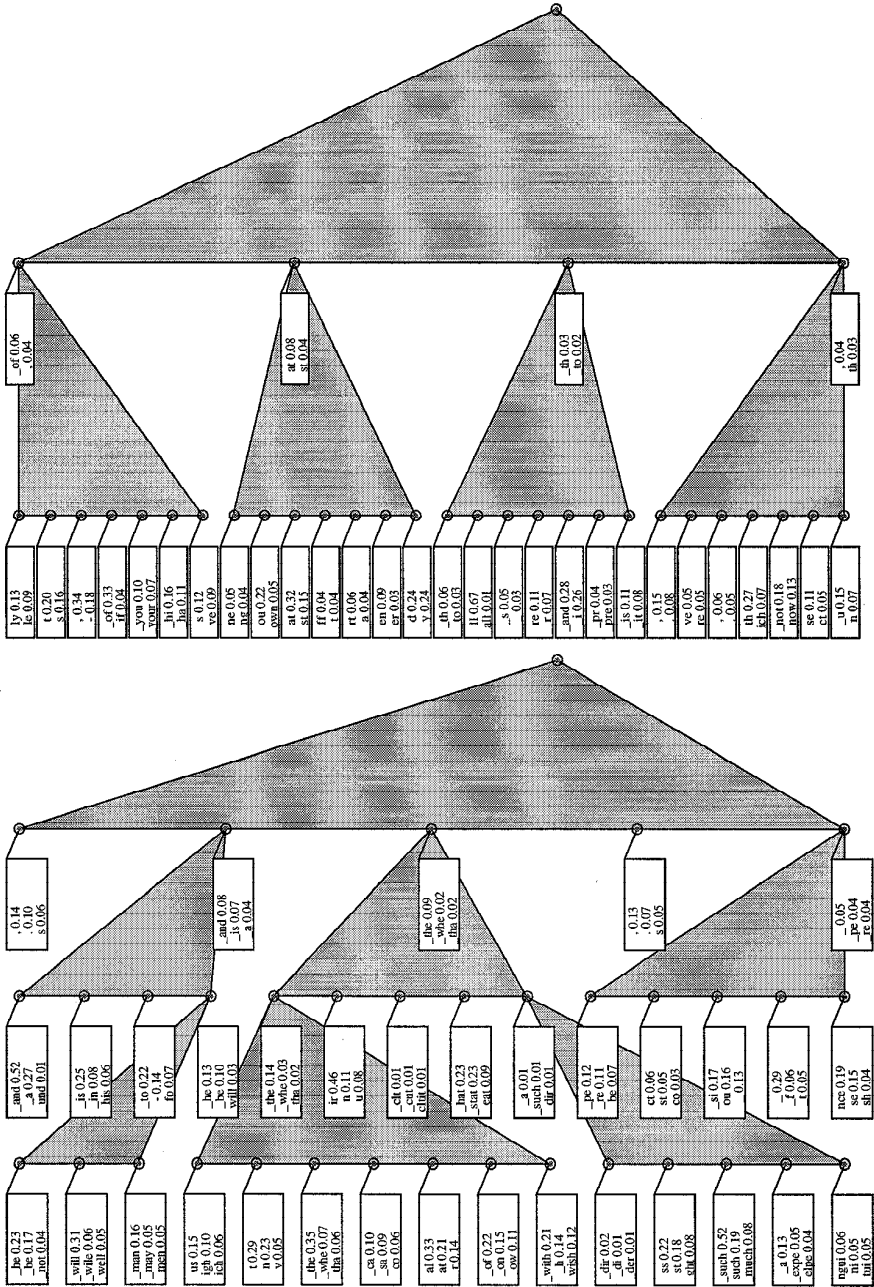


그림 3. 두 개의 서로 다른 HHMM에 대해 각 상태에서 생성될 가능성이 가장 높은 문자열입니다.

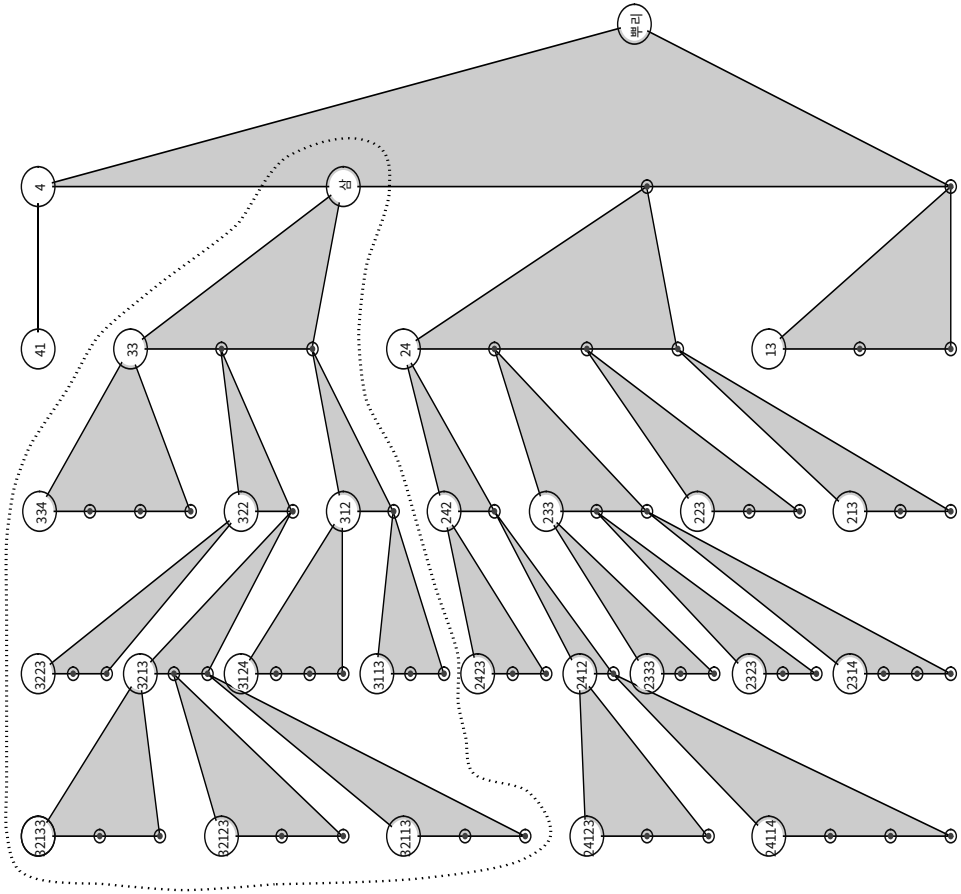


그림 4. 필기체 필기 얼룩 실험에 사용된 HHMM. 전체 모델은 완전한 단어에 대해 훈련되었습니다. 점선으로 표시된 하위 모델은 문자 조합의 발생 위치를 찾는 데 사용되었습니다. 일체 포함.

초서체로 쓰여진 단어에서 문자의 반복 및 문자 조합과 같은 작업에 사용되는 HHMM의 구조미디엄 유지하다그림 4에 나와 있습니다. 일반화된 Baum-Welch 알고리즘을 사용하여 HHMM의 매개변수를 추정했습니다. 그런 다음 훈련된 HHMM을 사용하여 필기체 필기체의 문자 조합을 나타내는 반복된 획을 식별했습니다.

결과 HHMM이 실제로 단어의 분포와 내부 구조를 학습했는지 확인하기 위해 HHMM에 일반화된 Viterbi 알고리즘을 사용하여 모터 제어 시퀀스의 다중 스케일 분할을 수행했습니다. 이러한 분할의 예시 결과는 그림 5에 나와 있습니다. 그림에서 필기체 w미디엄영형 1아르 자형나디엔테인,모터 제어 명령에서 재구성되어 계층적 분할과 함께 표시됩니다.

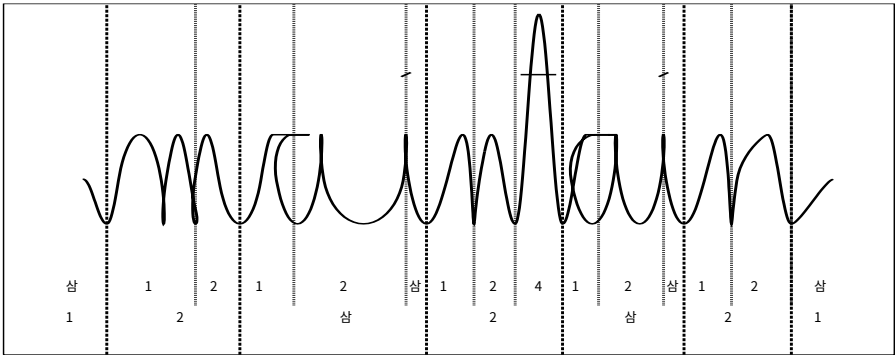


그림 5.wo의 계층적 세분화미디엄일 유지하다HHMM에 대한 Viterbi 알고리즘에 의해 얻어집니다. 그만큼 HHMM의 처음 두 레벨에 따른 시간 분할이 표시됩니다. 단어는 인코딩된 동적 표현에서 재구성되었습니다.

우리는 단어를 분할하기 위해 구조가 그림 4에 표시된 깊이 5의 훈련된 HHMM을 사용했습니다.

그림 5에서는 HHMM의 처음 두 계층에서 상태로의 시간 분할을 보여줍니다. 일반화된 Viterbi 알고리즘이 HHMM의 다른 상태를 다른 필기체 스트로크에 할당한 것은 그림에서 분명합니다. 또한 동일한 문자 조합을 나타내는 획을 생성하기 위해 동일한 상태가 일관되게 사용됩니다. 예를 들어, 상태삼첫 번째 계층에서 콤비나 생성을 담당합니다. ㅏtio.n 하위 상태는 lette를 더 분할합니다. ㅏ아르 자형나s를 하위 스트로크로: 첫 번째 하위 상태로 표시삼1b,y는 필기체 let의 첫 번째 부분을 생성합니다. ㅏt, etrhe 초 삼,2, 중간 부분을 생성합니다 (lette를 연결하는 관절 스트로크 ㅏrsand나),그리고 세 번째,삼삼, 편지를 생성나아르 자형. 유사한 현상이 HHMM의 다른 상태에서도 관찰될 수 있습니다.

음성 신호 및 필기 텍스트와 같은 시퀀스의 확률적 모델링에서 발생하는 일반적인 질문은 시퀀스를 구성하는 '자연스러운' 단위가 무엇인지입니다. 널리 사용되는 접근 방식은 이러한 단위를 수동으로 정의하는 것입니다(예: 음성 언어의 음소, 서면 텍스트의 문자 등). 이 접근 방식에는 몇 가지 단점이 있습니다. 수동 분할이 필요하고 시간적 상호 작용(예: 음성의 공동 조음) 연속 '단위' 사이. 우리는 이제 훈련된 HHMM의 하위 상태를 사용하여 위의 질문에 대한 부분적인 답변을 제공하는 대체 접근 방식을 제안하고 간략하게 시연합니다. HHMM의 자체 유사 구조로 인해 각 하위 상태를 자율 모델로 사용할 수 있습니다. 위에서 설명한 HHMM의 두 번째 수준에서 하위 상태를 사용하고 관찰 시퀀스의 각 하위 시퀀스에 대해 유도하는 확률을 계산했습니다. 또한 가능한 모든 기호에 대해 균일한 분포를 유도하는 간단한 HMM을 정의했습니다. 이 간단한 모델은 유

귀무 가설로서의 d,bserves 전체 HHMM에서 빼낸 하위 모델과 경쟁합니다. 후속 시퀀스의 확률 영향

\tilde{e} 하위 모드에서 생성 미디엄 귀무가설과 비교하여(- 대안에 대해 동일한 사전 합산) $\pi(\text{오})$ 이다(, $\frac{\pi(\text{오}|미디엄)}{\pi(\text{오}|유)+\pi(\text{오}|유)}$. 높은 값 $\pi(\frac{\text{삼}}{\text{오}})$ 빼낸 m에 해당하는 문자의 발생을 나타냅니다.

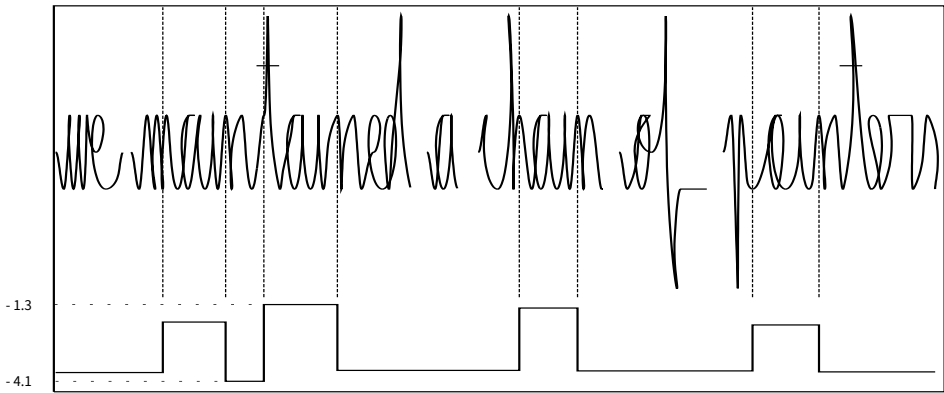


그림 6. *lette*의 발생을 발견한 λ -아르 자형나문장을 죄 우리는 포인터 체인을 유지했습니다. 이 문자 조합에 해당하는 하위 모델은 단어용으로 구축된 그림 5에 표시된 HHMM에서 추출되었습니다. 유지하다. 문자의 발생은 모델이 가능한 모든 기호에 대해 균일한 분포를 유도하는 귀무 가설과 경쟁하게 함으로써 발견됩니다.

따라서 이 값을 임계화하면 분할되지 않은 데이터에서 문자 조합의 위치를 식별하는 데 사용할 수 있습니다. HHMM을 사용한 문자 조합 스포팅의 예시 결과는 그림 6에 나와 있습니다. 그림에서 조건부 확률의 로그를 보여줍니다. $\pi(\text{문자열의 길이로 정규화}/\omega)$ 이 확률은

가능한 모든 시작 위치. 더 서브모 λ/ω 조합에 대응하는 델 λN 나, 그건, sta 에 뿌리를 둔 하위 모델삼그림 4에서 점선으로 표시된 t 는 wor 용으로 구축된 HHMM에서 빼낸 것입니다. 미디엄디유지. 분명히, 문자의 모든 발생일체 포함올바르게 위치했습니다. 레트의 조합영형이 자형나 rsin 단어포인터또한 높은 가능성을 받았습니다. 이것과 다른 모호함은 하위 모델 세트를 더 세분화하고 더 높은 수준의 확률적 언어 모델을 사용하여 해결할 수 있습니다.

5. 결론

계층적 숨겨진 Markov 모델은 복잡한 시퀀스 모델링에서 발생하는 두 가지 근본적인 문제에 대한 부분적인 해답을 제공하는 HMM의 일반화입니다. 첫째, HHMM은 간단한 Markov 프로세스의 단순성과 계산 용이성을 유지하면서 관찰 시퀀스에서 상대적으로 멀리 떨어져 있는 구조를 상호 연관시킬 수 있습니다. 둘째, 음성 자연어 및 기타 복잡한 시퀀스에서 흔히 발생하는 통계적 비균질성을 처리할 수 있습니다. 예를 들어, 숨겨진 Markov 모델은 최근 신경 생리학적 데이터를 분석하는 데 사용되었습니다(Gat et al., 1997). 계층적 숨겨진 Markov 모델은 녹음의 다양한 세그먼트에서 여러 수준의 피질 활동을 모델링하기 위한 더 나은 도구를 제공할 수 있습니다.

최대 우도 매개변수 추정 절차와 Viterbi 가장 가능성 있는 상태 디코딩은 둘 다 자연스럽게 이 풍부한 구조로 일반화됩니다. 그러나 여전히 누락된 구성 요소가 있습니다. HHMM에는 토폴로지를 조정할 수 있는 기능이 없습니다.

하위 모델의 자체 구성 병합 및 성장을 위해. HHMM의 몇 가지 자연스러운 일반화도 있습니다. 예를 들어 Bengio와 Frasconi(1995)가 입력-출력 HMM에 대해 도입한 프레임워크를 사용하면 계층적 HMM을 일반화하여 두 개의 서로 다른 알파벳을 통해 문자열 간의 입력-출력 매핑을 설명할 수 있습니다.

이 백서에 설명된 HHMM을 사용한 실험은 자연 복합 시퀀스에 대한 계층적 확률 모델을 더 잘 이해하기 위한 초기 단계입니다. 계승 은닉 마르코프 모델(Ghahramani & Jordan, 1997) 및 대체 매개변수 추정 기법(Singer & Warmuth, 1997)과 같은 다른 모델을 사용할 수 있습니다. 이러한 서로 다른 접근법 사이의 연결을 이해하고 계층적 확률 모델링의 공식적인 분석을 수행하는 것은 현재 진행 중인 중요한 연구 방향입니다.

부록

일반화된 Baum-Welch 알고리즘

독자에게 상기시키기 위해 기대치에서 4개의 경로 변수를 계산합니다. α , β , χ , 그리고 ξ 이는 주어진 HHMM에서 '앞으로', '뒤로', '아래로' 및 '위로' 확률적 전환에 비공식적으로 해당합니다. 변수 α 는 β 다음과 같이 확률이 유도되므로 상향식 방식으로 계산됩니다. γ 는 α 에 의존하는 β 가 있는 트리(하위 모델)에 속하는 하위 상태에서만 유지됩니다. γ 주어진 변수에서 α 그리고 β , 변수 χ 그리고 ξ 하향식으로 계산됩니다. 경로 변수의 유도를 단순화하기 위해 두 개의 보조 변수도 정의합니다. η , η^* 그리고 η^{back} . 이제 이러한 변수의 자세한 유도를 제공합니다.

정의:

$$\alpha(E_i, E_{i+K}, i, \gamma_{i+1}, \gamma_{i+1}) = \prod_{(E_{i+K}, \dots, E_{i+K+1}) \in \text{path}} \alpha(E_i, E_{i+K}, i, \gamma_{i+1}) \text{에 시작됨}$$

견적:

$$\begin{aligned} \alpha(t, t, q_i, \gamma_{i+1}) &= \prod_{(E_{i+K}, \dots, E_{i+K+1}) \in \text{path}} \alpha(E_i, E_{i+K}, i, \gamma_{i+1}) \\ \alpha(E_i, E_{i+K}, i, \gamma_{i+1}) &= \sum_{q_{i+1}} \alpha(E_i, E_{i+K}, i, q_{i+1}, \gamma_{i+1}) \prod_{j=1}^{K-1} \beta_{ij} \\ \alpha(t, t, q_i, \gamma_{i+1}) &= \prod_{(E_{i+K}, \dots, E_{i+K+1}) \in \text{path}} \alpha(E_i, E_{i+K}, i, q_i, \gamma_{i+1}) \prod_{j=1}^{K-1} \beta_{ij} \\ \alpha(E_i, E_{i+K}, i, \gamma_{i+1}) &= \sum_{q_{i+1}} \alpha(E_i, E_{i+K}, i, q_{i+1}, \gamma_{i+1}) \prod_{j=1}^{K-1} \beta_{ij} \\ \alpha(E_i, E_{i+K}, i, \gamma_{i+1}) &= \sum_{q_{i+1}} \alpha(E_i, E_{i+K}, i, q_{i+1}, \gamma_{i+1}) \prod_{j=1}^{K-1} \beta_{ij} \end{aligned}$$

$$+ \sum_{\substack{\text{타는 } / \\ \text{에스}=1}} \text{파이}_{\text{쿠티}}(\text{쿠티}) \alpha(E_i, E_{t+1} \text{ 케이}, \text{큐에스}, \text{쿠티}) \text{타는 } \text{쿠티}$$

정의:

$$\beta(E_i, E_{t+1} \text{ 케이}, \text{쿠티}, \text{쿠티}^{-1}) = \text{피}(\text{영형}_1 \dots \text{영형}_{\text{타 케이}} | q_{\text{타}} \text{에서 시작, 쿠티-1 파이에 마쳤다 } E_{t+1} \text{ 케이})$$

견적:

$$\begin{aligned} \beta(t, t, q_{\text{타}}, \text{쿠티}^{t-1}) &= \text{타는 } \text{쿠티}^{t-1} \text{ (영형타) 타는 끝} \\ \beta(E_i, E_{t+1} \text{ 케이}, \text{쿠티}, \text{쿠티}) &= \frac{1}{\text{쿠티}} \text{타는 } \text{쿠티} \text{ (영형타) } \sum_{\substack{\text{타는 } / \\ \text{제이}=1}} \text{타는 } \text{쿠티} \beta(E_{t+1}, E_{t+1} \text{ 케이}, \text{큐에스}, \text{쿠티}^{t-1}) \\ \beta(t, t, q_{\text{타}}, \text{쿠티}) &= \sum_{\substack{\text{타는 } / \\ \text{에스}=1}} \text{파이}_{\text{타는 } \text{쿠티}}(\text{쿠티}) \beta(t, t, q_{\text{타}}, \text{타는 } \text{쿠티}) \\ \beta(E_i, E_{t+1} \text{ 케이}, \text{쿠티}, \text{쿠티}) &= \sum_{\substack{\text{타는 } / \\ \text{에스}=1}} \text{파이}_{\text{타는 } \text{쿠티}}(\text{쿠티}) \beta(E_i, E_{t+1} \text{ 케이}, \text{큐에스}, q_{\text{타}}) \\ &= \sum_{\substack{\text{타는 } / \\ \text{제이}=1}} \text{타는 } \text{쿠티} \beta(E_{t+1}, E_{t+1} \text{ 케이}, \text{쿠티} \text{ 제이}, \text{쿠티}) \\ &+ \sum_{\substack{\text{타는 } / \\ \text{에스}=1}} \text{파이}_{\text{타는 } \text{쿠티}}(\text{쿠티}) \beta(E_i, E_{t+1} \text{ 케이}, \text{큐에스}, \text{쿠티}) \text{타는 } \text{쿠티} \end{aligned}$$

정의:

$$\eta_{\text{타는}}(E_i, \text{쿠티}, \text{쿠티}) = \text{피}(\text{영형}_1 \dots \text{영형}_{\text{타 케이}}, q_{\text{타}} \text{에서 시작} / \lambda)$$

견적:

$$\begin{aligned} \eta_{\text{타는}}(1, q_{\text{타}}, \text{쿠티}) &= \text{파이}_{\text{타는 } \text{쿠티}}(\text{쿠티}) \\ \eta_{\text{타는}}(E_i, \text{쿠티}, \text{쿠티}) &= \sum_{\substack{\text{타는 } / \\ \text{제이}=1}} \alpha(1, t-1, q_{\text{제이}}, \text{타는 } \text{쿠티}) \text{타는 } \text{쿠티} \quad (1 < E_i) \\ \eta_{\text{타는}}(1, q_{\text{타}}, \text{쿠티}) &= \eta_{\text{타는}}(1, q_{\text{타}}, \text{쿠티}) \text{파이}_{\text{타는 } \text{쿠티}}(\text{쿠티}) \\ \eta_{\text{타는}}(E_i, \text{쿠티}, \text{쿠티}) &= \sum_{\substack{\text{타는 } / \\ \text{제이}=1}} \eta_{\text{타는}}(E_i, \text{쿠티}, q_{\text{제이}}^2) \alpha(E_i, t-1, q_{\text{제이}}, \text{쿠티}) \text{타는 } \text{쿠티} \\ &+ \eta_{\text{타는}}(E_i, \text{쿠티}, q_{\text{타}}) \text{파이}_{\text{타는 } \text{쿠티}}(\text{쿠티}) \quad (1 < E_i) \end{aligned}$$

정의:

$$\eta_{\text{밖으로}}(E_l, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}) = \Pi(\overline{\text{큐}}_{\text{다-1}} \text{에 완료 } E_l \text{에 } E_{l+1} \cdots \text{영형 } E_l / \lambda)$$

전적:

$$\begin{aligned} \eta_{\text{밖으로}}(E_l, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}) &= \sum_{\substack{\text{제이}=1 \\ \text{제이}=1}}^{\overline{\text{큐}}_{\text{다-1}}} \beta(E_{l+1}, E_l, q_{\text{제이}}^2, \overline{\text{큐}}_{\text{다-1}}) \quad (E_l < E_l) \\ \eta_{\text{밖으로}}(E_l, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}^{-1}) &= \sum_{\substack{\text{제이}=1 \\ \text{제이}=1}}^{\overline{\text{큐}}_{\text{다-1}}} \beta(E_{l+1}, k, q_{\text{다제이}, \overline{\text{큐}}_{\text{다-1}}}^{-1}) \cdot \eta_{\text{밖으로}}(\text{제이}, \overline{\text{큐}}_{\text{다-1}}^{C_{l-1}}, q_{\text{다-1}}^{-1}) \\ &\quad + \sum_{\substack{\text{제이}=1 \\ \text{제이}=1}}^{\overline{\text{큐}}_{\text{다-1}}} \beta(E_{l+1}, k, q_{\text{다제이}, \overline{\text{큐}}_{\text{다-1}}}^{-1}) \cdot \eta_{\text{밖으로}}(\text{제이}, \overline{\text{큐}}_{\text{다-1}}^{C_{l-1}}, q_{\text{다-1}}^{-1}) \quad (E_l < E_l) \\ \eta_{\text{밖으로}}(E_l, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}^{C_{l-1}}) &= \sum_{\substack{\text{제이}=1 \\ \text{제이}=1}}^{\overline{\text{큐}}_{\text{다-1}}} \beta(E_{l+1}, k, q_{\text{다제이}, \overline{\text{큐}}_{\text{다-1}}}^{-1}) \cdot \eta_{\text{밖으로}}(\text{제이}, \overline{\text{큐}}_{\text{다-1}}^{C_{l-1}}, q_{\text{다-1}}^{-1}) \end{aligned}$$

정의:

$$\begin{aligned} \xi(E_l, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}) &= \Pi(\overline{\text{큐}}_{\text{다-1}} \text{에 } E_l \text{에 } E_{l+1} \text{에서 시작 } 1/\lambda, \text{오}) \\ &= \Pi(\text{영형 } 1 \cdots \text{영형 } E_l, q_{\text{다-1}} \text{ 나 } \rightarrow \overline{\text{큐}}_{\text{다-1}}, \text{영형 } E_{l+1} \cdots \text{영형 } E_l / \lambda, \text{오}) \end{aligned}$$

전적:

$$\begin{aligned} \xi(E_l, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}) &= \frac{\alpha(1, t, q_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}) \cdot \beta(E_{l+1}, E_l, q_{\text{다-1}}^2, \overline{\text{큐}}_{\text{다-1}})}{\Pi(\text{오}/\lambda)} \quad (E_l < E_l) \\ \xi(E_l, \overline{\text{큐}}_{\text{다-1}}^2, \overline{\text{큐}}_{\text{다-1}}^2, \overline{\text{큐}}_{\text{다-1}}^2) &= \frac{\alpha(1, E_l, q_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}) \cdot \beta(E_{l+1}, E_l, q_{\text{다-1}}^2, \overline{\text{큐}}_{\text{다-1}})}{\Pi(\text{오}/\lambda)} \\ \xi(E_l, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}^{-1}) &= \frac{1}{\Pi(\text{오}/\lambda)} \left[\sum_{\substack{\text{제이}=1 \\ \text{제이}=1}}^{\overline{\text{큐}}_{\text{다-1}}} \beta(E_{l+1}, \text{제이}, q_{\text{다제이}, \overline{\text{큐}}_{\text{다-1}}}^{-1}) \cdot \eta_{\text{밖으로}}(\text{제이}, \overline{\text{큐}}_{\text{다-1}}^{C_{l-1}}, q_{\text{다-1}}^{-1}) \right] \\ \xi(E_l, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}^{C_{l-1}}) &= \frac{1}{\Pi(\text{오}/\lambda)} \left[\sum_{\substack{\text{제이}=1 \\ \text{제이}=1}}^{\overline{\text{큐}}_{\text{다-1}}} \beta(E_{l+1}, \text{제이}, q_{\text{다제이}, \overline{\text{큐}}_{\text{다-1}}}^{-1}) \cdot \eta_{\text{밖으로}}(\text{제이}, \overline{\text{큐}}_{\text{다-1}}^{C_{l-1}}, q_{\text{다-1}}^{-1}) \right] \quad (E_l < E_l) \\ \xi(E_l, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}, \overline{\text{큐}}_{\text{다-1}}^{C_{l-1}}) &= \frac{1}{\Pi(\text{오}/\lambda)} \left[\sum_{\substack{\text{제이}=1 \\ \text{제이}=1}}^{\overline{\text{큐}}_{\text{다-1}}} \beta(E_{l+1}, \text{제이}, q_{\text{다제이}, \overline{\text{큐}}_{\text{다-1}}}^{-1}) \cdot \eta_{\text{밖으로}}(\text{제이}, \overline{\text{큐}}_{\text{다-1}}^{C_{l-1}}, q_{\text{다-1}}^{-1}) \right] \quad (E_l < E_l) \end{aligned}$$

정의:

$$\begin{aligned} \chi(E_t, \bar{q}_{t-1}, \bar{q}_{t-1}) &= \frac{\pi(\bar{q}_{t-1} \text{에서 시작} | \lambda, \omega)}{\bar{q}_{t-1}} \\ &= \pi(\text{영형}_t \cdots \text{영형}_{t-1}, \downarrow, \text{영형}_t \cdots \text{영형}_t | \lambda, \omega) \bar{q}_{t-1} \end{aligned}$$

견적:

$$\begin{aligned} \chi(1, q_{t-1}, \bar{q}_{t-1}) &= \frac{\pi(\omega | \bar{q}_{t-1}) \beta(1, E_t, q_{t-1}, \bar{q}_{t-1})}{\pi(\omega | \lambda)} \\ \chi(E_t, \bar{q}_{t-1}, \bar{q}_{t-1}) &= \frac{\eta_{\text{인애}}(E_t, \bar{q}_{t-1}, q_{t-2}) \pi(\omega | \bar{q}_{t-1})}{\pi(\omega | \lambda)} \\ &= \left[\sum_{\bar{q}_{t-1}} \beta(t, e, q_{t-1}, \bar{q}_{t-1}) \eta_{\text{밖으로}}(\text{전자}, q_{t-1}, q_{t-2}) \right] \quad (2 < t) \end{aligned}$$

이항화-리

부록 B

일반화된 Viterbi 알고리즘

독자에게 상기시키기 위해 각 sta(t 쿼이자형 t에스-1, qd t)세 가지 변수를 유지합니다.

- $\delta(E_t, E_t \text{ 케이}, \bar{q}_{t-1}, \bar{q}_{t-1})$ 가장 가능성 있는 상태 시퀀스가 생성될 가능성입니다.
영형_t...영형_{t+케이}에서 시작된 재귀 활성화에 의해서만 생성되었다고 가정합니다.
시간 단계 E/주에서 \bar{q}_{t-1} 그리고 종료 쿼티 d t 나로 돌아온 \bar{q}_{t-1} 시간 단계에서 E/ 케이.
- $\psi(E_t, E_t \text{ 케이}, \bar{q}_{t-1}, \bar{q}_{t-1})$ 활성화될 가능성이 가장 높은 상태의 인덱스입니다. \bar{q}_{t-1} 와이₁
~ 전에 \bar{q}_{t-1} 그러한 상태가 존재하지 않는 경우 영형_t(...영형_{t+케이}) 단독으로 생성되었습니다. \bar{q}_{t-1} 이 d t) 우리는 설정
 $\psi(E_t, E_t \text{ 케이}, \bar{q}_{t-1}, \bar{q}_{t-1}) = 0$.
- $\tau(E_t, E_t \text{ 케이}, \bar{q}_{t-1}, \bar{q}_{t-1})$ 는 시간 단계입니다. \bar{q}_{t-1} 를 가능성이 가장 높았다.
 \bar{q}_{t-1} 만약에 \bar{q}_{t-1} 전체 하위 시퀀스를 생성했습니다. $\psi_{\text{에스}}(E_t \text{ 이자형}, E_t \text{ 케이}, \bar{q}_{t-1}, \bar{q}_{t-1}) = E_t$.

표기법을 단순화하기 위해 함수를 정의합니다. $\pi(\omega | \lambda)$ 도가 매개변수가 함수인
에프 그리고 유한 se에프,

$$\begin{aligned} & \left(\right) \\ & \text{MAX}_{\omega \in \text{에스}} \{ \text{에프}(\omega) \} \text{디크에프} \text{최대} \{ \text{에프}(\omega) \}, \text{인수} \text{최대} \{ \text{에프}(\omega) \} \\ & \omega \in \text{에스} \end{aligned}$$

일반화된 Viterbi 알고리즘은 생산 상태 및 계산에서 시작합니다. δ, ψ 는 그리고
아래와 같이 상향식으로.

생산 상태 :

1. 초기화:

$$\delta(t, t, q_{t-1}, \bar{q}_{t-1}) = \pi(\omega | \bar{q}_{t-1}) \beta(\bar{q}_{t-1}, \text{영형}_t) \quad \psi(t, t, q_{t-1}, \bar{q}_{t-1}) = 0 \quad \tau(t, t, q_{t-1}, \bar{q}_{t-1}) = E_t$$

2. 재귀:

$$\begin{aligned} & \left(\delta(E_l, E_{l+1} \text{ 케이}, \overline{q}_{d-1}, \overline{q}_{d-1}), \psi(E_l, E_{l+1} \text{ 케이}, \overline{q}_{d-1}, \overline{q}_{d-1}^{l-1}) \right) = \\ & \quad \text{MAX}_{1 \leq \text{제이} \leq |\overline{q}_{d-1}|} \left\{ \delta(E_l, E_{l+1} \text{ 케이} - 1, q_{\text{제이}}^{d-1}) \vdash \overline{q}_{\text{제이}}^{d-1} \right. \\ & \quad \left. \left(\psi(E_l, E_{l+1} \text{ 케이}, \overline{q}_{d-1}, \overline{q}_{d-1}^{l-1}) \right) \right\} \\ & \tau(E_l, E_{l+1} \text{ 케이}, \overline{q}_{d-1}, \overline{q}_{d-1}) \vdash E_{l+1} \text{ 케이} \end{aligned}$$

내부 상태:

1. 초기화:

$$\delta(t, t, q_{d-1}, \overline{q}_{d-1}) = \text{최대}_{1 \leq \text{제이} \leq |\overline{q}_{d-1}|} \left\{ \text{파이}_{\overline{q}_{d-1}}(\overline{q}_{d-1} \delta(t, t, q_{d-1}^{d-1}, q_{\text{제이}}^{d-1} \text{ 아르 자형 끝})) \right\}$$

$$\psi(t, t, q_{d-1}, \overline{q}_{d-1}) = 0 \quad \tau(t, t, q_{d-1}, \overline{q}_{d-1}) = E_l$$

2. 재귀:

(가) $E_l' = E_l + 1, \dots, E_{l+1} \text{ 케이}$ 세트:

$$\begin{aligned} & \text{아르 자형} = \text{최대}_{1 \leq \text{아르 자형} \leq |\overline{q}_{d-1}|} \left\{ \delta(E_l', E_{l+1} \text{ 케이}, \overline{q}_{d-1}, q_{\text{아르 자형}}^{d-1}) \right\} \\ & \left(\Delta(E_l'), \Psi(E_l') \right) = \text{MAX}_{1 \leq \text{제이} \leq |\overline{q}_{d-1}|} \left\{ \delta(t, t'-1, q_{\text{제이}}^{d-1}, \overline{q}_{d-1}^{l-1}) \vdash \overline{q}_{\text{제이}}^{d-1} \right. \\ & \quad \left. \left(\psi(E_l', E_{l+1} \text{ 케이}, \overline{q}_{d-1}, \overline{q}_{d-1}^{l-1}) \right) \right\} \end{aligned}$$

(나) E_l 세트:

$$\Delta(E_l) = \text{파이}_{\overline{q}_{d-1}}(\overline{q}_{d-1}) \text{ 최대}_{1 \leq \text{아르 자형} \leq |\overline{q}_{d-1}|} \left\{ \delta(E_l, E_{l+1} \text{ 케이}, \overline{q}_{d-1}, q_{\text{아르 자형}}^{d-1}) \right\}$$

$$\Psi(E_l) = 0$$

(C) 가장 가능성이 높은 스위칭 시간 찾기:

$$\begin{aligned} & \left(\delta(E_l, E_{l+1} \text{ 케이}, \overline{q}_{d-1}, \overline{q}_{d-1}), \tau(E_l, E_{l+1} \text{ 케이}, \overline{q}_{d-1}, \overline{q}_{d-1}^{l-1}) \right) = \text{MAX}_{E_l' \leq E_l'' \leq E_{l+1} \text{ 케이}} \left(\Delta(E_l'') \right) \\ & \psi(E_l, E_{l+1} \text{ 케이}, \overline{q}_{d-1}, \overline{q}_{d-1}^{l-1}) = \psi(\tau(E_l, E_{l+1} \text{ 케이}, \overline{q}_{d-1}, \overline{q}_{d-1}^{l-1})) \end{aligned}$$

마지막으로 가장 가능성이 높은 상태 시퀀스의 확률은 다음과 같이 구합니다.

$$\left(\Pi_l, q_l \right) \text{ 마지막} = \text{MAX}_{\overline{q}_{d-1}} \left\{ \delta(1, E_l, q_{d-1}, \overline{q}_{d-1}) \right\},$$

그리고 가장 가능성 있는 상태 시퀀스 자체는 th를 스캔하여 찾을 수 있습니다. $\psi_{\text{eainsdts}} \tau$ 시작 ~에서 $\tau(1, E_l, q_l \text{ 마지막}, q_1)$ 그리고 $\psi(1, E_l, q_l \text{ 마지막}, \overline{q}_{d-1})$.

부록 C

표 C.1. 기호 및 변수 목록

상징	정의	부분
Σ	유한 알파벳	2
$\text{영형} = \text{영형}_1 \text{영형}_2 \dots \text{영형}_T \quad (\text{영형}_{L_t} \in \Sigma)$	관찰 순서	2
$D \in \{1, \dots, D\}$	계층 깊이	2
$\bar{q}_{L_t}^D$	그만큼 L_t -레벨의 t th 하위 상태 D -열	2
$\Pi_{\bar{q}^D} = \{\pi_D(\bar{q}_{L_t}^D) \} = \{\pi(\bar{q}_{L_t}^D q^D)\}$	초기 하위 상태 분포	2
$\iota_{\bar{q}^D} = \{\iota_{\bar{q}_{L_t}^D} \} = \{\pi(\bar{q}_{L_t}^D q_{L_t}^{D+1}) \}$	하위 상태 전이 확률	2
$b/\bar{q}^D = \{b/\bar{q}_{L_t}^D \text{케이}\} = \{\pi(\sigma_{\text{케이}} q_{L_t}^D)\}$	출력 확률 분포	2
$\lambda = \{\{\iota_{\bar{q}^D}\} \cap \{\bar{q}^D\} \}$	HHMM의 매개변수 세트	2
$\alpha(E, E_+ \text{케이}, \bar{q}_{L_t}, \bar{q}_{D-1})$	$\pi(\text{영형}_1 \dots \text{영형}_{T+} \text{케이}, q_{D-1}^D \text{파이에} \text{마쳤다 } E_+ \text{케이} / \bar{q}_{D-1})$ 에 시작됨	3.1
$\beta(E, E_+ \text{케이}, \bar{q}_{L_t}, \bar{q}_{D-1})$	$\pi(\text{영형}_1 \dots \text{영형}_{T+} \text{케이} q_{D-1}^D \text{에서 시작}, \bar{q}_{D-1} \text{파이에} \text{마쳤다 } E_+ \text{케이})$	3.1
$\xi(E, \bar{q}_{L_t}, \bar{q}_{D-1})$	$\pi(\text{영형}_1 \dots \text{영형}_T, q_{D-1}^D \rightarrow \bar{q}_{D-1}^D \text{케이}, \text{영형}_{T+1} \dots \text{영형}_T / \lambda)$	삼
$\gamma_{\text{연애}}(E, \bar{q}_{L_t}, \bar{q}_{D-1})$	$\sum_{\text{케이}=1}^{ q_{D-1}^D } \xi(E -1, q_{D-1}^D, \bar{q}_{D-1}) $	삼
$\gamma_{\text{백으로}}(E, \bar{q}_{L_t}, \bar{q}_{D-1})$	$\sum_{\text{케이}=1}^{ q_{D-1}^D } \xi(E, \bar{q}_{L_t}, \bar{q}_{D-1}, q_{D-1}^D) $	삼
$\chi(E, \bar{q}_{L_t}, \bar{q}_{D-1})$	$\pi(\text{영형}_1 \dots \text{영형}_{T+1}, \downarrow_{\bar{q}_{L_t}^D}, \text{영형}_T \dots \text{영형}_T / \lambda) \bar{q}_{L_t}^D$	삼
$\eta_{\text{연애}}(E, \bar{q}_{L_t}, \bar{q}_{D-1})$	$\pi(\text{영형}_1 \dots \text{영형}_{T+1}, q_{D-1}^D \text{에서 시작} / \lambda)$	ㅏ
$\eta_{\text{백으로}}(E, \bar{q}_{L_t}, \bar{q}_{D-1})$	$\pi(\bar{q}_{L_t}^D \text{파이에} \text{마쳤다 } E \text{케이}_{T+1} \dots \text{영형}_T / \lambda)$	ㅏ
$\delta(E, E_+ \text{케이}, \bar{q}_{L_t}, \bar{q}_{D-1})$	값 (δ), 상태 목록 ψ (), 전환 시간 $\tau()$	비
$\psi(E, E_+ \text{케이}, \bar{q}_{L_t}, \bar{q}_{D-1})$	가장 가능성이 높은 세대 $\text{영형}_T \text{에프} \dots \text{영형}_{T+} \text{케이}$	
$\tau(E, E_+ \text{케이}, \bar{q}_{L_t}, \bar{q}_{D-1})$	에 의해 시작 \bar{q}_{D-1} ~에 E 에 의해 종료 \bar{q}_{L_t} ~에 $E_+ \text{케이}$	

감사의 말

도움이 되는 의견을 주신 Yoshua Bengio와 Raj Iyer에게 감사드립니다.

참조

- Abe, N. & Warmuth, M. (1992). 확률에 의한 근사 분포의 계산 복잡성
오토마타. *기계 학습*, 9:205-260.
- Baldi, P., Chauvin, Y., Hunkapiller, T. & McClure, M. (1994). 생물학적 기본의 숨겨진 Markov 모델
시퀀스 정보. *큰 피조. Nat. Acd. 과학. (미국)*, 91(3):1059-1063.
- Baum, LE & Petrie, T. (1966). 유한 상태 Markov 체인의 확률 함수에 대한 통계적 추론.
수학 통계 연대기, 37.
- Bengio, Y. & Frasconi, P. (1995). 입출력 HMM 아키텍처. G. Tesauro, DS Touretzky 및 TK에서
린, 편집자, *신경 정보 처리 시스템의 발전, 미디어 피에스 7* ges 427-434. MIT 프레스, 케임브리지,
엄마.
- 커버, T. & 토마스, J. (1991) *이자형. 정보 테오의 요소. 라이와일리*.
- Dempster, AP, Laird, NM & Rubin, DB (1977). EM을 통한 불완전한 데이터의 최대 가능성
연산. *왕립통계학회지, 시리즈, 에스삼비* 9:1~38.
- Gat, I., Tishby, N. & Abeles, M. (1997). 동시에 기록된 셀의 은닉 마르코프 모델링
행동하는 원숭이의 연합 피질. *네스 이자형. twork: 신경계의 계산, 미디어 8* 에스:3, 페이지 297-322.
- Ghahramani Z. & Jordan, MI (1997). 팩토리얼 은닉 마르코프 모드. *미디어* 엘자, *피에스 씨, 하인 학습*, 29: 245-273.
- Gillman, D. & Sipser, M. (1994). 숨겨진 Markov 체의 추론 및 최소화. *피에스 씨, 하인 학습*, 29: 245-273.
안에 아르 자형에스 영형. 의 종파
158. 엘리넥, F. (1985) 숨겨진 Markov m을 사용한 강력한 품사 태깅. *영형 미디어* 엘자. *TJ 왓슨 리서치*
기술 레포. RT 엘리
- 네크, F. (1983). 텍스트 생성의 Markov 소스 모델링. *기술적 재* *피에스 RT, TJ 왓슨 리서치*
센터 기술 레포. RT
- 엘리넥, F. (1985). 음성 인식을 위한 자가 조직 언어 모델링. *피에스 RT* *에. 왓슨 연구센터*
기술 레포. RT
- Krogh, A., Mian, SI & Haussler, D. (1994). E.coli DNA에서 유전자를 찾는 Hidden Markov 모델. *NAR*,
22:4768-4778.
- Lari, K. & Young, SJ (1990). Inside-Outside를 이용한 확률적 문맥 자유 문법 추정
연산. *컴퓨터 음성 및 언어, 4* *로마자*
- Nag, R., Wong, KH & Fallside, F. (1985). Hidden Markov mod를 이용한 스크립트 인식. *피자형 아르 자형에스 영형에스 씨. 의 먹이*
음향 음성 및 신호 처리에 관한 국제 회의, 에스피에스 4 *로마자* 2074.
- Rabiner, LR & Juang, BH (1986). 숨겨진 Markov 모드 소개. *미디어 이자형에스. EASSP 매거진, 이자형 3* 번.
- Rabiner, LR (1989). 음성 인식에서 숨겨진 Markov 모델 및 선택된 응용 프로그램에 대한 자습서입니다.
IEEE 절차. 이자형
- Rissanen, J. (1986). Markov 사위 클래스의 문자열 복잡성. *미디어 이자형에스 이자형. 정보 거래*
이론, 32(4):526-532.
- Ron, D., 가수, Y. & Tishby, N. (1996). 기억 상실의 힘: 변수를 이용한 확률론적 오토마타 학습
메모리 길이. *기계 학습*, 25:117-149.
- 가수, Y. & Tishby, N.. (1994). 필기체 필기체의 동적 인코딩. *미디어 영형. 논리적 사이버네틱스, 에스 71* (3):227-
237.
- 가수, Y. & Warmuth, MK (1997). 엔트로피 기반 거리를 사용하는 숨겨진 Markov 모델의 학습 알고리즘
기능. MC Mozer, MI Jordan 및 T. Petsche에서 편집, *피에스 V, 신경 정보 처리의 ances*
시스템, 9641-647쪽. MIT Press, 케임브리지, 메사추세츠.
- Stolcke A. & Omohundro, SM (1994). 숨겨진 Markov 모델 유도를 위한 Best-first 모델 병합. *미디어 이자형에스 이자형. 니칼*
ICSI TR-94-00 신고, 삼
- Viterbi, AJ (1967). 경련 코드에 대한 오류 범위 및 점근적으로 최적인 디코딩 알고리즘.
Information Theo에 대한 IEEE 트랜잭션, 아르 자형 1 와 이 3:260-269.

1996년 2월 21일 접수 1997년 9월 25일 최

종원고 1997년 9월 29일 접수