論文 90-27-4-20

狀態의 占有時間 情報를 包含하는 Hidden Markov Model

(Hidden Markov Models Containing Durational Information of States)

曺 政 鎬* 洪 再 根* 金 秀 重*

(Jeong Ho Cho, Jae Keun Hong, and Soo Joong Kim)

要 約

Hidden Markov model(HMM)은 음성의 특징을 잘 표현하는 유용한 모델로서 다양한 음성 시스템에서 사용되고 있다. 이를 음성인식에 응용할 때는 발성시간에 상응하는 상태의 점유시간 정보를 모델에 포함시키는 것이 바람직하다.

이 논문에서는 left-to-right 모델에서 상태점유시간 정보를 포함하는 점유시간종속 HMM을 제안하였다. 제안한 모델의 파라미터 재추정식을 유도하였고 이 식의 수렴성도 확인하였다. 그리고 이 모델을 화자독립, 고립단어 인식 시스템에 적용하여 타당성을 확인하였다.

Abstract

Hidden Markov models(HMM's) have been known to be useful representations for speech signal and are used in a wide variety of speech systems. For speech recognition applications, it is desirable to incorporate durational information of states in the model which correspond to phonetic duration of speech segments.

In this paper we propose duration-dependent HMM's that include durational information of states appropriately for the left-to-right model. Reestimation formulae for the parameters of the proposed model are derived and their convergence is verified. Finally, the performance of the proposed models is verified by applying to an isolated word, speaker independent speech recognition system.

I. 序 論

지금까지 주로 사용된 音聲認識技法은 크게 패턴 인식방법[1-3] 및 hidden Markov model(HMM)을 이용한 방법[4-8]으로 나눌 수 있다. 패턴 인식방법은 입력음성의 特徵 벡터를 추출한 후, 이를 이미 저장되어있는 기준음성과 비교하여 類似度가 가장 높은 기준음성을 인식된 음성으로 하는 방법으로, 입력음성과 기준음성간의 時間整合을 위해 dynamic time warping(DTW) 알고리듬을 주로 사용한다. 이 방법은 훈련과정이 쉽고 인식률이 높으나, 話者獨立 認識時에는 계산량 및 메모리 필요량이 많다는 단점이 있다. 또한 연결음성인식 또는 연속음성인식등의 보다 어려운 문제로의 확장이 어렵다. 한편, HMM을 이용한 방법은 Markov chain의 確率函數를 이용하여 인식하는 방법으로 최근에 많이 연구되고 있다. 이 방법은 기준음성에 대한 HMM을 만든 후, 이들 모델로부터 입력음성의 觀測確率을 구하여 가장 높은 확률을 가

*正會員, 慶北大學校 電子工學科
(Dept. of Elec. Eng., Kyungpook Nat'l Univ.)
接受日字 : 1989年 8月 10日

지는 모델의 음성으로 인식한다. HMM은 Parameteric model이므로 단어 단위뿐만 아니라 音節, 半音節, 또는 音素單位로도 모델링이 가능하며 또한, 이의 저장에 필요한 메모리량 및 인식시의 계산량도 적은 장점이 있다. 음성은 발성구조가 시간적으로 변하여 발생된 신호이므로 보통 left-to-right 구조의 모델이 사용된다. 이와 같은 HMM에서 狀態의 占有時間(duration)은 음성 세그먼트의 발성시간 정보를 나타내므로 음성인식시 이를 고려할 필요가 있다. 1985년 Rabiner 等[8]은 상태의 점유시간정보를 인식시 고려함으로써 인식률이 개선됨을 보였으며, 1985년 Russell 等[9]은 狀態占有時間을 명확히 고려한 semi-Markov model을 제안하였다.

이 論文에서는 상태의 占有時間정보를 포함하는 HMM을 제안하고, 이 모델의 각 파라미터에 대한 再推定式을 유도하였으며, 이 식의 收斂性도 確認하였다. 또한, 訓練 및 認識 알고리듬을 기술하고, 이를 이용한 한국어 고립음 인식실험의 결과를 보였다.

Ⅱ. Hidden Markov model

Markov chain은 初期狀態確率벡터 $\mathbf{U}=(u_1, u_2, \cdots, u_N)^T$와 狀態遷移行列 $\mathbf{A}=[a_{i,j}]\ (1\le i, j\le N)$로 표현된다. N은 상태의 수이며 $a_{i,j}$는 현재의 상태 i에서 다음 상태 j로 천이할 확률을 나타낸다. 또 Markov chain에서 각 상태의 確率密度函數를 정의하는 파라미터의 집합을 $\mathbf{B}=\{b_j(\cdot)\}$로 표현하면, HMM은 $\lambda_h=(\mathbf{U}, \mathbf{A}, \mathbf{B})$로 나타낼 수 있다. Markov chain의 state sequence $\theta=(\theta_0, \theta_1, \theta_2, \cdots, \theta_T)$ (T는 sequence의 길이)가 주어질 때 observation sequence $S=(s_1, s_2, \cdots s_T)$의 관측확률은

$$f(S\mid\theta, \lambda_h)=\prod_{t=1}^{T} b_{\theta t}(s_t) \qquad (1)$$

로 표현된다. 또 모델 λ_h에서 S의 관측확률은

$$f(S\mid\lambda_h)=\sum_{all\theta} U_{\theta 0}\prod_{t=1}^{T} a_{\theta t-1,\theta t}b_{\theta t}(s_t) \qquad (2)$$

가 된다.

확률밀도함수 $b_j(s)$로는 다음과 같은 Gaussian autoregressive 밀도의 혼합형[8,10]을 사용하였다.

$$b_j(s)=\sum_{k=1}^{M} c_{j,k}b_{j,k}(s), \qquad (3)$$

여기서, M은 상태 j의 branch 수이다. 또 $c_{j,k}$는 상태 j의 k번째 branch의 weight이며

$$\sum_{k=1}^{M} c_{j,k}=1, \quad \text{for } j=1, 2, \cdots N \qquad (4)$$

을 만족해야 한다. $b_{j,k}(s)$는 상태 j의 k번째 branch의 기본 밀도함수로서

$$b_{j,k}(s)=(2\pi)^{-L/2}\exp\{-\frac{1}{2}\delta(s; a_{j,k})\} \qquad (5)$$

이며 이를 정의하는 파라미터는 LPC 自己相關函數(autocorrelation) $a_{j,k}$이다. 여기서 L은 observation vector s의 sample 수이며,

$$\delta(s;\ a_{j,k})=r_a(0)r(0)+2\sum_{i=1}^{P} r_a(i)r(i) \qquad (6)$$

이고, p는 AR 모델의 차수이다. 그리고, $r_a(i)$는 k번째 branch의 centroid에 해당하는 LPC의 자기상관함수이며, r(i)는 s의 이득정규화 자기상관함수(gain normalized autocorrelation)이다.

1. 상태점유시간(state duration)의 확률함수

음성인식에는 그림 1과 같은 left-to-right 구조의 HMM이 주로 사용된다. 여기서 상태 0 및 상태 N+1은 각각 음성의 시작과 끝을 나타낸다. 또한 초기상태 확률벡터 및 상태천이확률은 다음과 같은 조건을 만족시켜야 한다.

$$\mathbf{U}=[1, 0, 0, \cdots, 0]^T \qquad (7)$$

$$a_{i,i}+a_{i,i+1}=1, \quad \text{for } i=1, 2, \cdots, N \qquad (8)$$

이와같은 구조의 HMM에서, process가 시간 t에서 임의의 상태 j로 천이한 후, 시간 $t+\tau_j$에서 다른상태 i≠j로 천이할 확률, 즉 상태 j에서 占有時間이 τ_j일 확률 $D_j(\tau_j)$는

$$D_j(\tau_j)=a_{j,j}^{\tau_j-1}\sum_{\substack{i=1\\ i\neq j}}^{N} a_{j,i} \qquad (9)$$

가 된다. 그러므로 점유시간 τ_j가 커질수록, 상태 j에 머물 확률은 지수함수적으로 감소하게 된다. 그러나 실제 음성에 대한 狀態占有時間의 確率分布는 일반적으로 Gamma 分布函數에 가까우므로, 既存의 HMM으로는 이와 같은 점유시간의 분포를 제대로 나타내지 못함을 알 수 있다. 점유시간분포의 한 例로서, 단어"서울"에 대한 점유시간의 분포를 그림 2에 나타

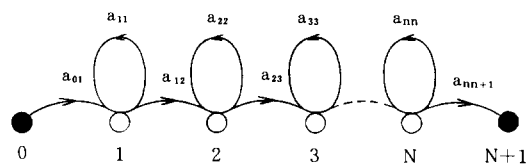


그림 1. left-to-right 구조의 HMM
Fig. 1. HMM of left-to-right structure.

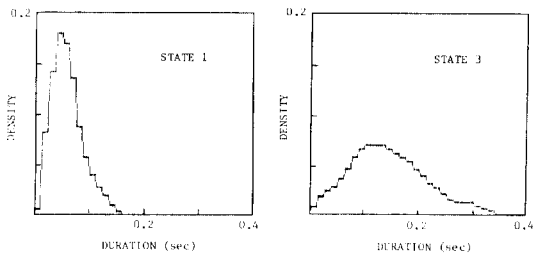


그림 2. 狀態占有時間 分布의 例示
Fig. 2. Illustration of the state duration distribution.

내었다. 이것은 50명의 成人男性이 단어"서울"을 15회씩 발음한 데이터로부터 5상태 HMM을 구성한 후 상태1과 3에서의 점유시간을 측정하여 구한 것이다. 전체적인 분포는 그림에서 보는 바와 같이 Gamma분포에 가깝게 나타남을 알 수 있다.

이와 같은 HMM에서 observation sequence S의 관측확률은

$$f(S \mid \lambda_h) = \sum_{\theta \in \Omega_S} \prod_{t=1}^{T} a_{\theta t-1, \theta t}\, b_{\theta t}(s_t)$$

$$= \sum_{\theta \in \Omega_S} \prod_{j=1}^{N} a_{j,j}^{\tau_j - 1}(1-a_{j,j}) \prod_{t=1}^{T} b_{\theta t}(s_t) \qquad (10)$$

로 나타낼 수 있다. 여기서 Ω_S는 left-to-right 구조에서 발생가능한 state sequence θ의 空間을 나타내며, τ_j는 θ를 따른 각 狀態別 占有時間을 나타낸다.

상태점유시간의 확률이 그림 2와 같은 任意의 分布를 가질 때, 이와 같은 분포를 나타내기 위한 狀態遷移行列은 다음과 같이 구할 수 있다. 먼저, 상태천이확률을 現在 狀態에 머문 시간 d에 따라 變化되도록 정의한다. 즉,

$$\hat{a}_{j,j}(d) \triangleq \text{prob}(\theta_{t+d}=j \mid \theta_{t+d-1}=\theta_{t+d-2}=\cdots=\theta_t=j, \theta_{t-1}=j-1, \theta, \lambda_h)$$
$$\hat{a}_{j,j}(0)=1 \qquad (11)$$
$$\hat{a}_{j,j+1}(d)=1-\hat{a}_{j,j}(d)$$

상태 j에서 점유시간의 확률을 $D_j(\tau_j)$라 할 때

$$\prod_{d=1}^{\tau_j-1} a_{j,j}(d)\,[1-\hat{a}_{j,j}(\tau_j)] = D_j(\tau_j) \qquad (12)$$

이므로 상태천이확률은

$$\hat{a}_{j,j}(d) = \frac{1-\sum_{n=1}^{d} D_j(n)}{1-\sum_{n=1}^{d-1} D_j(n)} \qquad (13)$$

로 된다. 본 논문에서는 observation sequence를 식(13)과 같은 상태천이확률을 가지는 HMM으로 부터 발생된 것으로 보고, 식(12)의 관계로부터 상태천이확률을 점유시간의 확률로 나타내었다. 또 점유시간의 확률을 나타내는 근사함수로는 Gamma 밀도함수 및 Gauss 밀도함수등의 2가지의 함수에 대해 연구하였다. 이들 함수는 점유시간 분포를 잘 나타낼 뿐만아니라 각각 2개씩의 파라미터만으로 정의되므로 모델貯藏時에도 간편하다. 점유시간의 확률분포 $D_j(\tau_j)$를 Gamma 밀도함수로 近似化할 경우

$$D_j(\tau_j) = \Gamma(\upsilon_j)^{-1}\, \eta_j^{\upsilon_j}\, \tau_j^{\upsilon_j-1} \exp(-\eta_j \tau_j) \qquad (14)$$

로 표현된다. 여기서 $\Gamma(\cdot)$는 Gamma 함수를 나타내며, υ_j 및 η_j는 Gamma 밀도함수를 정의하는 파라미터이다. 또 Gauss 밀도함수로 近似化할 경우에는

$$D_j(\tau_j) = (2\pi\sigma_j^2)^{-1/2} \exp\left[-\frac{(\tau_j-m_j)^2}{2\sigma_j^2}\right] \qquad (15)$$

이 되며, 여기서 m_j와 σ_j^2은 각각 상태 j에서 점유시간의 平均과 分散이다.

근사함수 $D_j(\tau_j)$를 정의하는 파라미터의 集合을 **D**로 나타내면, 본 論文에서 제안한 HMM은 $\lambda_d=(\mathbf{D}, \mathbf{B})$로 정의된다. 이 모델에서 **S**의 관측확률은 식 (10)과 (12)로부터

$$f(\mathbf{S} \mid \lambda d) = \sum_{\theta \in \Omega_S} \prod_{j=1}^{N} D_j(\tau_j) \prod_{t=1}^{T} b_{\theta t}(s_t) \qquad (16)$$

이 된다. 또한 branch sequence를 $K=(k_1, k_2, \cdots, k_T)$로 표기할 때

$$f(S, \theta \mid \lambda_d) = \prod_{j=1}^{N} D_j(\tau_j) \prod_{t=1}^{T} \left[\sum_{k=1}^{M} c_{\theta t,k}\, b_{\theta t,k}(s_t)\right]$$
$$= \sum_{k1=1}^{M} \sum_{k2=1}^{M} \cdots \sum_{kT=1}^{M} \left[\prod_{j=1}^{N} D_j(\tau_j) \prod_{t=1}^{T} b_{\theta t,kt}(s_t)\right] \cdot c_{\theta 1,k1}\, c_{\theta 2,k2} \cdots c_{\theta T,kT} \qquad (17)$$

으로 표현되며

$$f(S, \theta, K \mid \lambda_d) = \prod_{j=1}^{N} D_j(\tau_j) \left[\prod_{t=1}^{T} b_{\theta t,kt}(s_t) \cdot c_{\theta t,kt}\right] \qquad (18)$$

로 정의하면 식 (16)의 관측확률은

$$f(S \mid \lambda_d) = \sum_{\theta \in \Omega_S} \sum_{K \in \Omega_b} f(S, \theta, K \mid \lambda_d) \qquad (19)$$

가 된다. 여기서 Ω_b는 branch sequence의 공간을 나타낸다.

2. 파라미터 재추정 (parameter reestimation)

두 모델 λ_d 및 λ_d'의 補助函數 $Q(\lambda_d, \lambda_d')$를 다음과

$$Q(\lambda_d, \lambda'_d) \triangleq \sum_{\theta \in \Omega S} \sum_{K \in \Omega b} f(S, \theta, K \mid \lambda_d) \log f(S, \theta, K \mid \lambda'_d) \tag{20}$$

이미 발표된 여러 연구[11-13]에서, 이와 같은 보조함수에 대해, $Q(\lambda_d, \lambda'_d) \geqq Q(\lambda_d, \lambda_d)$이면 $f(S \mid \lambda'_d) \geqq f(S \mid \lambda_d)$, 즉 모델 λ'_d에서 S의 관측확률이 모델 λ_d에서의 관측확률보다 같거나 증가됨이 증명되어 있다. 단, 모델의 각 파라미터는 다음과 같은 log concavity를 만족해야 한다.

$$\frac{\partial^2}{\partial \lambda_d^2} \log [f(S \mid \lambda_d)] < 0 \tag{21}$$

각 파라미터에 대한 재추정식의 유도과정은 다음과 같다. 먼저

$$\log f(S, \theta, K \mid \lambda'_d) = \sum_{j=1}^{N} \log D'_j(\tau_j) + \sum_{t=1}^{T} \log b'_{\theta t, kt}(s_t) + \sum_{t=1}^{T} \log c'_{\theta t, kt} \tag{22}$$

로 분리되므로 식 (20)의 補助函數는 다음과 같이 나타낼 수 있다.

$$Q(\lambda_d \mid \lambda'_d) = \sum_{\theta} \sum_{K} f(S, \theta, K \mid \lambda_d) \left\{ \sum_{j=1}^{N} \log D'_j(\tau_j) + \sum_{t=1}^{T} \log b'_{\theta t, kt}(s_t) + \sum_{t=1}^{T} \log c'_{\theta t, kt} \right\} \tag{23}$$

$$= \sum_{j=1}^{N} Q_d(\lambda_d \mid D'_j) + \sum_{j=1}^{N} \sum_{k=1}^{M} Q_b(\lambda_d \mid b'_{jk}) + \sum_{j=1}^{N} Q_c(\lambda_d \mid c'_{jk})$$

여기서

$$Q_d(\lambda_d \mid D'_j) = \sum_{\theta} \sum_{K} f(S, \theta, K \mid \lambda_d) \log D'_j(\tau_j) \delta(\theta_t - j) = \sum_{\tau_j=1}^{T} \sum_{t=1}^{T-\tau_j} f(S, \tau_j \mid \lambda_d) \log D'_j(\tau_j) \tag{24}$$

이며

$$f(S, \tau_j \mid \lambda_d) = f(S, \theta_{t+\tau_j} = j+1, \theta_{t+\tau_j-1} = \theta_{t+\tau_j-2} = \cdots \theta_t = j, \theta_{t-1} = j-1 \mid \lambda_d) \tag{25}$$

이다. 또,

$$Q_b(\lambda_d \mid b'_{jk}) = \sum_{\theta} \sum_{K} f(S, \theta, K \mid \lambda_d) \sum_{t=1}^{T} \log b'_{\theta t, kt}(s_t) \delta(\theta_t - j) \delta(k_t - k) = \sum_{t=1}^{T} f(S, \theta_t = j, k_t = k \mid \lambda_d) \log b'_{jk}(s_t) \tag{26}$$

$$Q_c(\lambda_d \mid c'_{jk}) = \sum_{\theta} \sum_{k} f(S, \theta, K \mid \lambda_d) \sum_{t=1}^{T} \log c'_{\theta t, kt} \delta(\theta_t - j) = \sum_{k=1}^{M} \sum_{t=1}^{T} f(S, \theta_t = j, k_t = k \mid \lambda_d) \log c'_{jk} \tag{27}$$

이다. 각 파라미터의 재추정식은 $Q_d(\cdot)$, $Q_b(\cdot)$, 및 $Q_c(\cdot)$를 각각 λ'_d의 파라미터에 대해 최대화하여 구한다. 점유시간의 근사함수가 Gamma 밀도함수인 경우

$$\frac{\partial}{\partial \eta'_j} Q_d(\lambda_d \mid D'_j) \Big|_{\eta'_j = \bar{\eta}_j} = 0 \tag{28}$$

$$\frac{\partial}{\partial \upsilon'_j} Q_d(\lambda_d \mid D'_j) \Big|_{\upsilon'_j = \bar{\upsilon}_j} = 0 \tag{29}$$

로부터 파라미터 η_j와 υ_j에 대한 재추정식

$$\bar{\eta}_j = \frac{\sum_{t=1}^{T} f(S, \tau_j \mid \lambda_d) \upsilon_j}{\sum_{t=1}^{T} f(S, \tau_j \mid \lambda_d) \tau_j} \tag{30}$$

$$\frac{d\Gamma(\bar{\upsilon}_j)}{d\bar{\upsilon}_j} \Gamma(\bar{\upsilon}_j)^{-1} = \frac{\sum_{t=1}^{T} f(S, \tau_j \mid \lambda_d) \log(\tau_j)}{\sum_{t=1}^{T} f(S, \tau_j \mid \lambda_d)} + \log \bar{\eta}_j \tag{31}$$

가 된다. 여기서

$$\frac{d\Gamma(\bar{\upsilon}_j)}{d\bar{\upsilon}_j} \Gamma(\bar{\upsilon}_j)^{-1} = -\gamma + (1 - \frac{1}{\bar{\upsilon}_j}) + (\frac{1}{2} - \frac{1}{\bar{\upsilon}_j + 1}) + (\frac{1}{3} - \frac{1}{\bar{\upsilon}_j + 2}) + \cdots \tag{32}$$

로 전개되므로 $\bar{\upsilon}_j$를 구할 수 있으며, γ는 Euler 常數이다. 한편, 점유시간의 근사함수가 Gauss 밀도함수일 경우에는

$$\frac{\partial}{\partial m'_j} Q_d(\lambda_d \mid D'_j) \Big|_{m'_j = \bar{m}_j} = 0 \tag{33}$$

$$\frac{\partial}{\partial \sigma'_j} Q_d(\lambda_d \mid D'_j) \Big|_{\sigma'_j = \bar{\sigma}_j} = 0 \tag{34}$$

로부터 파라미터 mj와 σ_j^2의 재추정식

$$\bar{m}_j = \frac{\sum_{t=1}^{T} f(S, \tau_j \mid \lambda_d) \tau_j}{\sum_{t=1}^{T} f(S, \tau_j \mid \lambda_d)} \tag{35}$$

$$\bar{\sigma}_j^2 = \frac{\sum_{t=1}^{T} f(S, \tau_j \mid \lambda_d) \tau_j^2}{\sum_{t=1}^{T} f(S, \tau_j \mid \lambda_d)} - \bar{m}_j^2 \tag{36}$$

를 구할 수 있다.

관측확률밀도함수의 파라미터 $\bar{a}_{j,k}$는

$$\nabla_{b'_{j,k}} Q_b(\lambda_d, b'_{j,k}) \Big|_{b'_{j,k} = \bar{b}_{j,k}} = 0 \tag{37}$$

의 조건을 만족시키는

$$\bar{r}_{j,k}(i) = \frac{\sum_{t=1}^{T} f(S, \theta_t = j, k_t = k \mid \lambda_d) r_t(i)}{\sum_{t=1}^{T} f(S, \theta_t = j, k_t = k \mid \lambda_d)} \tag{38}$$

로부터 normal equation을 풀어서 구할 수 있고,

branch weight $c_{j,k}$의 재추정식도 $Q_c(\cdot)$를 최대화하여 구하면

$$\bar{c}_{j,k}=\frac{\sum_{t=1}^{T} f(S, \theta_t=j, k_t=k \mid \lambda_d)}{\sum_{t=1}^{T} f(S, \theta_t=j \mid \lambda_d)} \tag{39}$$

이다.[11]

3. 再推定式의 收斂性에 대한 確認

위에서 유도된 재추정식은 모델의 각 파라미터에 대해 log concave 해야 수렴하므로 이에 대한 확인이 필요하다. λ_d의 파라미터 중에서 $c_{j,k}$ 및 $b_{j,k}(s)$의 파라미터에 대한 log concavity는 이미 발표된 여러 연구[11-13]에서 證明되어 있으므로 $D_j(\tau_j)$의 파라미터에 대해서만 기술하기로 한다. Gamma 밀도함수의 파라미터 υ_j 및 η_j에 대해서는

$$\frac{\partial^2}{\partial \eta_j^2} \log [f(S \mid \lambda_d)]=-\frac{\upsilon_j}{\eta_j^2}<0,\ \upsilon_j>0,\ \eta_j>0 \tag{40}$$

$$\frac{\partial^2}{\partial \upsilon_j^2} \log [f(S \mid \lambda_d)]=-\sum_{k=0}^{\infty} \frac{1}{(\upsilon_j+k)}<0 \tag{41}$$

이므로 log concavity가 확인되며, 또 Gauss 밀도함수의 파라미터에 대해서는 다음과 같이 확인할 수 있다. 먼저 $g(y)=\exp(-y^2/2)$로 두면

$$D_j(\tau_j)=(2\pi\sigma_j^2)^{-1/2} g\left(\frac{\tau_j-m_j}{\sigma_j}\right) \tag{42}$$

로 표현할 수 있다. Baum 등[13]의 연구에서 $g(y)$가 log concavity를 가지면 $\partial Q_d(\lambda_d \mid D_j')/\partial m_j'=0$과 $\partial Q_d(\lambda_d \mid D_j')/\partial \sigma_j'=0$을 만족하는 재추정 $\bar{m}_j$와 $\bar{\sigma}_j$는 $Q_d(\lambda_d \mid D_j')$를 전체적 최대치(global maximum)로 하는 유일한 해가 됨이 증명되어 있다. $g(y)$의 log concavity는

$$\frac{\partial^2}{\partial y^2} \log g(y)=-1<0 \tag{43}$$

로부터 확인된다.

4. 前向-後向確率(forward-backward probability)을 이용한 재추정식의 계산

앞서 기술한 재추정식은 매우 복잡하여 계산상의 어려움이 있다. 그러나 다음과 같이 정의되는 前向確率과 後向確率을 이용하면 계산량과 복잡도를 크게 줄일 수 있다. 前向確率 $\alpha_t(j)$는 다음과 같이 정의한다.

$$\alpha_t(j) \triangleq \text{prob}(s_1, s_2, \cdots, s_t, \theta_t=j \ \theta_{t+1}=j+1 \mid \lambda_d)$$
$$=\sum_{k=0}^{t-1} \alpha_k(j-1) \left[\prod_{n=k+1}^{t} b_j(s_n)\right] D_j(t-k) \tag{44}$$

단,

$$\alpha_0(0)=1$$
$$\alpha_t(0)=0, \qquad t\neq 0$$
$$\alpha_0(j)=0, \qquad j\neq 0$$

또, 後向確率 $\beta_t(j)$는

$$\beta_t(j) \triangleq \text{prob}(s_{t+1}, s_{t+2}, \cdots, s_T \mid \theta_t=j-1, \theta_{t+1}=j, \lambda_d)$$
$$=\sum_{k=1}^{T-k} \left[\prod_{n=k+1}^{t+k} b_j(s_n)\right] D_j(k) \beta_{t+k}(j+1) \tag{45}$$

단,

$$\beta_T(N+1)=1$$
$$\beta_t(N+1)=0, \qquad t\neq T$$
$$\beta_T(j)=0, \qquad j\neq N+1$$

로 정의한다. 이를 이용하면 각 파라미터의 재추정을 위한 식 (30), (31), (35), (36), (38), 그리고 (39)의 계산에 필요한 수식은 다음과 같이 정리된다.

$$\sum_{t=0}^{T-\tau_j} f(S, \tau_j \mid \lambda_d)=\alpha_t(j-1) \left[\prod_{n=t+1}^{t+\tau_j} b_j(s_n)\right] D_j(\tau_j) \beta_{t+\tau_j}(j+1) \tag{46}$$

$$f(S \mid \lambda_d)=\alpha_T(N)$$
$$=\beta_0(1) \tag{47}$$

$$\sum_{t=1}^{T} f(S, \theta_t=j, k_t=k \mid \lambda_d) r_t(i)=\sum_{d=1}^{T} \sum_{t=0}^{T-d} \alpha_t(j-1) \left[\prod_{n=t+1}^{t+d} b_j(s_n)\right] D_j(d)$$
$$\cdot \frac{1}{d} \left[\sum_{n=t+1}^{t+d} \frac{c_{j,k}\, b_{j,k}(s_n)\, r_n(i)}{b_j(s_n)}\right] \cdot \beta_{t+d}(j+1) \tag{48}$$

$$\sum_{t=1}^{T} f(S, \theta_t=j, k_t=k \mid \lambda_d)=\sum_{d=1}^{T} \sum_{t=0}^{T-d} \alpha_t(j-1) \left[\prod_{n=t+1}^{t+d} b_j(s_n)\right] D_j(d)$$
$$\cdot \frac{1}{d} \left[\sum_{n=t+1}^{t+d} \frac{c_{j,k}\, b_{j,k}(s_n)}{b_j(s_n)}\right] \cdot \beta_{t+d}(j+1) \tag{49}$$

$$\sum_{t=1}^{T} f(S, \theta_t=j \mid \lambda_d)=\sum_{d=1}^{T} \sum_{t=0}^{T-d} \alpha_t(j-1) \left[\prod_{n=t+1}^{t+d} b_j(s_n)\right] D_j(d)$$
$$\cdot \beta_{t+d}(j+1) \tag{50}$$

V. 인식 알고리듬

제안한 모델은 상태의 점유시간에 따라 상태 천이 확률이 다르므로, 다음과 같이 기존의 Viterbi 알고리즘을 수정하여 사용한다. 여기서 $d_t(j)$는 시간 t일 때 상태 j의 점유시간을 나타내며, $I_t(j)$는 여기로 천이한 以前 상태를 나타내는 逆方向指示子(back-pointer)이다.

[수정된 Viterbi 알고리듬]

[Step 1] Initialization

$$\delta_1(1)=\log[b_1(s_1)]$$
$$d_1(1)=1$$

For $2 \le j \le N$

$\delta_1(j) = -\infty$

$d_1(j) = 0$

[Step 2] Recursion

For $2 \le t \le T$

$\delta_t(1) = \delta_{t-1}(1) + \log[b_1(s_t)]$

$d_t(1) = d_{t-1}(1) + 1$

$I_t(1) = 1$

For $2 \le j \le N$

If $\delta_{t-1}(j-1) + \log[D_{j-1}\{d_{t-1}(j-1)\}] > \delta_{t-1}(j)$

then

$\delta_t(j) = \delta_{t-1}(j-1) + \log[D_{j-1}\{d_{t-1}(j-1)\}] + \log[b_j(s_t)]$

$d_t(j) = 1$

$I_t(j) = j-1$

else

$\delta_t(j) = \delta_{t-1}(j) + \log[b_j(s_t)]$

$d_t(j) = d_{t-1}(j) + 1$

$I_t(j) = j$

endif

[Step 3] Termination

$\log P^* = \delta_T(N) + \log[D_N\{d_T(N)\}]$

$I_T^* = N$

[Step 4] Path backtracking

For $t = T-1, T-2, \cdots, 1$

$I_t^* = I_{t+1}(I_{t+1}^*)$

위의 알고리듬은 step 2에서 상태천이확률을 점유시간의 확률로 하는 점에서 기존의 Viterbi 알고리듬과 구분된다. 한편, step 4 의 경로 역추적과정은 인식시에는 사용되지 않으나, HMM의 훈련시 observation sequence를 상태별로 분할시킬 때 쓰인다.

Ⅲ. 實驗 및 結果

제안한 HMM은 어느단위의 音聲도 모델링이 가능하나, 여기서는 단어 단위의 HMM을 구성한 후, 이를 이용한 話者獨立 孤立單語認識 實驗을 통하여 그 性能을 조사하였다.

1. 音聲 데이터

認識對象單語는 數字音 10개 및 大都市名 10개로 하고 HMM의 訓練에는 30명의 成人話者(男 15명, 女 15명)가 각 단어를 3 회씩 발음한 1,800개의 데이터를 사용하였으며 각 단어별로 남·여 각각 하나씩의 모델을 구성하였다. 認識時에는 訓練에 참가하지 않은 20명의 화자(男 10명, 女 10명)가 각 단어를 2회씩 발음한 800개의 音聲을 사용하였다.

2. HMM의 訓練 및 認識

훈련 및 인식시 사용한 여러 常數는 다음과 같다.

프레임 size : 300 [samples]

프레임 이동 size : 100 [samples]

LPC 차수 : 8

branch의 수 : 5

狀態의 수 : 3-8

HMM의 訓練過程은 Juang 等[8]이 사용한 방법과 동일하나, 다만 모델의 초기 추정치는 다음과 같이 軌跡分割(trace segmentation)[14]을 이용하여 설정하였다. 먼저, observation sequence를 N개의 segment로 軌跡分割한 후, Gauss 밀도함수의 파라미터 m_j 및 σ_j^2는

m_j = segment j에 속한 observation의 평균 길이

σ_j^2 = segment j에 속한 observation 길이의 분산

으로 하였고, Gamma 밀도함수의 파라미터는

$$\eta_j = \frac{m_j}{\sigma_j^2}$$

$$\nu_j = \frac{m_j^2}{\sigma_j^2}$$

으로 정하였다. 또, $c_{j,k}$ 및 $a_{j,k}$의 초기값은 각 segment에 속한 observation을 K-means 알고리듬으로 M개의 cluster로 나눈 후,

$a_{j,k}$ = segment j의 k번째 cluster의 centroid에 해당하는 LPC 自己相關函數(autocorrelation)

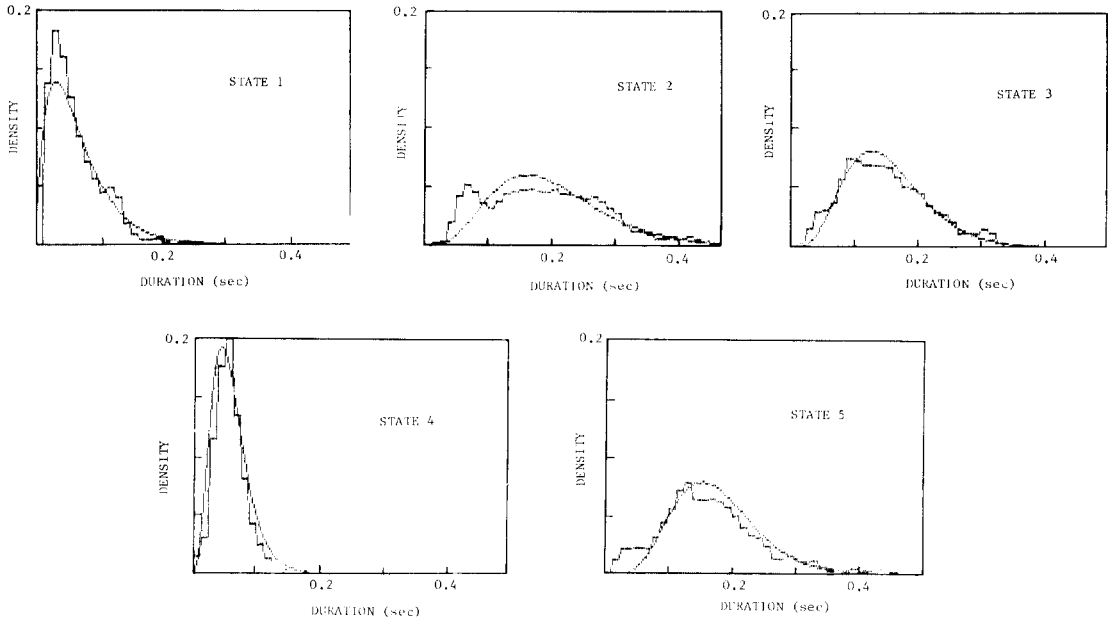
$$c_{j,k} = \frac{\text{segment j의 k번째 cluster에 속한 observation의 수}}{\text{segment j에 속한 observation의 수}}$$

로 하였다.

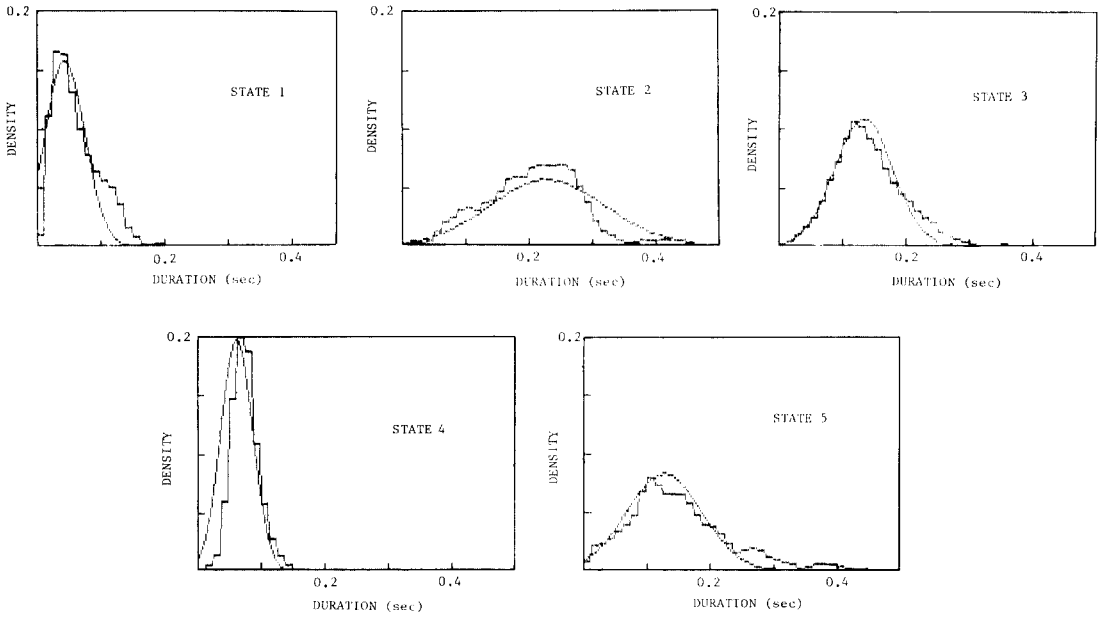
認識方法은 앞서 기술한 수정된 Viterbi 알고리듬을 사용하였으며, 각 단어 HMM에서의 觀測確率을 구하여 가장 높은 확률을 가지는 HMM의 단어를 인식된 단어로 하였다.

3. 狀態占有時間의 分布

그림 3 은 상태점유시간의 확률분포와 이의 근사함수를 비교한 한 예를 나타낸다. 여기에 사용된 음성은 50명의 남성화자가 단어 "서울"을 15회씩 발음한 것이며, 상태점유시간은 수정된 Viterbi 알고리듬을 이용하여 역추적(backtracking)하여 측정하였다. 그림 3(a) 및 3(b)는 각각 점유시간 밀도함수를 Gamma 밀도 및 Gauss 밀도로 한 HMM에서 점유시간의 측정치와 근사함수를 비교한 것이다. 그림에서 보는 바와 같이 점유시간의 측정치는 대략 Gamma 분포에 가깝게 나타났으며, Gamma 분포는 점유시간이 작은



(a)



(b)

그림 3. 단어 "서울"의 5상태 HMM에 대한 점유 시간분포의 측정치와 모델 밀도의 비교
(a) 모델 : HMM/Gamma (b) 모델 : HMM/Gauss

Fig. 3. Comparison of measured duration distribution (jagged contour) and model density (smooth contour) for a 5 states HMM of the word "Seoul".
(a) model : HMM/Gamma, (b) model : HMM/Gauss.

부분에서 비교적 큰 오차를 보였으며 Gauss 분포는 점유시간이 큰 부분에서 오차가 컸다.

4. 實驗結果 및 考察

본 논문에서 제안한 HMM을 이용한 고립단어 인식실험의 결과를 표 1에 요약하였다. 또, 기존의 HMM 및 DTW에 의한 오인식율은 표 2에 나타내었다. 기존의 모델로는 1985년 Juang 等이 제안한 mixture AR 모델[8]을 使用하였고, DTW에 의한 인식 방법은 1978年 Sakoe 등[2]이 제안한 방법을 사용했으며 기준단어당 template 수는 6개로 하였다. 제안한 모델 및 기존 모델의 오인식률은 전반적으로 상태의 수가 증가할수록 감소하였다. 상태의 수가 적을 경우, 제안한 모델은 기존의 모델에 비해 크게 감소된 오인식률을 보였으며, 상태의 수가 많아짐에 따라 기존 모델의 오인식률에 가까운 수치를 나타냈다. 기존의 HMM은 상태의 점유시간 정보를 충분히 반영하지 못하나, left-to-right 구조일 경우 상태의 수가 많아지면, 각 상태의 관측확률밀도함수 및 상태천이행렬로 인하여 음성세그먼트의 길이를 어느 정도 나타내므로 제안한 HMM에 가까운 오인식률을 나타내는 것으로 생각된다.

제안한 HMM에서 점유시간분포의 근사함수로 Gamma 밀도함수 및 Gauss밀도함수를 사용한 두 경우에 거의 같은 오인식률을 보였다. 상태점유시간은 항상 양이므로 전체적인 분포로 Gamma 밀도함수를 예상할 수 있으나, Gauss 밀도함수도 이에 근사화할 수 있고, 또한 observation sequence의 관측확률은 관측확률밀도함수의 영향도 크게 받는 관계로 전체적인 인식률면에서는 큰 차이가 없는 것으로 생각된다.

표 1. HMM/Gamma와 HMM/Gauss 단어인식기의 오인식률

Table 1. Error rates of HMM/Gamma and HMM/Gauss word recognizers.

	Error rate[%]					
No. of states	3	4	5	6	7	8
HMM/Gauss	5.0	5.3	3.9	3.8	4.1	4.0
HMM/Gamma	5.3	4.8	2.9	4.4	4.1	4.3

표 2. Mixture AR HMM과 DTW 단어인식기의 오인식률

Table 2. Error rates of mixture AR HMM and DTW word recognizers.

	Error rate[%]					
No. of states	3	4	5	6	7	8
Mixture AR HMM	7.9	8.0	5.8	5.5	5.1	4.3
DTW method	5.3					

Ⅳ. 結 論

본 논문에서는 音聲의 狀態占有時間 情報를 기존의 HMM에 통합시킨 새로운 형태의 HMM을 제안하고, 각 파라미터의 재추정식을 유도하였다. 또, forward, backward 확률을 이용한 再推定式의 計算方法 및 狀態占有時間을 고려한 認識 알고리듬도 기술하였다. 점유시간의 확률은 Gamma 밀도함수 및 Gauss 밀도함수로 근사화하였다. 이들 함수는 각각 2개의 파라미터로 점유시간의 확률분포를 나타낼 뿐만 아니라 기존 HMM의 狀態遷移行列을 대신하므로 모델을 저장하는데 필요한 메모리량이 적은 장점도 있다. 이 모델의 性能을 알아보기 위하여, 20개 단어를 대상으로 한 認識實驗 結果, 제안한 모델은 기존의 모델에 비해 전체적인 인식률면에서 개선된 성능을 보였으며, 특히 상태의 수가 5 이하인 경우 1/3-1/2정도 감소된 誤認識率을 나타내었다.

參 考 文 獻

[1] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-23, no. 1, pp. 67-72, Feb. 1975.

[2] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-26, no. 1, pp. 43-49, Feb. 1978.

[3] H. Ney, "The use of a one-stage dynamic programming algorithm for connected word recognition," *IEEE Trans. Acoust., Speech, and Signal processing*, vol. ASSP-32, no. 2, pp. 263-271, Apr. 1984.

[4] L.R. Rabiner, S.E. Levinson, and M.M. Sondhi, "On the application of vector quantization and hidden Markov models to speaker-independent, isolated word recognition," *B.S.T.J.*, vol. 62, no. 4, pp. 1075-1105, Apr. 1983.

[5] 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," *B.S.T.J.*, vol. 62, no. 4, pp. 1035-1074, Apr. 1983.

[6] L.R. Rabiner, B.H. Juang, S.E. Levinson, and M.M. Sondhi, "Recognition of isolated digits using hidden Markov models with continuous mixture densities," *AT&T Tech. J.*, vol. 64, no. 6, pp. 1211-1234, July-Aug. 1985.

[7] B.H. Juang, "On the hidden Markov model and dynamic time warping for speech recognition-a unified view," *AT&T Tech. J.*, vol. 63, no. 7, pp. 1213-1243, Sept. 1984.

[8] B.H. Juang and L.R. Rabiner, "Mixture autoregressive hidden Markov models for speech signals," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-33, no. 6, pp. 1404-1413, Dec. 1985.

[9] M.J. Russell and R.K. Moore, "Explicit modeling of state occupancy in hidden Markov models for automatic speech recognition," *Proc. ICASSP'84*, pp. 5-8, Tampa, FL, Mar. 1985.

[10] A.B. Poritz, "Linear predictive hidden Markov models and the speech signal," *Proc. ICASSP'82*, pp. 1291-1294, Paris, France, May 1982.

[11] B.H. Juang, "Maximum-likelihood estimation for mixture multivariate stochastic observations of Markov chains," *AT&T Tech. J.*, vol. 64, no. 6, pp. 1235-1249, July-Aug. 1985.

[12] L.A. Liporace, "Maximum likelihood estimation for multivariate observations of Markov sources," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 5, pp. 729-734, Sept. 1982.

[13] L.E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, no. 1, pp. 164-171, 1970.

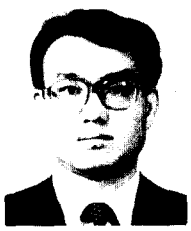
[14] M.H. Kuhn and H.H. Tomaschewski, "Improvements in isolated word recognition," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-31, no. 1, pp. 157-167, Feb. 1983.

著 者 紹 介

曺 政 鎬(正會員)

1959年 1月 3日生. 1981年 2月 경북대학교 전자공학과 졸업. 1986年 2月 경북대학교 대학원 전자공학과 공학석사 학위취득. 1990年 2月 경북대학교 대학원 전자공학과 공학박사 학위취득. 주관심분야는 음성인식 및 음성합성 등임.

●



洪 再 根(正會員)

1951年 1月 25日生. 1975年 2月 경북대학교 전자공학과 졸업. 1979年 2月 경북대학교 대학원 전자공학과 공학석사 학위취득. 1985年 2月 경북대학교 대학원 전자공학과 공학박사 학위취득. 1979年～1983年 경북공업전문대학 전자과 조교수. 1983年～현재 경북대학교 전자공학과 부교수. 주관심분야는 음성인식 및 음성합성 등임.

金 秀 重(正會員)

1941年 6月 25日生. 1962年 12月 인하대학교 전기공학과 졸업. 1966年 2月 인하대학교 대학원 공학석사 학위취득. 1979年 2月 인하대학교 대학원 전자공학과 공학박사 학위취득. 1966年 3月～1970年 12月 삼척공업전문대학 전임강사 조교수. 1976年 9月～1977年 1月 미국 SUNY at Bufflo교환조교수. 1980年 8月～1981年 8月 미국 University of Texas at Austin 연구교수. 현재 경북대학교 전자공학과 교수. 주관심분야는 광신호처리 및 패턴인식등임.