확률 조건을 만족하는 개선된 은닉 마르코프 모델

김 호 연[O], 조 성 배, 김 진 형
한국과학기술원 전산학과 및 인공지능연구센터

Refined Hidden Markov Models to Satisfy Probabilistic Constraints

Ho-yon Kim, Sung-Bae Cho and Jin H. Kim
Center for Artificial Intelligence Research and Computer Science Department, KAIST

요 약

음성인식이나 문자인식과 같이 시간에 따라 변화하는 입력열의 클래스를 결정하는 문제에 널리 응용되고 있는 은닉 마르코프 모델은 그 출력이 $P(y|\lambda)$라는 가정을 기초로 이루어 진다. 은닉 마르코프 모델의 출력이 $P(y|\lambda)$가 되려면 모델의 출력이 확률조건인 $\sum_y P(y|\lambda) = 1.0$을 만족해야 한다. 하지만, 일반적인 은닉 마르코프 모델은 이러한 조건을 만족시키지 못한다. 본 논문에서는 확률조건을 만족시키기 위한 모델의 제약조건을 제시하고, 이러한 조건이 있을 때 모델의 출력이 확률조건을 만족한다는 것을 증명하였다. 여기에서 제시된 제약조건하에서 구성된 은닉 마르코프 모델을 사용하면 출력값을 직접 확률값으로 간주할 수 있기 때문에 후처리 과정에서 사용되는 여러가지 확률값과 쉽게 결합할 수 있을 것이다.

1 서론

은닉 마르코프 모델(HMM, Hidden Markov Model)은 순차적으로 발생하는 입력신호의 시간적 변형을 모델링하는데 뛰어난 모델로 음성인식등의 분야에서 성공적으로 사용되어 왔으며[LEV82], [RAB88], 최근에는 영어나 한글 문자의 인식에도 적용되어[HA92], [SIN92] 좋은 결과를 보이고 있다. 이와같은 인식문제에 사용된 은닉 마르코프 모델은 주어진 모델이 입력 패턴을 생성할 조건부 확률값을 출력한다. 예를들면 주어진 모델이 λ이고 입력이 y라고 할 때, 모델의 출력은 $P(y|\lambda)$가 된다. 따라서, 모델의 출력 $P(y|\lambda)$는 확률조건인 $\sum_y P(y|\lambda) = 1.0$을 만족해야 한다.

은닉 마르코프 모델을 이러한 조건이 만족되도록 구성하지 않으면, 입력패턴에 가장 가까운 모델을 결정하는 것과 같이 클래스를 구분하는 인식문제에는 직접 적용할 수 없다는 문제가 있다. 주어진 입력에 가까운 클래스를 결정하기 위해서 $\max_\lambda P(\lambda|y) = \max_\lambda \{P(y|\lambda)P(\lambda)/P(y)\}$인 λ를 결정해야 하며, 이때 $P(y|\lambda)$는 은닉 마르코프 모델의 출력을 이용할 수 있다. 하지만, 은닉 마르코프 모델의 출력이 확률값이 아니라면, 이러한 수식을 적용할 수 없다. 만일 이값을 그대로 사용하면 $\sum_y P(y|\lambda)$ 값이 큰 모델이 더 높은 확률분포를 갖기 때문에 유리하게 되는 문제가 발생한다.

본 논문에서는 일반적인 은닉 마르코프 모델이 이러한 확률조건을 만족하지 않음을 보이고, 이러한 확률조건을 만족시키는 제약조건을 제시한다. 또, 은닉 마르코프 모델에서 확률적 정보를 얻으려면 구성된 모델이 본 논문에서 제안한 조건을 만족하도록 설계되어야 함을 보인다. 논문의 구성은 다음과 같다. 2장에서는 은닉 마르코프 모델의 기본 개념및 문제점을 설명한다. 다음 3장에서는 이러한 문제를 해결하기위한 제약된 모델을 제시하고 이것이 확률조건을 만족시킴을 증명한다.

2 은닉 마르코프 모델

2.1 정의

은닉 마르코프 모델은 상태노드(state)에서 출력을 내는 것과 상태노드를 전이(transition)할 때 출력을 내는 것으로 각각 정의할 수 있다. 이 두가지 정의는 동등하다는 것이 [CAS90]에 증명되어 있다. 이 논문에서는 상태노드를 전이할 때 출력을 내는 것으로 은닉 마르코프 모델을 정의하며 세부사항은 다음과 같다.

- N: 모델에서 상태의 수

† 이 논문은 노트패드 콘솔시움의 연구비 지원에 의한 것임

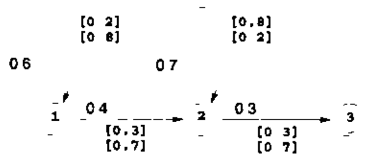

그림 1 left-right 모델

상태집합 $S = \{s_1, s_2, \ldots, s_N\}$
q_t 어떤 시간 t에서의 상태

- M 각 전이에서 관측가능한 심볼의 수
  심볼집합 $V = \{v_1, \ldots, v_M\}$
- $A = a_{ij}$ 상태 전이 확률 분포
  $a_{ij} = P(q_{t+1} = s_j | q_t = s_i)$ $\sum_j a_{ij} = 1$, $1 \le i, j \le N$
- $B = b_{ij}(k)$ 관측 심볼 확률 분포
  $b_{ij}(k) = P(v_k | q_t = s_i, q_{t+1} = s_j)$ $\sum_k b_{ij}(k) = 1$, $1 \le i, j \le N$ $1 \le k \le M$
- $\Pi = \pi_i$ 초기 상태 전이 확률 분포
  초기 상태 $t = 0$일 때 각 상태 s_i의 확률 분포
  $\pi_i = P(q_0 = s_i)$,
  $\sum_i \pi_i = 1$ $1 \le i \le N$

(그림 1)은 실제 응용에서 가장 널리 사용되고 있는 left-right 모델의 예이며 각 아크에 있는 숫자는 전이확률을 나타내고 괄호 안의 숫자는 출력확률을 나타낸다

2.2 확률적 모델로서의 문제점

서론에서 언급했듯이 은닉 마르코프 모델을 확률모델로 본다면 모델의 출력 $P(y|\lambda)$는 확률조건인 $\sum_y P(y|\lambda) = 1.0$을 만족해야 한다 하지만 일반적으로 은닉 마르코프 모델은 이러한 조건을 만족하지 않는다 예를들어 (그림 2)와 같은 모델에서, 가능한 모든 입력들에 대한 출력의 총합은 1.0이어야 한다 하지만 입력이 ab abb $abbb$ 일 경우 각각의 확률값은 $0.56 + 0.448 + 0.3584 \ge 1.3664 > 1.0$이 되므로 총 합이 1.0을 넘게 된다 따라서 이러한 모델의 출력은 $P(y|\lambda)$라고 말할 수 없다 최종 상태노드에서 다른 상태노드로 전이하도록 구성된 은닉 마르코프 모델은 이처럼 출력의 합이 1.0을 넘게 된다 다음 절에서는 은닉 마르코프 모델이 확률조건을 만족하기 위한 조건을 제시하고 이를 증명할 것이다

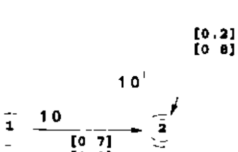

그림 2 확률조건을 만족하지 않는 은닉 마르코프 모델의 예

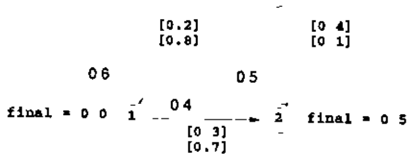

그림 3 제약된 은닉 마르코프 모델의 예

3 제약된 은닉 마르코프 모델

3.1 정의

은닉 마르코프 모델이 확률적 모델이 되기 위해서는 다음과 같은 조건이 부가되어야 한다 즉, 원래의 모델에서는 각 상태노드가 최종 상태(final state)가 되는데 아무런 제약이 없는데, 제약된 모델에서는 각 상태노드가 최종 상태가 되는 것을 (그림 3)과 같이 확률적으로 표현한다 이것을 다른 방식으로 설명하면 상태노드에서 다른 상태노드로 전이할 확률 뿐만 아니라 더이상 전이하지 않고 끝날 확률도 동시에 구해야 한다는 것을 의미한다

(그림 4-a)를 보면 각각의 상태노드가 최종 상태가 되는 것이 확률적으로 표현되어 있음을 알 수 있다 따라서 $P(y|\lambda)$를 구할 때 전이확률과 전이할 때 심볼이 발생할 확률 뿐만 아니라 그 상태노드가 최종 상태가 아닐 확률도 고려되어야 하며 모든 심볼에 대한 계산이 끝났을 때는 마지막 상태가 최종 상태일 확률이 고려되어야 한다는 것이다 이를 단순하게 null 전이가 있는 모델로 표현하면 (그림 4-b)에서와 같이 최종 상태를 따로 두고 각각의 상태에서 최종 상태가 될 확률로 null 전이하여 최종 상태로 전이하는 모델과 같다 이때 최종 상태에서는 전이가 없어야 한다 즉, 기존의 모델에서 최종 상태를 따로 두고 최종 상태에서는 전이가 없도

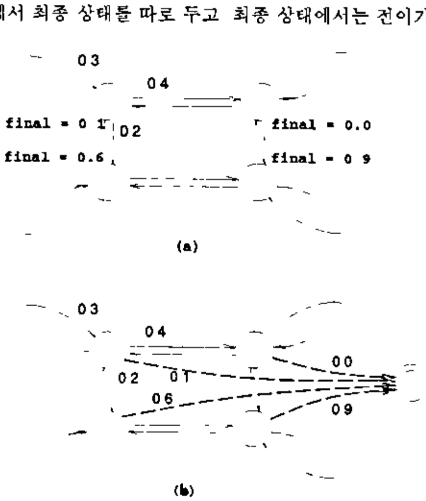

그림 4 제약된 은닉 마르코프 모델

록 한 모델과 같다. (이때, 각 상태에서 최종 상태를 포함한 다른 모든 상태로 전이할 확률의 합은 1.0이 되어야 한다.)

이것은 은닉 마르코프 모델이 다음과 같은 조건을 만족하도록 하여야 한다는 것을 의미한다.

$$\sum_j \sum_k \tau_{ijk} + f_i = 1.0 \quad \text{for each } i$$

단

- $\tau_{ijk} = a_{ij} b_{ij}(k)$: 상태 i에서 j로 출력 v_k를 내면서 전이할 확률
- f_i : 상태 i가 최종 상태가 될 확률
- π_i : 상태 i에서 시작할 확률

다음 절에서는 이러한 모델이 확률 조건을 만족한다는 것을 증명할 것이다.

3.2 확률 조건의 검증

앞 절에서 제안한 조건을 만족하는 은닉 마르코프 모델은 $\sum_y P(y|\lambda) = 1.0$을 만족한다는 것을 증명하기 위해서 먼저 다음과 같은 몇가지 사항을 정의한다.

[정의 1] obs(l)은 길이가 l인 모든 가능한 입력들의 집합이고, OBS(n)은 길이가 n보다 작거나 같은, 모든 가능한 입력들의 집합이다.

[정의 1]에 의하면 다음과 같은 성질이 성립된다.

[성질 1]

$$\begin{aligned} \mathrm{obs}(l) &= \{(v_1, v_2, v_3, \cdots, v_l) \mid \forall v_k \in V, 1 \le k \le l\} \\ \mathrm{OBS}(n) &= \bigcup_{i=0}^{n} \mathrm{obs}(i) \end{aligned}$$

[정의 2] $M_i(S)$는 i 상태에서 시작할 때($q_{t=0} = s_i$) 집합 S에 속한 모든 원소(입력)가 갖는 확률값의 총합이다.

[정의 2]에서 $\sigma_{ij} = \sum_k \tau_{ijk}$라고 하면 $M_i(\mathrm{OBS}(n))$는 다음과 같이 나타낼 수 있다.

$$M_i(\mathrm{OBS}(n)) = \begin{cases} f_i, & \text{if } n = 0 \\ \sum_j \sigma_{ij} M_j(\mathrm{OBS}(n-1)) + f_i & \text{if } n > 0 \end{cases} \tag{1}$$

[소정리 1]

$$M_i(\mathrm{OBS}(n)) = \sum \sigma_{ij} M_j(\mathrm{OBS}(n-1)) + f_i$$

$\sum_j \sigma_{ij} + f_i = 1.0$, $M_i(\mathrm{OBS}(0)) = f_i \le 1.0$, $1 \le i \le k$ 를 만족하는 $\lim_{n\to\infty} M_i(\mathrm{OBS}(n))$는 수렴한다.

증명

i) n이 1일 때 $M_i(\mathrm{OBS}(1)) = \sum_j \sigma_{ij} M_j(\mathrm{OBS}(0)) + f_i \le \sum_j \sigma_{ij} + f_i = 1.0$ 이므로 성립한다.

ii) $n < k$인 모든 n에 대해 성립한다고 가정하면, $M_i(\mathrm{OBS}(k-1)) \le 1.0$ 이다. 이를 이용하면 n이 k일 때, $M_i(\mathrm{OBS}(k)) = \sum_j \sigma_{ij} M_j(\mathrm{OBS}(k-1)) + f_i \le \sum_j \sigma_{ij} + f_i = 1.0$ 이므로 성립한다.

따라서 $M_i(\mathrm{OBS}(n)) \le 1.0$, for all $0 \le n < \infty$이므로

$$\lim_{n\to\infty} M_i(\mathrm{OBS}(n)) \le 1.0$$

이다. 정의에 의하면, 임의의 집합 S에 대해서 $M_i(\mathrm{S}) \ge 0.0$이고 $M_i(\mathrm{OBS}(n)) = M_i(\mathrm{OBS}(n-1)) + M_i(\mathrm{obs}(n)) \ge M_i(\mathrm{OBS}(n-1))$ 이다. 따라서 $M_i(\mathrm{OBS}(n))$은 n이 증가함에 따라 단조증가하며 bound되어 있으므로 수렴한다. ■

[성질 2]

$$M_i(S) = \sum_{y \in S} P(y|\lambda, q_{t=0} = s_i) \tag{2}$$

[성질 2]를 이용하면 $\sum_y P(y|\lambda)$는 다음과 같이 나타낼 수 있다.

$$\sum_y P(y|\lambda) = \lim_{n\to\infty} \sum_i \pi_i M_i(\mathrm{OBS}(n)) \tag{3}$$

따라서 $\sum_y P(y|\lambda) = 1.0$임을 보이기 위해서는 다음의 정리를 증명하면 된다.

[정리 1]

$$\lim_{n\to\infty} \sum_i \pi_i M_i(\mathrm{OBS}(n)) =$$
$$\lim_{n\to\infty} \sum_i \pi_i \left(\sum_j \sigma_{ij} M_j(\mathrm{OBS}(n-1)) + f_i\right) = 1.0$$

증명 [소정리 1]에 의하면 (식 1)는 $\lim_{n\to\infty} M_i(\mathrm{OBS}(n))$일 때 수렴하므로

$$\lim_{n\to\infty} M_i(\mathrm{OBS}(n)) = \lim_{n\to\infty} M_i(\mathrm{OBS}(n-1)) = X_i \tag{4}$$

로 나타낼 수 있다. (식 1)에 $\lim_{n\to\infty}$를 적용하면,

$$\lim_{n\to\infty} M_i(\mathrm{OBS}(n)) = \lim_{n\to\infty} \sum_j \sigma_{ij} M_j(\mathrm{OBS}(n-1)) + f_i \tag{5}$$

이므로 상태의 수가 k개 일 때 다음을 만족하는 연립방정식의 근을 구하는 문제로 볼 수 있다.

$$X_i = f_i + \sigma_{i1}X_1 + \sigma_{i2}X_2 + \cdots + \sigma_{ik}X_k \qquad (6)$$

단 $f_i + \sigma_{i1} + \sigma_{i2} + \cdots + \sigma_{ik} = 1.0$ and $\sigma_{ij} >= 0.0$ and $f_i > 0.0$ for all $1 \le i, j \le k$ 여기에서 [소정리 1]에서 알 수 있듯이 (식 5)의 $\lim_{n\to\infty} M_i(\mathrm{OBS}(n))$는 수렴된 하나의 근을 갖는다. 따라서 X_i도 하나의 근을 가져야 한다.

(식 6)을 각각의 i에 대해서 나타내면,

$$X_1 = f_1 + \sigma_{11}X_1 + \sigma_{12}X_2 + \cdots + \sigma_{1n}X_n$$
$$X_2 = f_2 + \sigma_{21}X_1 + \sigma_{22}X_2 + \cdots + \sigma_{2n}X_n$$
$$\vdots$$
$$X_n = f_n + \sigma_{n1}X_1 + \sigma_{n2}X_2 + \cdots + \sigma_{nn}X_n$$

단 $f_i + \sigma_{i1} + \cdots + \sigma_{in} = 1.0$ for all $1 \le i \le n$ 가 된다.

미지수가 n개인 n차 선형 연립방정식은 단 하나의 근을 갖거나 근이 없거나 무수한 근을 갖는다. (식 7)은 각 X_i에 1.0을 대입하면 성립함을 알 수 있다. 따라서 (식 7)은 적어도 하나 이상의 근을 갖음을 알 수 있다. 하지만 각 X_i는 bound되어 있고 수렴된 하나의 값이기 때문에 (식 7)은 $X_i = 1.0$ for all $1 \le i \le n$인 하나의 근을 갖는다. 이를 은닉 마르코프 모델의 관점에서 설명하면 다음과 같다. (식 7)만 본다면 계수에 따라 하나의 근이나 무수한 근을 가질 수 있다. 하지만 최종 상태에 도달할 확률이 없는 상태들만으로 이루어진 사이클이 없다는 가정을 하면 연립방정식 7을 소거법에 의해 풂음으로써 동일한 결과를 얻을 수 있다.

따라서, (식 4)에 의해 $\lim_{n\to\infty} M_i(\mathrm{OBS}(n)) = 1.0$ for all $1 \le i \le n$이므로 이를 (식 3)에 대입하면

$$\sum_y P(y|\lambda) = \lim_{n\to\infty} \sum_i \pi_i M_i(\mathrm{OBS}(n)) = \sum_i \pi_i X_i = \sum_i \pi_i = 1.0$$

이 성립한다. 따라서 제안된 조건을 만족하는 모델은 확률조건을 만족함이 증명된다. ■

4 결론및 토의

은닉 마르코프 모델이 확률적인 모델로 사용되기 위해서는 모든 가능한 입력에 대해 출력되는 확률값의 합이 1.0이 되어야 한다. 하지만 아무런 제약이 없는 일반적인 은닉 마르코프 모델을 사용하는 경우에는 이러한 조건이 만족되지 않기 때문에 은닉 마르코프 모델을 확률적 모델로 사용하는 것은 문제가 있다. 따라서 은닉 마르코프 모델이 확률적인 조건을 만족하도록 제약조건을 부가해야 할 것이다.

본 논문에서는 이러한 확률 조건을 만족시키기 위해서 최종 상태를 따로 두고 최종 상태에서는 밖으로 전이하거나 자신으로 전이하는 확률이 없도록 하여야 함을 주장하였다. 이를 다른 의미에서 본다면 각각의 상태가 최종 상태가 된다는 것도 확률적으로 표현되어 전이확률에 포함 되어야 한다고 할 수도 있다. 이러한 조건을 만족하는 은닉 마르코프 모델은 모든 입력에 대해 모델이 내는 출력의 합이 1.0을 만족한다는 것을 증명 하였다. 즉, 가능한 모든 입력들의 모든 경로확률의 합이 1.0이라는 것이다.

본 논문의 증명에서 유추할 수 있는 것은 은닉 마르코프 모델에서 하나의 경로확률만을 계산하는 것은 확률이 아니라는 것이다. 각 입력에 대해서 가능한 모든 상태 순서들의 출력 값을 더한 것이 확률조건을 만족하기 때문이다. 따라서, 경로의 확률을 최대화하는 은닉 마르코프 모델을 확률적으로 사용하는 것에 대해서는 앞으로 좀더 연구 되어야 할 것이다.

참고 문헌

[BAU70] L. E. Baum et al., "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains", *Ann. Math. Stat.* v.41, n.1, pp.164-171, 1970.

[CAS90] F. Casacuberta, "Some Relations Among Stochastic Finite State Networks Used in Automatic Speech Recognition", *IEEE Trans. on PAMI* v.12, n.7, Jul. 1990.

[EPH89] Y. Ephraim, A. Dembo, and L. R. Rabiner, "A minimum discrimination information approach for hidden Markov modeling", *IEEE Trans. Inform. Theory*, v.IT-35, n.5, pp.1001-1013, Sept. 1989.

[FOR78] G. D. Forney, "The Viterbi Algorithm", *Proc. IEEE*, v.61, pp.268-278, Mar. 1978.

[HA 92] 하진영, 김진형, "HMM을 이용한 온라인 영어 인식", 제6회 한국 인지과학회 춘계 학술발표논문집, pp. 264-274, 1992.

[LEV82] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition", *The Bell System Technical Journal*, v.62, n.4, pp.1035-1074, April 1983.

[RAB88] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Apllications in Speech Recognition", *Proc. of the IEEE*, v.77, n.2, pp.257-286, 1989.

[SIN92] 신봉기, 김진형, "통계적 방법에 의한 온라인 한글 필기 인식", 제4회 한글 및 한국어 정보처리 학술발표 논문집, pp. 533-542, 1992. 10.