

향상된 MDL 기법에 의한 음향모델의 최적화 연구

A Study on Improved MDL Technique for Optimization of Acoustic Model

조 훈 영*, 김 상 훈*

(Hoon-Young Cho*, Sang-Hun Kim*)

*한국전자통신연구원 소프트웨어연구부 음성언어정보연구부

(접수일자: 2009년 10월 22일; 채택일자: 2009년 11월 24일)

본 논문에서는 HMM 기반의 연속음성인식에서 음향모델의 최적화 기법을 논한다. 대부분의 음성인식 시스템에서 HMM 상태별로 동일한 개수의 가우시안 성분 (mixture component)을 사용해 왔다. 그러나, 음향 모델링에 사용되는 데이터 샘플의 개수는 HMM 상태별로 다르므로 이에 따른 최적화를 수행할 경우 모델 파라미터의 개수를 효과적으로 줄일 수 있을 뿐 아니라, 디코딩 단계에서 음성인식기의 속도 및 인식 성능 개선이 기대된다. 본 연구에서 제안한 방법은 기존에 알려진 MDL (minimum description length) 기반의 음향모델 최적화 방법에서 가우시안 성분들의 통합과정에 가우시안 성분의 가중치 정보 (mixture weight)를 반영하도록 개선하였다. 인식 실험 결과, 제안한 방법은 가우시안 성분의 가중치를 반영하지 않는 기존 방법에 비해 향상된 최적화 성능을 보임을 확인할 수 있었다.

핵심용어: 연속음성인식, 음향모델, 최적화, MDL

투고분야: 음성처리 분야 (2.5)

This paper describes optimization methods of acoustic models in HMM-based continuous speech recognition. Most of the conventional speech recognition systems use the same number of Gaussian mixture components for each HMM state. However, since the number of data samples available for each state is different from each other, it is possible to reduce the overall number of model parameters and the computational cost at the decoding step by optimizing the number of Gaussian mixture components. In this study, we introduced the Gaussian mixture weight term at the merging stage of Gaussian components in the minimum description length (MDL) based acoustic modeling optimization. Experimental results showed that the proposed method can obtain better ASR accuracy than the previous optimization method which does not consider the Gaussian mixture weight term.

Keywords: Continuous Speech Recognition, Acoustic Model, Optimization, Minimum Description Length

ASK subject classification: Speech Signal Processing (2.5)

I. 서론

근래의 음성인식은 프로세서의 고속화, 메모리 양의 증가, 병렬 처리 기법, 음성언어 자원의 양적 증가 등으로 인해 인식 성능이 지속적으로 향상되고 있다. 음성인식 기술은 서버용 음성인식 기술과 단말용 음성인식 기술로 구분해 볼 수 있는데, 서버용 음성인식 기술은 수 백 시간이 넘는 분량의 음성 데이터를 음향 모델의 학습에 사용

하고, 수 천만 어절 이상의 텍스트 코퍼스로 언어모델을 훈련하며, 세 개 이상의 음성인식기를 동시에 수행하여 그 결과를 조합하는 등 활용 가능한 모든 컴퓨팅 자원을 이용하여 음성인식 성능을 최대로 높이고자 한다. 이와 달리 단말용 음성인식 기술은 프로세서 속도, 저장 공간, 배터리 용량 등에 제약이 있는 상황에서 서버급 인식기 및 인식기 개발에 활용된 자원을 최적화하여 일정 수준 이상의 인식 성능 및 속도를 확보하는 것을 목표로 한다. 이러한 목표를 달성하기 위해서는 특징추출, 모델학습, 디코딩, 후처리 등 음성인식 전 과정을 통해 최적화가 이루어져야 한다 [1-5].

책임저자: 조 훈 영 (hycho@etri.re.kr)

대전시 유성구 가경동 161번지 한국전자통신연구원

음성언어정보연구부 자동통역연구팀

(전화: 042-860-6591; 팩스: 042-860-4889)

본 논문은 음성인식 과정의 전체 연산량 중에서 절반 이상의 비중을 차지하는 음향적 우도값 (acoustic likelihood score) 계산을 최소화하기 위한 음향 모델의 최적화에 중점을 둔다 [3]. HMM 기반의 음성인식에서 음향 모델을 최적화하기 위해 주어진 데이터에 대해 우선 충분히 많은 수의 모델 파라미터를 이용하여 음향 모델을 학습한 뒤, 각 HMM 상태의 가우시안 성분 (Gaussian mixture component)들을 점차적으로 줄여나가는 방법에 대한 연구들이 진행되어 왔다. 이 방식들은 확률적 분포가 가장 유사한 두 개의 가우시안 성분들을 점진적으로 통합하여 이진 트리를 구성한 뒤, 트리를 적정 수준에서 가지치기 (pruning)하여 최적의 음향 모델을 생성해 낸다. 이 때, 두 가우시안 성분 간의 거리 척도로서 K-L 거리 (Kullback-Leibler divergence), Bhattacharyya 거리, 가우시안 성분들의 가중치 (mixture weight)의 합 등이 연구되어 왔으며, K-L 거리에 기반한 방식이 상대적으로 좋은 성능을 나타낸다고 보고된 바 있다 [3-5][7, 8]. 또한, 가우시안 성분들 간의 K-L 거리를 계산하는 과정에서 가우시안 성분의 가중치를 반영할 경우 더욱 향상된 최적화 성능을 얻을 수 있었다 [3].

가우시안 성분의 가중치는 K-L 거리 계산에서 반영하는 것 이외에도, 두 개의 가우시안 성분을 통합하는 단계에서도 적용될 수 있으나, 기존의 방법들은 각각의 가우시안 성분들에 해당하는 데이터 샘플들의 수가 동일하다고 가정하고 가중치를 고려하지 않았다 [3, 4]. 본 연구에서는 이를 수정하여 가우시안 성분들을 통합하는 단계에서 가우시안 성분의 가중치를 반영할 수 있도록 개선하였다.

본 논문의 2장에서는 최소 서술 길이에 기반한 음향 모델의 최적화 과정을 간략히 정리하고, 기존의 가중 K-L 거리를 설명한다. 제 3장에서는 두 근접한 가우시안 성분들을 통합하는 방법과 관련하여 제안한 방법을 기술하며, 제 4장에서 HMM 기반의 음향 모델 최적화에서 최소 서술 길이 척도에 대해 보다 자세히 설명하기로 한다. 제 5장에서 실험 및 결과를 기술하고, 제 6장에서 결론을 맺는다.

II. MDL에 기반한 음향 모델 최적화

최소 서술 길이 (minimum description length; MDL)에 기반한 최적화 방법은 각각의 HMM 상태를 구성하는 가우시안 성분들에 대해 유사한 성분들끼리 통합하여 새

로운 가우시안 성분을 만들고 이를 부모 노드화하는 과정을 반복하여 HMM 상태별로 이진 트리를 구성한다. 그 다음 단계로는 생성된 이진 트리의 뿌리 노드로부터 하향 (top-down)으로 진행하면서 각 노드에서 자식 노드들의 노드 분리 (split)에 따른 서술 길이 값을 계산하고, 분리 전과 분리 후의 서술 길이 값을 비교하여 분리 후의 서술 길이 값이 더 큰 경우에는 해당 노드에서 멈춘다. 이 과정을 음향 모델의 모든 상태들에 대해 수행함으로써 최적화된 음향 모델을 생성한다. 이러한 최적화 방식에서 주된 이슈는 두 개의 가장 근사한 가우시안 성분들을 찾기 위한 거리 척도, 가우시안 성분들을 통합하는 방법 및 최소 서술 길이 척도이다. 본 장에서는 가우시안 성분들 간의 거리 척도에 대해 살펴보고, 다음 장에서 가우시안 성분들의 통합 방법 및 MDL 척도에 대해 보다 자세히 기술하기로 한다.

연속 HMM 기반의 음성인식에서 임의의 상태를 s 라고 하고, 이 상태 s 에서 가우시안 성분들의 개수가 M 개라고 할 때, 상태 s 에서 관측벡터 \mathbf{x} 의 출력 확률 또는 우도 (likelihood)는 다음과 같이 계산된다.

$$\Pr(\mathbf{x} | s) = \sum_{m=1}^M G_m(\mathbf{x}) = \sum_{m=1}^M w_m \cdot g_m(\mathbf{x}) = \sum_{m=1}^M w_m \cdot N(\mathbf{x}; \boldsymbol{\mu}_m, \boldsymbol{\sigma}_m) \quad (1)$$

위 식 (1)에서 w_m 는 가우시안 성분의 가중치 (mixture weight)이며, $\boldsymbol{\mu}_m, \boldsymbol{\sigma}_m$ 은 각각 m 번째 가우시안 성분의 평균 및 대각 공분산 행렬을 뜻한다. 임의의 두 가우시안 성분은 $G_p(X), G_q(X)$ 라고 하면, 이 두 확률분포 사이의 유사도를 측정하는 대표적인 방법 중의 하나로서 K-L 거리 척도는 다음과 같이 정의된다.

$$d_{KL}(G_p(x), G_q(x)) = \int g_p(x) \log \frac{g_p(x)}{g_q(x)} dx + \int g_q(x) \log \frac{g_q(x)}{g_p(x)} dx \quad (2)$$

또, 여기에 가우시안 성분의 가중치를 추가적으로 고려할 경우인 가중 K-L 거리 척도는 다음과 같다 [3].

$$\begin{aligned} d_{wKL}(G_p(x), G_q(x)) &= \int w_p g_p(x) \log \frac{w_p g_p(x)}{w_q g_q(x)} dx + \int w_q g_q(x) \log \frac{w_q g_q(x)}{w_p g_p(x)} dx \\ &= \frac{1}{2} \sum_{d=1}^D \left[(w_q - w_p) \log 2\pi\sigma_p^2(d) + (w_p - w_q) \log 2\pi\sigma_q^2(d) \right] \end{aligned}$$

$$\begin{aligned}
 & + \frac{w_p \{ \sigma_p^2(d) + (\mu_p(d) - \mu_q(d))^2 \}}{\sigma_q^2(d)} \\
 & + \frac{w_q \{ \sigma_q^2(d) + (\mu_q(d) - \mu_p(d))^2 \}}{\sigma_p^2(d)} \Bigg\} \\
 & + D \{ (w_p - w_q) \log w_p + (w_q - w_p) \log w_q \\
 & - 0.5(w_p + w_q) \}
 \end{aligned} \tag{3}$$

식 (3)에서 D 는 특징벡터의 차수를 뜻한다. 여러 개의 가우시안 성분들 중 가장 근사한 두 개를 찾아서 하나로 통합하는 단계에서 중요한 점은 통합 이전과 통합 이후의 우도 값의 변화가 가장 작아야 한다는 것이며, 이 측면에서 볼 때, 식 (3)의 가중 K-L 거리가 가우시안 성분의 가중치를 고려하지 않는 식 (2)의 K-L 거리에 비해 더 나은 성능을 보였으므로 본 연구에서는 가중 K-L 거리를 적용하였다 [3].

III. 가우시안 성분 통합 방법

앞 장에서 HMM 상태별로 생성한 트리에서 노드 k 의 자식 노드들을 구성하는 M_k 개 (이진 트리의 경우 $M_k = 2$)의 가우시안 성분들에 대해서 i 번째 차원의 평균 및 공분산 값을 각각 $\mu_m^{(k)}(i)$ 및 $\sigma_m^{(k)}(i)$ 라고 할 때, 이들을 통합하여 생성된 노드 k 의 평균 벡터 및 대각 공분산 행렬 값은 다음 식 (4), (5)와 같이 계산된다.

$$\mu_k(i) = \frac{1}{M_k} \sum_{m=1}^{M_k} E(x_m^{(k)}(i)) = \frac{1}{M_k} \sum_{m=1}^{M_k} \mu_m^{(k)}(i), \tag{4}$$

$$\begin{aligned}
 \sigma_k^2(i) &= \frac{1}{M_k} \sum_{m=1}^{M_k} E((x_m^{(k)}(i) - \mu_k(i))^2) \\
 &= \frac{1}{M_k} \left[\sum_{m=1}^{M_k} \sigma_m^{(k)2}(i) + \sum_{m=1}^{M_k} \mu_m^{(k)2}(i) - M_k \mu_k^2(i) \right] \tag{5}
 \end{aligned}$$

위 식에서 $x_{in}^{(k)}(i)$ 는 노드 k 에서 발생한 샘플 데이터 벡터의 i 번째 값을 뜻한다. 식 (4) 및 (5)와 같은 기존의 통합 방법은 각 가우시안 성분으로부터 발생된 학습 데이터 샘플의 개수가 동일하며 따라서 가우시안 성분의 가중치가 모두 동일하다고 가정하는 방식이다. 그러나, 실제로는 HMM의 각각의 상태를 구성하는 GMM (Gaussian mixture model)에서 각 가우시안 성분에 해당하는 데이터 샘플의 수는 상이하므로 가중치 값은 모두 다른 값을 가지며, 이를 모델 통합 시에 반영할 필요가 있다. 따라서

본 연구에서는 이를 반영하여 식 (6)–(8)에 해당하는 새로운 통합 방식을 제안한다.

A. 가중치 정규화:

$$\alpha_m^{(k)} = \frac{w_m^{(k)}}{\sum_{m=1}^{M_k} w_m^{(k)}} \tag{6}$$

B. 평균 통합:

$$\mu_k(i) = \sum_{m=1}^{M_k} \alpha_m^{(k)} E(x_m^{(k)}(i)) = \sum_{m=1}^{M_k} \alpha_m^{(k)} \mu_m^{(k)}(i) \tag{7}$$

C. 공분산 통합:

$$\begin{aligned}
 \sigma_k^2(i) &= \sum_{m=1}^{M_k} \alpha_m^{(k)} E[(x_m^{(k)}(i) - \mu_k(i))^2] \\
 &= \sum_{m=1}^{M_k} \alpha_m^{(k)} E[x_m^{(k)2}(i) - 2x_m^{(k)}(i)\mu_k(i) + \mu_k^2(i)] \\
 &= \sum_{m=1}^{M_k} \left[\alpha_m^{(k)} E[x_m^{(k)2}(i)] - 2\mu_k(i)\mu_k(i) + \sum_{m=1}^{M_k} \alpha_m^{(k)} E[\mu_k^2(i)] \right] \\
 &= \sum_{m=1}^{M_k} \left[\alpha_m^{(k)} \sigma_m^{(k)2}(i) + \alpha_m^{(k)} \mu_m^{(k)2}(i) \right] - \mu_k^2(i) \tag{8}
 \end{aligned}$$

제안한 방법은 노드 k 의 자식 노드들에 해당하는 가우시안 성분들의 가중치들에 대한 상대적 중요도를 총합이 1이 되도록 정규화하고, 이를 다시 평균과 공분산 통합에 적용한다. 식 (4)와 식 (7) 및 식 (5)와 식 (8)을 각각 비교해 보면, 기존의 통합 방법은 제안한 방법에서 $\alpha_m^{(k)} = 1/M_k = const.$ 인 특수한 경우로 볼 수 있다.

IV. 최소 서술 길이 척도

다량의 음성 데이터를 이용하여 음향 모델의 파라미터를 학습함에 있어 데이터에 비해 모델의 파라미터 수가 너무 적을 경우에는 데이터를 모델이 잘 표현하지 못하게 되며, 이와 반대의 경우에는 데이터의 부족으로 모델 파라미터 추정치의 정확도가 떨어지게 된다 [4–6]. 최소 서술 길이는 주어진 데이터에 대해 최적의 모델 파라미터들을 얻을 수 있는 최적화 척도로서, 식 (9)는 주어진 데이터 집합 $x^N = x_1, \dots, x_N$ 에 대한 GMM 모델 h 의 최소 서술 길이를 나타낸다 [4][9].

$$DL_h(x^N) = -\log P_{\hat{\theta}(h)}(x^N) + \alpha \cdot \frac{Y_h}{2} \log N + \log l \tag{9}$$

이 식에서 첫째 항은 일련의 데이터 x^N 에 대한 모델 h 의 코드 길이를 의미한다. 여기서 $\hat{\theta}(h)$ 는 최대 우도 기법으로 추정된 모델 h 의 파라미터값들을 나타내며, HMM 기반의 음향 모델에서는 가우시안 성분들의 평균 벡터와 공분산 행렬에 해당한다. 모델의 복잡도가 증가할수록 주어진 데이터에 대한 확률값이 증가하므로 첫째 항은 점점 작은 값을 갖게 된다. 둘째 항은 모델 h 의 복잡도 및 데이터 샘플의 개수 N 과 관련된 항으로서 γ_h 는 모델 h 의 파라미터들의 개수를 의미하고, α 는 둘째 항의 기울기를 결정하는 변수로서 0보다 큰 값을 갖는다. 둘째 항은 모델 h 의 복잡도가 증가할수록 값이 증가하며 모델 파라미터 수의 증가에 대한 패널티로 작용한다. 셋째 항은 모델 h 를 선택하는 데 있어 필요한 코드 길이로서 여기서는 상수로 가정한다.

모델이 복잡해질수록 식 (9)에서 첫째 항의 값은 감소하고, 둘째 항의 값은 증가하기 때문에, 모델 파라미터를 증가시키에 따라 특정 지점에서 최소값을 갖게 된다. MDL 기법은 이 지점에서의 모델, 즉, 표현 길이 DL_h 이 가장 작은 모델 h 를 최적 모델로서 선정하게 된다.

V. 실험 및 결과

음향 모델 최적화를 위한 기존 방법 및 제안한 방법을 평가하기 위해 여행영역 영어문장 음성인식 태스크에 대해 인식 실험을 수행하였다. HMM 모델은 WSJ (Wall Street Journal) DB를 포함한 총 330시간 분량의 음성 데이터에 대해 HTK Toolkit을 사용하여 학습하였다. 음성 데이터는 모두 16 kHz 샘플링된 데이터이며, 특징벡

터로 39차의 MFCC (Mel frequency cepstral coefficient)를 사용하였다. 각 HMM은 5개의 상태를 left-to-right 토폴로지로 구성하였으며, 상태 클러스터링 등의 과정을 거쳐 최종적으로 문맥 종속 삼음소 모델 (triphone)들을 생성하였다.

기본 시스템 (baseline)의 음향모델에서 서로 다른 HMM 상태의 수는 4554개, 물리적 삼음소 HMM 모델 개수는 17159개이며, 각 상태별로 16개의 가우시안 성분을 사용하므로 음향모델 전체의 가우시안 성분 개수는 72864개이다. 인식 실험을 위해 여행영역 텍스트 77만 6천 문장으로 3-gram 언어 모델을 생성하였다. 음성 인식 실험용 디코더는 HTK 3.4.1의 HDecode를 이용하였고, 인식 실험을 위한 데이터는 실험실 환경에서 수집된 여행 영역 평가용 200문장 음성 데이터를 사용하였다.

그림 1은 전체 HMM 상태들에서 관측된 학습 데이터들의 개수를 나타내며, 가로축의 레이블에는 일부 HMM 상태들의 식별자만을 표시하였다. 그림에서 각각의 상태들에서 관측되는 학습 데이터 개수는 매우 상이하며, 이에 맞추어 각 HMM 상태는 보다 최적화되어야 함을 알 수 있다.

표 1은 HMM 상태별로 가우시안 성분의 개수가 변함에 따른 인식 성능을 나타낸다. 기본 시스템 (baseline)은 음향 모델 학습 과정에서 가우시안 성분들을 증가시키는 과정에서 얻은 음향 모델을 사용하였다. 표에서 GCR (Gaussian component reduction)은 MDL은 적용하지 않고, 단지 제 2, 3장에서 설명한 가우시안 성분들 간의 거리 측정 및 제한한 통합 기법을 사용하여 각 HMM 상태별로 가우시안 성분들을 16개로부터 4개로 점차적으로 통합해 나간 경우로서 각 상태들은 동일한 개수의 가우시안

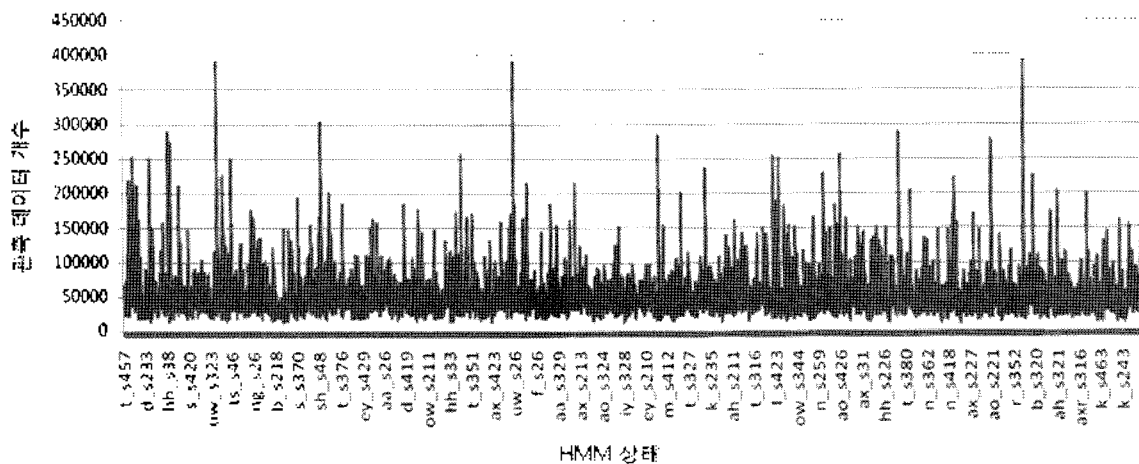


그림 1. HMM 상태별로 관측된 특징 벡터의 개수
 Fig. 1. Number of observed feature vectors at each HMM state.

표 1. GCR에 의한 음향 모델 최적화 성능

Table 1. Performance of acoustic model optimization by Gaussian component reduction (GCR).

음향 모델 상대적 크기	HMM상태별 mixture 개수	Gaussian 성분의 전체 개수	Baseline	GCR (weighted merge)	
				재학습 0회	재학습 3회
1	16	72864	81.51	-	-
3/4	12	54648	79.45	81.29	81.83
1/2	8	36432	69.47	78.69	79.83
1/4	4	18126	62.26	70.50	71.69

표 2. 기존[3, 4] 및 제안한 가우시안 통합 방식의 성능 비교 및 MDL의 적용에 따른 인식 성능

Table 2. Optimization performance using the previous and the proposed Gaussian component merging methods with the application of MDL criterion.

음향 모델 상대적 크기	Gaussian 성분의 전체 개수	MDL (conventional merge)		MDL (weighted merge)	
		재학습 0회	재학습 3회	재학습 0회	재학습 3회
1	73K	-	-	-	-
3/4	55K	80.97	81.29	80.98	81.62
1/2	36K	78.47	79.56	79.12	79.28
1/4	18K	72.89	74.67	74.02	75.16

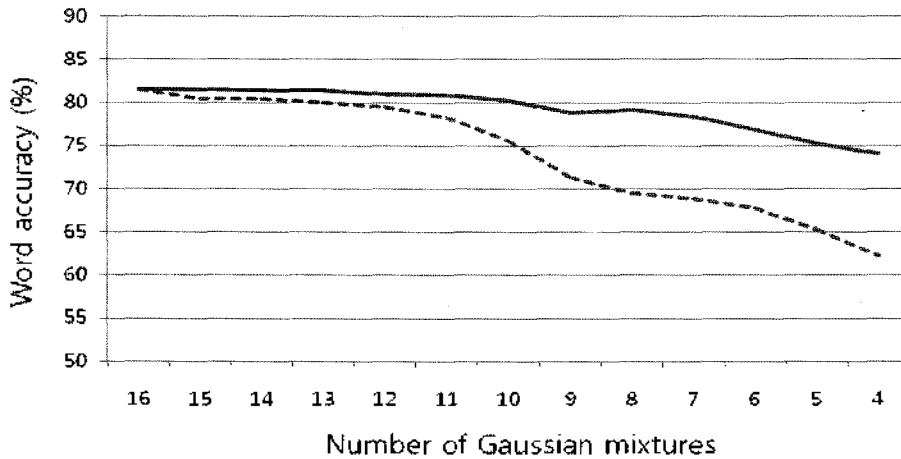


그림 2. 기본 시스템과 개선된 MDL의 성능 비교: 점선은 기본 시스템의 단어 인식률, 실선은 개선된 MDL의 단어 인식률.

Fig. 2. ASR performance of the baseline (dotted line) and the improved MDL method (solid line).

성분들을 갖는다. GCR 방법은 음향 모델 크기가 기본 시스템의 75% 수준까지는 인식 성능의 차이가 거의 없으며, 기본 시스템의 50% 수준에서도 약 1.7%의 성능 저하가 발생하는 정도였다. 가우시안 성분이 12인 경우, GCR에 의해 통합한 뒤 3회의 재학습으로 모델 파라미터를 재조정할 경우 기본 시스템에 비해 오히려 성능이 증가하기도 함을 볼 수 있다.

표 2는 제 4장에서 기술한 MDL 계산식을 적용한 MDL 최적화 기법의 성능을 보이며, 이 경우 표 1과 달리 HMM 상태들은 서로 상이한 개수의 가우시안 성분들을 가질 수 있다. 이 표에서는 HMM 상태 내의 가우시안 성분들을 통합하는 방식에 있어서 기존의 방식 (conventional merge)과 제안한 방법 (weighted merge)을 비교하였다.

표에서 제안한 방법이 기존 방식에 비해 더 향상된 결과를 나타냄을 알 수 있으며, 특히 감축된 모델의 크기가 작을수록 더욱 효과적임을 알 수 있었다. 음향 모델을 위한 학습 데이터가 더 이상 재공되지 않는 상황에서는 재학습을 수행할 수 없으며, 이는 표에서 재학습 0회에 해당한다. 학습 데이터가 재공되는 경우, 수 차례의 재학습을 통해 MDL에 의해 통합된 모델 파라미터를 재조정할 경우 성능이 더욱 향상됨을 볼 수 있다. 제안한 가우시안 통합 방법은 가우시안 성분의 가중치에 포함된 통계적 정보를 반영함으로써 학습 데이터가 주어지지 않는 경우에도 통합된 모델의 평균 및 공분산을 보다 정확하게 추정할 수 있었다고 판단된다.

마지막으로 그림 2는 가우시안 성분의 개수가 16개에

서 4개로 감축됨에 따른 인식 성능의 추세를 나타낸다. 개선된 MDL 기반의 음향 모델 감축 기법은 모델의 크기가 50 % 수준일 경우까지에 대해 약간의 성능 저하가 발생할 뿐이며, 감축된 모델의 크기가 25 % 수준일 경우 기본 시스템에 비해 10 % 이상의 성능 개선을 얻을 수 있었다.

VI. 결론

본 논문에서는 기존의 MDL 기반의 음향 모델 감축 기법을 살펴보고, 유사도가 높은 가우시안 성분들을 통합함에 있어 가우시안 성분의 가중치 (mixture weight)를 반영할 수 있도록 기존의 통합 방식을 개선하였다. 모든 HMM 상태가 동일하게 16개의 가우시안 성분을 갖는 기존 음성인식 시스템과 비교한 결과, 개선된 MDL 기반의 모델 감축 기법은 음향 모델의 크기가 기존 시스템의 50 % 수준일 경우까지에 대해서도 성능 저하가 크지 않았으며, 감축된 음향 모델 크기가 25 % 수준일 경우, 동일 크기의 기존 시스템에 비해 10 % 이상의 성능 개선을 얻을 수 있었다.

감사의 글

본 연구는 지식경제부의 IT성장동력기술개발사업 중 휴대형 한/영 자동통역 기술개발 과제 (2009-S-019-02)에서 수행되었습니다.

참고 문헌

1. E. Bocchieri, "Vector quantization for the efficient computation of continuous density likelihoods," in *Proc. ICASSP*, pp. 692-695, 1993.

2. G.-J. Jung, H.-Y. Cho, and Y. H. Oh, "Data-driven subvector clustering using the cross-entropy method," in *Proc. ICASSP*, pp. 977-980, 2007.

3. A. Ogawa and S. Takahashi, "Weighted Distance Measure for Efficient Reduction of Gaussian Mixture Components in HMM-based Acoustic Model," in *Proc. ICASSP*, pp. 4173-4176, 2008.

4. K. Shinoda and K. Iso, "Efficient reduction of Gaussian components using MDL criterion for HMM-based speech recognition," in *Proc. ICASSP*, vol. 1, pp. 869-872, 2002.

5. K. Shinoda and T. Watanabe, "MDL-based context-dependent subword modeling for speech recognition," *J. Acoust. Soc. Jpn. (E)*, vol. 21, no. 2, pp. 99-102, 2000.

6. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd Ed., Wiley Interscience, 2000.

7. G. F. G. Yared, F. Violaro and L. C. Sousa, "Gaussian elimination algorithm for HMM complexity reduction in continuous speech recognition system," in *Proc. INTERSPEECH*, pp. 377-380, 2005.

8. P. L. Dognin, J. R. Hershey, V. Goel, and P. A. Olsen, "Refactoring acoustic models using variational density approximation," in *Proc. ICASSP*, pp. 4473-4476, 2009.

9. J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Trans. IT* 30, pp. 629-636, 1984.

저자 약력

• 조 훈 영 (Hoon-Young Cho)

2006년 ~ 현재 ETRI 음성언어정보연구부 선임연구원
 2004년 ~ 2005년 LG전자기술원 선임연구원
 2003년 ~ 2004년 미국 UCSD Post Doc.
 1998년 ~ 2002년 KAIST 전산학 박사
 1996년 ~ 1997년 KAIST 전산학 석사
 1991년 ~ 1995년 KAIST 전산학 학사
 ※ 주관심분야: 잡음처리, 음성인식, 입체음향, 멀티미디어 검색

• 김 상 훈 (Sang-Hun Kim)

1992년 ~ 현재 ETRI 음성언어정보연구부 책임연구원
 2007년 ~ 2008년 미국 UCSD 방문연구원
 1998년 ~ 2003년 Univ. of Tokyo 전기공학 박사
 1990년 ~ 1992년 KAIST 전기공학 석사
 1986년 ~ 1990년 연세대학교 전기공학 학사
 ※ 주관심분야: 음성합성, 음성인식, 자동통역