

화자공간모델 진화에 근거한 연속밀도 은닉 마코프모델의 온라인 적응

김동국, 김영준, 김현우, 김남수
서울대학교 전기공학부

Online Adaptation of Continuous Density Hidden Markov Models Based on Speaker Space Model Evolution

Dong Kook Kim, Young Joon Kim, Hyun Woo Kim and Nam Soo Kim

School of Electrical Engineering
Seoul National University, Seoul, Korea

{dkkim11, kimtwins, andy24, nkim}@snu.ac.kr

ABSTRACT

본 논문에서 화자공간모델 evolution에 기반한 continuous density hidden Markov model (CDHMM)의 online 적응에 대한 새로운 기법을 제안한다. 학습화자의 a priori knowledge을 나타내는 화자공간모델은 factor analysis (FA) 또는 probabilistic principal component analysis (PPCA)와 같은 은닉변수모델(latent variable model)에 의해 효과적으로 나타내어진다. 은닉변수모델은 화자공간모델뿐 아니라 CDHMM 파라미터의 a joint prior 분포를 표시함으로써, maximum a posteriori (MAP) 적응기법에 직접 적용되어진다. 화자공간모델의 hyperparameters와 CDHMM 파라미터를 동시에 순차적으로 적응하기 위해 quasi-Bayes (QB) 추정 기술에 기반한 online 적응기법을 제안한다. 연속숫자음 인식과 관련된 화자적응 실험을 통해 제안된 기법은 적은 적응데이터에서 좋은 성능을 나타내며, 데이터가 증가함에 따라 성능이 지속적으로 증가함을 보여준다.

제 1 절 서론

Continuous density hidden Markov model (CDHMM)을 위한 많은 online 적응기법들은 학습과 테스트 환경사이에 존재하는 불일치(mismatch)을 줄이기 위해 연구되어 왔다 [1]. 이런 기법들은 많은 적응 데이터를 모으지 않고 인식시스템이 변화하는 환경에 계속해서 적응하도록 한다 [1].

Huo and Lee [2]는 a prior evolution 과정을 통해 CDHMM 파라미터와 hyperparameters를 동시에 순차적으로 적응하는 quasi-Bayes (QB) 학습 기법을 제안하였다. 또한 모든 CDHMM mean 벡터가 a prior 분포로 correlated 되어 있는 CDHMM 파라미터에 대해 QB 학습기법을 확장하였다 [3]. 최근에는 multiple-stream prior evolution and posterior pooling 이라 불리는 새로운 적응기법을 다양한 량의 적응데이터에 대해 인식기의 성능을 향상시키기 위해 제안하였다 [4]. Chien [5]는 적은량뿐 아니라 많은량의 적응데이터에 대해 성능을 향상시키기 위해 파라미터에 대한 간단한 transformation을 적용하고 이 transformation 파라미터에 대한 prior probability density function (pdf)을 가정하므로 online transformation-based QB 적응기법을 제안하였다.

최근에 본 연구실에서는 eigenvoice [7] 적응기법을 maximum a posteriori (MAP)기반 적응기법으로 확장하기 위해 probabilistic principle component analysis (PPCA) [6]에 기반한 rapid 화자적응 기법을 제안하였다 [8]. 확률밀도 모델이 기존 PCA [9]에 포함된 PPCA는 expectation maximization (EM)[10] 알고리즘에 기반하여 주요 화자공간모델을 구한다. 이러한 PPCA 모델은 화자공간 모델 뿐 아니라 모델

파라미터에 대한 a priori 분포를 제공하여 MAP [11] 추정기법을 쉽게 적용할 수 있도록 한다 [8].

본 논문에서는 PPCA기반 기법을 factor analysis (FA) [12]와 같은 은닉변수모델로 확장하고 화자공간모델 evolution에 기반한 CDHMM mean 파라미터에 기반한 새로운 online 적응기법을 제안한다. 순차적 적응실험 결과를 통해 제안된 기법은 적은 적응 데이터에 대해서도 좋은 성능이 보일뿐 아니라 데이터 크기가 많음에 따라 지속적인 성능 향상을 이룬다.

제 2 절 은닉변수모델

여기서 사용된 은닉변수 모델 2가지는 FA [12]와 PPCA [6] 모델이다. D 차원 벡터를 갖는 관측데이터 y 가 $P (\ll D)$ 차원 은닉변수 v 와 다음과 같은 같은 관계식을 갖는다.

$$y = Wv + \bar{y} + \epsilon. \quad (1)$$

여기서 $W = [w_1, w_2, \dots, w_P]$ 는 관측 데이터의 subspace을 표현하는 $D \times P$ matrix 이고, \bar{y} 는 y 의 mean이고 ϵ 는 v 에 독립적인 Gaussian random noise 이다. 일반적으로 은닉변수는 unit variance을 갖는 독립적인 Gaussain, $p(v) \sim \mathcal{N}(0, I_P)$ 으로 정의되고, 여기서 I_P 는 $P \times P$ identity matrix이다. FA와 PPCA의 주요 차이점은 noise 모델이 어떻게 정의되는가에 따라 다르다. FA모델에서는 noise가 diagonal covariance matrix Λ 을 갖는 Gaussian, $p(\epsilon) \sim \mathcal{N}(0, \Lambda)$ 으로 정의된다. 한편 PPCA는 noise covariance matrix가 isotropic, 즉 $\Lambda = \sigma^2 I_D$ 으로 정의되는데, 여기서 I_D 는 $D \times D$ identity matrix이다. 이런 가정하에서 관측벡터는 $p(y) \sim \mathcal{N}(\bar{y}, \Lambda + WW^T)$ 와 같은 normally 분포되어진다. v 가 주어진 경우 y 의 pdf는 다음과 같다.

$$p(y|v) \sim \mathcal{N}(Wv + \bar{y}, \Lambda). \quad (2)$$

Bayes' 정리에 의하면, 관측벡터 y 가 주어진 경우 은닉변수 v 의 posteriori 분포는 다음과 같이 유도된다.

$$p(v|y) \sim \mathcal{N}(M^{-1}W^T\Lambda^{-1}(y - \bar{y}), M^{-1}). \quad (3)$$

여기서 $M = (I_P + W^T\Lambda^{-1}W)^{-1}$ 이다. 관측벡터열 $Y = \{y_1, \dots, y_T\}$ 가 주어진 경우, 은닉변수모델은 은닉변수열 $V = \{v_1, \dots, v_T\}$ 을 추정하고 maximum likelihood (ML) criterion에 따라 최적의 모델 파라미터 $\hat{\phi} = \{\hat{W}, \hat{\bar{y}}, \hat{\Lambda}\}$ 을 구한다. 은닉변수 $\{v_t\}$ 을 hidden된 것으로 생각하므로 반복적으로 파라미터를 갱신하는 EM 알고리즘을 적용할 수 있

다. $\phi^{(n)}$ 을 n 번째 iteration에서 얻어진 파라미터 값이라 하자. 그러면 새로운 파라미터 값 ϕ 은 다음과 같이 얻어진다.

$$\hat{\phi} = \underset{\phi}{\operatorname{argmax}} E \left[\log p(\mathbf{Y}, \mathbf{V}|\phi) | \mathbf{Y}, \phi^{(n)} \right]. \quad (4)$$

제 3 절 화자공간모델 진화에 기반한 online 적용

3.1. 화자공간모델

K mixture components을 갖는 N -state CDHMM, $\lambda = \{\lambda_j\} = \{w_{jk}, \mu_{jk}, \Sigma_{jk}\}, j = 1, \dots, N, k = 1, \dots, K$ 을 생각해 보자. 관측벡터 \mathbf{x} 의 상태 관측 pdf는 multivariate Gaussians mixture로 정의되어 진다.

$$p(\mathbf{x}|\lambda_j) = \sum_{k=1}^K w_{jk} \mathcal{N}(\mathbf{x}; \mu_{jk}, \Sigma_{jk}) \quad (5)$$

여기서 w_{jk} 는 state j 에서 mixture component k 에 대한 weight이고 μ_{jk} 는 d 차원 mean 벡터이고, Σ_{jk} 는 $d \times d$ covariance matrix 이다.

$\{\mu_1, \dots, \mu_R\}$ 을 주어진 학습 데이터로부터 얻어진 R 개의 잘 학습된 speaker dependent (SD) 모델이라 하자. 여기서 $\mu_r = [\mu_{r,00}^T, \dots, \mu_{r,NK}^T]^T$ 는 r 번째 화자모델로부터 얻어진 Gaussian mean 벡터를 모아서 만든 D 차원의 super-vector이다. $\mu_{r,jk}$ 는 r 번째 화자 CDHMM의 j 번째 state에 있는 k 번째 Gaussian mean 벡터를 나타낸다. SD 모델 집합, $\{\mu_1, \dots, \mu_R\}$ 이 다음과 같이 파라미터 $\phi = \{W, \bar{\mu}, \Lambda\}$ 을 갖는 은닉변수 모델에 의해 발생한다고 가정하자.

$$\boldsymbol{\mu} = \sum_{i=1}^P w_i v_i + \bar{\boldsymbol{\mu}} + \boldsymbol{\epsilon} \quad (6)$$

여기서 $\bar{\boldsymbol{\mu}}$ 는 super-vectors의 mean이고, w_i 는 \mathbf{W} 의 i 번째 열을 나타내는 벡터이고 그리고 $\mathbf{v} = [v_1, \dots, v_P]^T$ 는 \mathbf{W} 의 P 개의 열벡터에 의해 span되어지는 화자공간내에서 한 점을 나타내는 weight이다. 그러면 화자공간모델 파라미터 ϕ 는 반복적인 EM 알고리즘에 의해 구해진다. 일단 파라미터 집합 ϕ 가 주어지면, $\boldsymbol{\mu}$ 에 대한 a prior pdf를 다음과 같이 구할 수 있다.

$$g(\boldsymbol{\mu}|\phi) \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \Lambda + \mathbf{W}\mathbf{W}^T). \quad (7)$$

이식을 통해 은닉변수 모델은 다른 음성단위간의 상관관계에 대한 정보뿐 아니라 CDHMM mean 파라미터와 관련된 prior pdf를 표현한다.

은닉변수 \mathbf{v} 가 주어진 경우, 화자모델 $\boldsymbol{\mu}$ 는 noise 때문에 화자공간내에서 한 점으로 나타내지지 않고, 하나의 random 변수가 된다. 은닉변수 \mathbf{v} 를 추정하는 것은 직접적으로 적용된 모델을 나타내는 것 대신 화자모델의 a priori pdf를 발생한다. 이런 prior pdf는 특정화자에 대한 추정된 prior knowledge을 표현하여 MAP기반 화자 적용기법을 사용하도록 한다.

3.2. 화자공간모델을 위한 QB 학습

여기서 QB 학습에 대한 기본 개념과 식에 대해 알아본다 [1]-[5]. $\mathcal{X}^n = \{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n\}$ 을 CDHMM 파라미터, λ 를 update 하기 위해 순차적으로 모아진 n 개의 독립적인 관측 데이터의 집합이라 하자. λ 의 a posteriori pdf에 대한 recursive한 표현식은 다음과 같이 주어진다.

$$p(\lambda|\mathcal{X}^n) = \frac{p(\mathcal{X}_n|\lambda) \cdot p(\lambda|\mathcal{X}^{n-1})}{\int p(\mathcal{X}_n|\lambda) \cdot p(\lambda|\mathcal{X}^{n-1}) d\lambda}. \quad (8)$$

이는 주어진 파라미터 λ 의 recursive한 Bayesian 추정을 할 수 있는 기본적인 식을 나타낸다. 그러나 이와 같은 recursive한 Bayesian 추정기법을 실현하기는 매우 어렵다. 이러한 문제를 해결하기 위해 QB학습이라 불리는 기법이 [2]에서 제안되었다. QB과정은 recursive한 Bayes 학습과정에서 true posterior density $p(\lambda|\mathcal{X}^n)$ 을 tractable하고 가장 가까운 parametric prior density $g(\lambda|\phi^{(n)})$ 로 두 densities가 같은 mode을 갖도록 근사화 시킨다. 여기서, $\phi^{(n)}$ 은 \mathcal{X}_n 을 관측한 후에 update된 hyperparameters을 나타낸다.

시간 n 순간에 관측벡터 집합 $\mathcal{X}_n = \{\mathbf{x}_1^{(n)}, \dots, \mathbf{x}_{T_n}^{(n)}\}$ 와 근사적인 prior pdf $g(\lambda|\phi^{(n-1)})$ 가 주어졌다고 가정하자. λ 는 hyperparameter $\phi^{(n-1)}$ 와 은닉변수 \mathbf{v} 를 갖는 식 (6)에 의해 주어진 모델을 통해 발생했다고 가정했기 때문에 λ 에 대한 complete-data likelihood는 쉽게 정의 될 수 있다. $(\mathcal{X}_n, \mathbf{S}_n, \mathbf{L}_n)$ 을 \mathcal{X}_n 에 대한 complete-data라 하자. 여기서 $\mathbf{S}_n = \{s_i^{(n)}\}$ 는 상태열을 나타내고, $\mathbf{L}_n = \{l_i^{(n)}\}$ 는 mixture component 열을 나타낸다. 그러면 EM 과정을 반복함으로 λ 에 대한 근사적 MAP 추정치 $\lambda^{(n)}$ 을 다음과 같이 구할 수 있다.

E-step: 보조함수를 계산하라.

$$R(\lambda|\hat{\lambda}_{m-1}) = E \left[\log p(\mathcal{X}_n, \mathbf{S}_n, \mathbf{L}_n | \lambda) + \log g(\lambda, \mathbf{v} | \phi^{(n-1)}) | \mathcal{X}_n, \hat{\lambda}_{m-1} \right]. \quad (9)$$

M-step: 파라미터를 선택하라.

$$\hat{\lambda}_m = \underset{\lambda}{\operatorname{argmax}} R(\lambda|\hat{\lambda}_{m-1}) \quad (10)$$

여기서 $m = 1, \dots, M$ 이고, M 은 $\hat{\lambda}_0 = \lambda^{(n-1)}$ 의 관계식을 갖고 수행된 반복 전체 횟수이다. 위의 EM 과정을 반복적으로 수행함으로, 일련의 근사적 pdf $g(\lambda|\phi^{(n)})$ 의 모드가 true posterior pdf $p(\lambda|\mathcal{X}_n)$ 의 모드에 근접하게 된다. 마지막 EM 과정에서 hyperparameters $\phi^{(n)}$ 집합은 다음식이 만족하도록 구해진다.

$$g(\lambda|\phi^{(n)}) \propto \exp \left\{ R(\lambda|\hat{\lambda}_{M-1}) \right\}. \quad (11)$$

그리고 CDHMM 파라미터는 $g(\lambda|\phi^{(n)})$ 의 모드를 취하므로 update 되어 진다.

3.3. 화자공간모델 진화

이 논문에서는 mixture weights와 covariance matrices는 적용되지 않고 고정되었다고 가정하고 CDHMM의 mean 벡터에 대한 prior evolution인 경우만을 고려한다. 식 (6)에 있는 $\boldsymbol{\mu}$ 에 대한 화자공간모델에 대한 가정하에서 expectation 단계에 있는 보조함수는 다음과 같이 전개된다.

$$R(\lambda|\hat{\lambda}_{m-1}) = \sum_{\mathbf{S}_n} \sum_{\mathbf{L}_n} p(\mathbf{S}_n, \mathbf{L}_n | \mathcal{X}_n, \hat{\lambda}_{m-1}) \cdot \log p(\mathcal{X}_n, \mathbf{S}_n, \mathbf{L}_n | \lambda) + E \left[\log p(\lambda | \mathbf{v}, \phi^{(n-1)}) p(\mathbf{v}) | \hat{\lambda}_{m-1} \right]. \quad (12)$$

그러면 (2)와 (5) 식에 기반하여 다음과 같이 유도된다.

$$R(\lambda|\hat{\lambda}_{m-1}) = \sum_{i=1}^{T_n} \sum_{j=1}^N \sum_{k=1}^K \gamma_{\epsilon}(j, k) \left[-\frac{1}{2} (\mathbf{x}_i^{(n)} - \boldsymbol{\mu}_{jk})^T \Sigma_{jk}^{-1} (\mathbf{x}_i^{(n)} - \boldsymbol{\mu}_{jk}) \right] + E \left[-\frac{1}{2} (\boldsymbol{\mu} - \mathbf{W}^{(n-1)} \mathbf{v} - \bar{\boldsymbol{\mu}}^{(n-1)})^T \Lambda^{-1, (n-1)} (\boldsymbol{\mu} - \mathbf{W}^{(n-1)} \mathbf{v} - \bar{\boldsymbol{\mu}}^{(n-1)}) | \hat{\lambda}_{m-1} \right] \quad (13)$$

여기서 $\gamma_t(j, k) = P(s_t^{(n)} = j, l_t^{(n)} = k | \mathcal{X}_n, \hat{\lambda}_{m-1})$ 는 관측열 \mathcal{X}_n 가 주어졌을 경우 시간 t 에서 state j 와 mixture component k 에 있을 posterior 확률이다. (11)식으로부터 $g(\lambda|\phi^{(n)})$ 는 다음과 같이 표현된다 [13].

$$g(\lambda|\phi^{(n)}) \propto C \cdot \exp \left\{ -\frac{1}{2}(\mu - m)^T \Phi^{-1}(\mu - m) \right\} \quad (14)$$

여기서 C 는 정규화 상수이고,

$$\begin{aligned} m &= \Phi \Lambda^{-1, (n-1)} (\mathbf{W}^{(n-1)} \hat{v} + \bar{\mu}^{(n-1)}) + \Phi \mathbf{C} \mathbf{S}^{-1} \mu_{ML} \\ \Phi &= (\Lambda^{-1, (n-1)} + \mathbf{C} \mathbf{S}^{-1})^{-1} \end{aligned} \quad (15)$$

이다. 여기서 $\mathbf{S} = \text{diag}(\Sigma_{11}, \dots, \Sigma_{NK})$ 와 $\mathbf{C} = \text{diag}(c_{11}, \dots, c_{NK})$ 이고, $c_{jk} = \sum_t \gamma_t(j, k)$ 는 사용된 적응 데이터가 j 번째 state에서 k 번째 Gaussian의 count이다. 그리고

$$\begin{aligned} \hat{v} &= E[v | \hat{\lambda}_{M-1}] \\ &= (\mathbf{I} + \mathbf{W}^T \Lambda^{-1, (n-1)} \mathbf{W}^{(n-1)})^{-1} \cdot \\ &\quad \mathbf{W}^T \Lambda^{-1, (n-1)} (\hat{\mu}_{M-1} - \bar{\mu}^{(n-1)}) \end{aligned} \quad (16)$$

이고, $\mu_{ML} = [(\mu_{11})_{ML}, \dots, (\mu_{NK})_{ML}]^T$ 이다. 여기서 $(\mu_{jk})_{ML} = (\sum_t \gamma_t(j, k) x_t^{(n)}) / c_{jk}$ 이다.

(14)식은 $\bar{\mu} = m$ 와 $\Lambda + \mathbf{W} \mathbf{W}^T = \Phi$ 을 갖는 (7)식에 있는 $g(\mu|\phi)$ 와 같은 분포형태에 속한다. 여기서 hyperparameter \mathbf{W} 가 evolve 하지 않고 (즉, prior evolution 동안 고정됨) Λ 에 대한 prior evolution은 Φ 에 의해 근사화된다고 가정한다. 그러면 $g(\lambda|\phi^{(n)})$ 는 다음과 같이 hyperparameters $\phi^{(n)}$ 을 갖는 것으로 표현된다.

$$\begin{aligned} \bar{\mu}^{(n)} &= m \\ \Lambda^{(n)} &= \Phi \\ \mathbf{W}^{(n)} &= \mathbf{W}^{(n-1)} \end{aligned} \quad (17)$$

그리하여 화자공간모델은 위와 같이 evolution 된다. PPCA에 대한 prior evolution을 위해 $\bar{\mu}^{(n)} = m$ 와 $\Lambda^{(n-1)} = \sigma^{2, (n-1)} \mathbf{I}$, 여기서 $\sigma^{2, (n)} = \text{trace}\{\Phi\} / D$ 로 FA와 비슷하게 얻어진다. 화자공간 모델 evolution 과정을 끝낸 후에 적용된 CDHMM mean 벡터 $\mu^{(n)}$ 는 evolved prior pdf의 모드를 취하므로 다음과 같이 얻어진다.

$$\mu^{(n)} = m. \quad (18)$$

적용된 파라미터 $\mu^{(n)}$ 는 화자공간내에서 추정된 prior 화자모델과 새로운 데이터에 대한 ML 추정값을 간단히 보간함으로 얻어진다. 은닉변수모델에 의한 화자공간모델 evolution과 적용된 mean 파라미터에 대한 개념은 그림 1에 나타내었다.

제 4 절 실험 및 결과

4.1. 실험 환경

제한된 online 적용기법의 성능은 연속어 한국 숫자음 인식에서 여러가지 supervised 적용실험을 통해 평가되었다. 기본 인식시스템을 구성하기 위한 모든 학습 및 테스트 데이터는 매우 조용한 환경에서 녹음되었다. 학습데이터는 105명의 화자를 사용하여 만들었고, 테스트 데이터는 35명을 대상으로 하였다. 각 화자는 3 ~ 7 숫자로 구성된 30 ~ 40 문장을 발음하였다. 각 숫자는 skip이 없는 7개 상

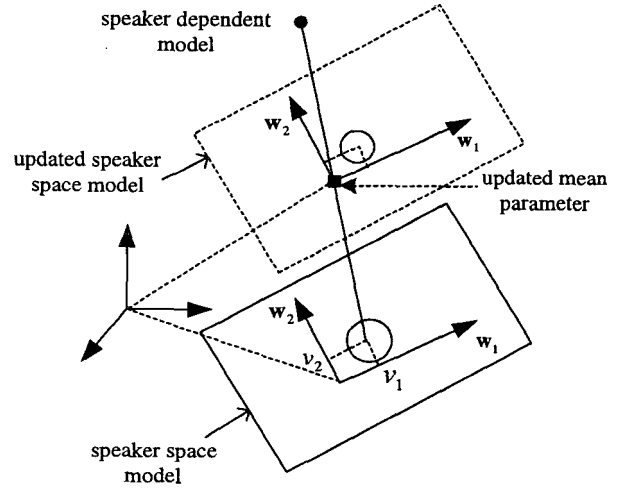


그림 1: 은닉변수모델에 의한 화자공간모델 진화 ($D = 3, P = 2$).

태, 3개 목음은 1개 상태 left-to-right HMM으로 구성되었다. CDHMM 파라미터는 각 state에 있는 Gaussian components 수를 1 ~ 2개로 변화시키면서 학습하였다. 음성신호는 8 kHz로 샘플링 되었고 10 ms 마다 30 ms 프레임으로 분할되었다. 각 음성 프레임은 12차 mel-frequency cepstral 계수와 1차 미분값을 이용하여 총 24차 특징벡터로 변환되었다. 인식실험에서는 화자적응을 위해 각 화자로부터 1 ~ 10 문장 (1 ~ 20 sec)을 사용하였고, 나머지 문장을 이용하여 인식을 수행하였다. 모든 화자적응 과정은 각 데이터에 대한 정확한 transcription 이용한 supervised 형태로 수행하였다. 화자독립 시스템은 한개 Gaussian과 two mixture Gaussians에 대해 87.58%와 89.6%의 단어 인식률을 각각 나타내었다.

은닉변수모델에 의한 화자공간모델을 얻기위해서, 먼저 모든 학습화자로부터 음성을 사용하여 화자독립 모델을 학습 시켰다. 105개의 SD CDHMM을 얻기 위해 한개의 Gaussian mixture에 대해서는 ML 학습을 시켰고, two mixture Gaussian에 대해서는 MAP기반 적응 알고리즘 사용하였다. 각 SD 모델의 mean 벡터 모두를 연결하여 하나의 supervisor를 만들었다. Mean supervisor는 $D = \{11 \times 24 + 3 \times 24\} \times \{1, 2\} = \{1920, 3840\}$ 차원의 크기를 갖는다. 은닉변수모델을 적용하기 전에 각 supervectors는 절대값이 큰 벡터가 분석과정에서 dominant하게 되는 것을 방지하기 위해 표준편차로 각 값을 정규화하였다 [7] [9]. EM 알고리즘을 사용하여 각 mean supervisor에 대한 $P = 10$ 차원의 은닉변수에 대한 파라미터 $\{\bar{\mu}, \mathbf{W}, \Lambda\}$ 값을 추정하였다.

4.2. Batch 적용결과

먼저 다음과 같이 4개의 다른 방법을 사용하여 supervised batch 화자적응실험을 수행하였다. 1) 기존 MAP 적용기법, 2) eigenvoice 기법, 3) PPCA 기반 적용기법, 4) FA기반 적용기법. 표 I과 II는 각각 한개의 Gaussian과 two mixture Gaussian CDHMM에 대한 성능을 나타낸다. 표는 적용 문장단어수에 따라 단어인식률을 표시하였다. 여기서 "EV"는 eigenvoice 기법을 나타내고 "PPCA"와 "FA"는 각각 PPCA와 FA 기반 기법을 나타낸다.

Eigenvoice 기법은 적은 적용 데이터 크기에서는 매우 잘 동작하지만 많은 데이터가 주어졌을 경우 성능이 개선되지 못했다. 반면 PPCA와 FA기반 적용기법은 rapid 성질과 지속적인 적응 성질을 동시에 지니고 있다. 결과적으로 제안된 기법은 매우 적은 적용 데이터뿐 아니라 많은 량의 데이터가 주어지는 경우 효율적인 화자적응을 수행한다.

Table I
한개 Gaussian 인식기의 batch 적응실험에 대한 단어 인식률 (%)

no. of sent.	1	2	4	6	8	10
MAP	87.48	88.04	88.47	88.89	89.47	89.97
EV	88.62	89.91	90.26	90.28	90.26	90.24
PPCA	89.76	90.20	90.22	90.53	90.63	90.95
FA	89.76	90.28	90.43	90.76	90.97	91.38

Table II
두개 Gaussian 인식기의 batch 적응실험에 대한 단어 인식률 (%)

no. of sent.	1	2	4	6	8	10
MAP	89.52	89.8	89.97	90.43	90.88	91.28
EV	91.05	91.13	91.40	91.51	91.63	91.71
PPCA	91.01	91.44	91.90	92.05	92.36	92.79
FA	90.53	91.13	91.82	92.03	92.25	92.71

4.3. Online 적응결과

그림 2는 다양한 적응 데이터에 따른 PPCA와 FA 기반 online 적응기법의 성능을 나타낸다. Online 적응을 위해 파라미터는 각 문장에 대해서 update 되었다. 그림 2로 부터 제안된 기법의 online 버전은 batch PPCA와 FA 적응기법과 비슷한 빠른 적응 성능을 갖고 있으며, 데이터가 증가함에 따라 좋은 수렴성질도 갖고 있다. 실험 결과를 통해 은닉변수 모델에 기반한 online 및 batch 적응기법은 둘 다 매우 잘 동작함을 알 수 있다.

제 5 절 결론

화자공간모델 진화에 기반한 CDHMM 파라미터의 순차적으로 적응하기 위한 기법을 제안하였다. CDHMM의 mean 벡터와 관련된 화자공간모델과 그에 대한 prior pdfs을 얻기 위해 은닉변수모델을 적용하였다. Recursive한 QB 학습 알고리즘을 화자공간모델을 사용한 online 화자 적응에 확장하였다. Online 화자 적응 실험을 통해 제안된 기법은 적은량뿐 아니라 많은량의 데이터에 주어진 화자 적응에 매우 효과적이라 결론을 얻을 수 있다.

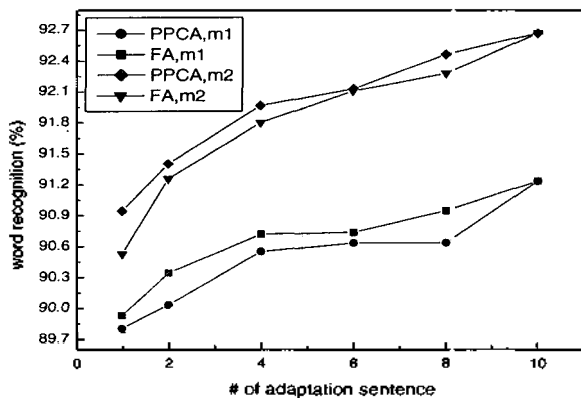


그림 2: 한개 (m1)와 두개 (m2) Gaussian 인식기의 online 적응실험에 대한 단어 인식률.

제 6 절 REFERENCES

- [1] C.-H. Lee and Q. Huo, "On adaptive decision rules and decision parameter adaptation for automatic speech recognition", Proc. of IEEE, vol. 88, pp.1241-1269, Aug. 2000.
- [2] Q. Huo and C.-H. Lee, "On-line adaptive learning of the continuous density hidden Markov model based on approximate recursive Bayes estimate," IEEE Trans. Speech and Audio Proc., vol. 5, pp. 161-172, Mar. 1997.
- [3] —, "On-line adaptive learning of the correlated continuous density hidden Markov models for speech recognition," IEEE Trans. Speech and Audio Proc., vol. 6, pp. 386-397, July 1998.
- [4] Q. Huo and B. Ma, "Online adaptive learning of continuous-density hidden Markov models based on multiple-stream prior evolution and posterior pooling," IEEE Trans. Speech and Audio Proc., vol. 9, pp. 388-398, May 2001.
- [5] J.-T. Chien, "On-line hierarchical transformation of hidden Markov models for speech recognition," IEEE Trans. Speech and Audio Proc., vol. 7, pp. 656-667, Nov. 1999.
- [6] M. Tipping and C. Bishop, "Mixtures of probabilistic principal component analysers," Neural Computation, vol. 11, no. 2, pp. 443-482, 1999.
- [7] R. Kuhn, J.-C. Junqua, P. Nguyen and N. Niedzielski, "Rapid speaker adaptation in eigenvoice space", IEEE Trans. Speech and Audio Proc., vol. 8, pp. 695-707, Nov. 2000.
- [8] D. K. Kim and N. S. Kim, "Bayesian speaker adaptation based on probabilistic principal component analysis", Proc. of ICSLP, pp. 734-737, 2000.
- [9] I. T. Jolliffe, Principal Component Analysis, Springer-Verlag, 1986.
- [10] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society, vol. 39, pp. 1-38, 1977.
- [11] J. L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate gaussian mixture observations of Markov chains", IEEE Trans. Speech and Audio Proc., vol. 2, pp. 291-298, Apr. 1994.
- [12] D. Rubin and D. Thayer, "EM algorithms for factor analysis," Psychometrika, vol. 47, pp. 69-76, 1982.
- [13] G. Zavaliagos, "Maximum a posteriori adaptation techniques for speech recognition," Ph.D. dissertation, Northeastern Univ., Boston, MA, 1995.