한국음향학회지 제22권 제3호 pp. 188~193 (2003)

가중 훈련을 이용한 화자 적응 시스템의 향상

Improvements in Speaker Adaptation Using Weighted Training

장 규 철*, 우 수 영*, 진 민 호*, 박 용 규*, 유 창 동*
(Gyu-Cheol Jang*, Soo-Young Woo*, Min-Ho Jin*, Yong-Kyu Park*, Chang D. Yoo*)

*한국과학기술원 전자전산학과

(접수일자: 2002년 11월 25일; 수정일자: 2003년 2월 4일; 채택일자: 2003년 2월 21일)

이전의 여러 가지 화자 적응을 위한 모델 적응 방법은 훈련 환경과 테스트 환경의 불일치를 보상하기 위한 방법으로 적응 데이터의 테스트 환경에서의 분포를 고려하지 않은 보상 방법이었다. 적은 적응 데이터에 대해서 보상을 극대화하기 위한 파라미터 변환 방법들은 고르지 못한 적응 데이터에 의해 시스템의 성능이 저하 될 가능성이 있다. 즉, 데이터가 적을 경우에는 적응 데이터의 분포가 적응 결과에 중대한 영향을 미치게 된다. 적은데이터에 대해서도 높은 인식률 향상을 가져오기 위한 supervised 훈련 과정을 구조적 사후확률 최대화 (SMAP: Structural Maximum a Posterior) 알고리듬에 적용하였다. 제안된 가중치 SMAP (Weighted SMAP) 알고리듬과 SMAP알고리듬을 TIDIGITS 코퍼스를 사용해서 비교해 보았다. 제안된 WSMAP은 적은 양의 데이터에 대해서 SMAP보다 좋은 성능을 나타내었다. 환경 적응에 적응 데이터의 분포를 고려하는 이와 같은 방법은 다른 적응 알고리듬에도 적용될 수 있다.

핵심용어: 화자 적응, 구조적 사후확률 최대화 알고리듬 (SMAP), 가중치 SMAP (WSMAP)
투고분야: 음성처리 분야 (2.5)

Regardless of the distribution of the adaptation data in the testing environment, model-based adaptation methods that have so far been reported in various literature incorporates the adaptation data undiscriminatingly in reducing the mismatch between the training and testing environments. When the amount of data is small and the parameter tying is extensive, adaptation based on outlier data can be detrimental to the performance of the recognizer. The distribution of the adaptation data plays a critical role on the adaptation performance. In order to maximally improve the recognition rate in the testing environment using only a small number of adaptation data, supervised weighted training is applied to the structural maximum a posterior (SMAP) algorithm. We evaluate the performance of the proposed weighted SMAP (WSMAP) and SMAP on TIDIGITS corpus. The proposed WSMAP has been found to perform better for a small amount of data. The general idea of incorporating the distribution of the adaptation data is applicable to other adaptation algorithms.

Keywords: *Speaker adaptation, SMAP, WSMAP*
ASK subject classification: *Speech signal processing (2.5)*

I. 서 론

음성 인식기 시스템 (Automatic Speech Recognizer)을 실제로 사용하는데 있어서 힘든 문제 중 하나는 훈련 (Training) 환경과 테스팅 (Testing) 환경의 불일치 (mismatch)로 인해 시스템의 성능이 저하된다는 것이다. 이와 같은 불일치를 보상하기 위해서 여러 가지 방법들이 연구되어 왔다. 일련의 연구들은 크게 두 가지로 구분할 수 있다. 음성 특징 벡터를 추출하는 과정의 개선에서부터 인식률을 보상하고자 하는 특징 보상 (feature compensation) 방법[1,8], 소수의 적응 데이터를 기반으로 테스트 환경에 적응된 새로운 모델을 만드는 모델 적응 (model adaptation) 방법[2-7]이 있다. 본 논문에서는 이중 모델 적응 방법을 사용하여 불일치를 보상하는 방법에 관한 연구를 하였다.

모델 적응은 크게 직접 모델 적응과 간접 모델 적응의 두 가지로 분류할 수 있다. 직접 모델 적응은 Baysian

교신저자: 장규철 (tupp@mail.kaist.ac.kr)
305-701 대전시 유성구 구성동 373-1
한국과학기술원 전자전산학과 Multimedia Processing Lab.
전화: 042-869-5470; 팩스: 042-862-0559)

추정 기반의 적응 기법[2]으로, 적응 자료의 수가 많아지면 유사도 최대화 추정치 (Maximum Likelihood Estimator)에 근사적으로 수렴하지만 적응 자료의 수가 적은 경우 성능 향상이 제한적이다. 그리고 모델 파라미터의 사전 확률 (priori density)의 결정이 어려우며, 파라미터의 수가 많아지면 인식률의 향상이 매우 느려지는 단점을 가지고 있다.

간접 모델 적응은 파라미터 변환 (parameter transformation) 기반의 적응 기법[3,4]으로, 적응 자료의 양이 커지면 화자 독립 시스템에 수렴 여부를 보장하지는 못한다. 간접 모델 적응에서는 적은 양의 적응 데이터만 가지고서도 높은 성능 향상을 얻기 위해 여러 개의 파라미터를 묶어서 파라미터들의 자유도를 떨어뜨린다. 그렇기 때문에 적은 양의 적응 자료에 대해서는 인식률의 향상이 높지만, 자료의 크기가 커지면 한계를 드러내게 된다.

위의 두 가지 적응 방식의 단점을 상호 보완하기 위해서 두 가지 적응 방법을 접목한 적응 알고리듬도 연구되었다[5-7]. 적응 자료가 적을 때는 큰 성능 적응 효과를 얻을 수 있으면서, 데이터가 많을 때에는 근사적 수렴 성질을 가지는 적응 방식이다. 이중에서 구조적 사후확률 최대화 알고리듬은[7] 모델 변환 기반의 사후 확률 최대화 기법을 사용한 알고리듬이다. 앞에서 언급한 바와 같이 빠른 적응 속도와 좋은 근사적 성질을 가지고 있지만, 적은 양의 적응 데이터가 사용될 경우에 적응에 좋지 않은 데이터에 의해 적응 시스템의 성능이 나빠지는 경우가 있다. 이와 같은 적응 데이터의 분포에 대한 의존도를 줄이기 위해서 본 논문에서는 각 적응 데이터를 유사도를 이용한 신뢰도를 기반으로 얻어진 신뢰 가중치로 해당 적응 데이터의 확률 값이 강조되는 가중 훈련 방법을 적용하였다.

본 논문의 구성은 다음과 같다. II에서는 가중 훈련 구조에 대해서 살펴보고, III에서는 가중 SMAP이 제시될 것이며, IV에서는 TIDIGITS[12]를 사용한 성능평가, 마지막으로 V에는 결론을 맺을 것이다.

II. 가중 적응 방법 (Weighted Adaptation)

화자 적응 시스템의 목표는 제한된 적응 자료를 이용하여 가능한 높은 성능 향상을 이끌어내는 것과 동시에 많은 데이터에 대해서 유사도 최대화 추정치 (MLE: Maximum Likelihood Estimate) 추정치를 얻는 것이다. 위에서 언급한 바와 같이 모델에 대한 사후 확률 최대화

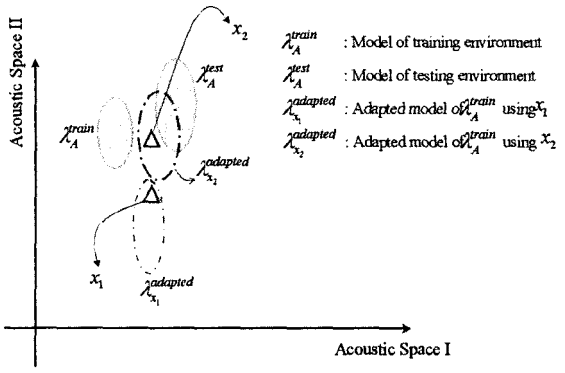

그림 1. 적응 자료가 적응에 미치는 영향
Fig. 1. Influence of adaptation data on adaptation model.

(MAP: Maximum a Posterior)를 이용한 적응 알고리듬과 파라미터 변환 방법을 사용하여 적은 양의 자료를 최대한 효율적으로 이용하기 위한 연구가 많이 있어 왔다[2-7]. 이 연구들에서 해당 적응 자료의 불일치 정도에 대한 신뢰도는 모두 그룹 단위로 결정된다는 점에 주목할 필요가 있다. 즉, 적응에 사용된 데이터의 양이 적으면 적을수록 각각의 적응 데이터가 차지하는 중요성은 커지게 된다. 왜냐하면 각각의 데이터의 성향이 모델 파라미터의 많은 부분에 대해서 영향을 미치기 때문이다.

그림 1은 적응 기반의 outlier인 적응 데이터 x_1가 적응 모델에 어떠한 영향을 미치는지를 보여준다. x_1을 사용하여 적응된 모델 λ_{x1}^{adapt}는 훈련 환경이 올바로 적용된 모델 λ_A^{test}와 많은 차이가 나게 된다. 그러나 적응 데이터 x_2에 의해서 적응된 모델 λ_{x2}^{adapt}는 λ_A^{test}와 가깝게 얻어진다. 이러한 이유로 각각의 훈련 데이터는 차별적으로 적응에 이용될 필요가 있다. 이 각각의 인식 단위의 훈련 데이터, 즉 훈련 토큰 (training token)에 테스트 레퍼런스들에 대한 유사도 확률을 기반으로 한 가중치를 줌으로서 테스트 환경에서의 데이터 근사적으로 적용할 수 있다.

테스트 환경에서의 음성 자료의 분포는 음성 자료들의 신뢰도로서 표현될 수 있다. 각 훈련 토큰의 신뢰도는 그 데이터에 가중치를 주는데 사용될 수 있다[9]. 반복적 적응 과정이 진행됨에 따라서 측정되는 신뢰도는 더더욱 실제 테스트 환경에 가까워진다.

2.1. 가중치에 대한 수렴성 (Convergence with confidence weighting)

적응 데이터에 신뢰 가중치를 가하는 방법을 사용함으로서 각 적응 데이터의 신뢰도를 적응 알고리듬에 적용하고자 한다. 우선 각 훈련 토큰에 가중치가 곱해짐에 따라

서 기대치 최대화 (EM: Expectation and Maximization) 알고리듬의 수렴성이 유지되는지부터 검증할 필요가 있다. 기존의 HMM (Hidden Markov Model) 모델을 λ, 업데이트된 HMM 모델을 λ'이라고 했을 때, λ의 유사도 최대화 추정치를 구하기 위한 보존 (auxiliary) 함수는 다음과 같이 표현된다[10].

$$Q(\lambda,\lambda')=\frac{1}{P(X|\lambda)}\sum_{all\ s}P(X,s|\lambda)\log P(X,s|\lambda') \quad (1)$$

여기서 X, s는 훈련 토큰, 상태 나열 (state sequence)을 각각 의미한다. 위의 보존 함수는 λ가 업데이트됨에 따라서 증가함을 보이면 수렴성이 증명된다. Baum[10]에 의해서 알려진 바를 이용해 업데이트된 모델값 λ'에 대하여 다음과 같이 보존 함수의 증가를 증명할 수 있다.

$$\begin{aligned}&Q(\lambda,\lambda')-Q(\lambda,\lambda)\\&=-\frac{1}{P(X|\lambda)}\sum_{all\ s}P(X,s|\lambda)\log\frac{P(X,s|\lambda')}{P(X,s|\lambda)}\\&\le\log\frac{\sum_{all\ s}P(X,s|\lambda')}{\sum_{all\ s}P(X,s|\lambda)}=\log\frac{P(X|\lambda')}{P(X|\lambda)}\end{aligned} \quad (2)$$

Arslan[9]에서와 같이 많은 적응 데이터 N에 대해서 확장시키면, 다음 식과 같다.

$$\frac{1}{N}\sum_{n=1}^{N}\log\frac{P(X|\lambda')}{P(X|\lambda)}\ge Q(\lambda,\lambda')-Q(\lambda,\lambda) \quad (3)$$

우리는 각 훈련 토큰에 따라서 확률에 가중치를 주려고 한다. n번째의 훈련 토큰에 대한 기존의 가중치, 업데이트된 가중치를 w_n, w'_n이라고 할 때, 이 가중치는 훈련 토큰과 사용되는 모델 파라미터에 따라서 달라질 것이므로 $w_n=\omega_n(X_n,\lambda)$와 같이 관찰 데이터와 모델 파라미터의 함수로 나타낼 수 있다. 식 (3)의 유사도 확률에 각각 n번째 적응 데이터에 대한 가중치 값 $\omega_n'(X,\lambda')$과 $\omega_n(X,\lambda)$을 곱해주면 다음과 같은 식을 얻는다.

$$\begin{aligned}&\frac{1}{N}\sum_{n=1}^{N}\log\frac{\omega_n'P(X|\lambda')}{\omega_nP(X|\lambda)}\\&=\frac{1}{N}\sum_{n=1}^{N}\log\frac{P(X|\lambda')}{P(X|\lambda)}+\frac{1}{N}\sum_{n=1}^{N}\log\frac{\omega_n'}{\omega_n}\\&\ge Q(\lambda,\lambda')-Q(\lambda,\lambda)\end{aligned} \quad (4)$$

즉 위의 수식에서 $\frac{1}{N}\sum_{n=1}^{N}\log\frac{\omega_n'}{\omega_n}\le 0$을 만족하는 경우에, 보존 함수가 증가함에 따라 $P(O,\lambda')$가 증가함을 증명할 수 있다. 본 논문에서는 위의 모델 수렴을 위한 충분 조건을 만족시키는 가중치를 인위적으로 발생시키지 않고, 수식 (6)의 가중치를 실험에 사용하였을 때 수렴성을 만족함을 실험적으로 보였다.

2.2. 신뢰 가중치 (Confidence weight)

각 훈련 토큰의 유사도 비에 의한 신뢰도는 다음과 같이 표현할 수 있다.

$$C_n^{(i)}=\frac{P(X_n^{(i)}|\lambda_i)}{(\Pi_{j=1,j\ne i}^{N}P(X_n^{(i)}|\lambda_j))^{1/N-1}} \quad (5)$$

$X_n^{(i)}$는 i번째의 모델에 해당하는 n번째의 음성 데이터를 의미한다. 위의 수식은 하위 인식 단위에 대해서 계산될 수 있지만 개념을 간단히 하기 위해서 훈련 토큰 단위의 신뢰도로서 표현한 것이다. 위와 같은 유사도비를 이용한 신뢰 척도로서 가중치 척도를 나타내는 방법은 여러 가지가 있을 수 있다. 본 논문에서는 다음과 같은 가중치 척도를 사용하였다.

$$\begin{aligned}w_n^{(i)}&=\alpha+\exp(-|\ln(P(X_n^{(i)}|\lambda_i)\\&\quad-\ln(P(X_n^{(i)}|\lambda_j))+\gamma|)\end{aligned} \quad (6)$$

위에서 λ_j는 훈련 토큰 $X_n^{(i)}$에 대해서 가장 높은 유사도를 가지는 모델이다. α는 가중치 값의 최소값을 의미하고 γ는 적응 데이터를 강조할 것인가의 여부를 결정하는 인자이다. 본 논문에서 α값으로 0.2를 γ값으로 1을 사용하였다. 이것은 Arslan[9]와 Juang[11]이 사용한 척도와 유사하다.

III. 가중적 Bayes 적응 (WSMAP)

서론에서 언급한 바와 같이 기존의 직접 적응 알고리듬과 간접 적응 알고리듬은 데이터양이 적을 때 시스템 성능의 향상이 좋지 못하거나, 데이터양의 많아질 때 수렴 상태가 좋지 못하다는 단점을 가지고 있었다. 이러한 단점을 해결하기 위해 계층적인 트리 구조를 이용하는 알고리듬이 Shinoda[7]에 의해 제안되었다.

구조적 사후확률최대화 (SMAP) 알고리듬은 사전확률을 올 결정함과 HMM 내의 가우시안 믹스쳐의 묶음의 불일치를 추정하는데 있어서 효과적인 방법을 제시하였다. 3.1장에서는 파라미터 변환 기반의 모델 적응 방법인 SMAP에서 사용되는 파라미터 그룹화를 위한 모델 파라미터의 트리 구조 생성법과 관련 정의에 대해 살펴보고,

3.2장에서는 기존의 SMAP알고리듬에 관련하여 제안된 WSMAP에 대해서 살펴보도록 하겠다.

3.1. 트리 구조

연속 밀도 은닉 마코프 모델 (Continuous density hidden Markov model)의 파라미터들을 노드 단위로 묶어서 적응을 수행하여 적응 자료의 수가 적은 경우의 간접 모델 적응의 효과를 얻기 위해 트리 구조를 생성한다. 이때 각 트리 노드의 원소는 가우시안 믹스쳐가 되며, 각 노드를 대표하는 분포는 가우시안 믹스쳐들의 분포로서 구할 수 있다.

3.1.1. 가우시안 원소간의 거리

가우시안 원소 (Gaussian component)들을 분류하여 노드별로 할당하기 위해, 각각의 가우시안 원소간의 거리를 쿨백-레이블러 발산(Kullback-Leibler divergence)의 합으로 정의한다. 두 가우시안 원소 $g_m(\cdot)$, $g_n(\cdot)$ 간의 거리 $d(n,m)$은 다음과 같다.

$$d(m,n)=\int g_m(x)\log\frac{g_m(x)}{g_n(x)}dx+\int g_n(x)\log\frac{g_n(x)}{g_m(x)}dx$$
$$=\sum_i[\frac{\sigma_m^2(i)-\sigma_n^2(i)+(\mu_n(i)-\mu_m(i))^2}{\sigma_n^2(i)}$$
$$+\frac{\sigma_n^2(i)-\sigma_m^2(i)+(\mu_m(i)-\mu_n(i))^2}{\sigma_m^2(i)}] \quad (7)$$

$\mu_m(i)$는 가우시안 믹스쳐 $g_m(\cdot)$의 평균 벡터 μ_m의 i번째 성분이고, $\sigma_m(i)$는 공분산 행렬 Σ_m의 i번째 대각선 성분이다.

3.1.2. 노드 확률 분포 함수

트리 구조상의 각각의 노드에는 여러 개의 가우시안 원소가 포함되고, 트리 구조 적응을 위해 각각의 노드는 이를 대표하는 하나의 분포를 필요로 하게 된다. 이를 위해 M_k개의 가우시안 믹스쳐를 원소로 가지는 k레벨에서의 한노드, $\{g_m^{(k)}(X)=N(X|\mu_m^{(k)},\Sigma_m^{(k)}:m=1,\ldots,M_k\}$에 대해서 노드 분포를 다음과 정의한다.

$$\mu_k(i)=\frac{1}{M_k}\sum_{m=1}^{M_k}E(x_m^{(k)}(i))=\frac{1}{M_k}\sum_{m=1}^{M_k}\mu_m^{(k)}(i) \quad (8)$$

$$\sigma_k^2(i)=\frac{1}{M_k}[\sum_{m=1}^{M_k}E(((x_m^{(k)}(i)-\mu_k(i))^2)]$$
$$=\frac{1}{M_k}[\sum_{m=1}^{M_k}\sigma_m^{2(k)}(i)+\sum_{m=1}^{M_k}\mu_m^{(k)2}(i)-M_k\mu_k^2(i)] \quad (9)$$

$x_m^{(k)}$는 가우시안 분포 함수 $g_m^{(k)}$에서의 관찰 벡터 (observation vector)를 나타낸다.

3.1.3. 트리 구조 결정 알고리듬

3.1.1절과 3.1.2절에서 정의한 가우시안 원소간의 거리와, 노드 확률분포함수를 이용, 트리 구조를 구성한다. 트리 구조는 탑다운 방식으로 각 노드의 자식 노드를 k-means알고리듬을 통해 구성하는 방식으로 진행된다. K-means를 위한 초기값은 최소최대 (minimax)알고리듬을 통해서 노드 내의 가우시안 원소들 중에서 선택한다[7]. 그리고 불필요한 노드의 세부화를 막기 위해서 적당한 개수의 가우시안 원소들을 가질 수 있도록 알고리듬을 구성하였다.

3.2. 구별적인 구조적 Bayes 적응 (WSMAP)

3.2.1. 가우시안 분포의 정규화 (Normalization of Gaussian distributions)

트리 구조를 이용한 적응 알고리듬을 적용하기 위해서 우리는 정규화된 관찰 벡터를 생성하고 그에 따른 정규화된 가우시안 분포를 구하였다. 왜냐하면 믹스쳐 구성원의 집합에 대한 불일치의 경향을 알아내야 트리 구조에 적용할 수 있기 때문이다. 즉, 믹스쳐 m의 파라미터 θ_m을 이용, n번째 훈련 토큰 X_n의 t번째 관찰데이터 x_{nt}를 t와 믹스쳐 원소 m에 대해 다음과 같이 변환하여 y_{nmt}라는 벡터를 생성한다.

$$y_{nmt}=\Sigma_m^{-1/2}(x_{nt}-\mu_m) \quad (10)$$

T를 데이터의 총 프레임 길이라고 할 때, $Y_{nm}=\{y_{nm1},y_{nm2},\ldots,y_{nmT}\}$의 분포를 통해 훈련 환경($\theta_m$)과 테스트 환경($\hat{\theta}_m$) 사이의 차이를 알 수 있다. 훈련과 테스트 환경의 불일치가 존재하지 않는다면, x_t는 θ_m의 분포를 따라야 하므로, Y_{nm}은 표준 정규 분포 $N(Y|\vec{0},I)$를 따라야 한다. 불일치가 존재하는 경우에는 이 불일치에 의해 Y_{nm}은 $N(Y|\nu,\eta)$의 형태로 표현되고, ν는 불일치에 의한 믹스쳐 원소의 평균의 이동를 는 분산의 크기 변화를 나타낸다. 즉, 우리는 노드의 확률분포를 나타내는 파라미터 (ν,η)를 통해서 노드 내의 합산적 불일치를 표현할 수 있게 된다.

M_k개의 믹스쳐 원소를 갖는 트리의 k번째 층의 노드 집합 $G_k=\{g_1,\ldots,g_m,\ldots,g_{M_k}\}$에 기대치 최대화 알고리듬을 적용하되 정규화된 데이터를 이용하여 훈련 토

큰어 따른 신뢰가중치 w_n를 적용하면 다음과 같은 새로운 추정치를 얻을 수 있다.

$$\tilde{\nu}_k = \frac{\sum_{n=1}^{N} w_n \sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_{nmt} y_{nmt}}{\sum_{n=1}^{N} w_n \sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_{nmt}}, \quad (11)$$

$$\tilde{\eta}_k = \frac{\sum_{n=1}^{N} w_n \sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_{nmt} (y_{nmt} - \tilde{\nu}_k)(y_{nmt} - \tilde{\nu}_k)^t}{\sum_{n=1}^{N} w_n \sum_{t=1}^{T} \sum_{m=1}^{M} \gamma_{nmt}} \quad (12)$$

위에서 N은 사용된 총 데이터 개수이고, $\gamma_{nmt} = P(m_t = m | X_n, \lambda)$이다.

3.2.2. 계층적 트리 구조를 이용한 사후 확률 최대화 추정치

사후 확률 최대화와 같은 추정자를 이용하여 화자 적응을 위한 추정치를 구할 때 어려운 문제 중 하나는 사전 확률 분포를 결정하는 문제이다. 모든 파라미터를 동일한 사전확률로 가정하는 것은 각 HMM 파라미터들의 특성을 반영할 수 없다는 단점이 있다. 반면에 각각의 파라미터마다 사전확률을 정의해 준다는 것은 너무 번거롭고 어려운 일이다. 이러한 문제를 해결하기 위해서 계층적 트리 구조를 이용한다. 즉 부모 노드의 사전 정보로부터 자식 노드의 사전 정보의 파라미터를 가정하는 구조이다. 부모 노드의 노드 파라미터를 상속받아서 현재 노드의 사전 확률의 파라미터로 사용한다[7].

트리 구조의 k번째 레벨에 있는 노드의 분포 (ν_k, η_k)의 사후확률 최대화 추정치인 는 $(\hat{\nu}_k, \hat{\eta}_k)$ $k-1$번째 레벨의 노드의 분포 $(\hat{\nu}_{k-1}, \hat{\eta}_{k-1})$을 이용하여 다음과 같이 구할 수 있다.

$$\hat{\nu}_k = \frac{\Gamma_k \tilde{\nu}_k + \tau_k \hat{\nu}_{k-1}}{\Gamma_k + \tau_k} \quad (13)$$

$$\hat{\eta}_k = \frac{\hat{\eta}_{k-1} + \Gamma_k \tilde{\eta}_k + \frac{\tau_k \Gamma_k}{\tau_k + \Gamma_k} (\tilde{\nu}_k - \hat{\nu}_{k-1})^t (\tilde{\nu}_k - \hat{\nu}_{k-1})}{\Gamma_k + \xi_k} \quad (14)$$

여기서 Γ_k는 $\Gamma_k = \sum_{n=1}^{N} w_n \sum_{t=1}^{T} \sum_{m \in C_k} \gamma_{nmt}$으로, $(\tilde{\nu}_k, \tilde{\eta}_k)$는 (ν_k, η_k)의 ML 추정치로 정의하였다. 기존의 알고리즘인 SMAP에서는 가중치 w_n이 없이 데이터의 가중치가 모두 동일하다. 제안된 SMAP알고리듬에서 가중치 w_n은 데이터로부터 유도된 불일치의 정도를 강조시키므로 적용되는 데이터마다 다르게 강조된다. 그리고 τ_k와 ξ_k는 두 파라미터들의 하이퍼 파라미터 (hyperparameter)로서[7] 트리내의 각 노드에 관계없이 동일하다고 가정하였다. 본 논문에서는 $\tau_k = 0.1$, $\xi_k = 1$의 값을 사용하였다. 위의 수식에서 $\hat{\nu}_0 = \vec{0}$, $\hat{\eta}_0 = I$로 가정하였다. $\overline{\mu}_m$와 $\overline{\Sigma}_m$을 가우시안 믹스쳐 원소 $g_m(\cdot)$와 관련된 평균벡터와 공분산행렬이라고 할 때, K번째 레벨의 $(\hat{\nu}_K, \hat{\eta}_K)$을 이용하여 가우시안 믹스쳐의 MAP 추정값 $\hat{\mu}_m$와 $\hat{\Sigma}_m$을 다음의 식을 통해서 근사적으로 구할 수 있다.

$$\hat{\mu}_m = \overline{\mu}_m + (\overline{\Sigma}_m)^{1/2} \hat{\nu}_K \quad (15)$$

$$\hat{\Sigma}_m = \overline{\Sigma}_m^{1/2} \hat{\eta}_K (\overline{\Sigma}_m^{1/2})^t \quad (16)$$

IV. 실험 및 결과고찰

이 논문에서 제시된 구별적인 구조적 Bayes 적응 (DSMAP) 알고리듬을 검증하게 위해 TIDIGITS[12]을 이용하여 인식 실험을 수행하였다. 특징 벡터는 30 ms의 프레임을 10 ms씩 이동시켜 얻은 13차 MFCC를 이용하였다. 여성 화자의 발화 모델을 여성 화자의 발화로 테스트, 환경의 불일치가 없는 경우에는 98.46%의 인식률을 얻을 수 있었다. 이 모델을 남성 화자의 발화로 테스트하여 환경의 불일치를 유도한 경우에는 80.38%의 인식률을 얻었다.

환경의 불일치를 구현하기 위해 남자 55명의 데이터로 남자를 위한 모델을 만들고 이를 여자 20명의 데이터로 모델을 적응시킨 후 또 다른 여성 56명의 데이터로서 테스트하는 실험을 하였다. 한 사람당 발화 횟수는 22회이다. TREE[7], SMAP 그리고 WSMAP을 이용하여 supervised 화자 적응 실험한 결과가 표 1이다. 이 결과는 적응 데이터가 늘어남에 따라서 SMAP과 WSMAP이 화자 독립 시스템에 수렴해 하는 것을 보여주고 있으며, WSMAP이 SMAP보다 약 3-5%정도의 낮은 오류율을 가진다. 이것은 WSMAP이 MAP보다 적은 양의 적응 데이터를 효율적으로 사용하고 있음을 보여준다. 이 결과는 총 4개의 레벨과 각 노드당 3개의 가지를 가지는 트리 구조를 이용한 적응 결과이지만 다른 트리 구조에서도 거의

표 1. TREE, SMAP 그리고 WSMAP을 이용하여 supervised 적응한 인식 결과

Table 1. Recognition rate obtained with supervised adaptation done with TREE, SMAP, and WSMAP.

적응 데이터 수	TREE	SMAP	WSMAP
Baseline	80.38	80.38	80.38
20	83.57	85.08	86.04
40	86.68	90.27	91.54
80	91.39	94.49	94.90
300	91.39	97.29	97.21

동일한 결과를 얻을 수 있었다.

V. 결 론

새롭게 제시된 WSMAP은 각각의 적응 자료의 실제 환경과의 신뢰도를 고려하여 이 신뢰도에 따라 각각의 적응 자료를 다른 비율로 적응에 이용함으로써, 기존의 SMAP 알고리듬에 비해 적응 데이터가 늘어남에 따라 실제 환경에의 적응이 빠른 결과를 보인다. 실제 supervised 실험 결과에서도 SMAP과 WSMAP은 모두 테스트 환경에 대한 인식률에 수렴하나 적응 자료의 수가 적은 경우 WSMAP이 보다 빠른 인식률의 향상을 보임을 확인할 수 있다. 환경 적응 시에 적응데이터의 분포를 고려하는 이와 같은 방법은 다른 적응 알고리듬에도 적용될 수 있다.

감사의 글

이 논문은 한국과학재단이 지원한 목적기초연구로 (과제번호 R01-2000-000-00259-0 (2002)) 얻은 연구 결과의 하나이며 이에 고마움을 나타냅니다.

참 고 문 헌

1. F. H. Liu, A. Acero and R. Stern, "Efficient joint compensation of speech for the effects of additive noise and linear filtering," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, I-257- I-260, March, 1992.
2. Q. Huo and C.-H. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of markov chains," *IEEE Trans. Speech Audio Processing*, 2, 291-298, 1994.
3. S. Furui, "Unsupervised speaker adaptation method based on hierarchical spectral clustering," *IEEE Trans. Acoust., Speech, Signal Processing*, 37, 1923-1930, December, 1989.
4. C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous-density hidden markov models," *Comput. Speech Lang.*, 9, 171-185, 1995.
5. J.-I. Takahashi and S. Sagayama, "Vector-field smoothed Bayesian learning for incremental speaker adaptation," *Proc. ICASSP-95*, Detroit, 696-699, MI, 1995.
6. O. Siohan, C. Chesta and C.-H. Lee, "Hidden Markov model adaptation using maximum a posteriori linear regression," *Proc. Workshop Robust Methods for Speech Recognition in Adverse Conditions*, 147-150, Tampere, Finland, 1999.
7. K. Shinoda and C.-H. Lee, "Structural MAP speaker adaptation using hierarchical priors," *Proc. IEEE Workshop Speech Recognition Understanding*, 1997.
8. C.-H. Lee, "On stochastic feature and model compensation approaches to robust speech recognition," *Speech Commun.*, 25, 29-47, 1998.
9. L. M. Arslan and J. H. L. Hansen, "Selective training for hidden markov models with applications to speech classification," *IEEE Trans. Speech and Audio Processing*, 7 (1), 46-54, JAN. 1999.
10. L. E. Baum and J. A. Eagon, "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology," *Bull. Amer. Math. Soc.*, 73, 360-363, 1967.
11. B.-H. Juang and S. Katagiri, "Discriminative learning for minimum error classification," *IEEE Trans. Signal Processing*, 40, 3043-3054, 1992.
12. R. G. Leonard, "A database for speaker-independent digit recognition," *ICASSP*, San Diego California, 3, 42, 1984.

저자 약력

● 장 규 철 (Gyu-Cheol Jang)

2001년 2월: 한국과학기술원 전기 및 전자공학과 졸업 (학사)
2001년 3월~현재: 한국과학기술원 전기 및 전자공학과 석사과정
※ 주관심분야: 신호처리, 음성인식

● 우 수 영 (Soo-Young Woo)

2001년 2월: 한국과학기술원 전기 및 전자공학과 졸업 (학사)
2001년 3월~현재: 한국과학기술원 전기 및 전자공학과 석사과정
※ 주관심분야: 음성인식, 음성신호처리

● 진 민 호 (Min-Ho Jin)

2002년 2월: 한국과학기술원 전기 및 전자공학과 졸업 (학사)
2002년 3월~현재: 한국과학기술원 전기 및 전자공학과 석사과정
※ 주관심분야: 음성인식, 화자인식

● 박 용 규 (Yong-Kyu Park)

1985년 2월: 한양대학교 전기과 졸업 (공학사)
1987년 8월: 한국과학기술원 전자전산학과 졸업 (공학석사)
1996년 8월: 한국과학기술원 전자전산학과 졸업 (공학박사)
1987년~2000년: 한국전기통신공사 선임연구원
2000년~2002년: 맨투머신(주) 대표이사
2002년 1월~현재: 한국과학기술원 연구교수

● 유 창 동 (Chang D. Yoo)

한국음향학회지 제20권 제3호 참조