

이웃 정보에 기초한 반모델을 이용한 발화 검증*

윤영선(한남대)

<차 례>

1. 서론	4.3. 이웃 모델을 이용한 반모델의 구축
2. 베이스 인수 및 우도비 테스트	5. 실험 및 결과
3. 이웃 모델	5.1. 이웃 모델의 결합
3.1. 배경	5.2. 사전 확률의 적용
3.2. 정의	5.3. 이웃 모델 결합에 의한 반모델의 구축
3.3. 모델 간의 비교	5.4. 분석
4. 대안 가설 모델의 구축	6. 결론
4.1. 반모델과 이웃 모델의 비교	
4.2. 대안 가설 모델의 구축	

<Abstract>

Utterance Verification Using Anti-models Based on Neighborhood Information

Young-Sun Yun

In this paper, we investigate the relation between Bayes factor and likelihood ratio test (LRT) approaches and apply the neighborhood information of Bayes factor to building an alternate hypothesis model of the LRT system. To consider the neighborhood approaches, we contemplate a distance measure between models and algorithms to be applied. We also evaluate several methods to improve performance of utterance verification using neighborhood information. Among these methods, the system which adopts anti-models built by collecting mixtures of neighborhood models obtains maximum error rate reduction of 17% compared to the baseline, linear and weighted combination of neighborhood models.

* Keywords: Utterance verification, Anti-model, Neighborhood method, Bayes factor, Likelihood ratio test.

* 본 연구는 2008년도 한남대학교 학술연구조성비 지원(과제번호 2008A055)과 정보통신부 및 정보통신연구진흥원의 IT성장동력기술개발사업의 일환으로 수행하였음. [2008-S-036-01, 신성장동력산업용 대용량/대화형 분산/내장처리 음성인터페이스 기술 개발]

1. 서 론

음성 인식 기술의 발달과 상용화에 대한 열망으로 많은 분야에서 음성 인식 결과의 신뢰성에 대해 평가하고자 하는 연구가 진행되고 있다. 신뢰도 측정을 통하여 음성 구간의 여부를 판단한다든지, 비인식 대상 어휘(out-of-vocabulary)를 검출 또는 기각, 잠재적인 음성 인식 오류의 검출 등 다양한 분야에서 연구가 진행되고 있으며, 이 결과는 음성 이해(speech understanding) 또는 담화(dialogue) 등의 연구에 대한 기초 자료로 사용된다.

음성 인식의 신뢰성을 높이는 연구는 크게 신뢰도 척도(confidence measure: CM)와 발화 검증(utterance verification: UV)이라는 이름으로 연구가 진행되어 왔다. CM은 주어진 음성 특징 X 에 대한 음성 인식의 결과 \hat{W} 에 대해 $p(\hat{W}|X)$ 로 표현되는 사후 확률(posterior probability)을 계산하는 방법에 기초를 두고 있으며, 발화 검증의 문제는 통계적 가설 검증(statistical hypothesis testing problem) 문제로 인식되고 있다. 가설 검증의 문제를 통계학에서 정의한 Neyman-Pearson 보조 정리(lemma)에 따르면 특정 조건에서는 우도비 테스트(likelihood ratio test: LRT)가 최적의 해법이 될 수 있다고 하였다. 그러나 LRT 방식을 UV에 적용할 경우 실제 데이터의 분포를 모르기 때문에 반모델(anti-model)이라고도 불리는 대안 가설(alternative hypothesis)을 정확히 모델링하기가 어렵다. 다양한 방법의 대안 가설 또는 반모델을 모델링하는 연구가 진행되어 오고 있으나, 변별 학습(discriminative training)에 의해 구축된 대안 가설이 가장 좋은 성능을 보인다고 알려져 있다. 또 다른 연구 방법으로는 발화 검증 문제를 베이스(Bayes) 통계 이론 관점에서 모델링한 베이스 인수(Bayes factor) 연구 방법이 있다[1]. LRT 방식이 가설을 구성하는 사전 확률(prior probability)들이 동일하다는 가정 하에 모델간의 우도 값을 비교하는 것이라면, 베이스 인수 방식은 사전 확률의 기여도를 반영하고 실제 모델(true model, null hypothesis)과 대안 가설 모델이 확률적 odds¹⁾ 관계에 있다고 가정하여 발화 검증 문제에 접근한 방식이다. 따라서 LRT 방식은 사전 확률, 실제 모델과 대안 가설 모델의 관계에 따라 베이스 인수 연구 방법의 단순화된 경우로 여겨진다[2].

국내의 발화 검증 방법에 대한 연구는 거의 대부분 LRT 방식 중심으로 연구되고 있기 때문에 본 논문에서는 베이스 인수 방식에서 사용되는 이웃 모델(neighborhood model) 방식을 이용하여 LRT 방식의 성능향상을 꾀하였다. 그 기초 연구로서 베이스 인수 방식에서 LRT 방식의 범위를 제한하였던 가정을 완화 또는 변형시켜 각 조건에 따른 성능 변화와 영향을 살펴보고자 하였으며, LRT 방식의 반모델의 구축을 베이스 인수 방식에서 적용하였던 대안 가설 모델로 변경하여

1) 사건 w 에 대해 $p(w)/(1-p(w))$ 의 관계를 가지는 경우

특성을 살펴보았다. 본 논문에서는 베이스 방식에서 이웃 모델을 결정하는 비교 척도를 제안하고, 결정된 이웃 모델의 결합 방식에 따른 성능을 확인하였다. 또한 단순히 베이스 인수 방식을 모델간의 비교 방식에 적용하였을 경우 기존 시스템에 비하여 성능이 저하되기 때문에, 대안 가설 모델로서 이웃모델에서 직접 반모델을 구축하는 방법을 제안하여 성능이 향상됨을 확인할 수 있었다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 발화 검증에서 널리 사용되는 베이스 인수 방식과 우도비 테스트의 상관관계를 조명하고, 3장에서는 베이스 인수 방식에서 사용되는 이웃 모델의 정의와 모델간 비교 알고리즘을 살펴본다. 4장에서는 실제 가설과 대비되는 대안 가설 모델의 연구를 살펴보고, 발화 검증 성능을 향상시키기 위한 대안 가설 모델의 구축 노력과 반모델 방식을 제안한다. 제안된 방법의 유효성을 판단하기 위한 실험 및 결과를 5장에서 정리하며, 마지막으로 본 연구의 요약 및 향후 연구에 대해 6장에서 기술한다.

2. 베이스 인수 및 우도비 테스트

발화 검증은 일반적으로 주장 또는 참에 해당되는 실제 가설 H_0 와 사칭 또는 거짓에 해당되는 대안 가설 H_1 에 대한 통계적 가설 검증의 형태로 간주된다. 전통적으로 베이스 방식이 아닌 일반적인 조건에서는 Neyman-Pearson 보조정리를 이용하여 식 (1)의 조건을 만족하면 실제 가설 H_0 을 거절하는 LRT를 최적의 검증 방법으로 제시하고 있다.

$$\Lambda(\mathbf{X}) = \frac{p(\mathbf{H}_0|\mathbf{X})}{p(\mathbf{H}_1|\mathbf{X})} \leq k \quad (1)$$

식 (1)을 베이스 정리를 이용하여 정리하면,

$$\frac{p(\mathbf{H}_0|\mathbf{X})}{p(\mathbf{H}_1|\mathbf{X})} = \frac{p(\mathbf{X}|\mathbf{H}_0)p(\mathbf{H}_0)}{p(\mathbf{X}|\mathbf{H}_1)p(\mathbf{H}_1)} \quad (2)$$

가 된다. 즉, 실제 가설과 대안 가설에 대한 사후 확률 값은 사전 확률 비에 우도비의 가중치를 곱한 것과 같게 된다. 이 우도비의 가중치를 베이스 인수(Bayes factor)라 한다.

$$B_{01} = \frac{p(\mathbf{X}|\mathbf{H}_0)}{p(\mathbf{X}|\mathbf{H}_1)} \quad (3)$$

다르게 해석한다면 베이스 인수는 사후 확률 값을 사전 확률비로 나눈 것을 의미하기도 한다. 따라서 사전 확률비가 제공된다면 베이스 인수는 화자 인식이나 발화 검증에 직접 사용될 수 있다[2].

식 (3)을 전통적인 통계적 패턴 인식의 문제로 본다면, 실제 가설 H_0 는 평가 데이터 \mathbf{y} 가 H_0 에 해당하는 클래스 C_0 에 속하고, 대안 가설 H_1 은 평가 데이터 \mathbf{y} 가 H_1 에 대응되는 클래스 C_1 에 속한다고 모델링하면 가설 검증 문제는 다음과 같이 표현될 수 있다.

$$B_{01} = \frac{p(\mathbf{y}, \mathbf{X}_0, \mathbf{X}_1 | H_0)}{p(\mathbf{y}, \mathbf{X}_0, \mathbf{X}_1 | H_1)} = \frac{p(\mathbf{y}, \mathbf{X}_0 | C_0)p(\mathbf{X}_1 | C_1)}{p(\mathbf{X}_0 | C_0)p(\mathbf{y}, \mathbf{X}_1 | C_1)} \quad (4)$$

여기서 \mathbf{X}_0 와 \mathbf{X}_1 은 각각 클래스 C_0 과 C_1 에 속하는 학습 데이터이다. 일반적으로 식 (4)를 가우스 분포로 가정하면 확률 밀도 $p(\mathbf{X} | C_k)$ 는 전체 모델 공간에서의 밀도 함수의 적분으로 구할 수 있다.

$$p(\mathbf{X} | C_k) = \int p(\mathbf{X} | \lambda_k) p(\lambda_k) d\lambda_k \quad (5)$$

여기서 λ_k 는 C_k 를 표현하는 모델 변수를 의미한다. 식 (5)를 식 (4)에 결합하면, 다음과 같은 베이스 인수식을 구할 수 있다[3].

$$B_{01} = \frac{\int p(\mathbf{y}, \mathbf{X}_0 | \lambda_0) p(\lambda_0) d\lambda_0 \cdot \int p(\mathbf{X}_1 | \lambda_1) p(\lambda_1) d\lambda_1}{\int p(\mathbf{X}_0 | \lambda_0) p(\lambda_0) d\lambda_0 \cdot \int p(\mathbf{y}, \mathbf{X}_1 | \lambda_1) p(\lambda_1) d\lambda_1} \quad (6)$$

따라서 베이스 인수를 계산하기 위해서는 각 모델의 사전 확률 $p(\lambda_0)$ 와 $p(\lambda_1)$ 가 반드시 필요하나 최적의 방법이 없기 때문에 일반적으로 근사 방법을 이용하여 사전 확률을 계산한다.

반면 베이스 인수를 구하는 모든 데이터가 독립이라고 가정하면,

$$\begin{aligned} B_{01} &= \frac{p(\mathbf{y}, \mathbf{X}_0 | C_0)p(\mathbf{X}_1 | C_1)}{p(\mathbf{X}_0 | C_0)p(\mathbf{y}, \mathbf{X}_1 | C_1)} \\ &= \frac{p(\mathbf{X}_0 | C_0)p(\mathbf{y} | C_0)p(\mathbf{X}_1 | C_1)}{p(\mathbf{X}_0 | C_0)p(\mathbf{y} | C_1)p(\mathbf{X}_1 | C_1)} \\ &= \frac{p(\mathbf{y} | C_0)}{p(\mathbf{y} | C_1)} = \Lambda_{01} \end{aligned} \quad (7)$$

가 되어 Neyman-Pearson 보조 정리로 변환된다. 즉, 식 (7)에서 실제 가설 H_0 와 대안 가설 H_1 에 해당되는 클래스를 표현하는 모델에 대한 사전 확률이 동일하다면 LRT가 된다.

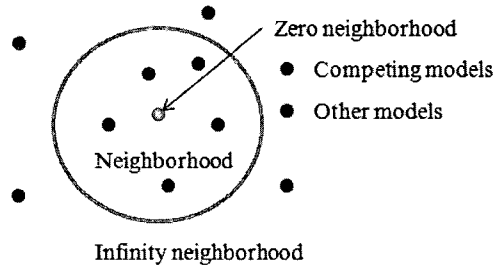
3. 이웃 모델

3.1. 배경

통계적 방식을 이용한 패턴인식의 대표적인 방법으로 사용되는 Neyman-Pearson 보조정리는 실제 가설과 대안 가설에 대한 우도 값의 비로 실제 가설을 인정할 것인지 또는 기각할 것인지를 결정한다. 따라서 실제 가설의 기각 성능은 실제 가설에 대응되는 대안 가설을 어떻게 모델링할 것인가에 따라 좌우된다. 일반적으로 대안 가설에 해당되는 모델 집합은 반모델(anti-model)이라고도 불리며 통계적 가설 검증의 성능을 높이기 위한 방법으로 다양한 구축방법이 연구되고 있다. 본 논문에서는 이해를 명확하게 하기 위하여 반모델은 단일 모델 방식에 의하여 대안 가설 모델을 구축하는 것을 말하며, 대안 가설 모델은 복수개의 모델을 서로 다른 방법에 의하여 결합하는 것을 지칭하도록 한다. 반모델을 구축하는 대표적인 방법은 화자 식별에서 적용하던 방법으로 모든 화자 정보를 포함한 universal background model (UBM) 방법이 있다. 비슷한 방식으로 발화 검증에서 모든 대상 어휘에 대한 모델의 혼합 방식을 통하여 반모델을 구축한다. 또 다른 방법으로는 초기에 핵심어 검출 방식에 적용되었던 필러(filler) 방법으로 인식 단계에서 인식 대상 어휘와 비 인식 대상 어휘를 혼합하여 최상위 우도를 갖는 모델로 인식하는 방법이다. 이들 반모델의 특징은 실제 가설 H_0 와 독립적으로 대안 가설 H_1 에 해당되는 인식모델들을 결합하여 모델링하거나 또는 개별적으로 모델링한다. 그러나 베이스 인수 방식에서는 실제 가설에 대응되는 모델과 대안 가설에 대응되는 모델들은 확률적으로 odds의 관계가 있다고 보고 실제 가설과 대안 가설에 포함된 모델들을 선택한다. 대안 가설 모델을 구축하는 대표적인 방법으로 모델들의 분포에서 거리에 따라 단계적으로 범위를 고려하여 이웃 모델을 구하고, 이 모델들을 결합하는 방식이 사용되고 있다[1][3]. 기존의 연구 중에서 이웃모델과 유사하게 실제 가설과 혼동되는 대안 가설을 설정하여 LRT를 진행하기도 하나, 이는 언어 정보(linguistic information)나 혼동 행렬(confusion matrix)을 이용하여 고정된 모델 집합을 결정한 후 대안 가설 모델을 구축하기 때문에[5], 음향학적 거리를 이용하여 대안 가설 모델을 구축하는 이웃 모델과는 차이가 있다.

3.2. 정의

이웃 모델은 HMM의 모델 공간에서 N개의 서로 다른 인식 모델이 존재한다고 하였을 때 실제 가설을 표현하는 모델의 주위에 있는 모델들을 단계별로 모델링하는 방법이다[1]. <그림 1>에서 보는 바와 같이 실제 가설에 해당되는 모델은 제로 이웃(zero neighborhood, 이웃이 없음)으로 규정하고 일정 거리 안에 있는 모델



<그림 1> 모델 공간에서의 이웃 모델의 정의

들은 이웃 모델(neighborhood)로 나머지는 무한 이웃(infinity neighborhood)으로 규정한다.

실제 모델(실제 가설을 표현하는 모델)과 이웃 모델들과의 비교는 널리 알려진 유클리드 거리(Euclidean distance)나 켈백-라이블러 발산(Kullback-Leibler divergence: KLD)을 이용할 수 있다. 기존의 연구[3]에서는 유클리드 거리를 이용하여 모델간의 거리를 계산하고 이웃 모델을 정의하였으나, 본 연구에서는 확률 이론과 정보 이론에서 널리 사용되는 두 확률 분포사이의 차이를 측정하는 KLD를 채택하였다. KLD는 실제 확률 분포 P 와 임의의 확률 분포 Q 와의 차이를 모델링하고 있지만, 대칭이 아니기 때문에 거리라는 용어보다는 발산(divergence)이라는 용어를 사용한다. 일반적으로 P 는 이론적인 모델, 또는 정확하게 계산된 확률 분포를 의미하며, Q 는 실제 데이터로부터 측정된 관측 값 또는 P 의 근사 값을 나타낸다. 분포 P 와 Q 가 주어지면 KLD $D_{KL}(P||Q)$ 는 일반적으로 식 (8)과 같이 정의되며, P 와 Q 의 결합 엔트로피(entropy)에서 P 의 엔트로피 값을 빼는 의미이기 때문에 항상 0보다 크거나 같게 된다.

$$\begin{aligned}
 D_{KL}(P||Q) &= \int p(x) \log \frac{p(x)}{q(x)} dx \\
 &= - \int p(x) \log q(x) dx + \int p(x) \log p(x) dx \\
 &= H(P, Q) - H(P) \geq 0
 \end{aligned} \tag{8}$$

여기서 P 와 Q 는 가우스 분포를 따르며 $P \sim N(\mu_1, \Sigma_1)$ 이고 $Q \sim N(\mu_2, \Sigma_2)$ 로 표현된다고 하면, d 차원 가우스 분포에 대한 KLD는 다음과 같이 정의될 수 있다.

$$\begin{aligned}
 D_{KL}(P||Q) &= \frac{1}{2} \left(\log \frac{|\Sigma_2|}{|\Sigma_1|} + \text{Tr}(\Sigma_2^{-1} \Sigma_1) \right. \\
 &\quad \left. + (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2) - d \right)
 \end{aligned} \tag{9}$$

위 식에서 알 수 있듯이 KLD는 비대칭 척도이기 때문에 일반적인 거리 개념

으로 사용할 수 없다. 그러나 베이스 인수 방식에서는 실제 가설 H_0 를 표현하는 모델 λ_0 에 대응되는 이웃 모델들을 결정하는데 사용되기 때문에 대칭 개념의 거리 척도 대신 KLD를 사용할 수 있다.

3.3. 모델 간의 비교

다음으로 고려할 사항은 KLD를 이용하여 모델간의 거리를 측정하고자 할 때, 모델을 구성하는 상태(state)와 상태를 구성하는 혼합 가우스 모델(mixture Gaussian model)의 비교 방법이다. 두 모델들은 상태수가 같거나 다를 수도 있기 때문에, 원칙적으로 동적 프로그래밍 방식에 의하여 상태 비교를 하여야 한다. 그러나, 상태 간의 대응관계를 살펴보기 위해 동일한 수의 상태를 갖는 모델들에 상태 정합을 실행한 결과, 거의 모든 모델에서 동일한 상태 번호가 정합되는 것을 알 수 있었다. 따라서 본 연구에서는 상태수가 같다면 동일한 상태끼리 비교하고, 다르다면 동적 정합 방법을 따른다. <표 1>에서 제시된 알고리즘은 각 인식 모델의 구조가 left-to-right의 형태를 가지며 건너뛰기(skip) 링크가 없을 경우를 가정한다.

<표 1> 상태 수에 따른 상태 비교 알고리즘

```

// StateNo( $\lambda_k$ ) means the number of states of  $\lambda_k$ 
if StateNo( $\lambda_1$ ) == StateNo( $\lambda_2$ ) then
    return  $D_{state}(StateNo(\lambda_1), StateNo(\lambda_2), F)$ ;
else
    return  $D_{state}(StateNo(\lambda_1), StateNo(\lambda_2), T)$ ;
function  $D_{state}(s_1, s_2, flags)$ 
begin
//  $C(s_1, s_2)$  means the distance or divergence function between states
if not flags then
    return  $D_{state}(s_1-1, s_2-1, flags) + C(s_1, s_2)$ ;
else
    return min {  $D_{state}(s_1-1, s_2, flags)$ ,  $D_{state}(s_1-1, s_2-1, flags)$ ,
                 $D_{state}(s_1, s_2-1, flags)$  } +  $C(s_1, s_2)$ ;
endif
end function
    
```

비교하고자 하는 상태 쌍이 결정되면, $C(s_1, s_2)$ 에 의해 상태 간의 거리 또는 발산 정도를 계산하여야 한다. 그러나 혼합 밀도는 상태와 달리 순서의 개념이 없기 때문에 모든 경우의 비교에 대하여 값을 계산하고 그 중에서 최소값을 거리 값으로 선택한다. 이 방법을 선택하더라도 각 상태에서의 혼합 밀도 수가 달라지는 경우 비교에 어려움이 생기기 때문에 <표 2>와 같은 발견법적인 방법을 제안

<표 2> 상태 모델의 발산 거리 계산

```

function C( $s_1, s_2$ )
begin
  forwardKLD = backwardKLD = 0
  for all mixtures  $m_i$  of  $s_1$ 
    for all mixtures  $m'_j$  of  $s_2$ 
      forwardKLD +=  $\min D_{KL}(m_i, m'_j)$ 
  forwardKLD /= MixtureNo( $s_1$ )
  for all mixtures  $m'_j$  of  $s_2$ 
    for all mixtures  $m_i$  of  $s_1$ 
      backwardKLD +=  $\min D_{KL}(m_i, m'_j)$ 
  backwardKLD /= MixtureNo( $s_2$ )
  divergence = (forwardKLD + backwardKLD)/2;
  return divergence;
end function

```

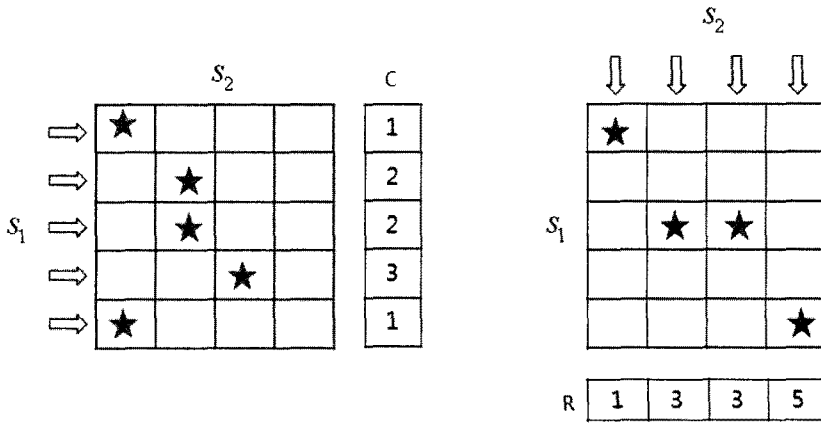
한다.

<표 2>의 발견법적인 비교에서는 실제 모델의 상태 s_1 과 임의 모델에서의 상태 s_2 의 비교를 가정한다. 혼합 밀도의 비교는 KLD를 사용하였으며, 비대칭이기 때문에 실제 모델과 임의 모델의 비교 순서는 바뀌지 않는다. 먼저 상태 s_1 의 모든 혼합 밀도에 대해 s_2 에서 가장 가까운 혼합 밀도와의 KLD를 더하여 이를 *forwardKLD*라 한다. 다음으로는 s_2 의 모든 혼합 밀도에 대해 s_1 에서 가장 가까운 혼합 밀도와의 KLD를 더하여 이를 *backwardKLD*라 하며, 전체 발산 거리는 *forwardKLD*와 *backwardKLD*의 평균으로 정의한다. *forwardKLD*는 실제 모델의 혼합 밀도에 대해 임의의 비교 밀도 중 가장 가까운 혼합 밀도를 선택하는 것이며, *backwardKLD*는 임의의 비교 밀도에 가장 가까운 실제 모델의 혼합 밀도를 선택하여 더한 값이다. 여기에서 *forwardKLD*와 *backwardKLD*의 KLD를 구하는 순서는 변하지 않고 다만 대응되는 혼합 밀도를 다르게 정의한 것이다.

*forwardKLD*와 *backwardKLD*의 의미를 그림을 통하여 표현한다면 <그림 2>와 같다. 그림에서 *C*와 *R*은 열과 행에서 가장 가까운 혼합 밀도의 인덱스를 보관하는 역할을 한다.

4. 대안 가설 모델의 구축

본 장에서는 발화 검증에서 실제 가설 모델과 대응되는 대안 가설 모델의 구현 방법을 LRT 방식과 베이스 인수 방법 측면에서 비교한다. 또한 베이스 인수



$$forwardKLD = \frac{1}{N_{s_1}} \sum_{m_i \in s_1} D_{KL}(m_i \| m'_{C(i)}) \quad backwardKLD = \frac{1}{N_{s_2}} \sum_{m_j \in s_2} D_{KL}(m_{R(j)} \| m'_j)$$

<그림 2> forwardKLD와 backwardKLD의 개념적 비교

방식의 이웃 모델을 이용하여 대안 가설 모델의 가능한 구축 방법을 설명하고, 마지막으로 이웃 모델을 이용하여 반모델을 구축하는 방법을 제안한다.

4.1. 반모델과 이웃 모델의 비교

반모델을 표현하는 방법으로 전체 혼합 밀도 방식이 널리 사용되고 있다. 이 방법은 대안 가설 모델을 구축하기 위하여 전체 모델들의 혼합 밀도들을 단일의 모델로 재구축하는 방식이기 때문에, 별도의 비 인식 단어 모델의 학습 없이 쉽게 모델을 재구성할 수 있다는 장점이 있다. 전체 혼합 밀도 방식은 <표 3>과 같이 쉽게 구할 수 있다.

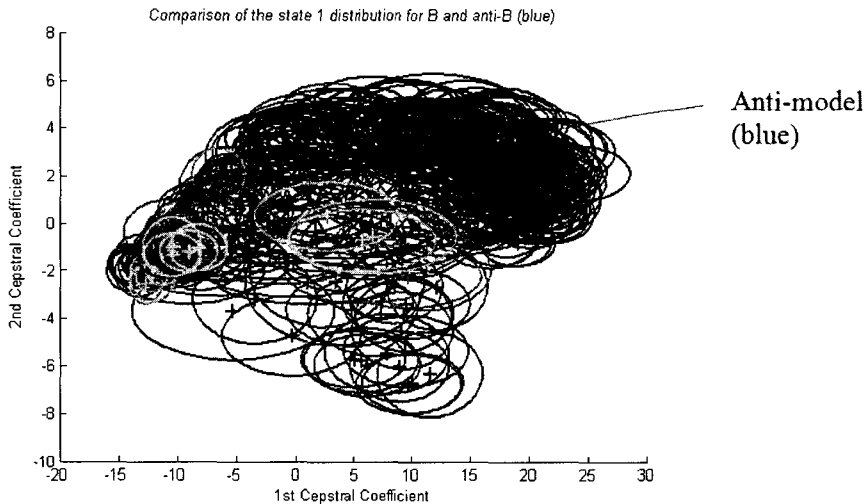
이상과 같이 구축한 반모델의 분포를 이웃 모델의 분포와 비교하여 <그림 3>과 <그림 4>에 각각 도식화하였다. 반모델에서 상태별 전체 혼합 밀도의 수는 약 315개 정도이며, 이웃모델의 경우 20개의 모델로 이웃을 정의하였다(실험에 사용된 상태별 혼합 밀도의 수는 7개이므로, 140개의 혼합 밀도). <그림 3>과 <그림 4>는 동일 구조를 갖는 인식 모델의 같은 상태 번호에 대해 표시하였다. 반모델의 경우, 해당 상태의 모든 혼합 밀도를 수집하여 단일의 모델로 만들었기 때문에, 개별 모델 구조들이 보존된 이웃 모델 방식과 직접적인 비교는 어렵지만 경향을

<표 3> 전체 혼합 밀도 방식을 이용한 대안 가설 모델 구축

```

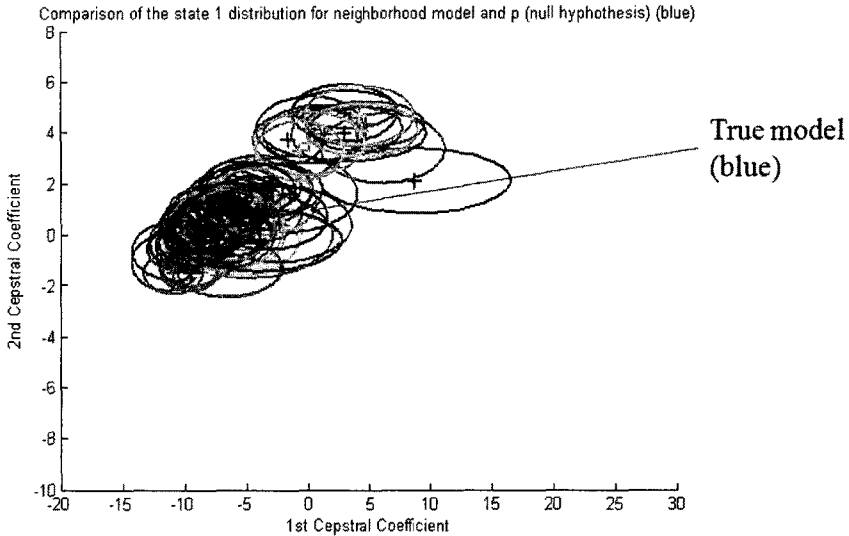
function buildAntiModel
  for all phone models  $\lambda_k$ 
    for all states  $s_i$  of given model  $\lambda_k$ 
      for all mixtures  $m_j$  with weight  $w_j$  of given state  $s_i$ 
        merge  $w_j m_j$  to the same state of anti-model  $AM_i$ 
        sum all weights  $\text{sum}_i += w_j$ 
      end for all mixtures
    end for all states
  end for all models
  for all mixture weights  $w_j$  of each state  $i$  of anti-model
     $\hat{w}_j = \frac{w_j}{\text{sum}_i}$ 
  end for
end buildAntiModel

```



<그림 3> 전체 혼합 밀도 방식을 이용한 대안 가설 모델(반모델)의 분포

파악할 수는 있다. <그림 3>을 살펴보면 반모델의 혼합 밀도 분포는 인식 모델의 분포와 독립적이기 때문에 동일한 그룹에 속하는 인식 모델들은 동일한 반모델 분포를 가진다. 반면에 <그림 4>의 이웃 모델의 경우, 인식 모델에 따라 이웃 모델이 달라지기 때문에 혼합 밀도의 분포는 다른 분포를 갖는다. 일반적으로 이웃 모델의 경우 이웃의 수를 ∞ 로 하는 경우 전체 혼합 밀도와 같다고 생각하는 경



<그림 4> 실제 모델과 대응되는 이웃 모델의 분포

우가 많다. 그렇지만, 이웃 모델의 경우에는 대안 가설 모델을 구성하는 과정에서 원래 인식 모델의 정보가 사라지지 않지만, 전체 혼합 밀도의 경우에는 모델의 상태간의 천이 정보 또는 그룹 정보가 사라진다는 점에서 차이가 있다.

이웃모델을 이용한 대안 가설 모델의 분포와 전체 혼합 밀도를 이용한 반모델의 비교 요약은 <표 4>와 같이 정리할 수 있다.

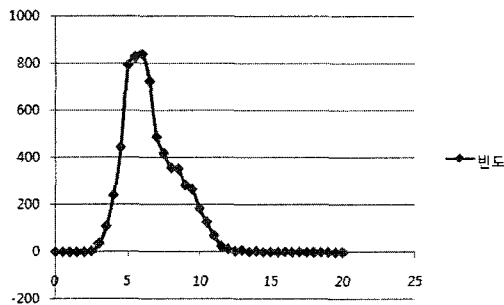
<표 4> 전체 혼합 모델과 이웃 모델의 비교

	전체 혼합 밀도 방식	이웃 모델 방식
차이점	<ul style="list-style-type: none"> - 인식 모델에 상관없이 반모델의 분포 동일 - 상태간의 천이 정보는 전체 혼합 모델의 구축단계에서 모두 사라짐 	<ul style="list-style-type: none"> - 인식 모델에 따라 이웃 모델의 분포 변함 - 인식 모델의 구조가 유지되기 때문에, 상태간의 천이는 인식모델에 따라 제약을 받음

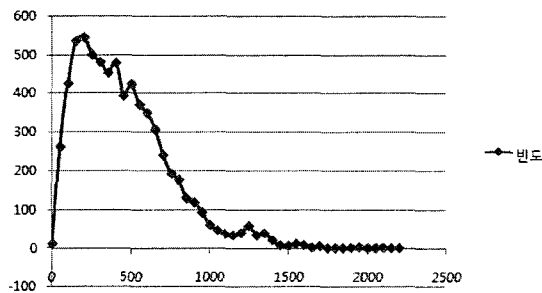
4.2. 대안 가설 모델의 구축

인식 모델(대상 모델)이 주어지면, 그 모델에 대한 이웃 모델을 거리에 따라 계산할 수 있으며, 이웃 모델의 수는 정해진 수 또는 발산 범위에 따라 선정할 수 있다. 본 연구에서는 이웃 모델의 수를 정해진 수만큼 고정을 시켰는데, 그렇지

얇은 경우 발산 범위에 따라 분포가 많은 차이를 보이기 때문이다. <그림 5>는 트라이폰 모델들에 대한 KLD 값의 분포를 보이고 있다. 전체 모델들을 비교한 경우 KLD의 값 대부분이 5~7 사이에 있기 때문에 이웃 모델을 선택하는데 기초가 될 수 있다. 다음으로 <그림 6>의 경우, KLD 값이 4보다 작거나 같은 경우에 각 인식 모델들에 대한 이웃 모델의 수에 대해 통계를 구한 것이다. 그림에서 보는 바와 같이, 각 인식 모델에 대한 이웃 모델의 수는 차이가 심하여 본 연구에서는 이웃 모델을 결정할 때 KLD 값을 직접 이용하지 않고 이웃 모델의 수를 미리 결정하여 그 수만큼의 이웃 모델을 선택하여 사용하였다.



<그림 5> 트라이폰 모델에 대한 KLD 분포도



<그림 6> $KLD \leq 4.0$ 인 경우 이웃 모델의 수 분포도

모델의 수에 따라 이웃 모델이 선택되면 이웃 모델을 이용하여 대안 가설 모델을 구축하여야 한다. 이웃 모델의 경우에는 전체 혼합 밀도와 달리 인식 모델 중에서 KLD 값에 따라 모델을 선택하기 때문에 모델의 구조는 변경되지 않아 대안 가설 모델의 우도 값을 계산하는 과정만 제시한다. 대안 가설 모델의 우도 값을 구하는 방법에 따라 균일하게 평균을 이용하는 방법(균일 결합 방법)과 KLD

값에 의하여 가중치를 주는 방법(가중 결합 방법)으로 구할 수 있다. 즉, 식 (7)에서 분모에 대응되는 $p(\mathbf{y}|C_1)$ 를 구하는 방법이다.

일반적으로 베이스 인수 방식을 적용하기 위해서는 실제 모델과 대안 가설 모델에 대해 사후 확률 값을 적용한 가설 검증 단계를 거쳐야 하기 때문에, 사후 확률 값을 구하기 위해서 식 (5)와 같이 모델에 대한 사전 확률을 먼저 계산하여야 한다. 사전 확률은 앞에서 분류한 두 가지 방법, 즉 균일 결합 방법과 KLD에 의한 가중 결합 방법에 각각 적용할 수 있다. 본 절에서는 이웃 모델을 결합하는 두 가지 방법에 대해 설명하고, 사전 확률을 적용하기 위한 방법을 소개한다.

4.2.1. 균일 결합

이웃 모델이 선택되면 이웃 모델들의 기여도를 반영하지 않고 단순히 각 모델에 대한 주어진 입력 값 \mathbf{y} 의 우도 값을 더한 후, 모델들의 수로 나눈 값으로 대안 가설 모델의 우도로 결정하는 방식이다. 즉,

$$p(\mathbf{y}|C_1) = \frac{1}{N} \sum_{i=1}^N p(\mathbf{y}|\lambda_i). \quad (10)$$

4.2.2. KLD 값에 의한 가중 결합

4.2.1항과 달리 이웃 모델의 기여도를 반영하기 위하여 본 논문에서는 KLD 값을 사용하였다. KLD 값은 가까운 모델에 대해서는 작은 값을, 멀리 떨어진 모델에 대해서는 큰 값을 나타내기 때문에 직접적으로 적용하기 힘들어, 모든 값들을 더한 값으로 나누어 정규화 값을 구하고, 1에서 그 값을 빼어 가중치로 사용하였다.

$$p(\mathbf{y}|C_1) = \sum_{i=1}^N \hat{w}_i \cdot p(\mathbf{y}|\lambda_i)$$

$$\hat{w}_i = 1 - \frac{w_i}{\frac{1}{N} \sum_{j=1}^N w_j} \quad (11)$$

4.2.3. 사전 확률의 계산

일반적으로 많이 사용되는 우도비 검증 방법과 달리 베이스 인수 방식은 각각의 클래스 (실제 모델과 대안 가설 모델)에 대해 사전 확률을 계산하여야 한다. 그러나 사전 확률은 정확한 값을 추정하기 어렵기 때문에 많은 연구 방법에서 근사

추정법을 사용하고 있다. 대표적인 방법은 점진적 학습 방법(incremental learning method)을 통하여 사전 확률을 추정하는 방법이다.

$$p(\lambda_k) = p(\lambda_k | \mathbf{X}_k^{t-1}) \quad (12)$$

위 식에서 $\mathbf{X}_k^{t-1} = \{\mathbf{X}_{k,1}, \dots, \mathbf{X}_{k,t-1}\}$ 를 의미하며, 모델 λ_k 에 대한 현재 시간 t 이전의 관측 벡터를 나타낸다. 이와 같은 해석에 의하여 식 (5)는 다음 식과 같이 정리된다.

$$\begin{aligned} p(\mathbf{X} | C_k) &= \int p(\mathbf{X} | \lambda_k) p(\lambda_k | \mathbf{X}_k^{t-1}) d\lambda_k \\ &= \prod_{t=1}^T \int p(x | \lambda_k) p(\lambda_k | \mathbf{X}_k^{t-1}) d\lambda_k \end{aligned} \quad (13)$$

위 식에서 \prod 안의 적분 값에 대해 고려하면

$$\int p(x | \lambda) p(\lambda | \mathbf{X}) d\lambda = \sum_{i=1}^N w_i \int p(x | \mu_i) p(\mu_i | \mathbf{X}) d\mu_i \quad (14)$$

로 되며, 점진적 학습 방법을 적용하면 다음과 같이 구할 수 있다.

$$\begin{aligned} &\int p(x | \mu_i) p(\mu_i | \mathbf{X}) d\mu_i \\ &= \prod_{d=1}^D \sqrt{\frac{\tau_i}{2\pi\delta_{id}^2(\tau_i+1)}} \exp\left\{-\frac{\tau_i(x_{id} - m_{id})^2}{2(\tau_i+1)\delta_{id}^2}\right\} \end{aligned} \quad (15)$$

식 (15)의 사전 확률을 구성하는 매개 변수는 MAP 변경식과 같다.

$$\begin{aligned} m'_i &= \frac{\tau_i m_i + P(i|x)}{\tau_i + P(i|x)} \\ \tau'_i &= \tau_i + P(i|x) \end{aligned} \quad (16)$$

여기서 τ'_i 과 m'_i 는 벡터 x 를 관측한 후 변경되는 매개 변수를 의미하며, $P(i|x)$ 는 관측 벡터를 생성하는 사후 확률 값을 나타낸다.

$$P(i|x) = \frac{w_i N(x_i | \mu_i, \Sigma_i)}{p(x|\lambda)} \quad (17)$$

즉, 베이스 인수 방식은 점진적 학습 방법에 의하여 구현될 수 있으며, 각 클래스의 불확실성을 보상하기 위하여 분산 값을 조정하는 것으로 해석할 수 있다[2].

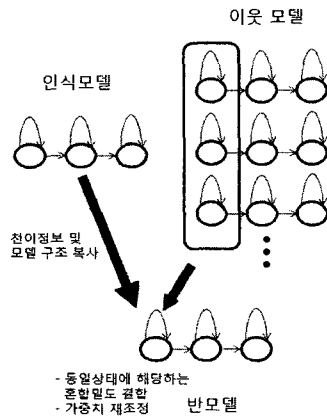
그러나 본 연구에서는 인식 모델을 훈련하는데 사용되었던 원 음성 데이터나 특징 벡터, 그리고 대응되는 텍스트 정보들을 이용하기 어려워, 어휘 사전에 포함된 텍스트 정보를 이용하여 인식 모델의 발생 빈도를 계산하여 사전 확률의 근사

치로 사용하였다.

4.3. 이웃 모델을 이용한 반모델의 구축

지금까지는 이웃 모델을 이용한 대안 가설 모델의 구축에 대해 살펴보았다. 엄격히 말하면 대안 가설 모델의 구축보다는 대안 가설 모델에 해당되는 우도 값을 계산하는 과정으로 보는 것이 정확할 것이다. 이와 같이 이웃 모델들을 선택한 후, 대안 가설 모델에 해당하는 우도 값을 계산하는 과정은 기존 모델의 구조 (topology)를 그대로 유지한 채 각 모델의 우도 값을 결합하여 최종 대안 가설 모델의 우도 값을 결정하였다. 이 방법은 기존 모델의 한계 또는 취약점이 그대로 대안 가설 모델에 반영된다는 문제점이 존재한다. 즉, 특정 모델의 우도 값이 비이상적인 값을 가질 경우 그 값이 그대로 대안 가설 모델의 우도 값에 반영된다는 것이다.²⁾

본 절에서는 이웃 모델의 혼합 밀도를 상태별로 결합한 단일의 반모델을 구축하는 방법을 제안한다. 이웃 모델을 반모델로 구축할 경우, 각 이웃 모델에서의 상태간 천이 정보는 사라지고 각 상태의 혼합 밀도 정보만이 이용된다. 현재는 이웃 모델과 실제 가설의 모델의 상태 수는 동일하다고 가정을 한다. 이웃 모델의 혼합 밀도를 결합하는 개념도는 <그림 7>과 같다.



<그림 7> 이웃 모델을 이용한 반모델 구축

2) 예를 들어 특정 음소 모델의 인식 값이 비이상적으로 큰 경우, 대안 가설 모델의 값도 같이 커지게 됨, 5.4절 참조

5. 실험 및 결과

본 장에서는 LRT 방식에서 대안 가설을 표현하는 반모델 방식을 베이스 인수 방식에서 사용되는 이웃 모델로 변경하여 성능 변화를 살펴본다. 베이스 인수 방식에서의 사전 확률을 구하는 최적의 해법이 존재하지 않기 때문에, LRT와의 상관관계를 조사하고 LRT에 적용되었던 가정을 완화시켜 베이스 인수 방식을 구현하기 위한 방법 중의 하나로 반모델 방식을 변경하였다. 이웃 모델 방식은 실제 가설 모델의 주변에 존재하기 때문에 인식과정에서 이웃모델들의 평균 우도 값은 실제 가설 모델보다 작을 것으로 예상되지만 실제 환경에서는 인식 모델의 분포와 밀접한 관계를 가지기 때문에, 그 값의 변화를 살펴보고 성능을 개선할 수 있는 결합 방법을 찾고자 한다.

본 논문의 실험 환경은 [4]의 환경과 동일하다. [4]에서는 인식용 음향 모델을 800명의 화자로부터 발생된 약 8만 여의 고립단어 음성 데이터베이스를 사용하였으며, 전통적인 39차 특징벡터와 캡스트럼 평균 정규화를 거친 모델당 3개의 상태와 7개의 혼합 밀도 분포를 갖는다. 발화 검증 실험을 위해서는 훈련에 사용되지 않는 1,000 단어 규모의 실험용 고립단어 음성 데이터베이스를 이용하였다. 평가 기준 시스템은 반모델을 전체 혼합 밀도를 이용하여 구하였으며, <표 5>와 같은 결과를 가진다. 표에서 EER은 equal error rate의 약자로서 오류 거절(false rejection)과 오류 수락(false acceptance)의 값이 같은 경우를 말한다. 오류 거절은 음성 인식 시스템이 정확하게 인식된 결과를 신뢰도가 낮다고 거절하는 경우이며, 오류 수락은 잘 못 인식된 결과를 신뢰도가 높다고 판단하여 제대로 된 인식 결과로 판단하는 경우를 말한다.

<표 5> 평가 기준 시스템의 성능

반모델 형태	평균 방식	EER (%)	FR (%)	FA (%)
-	WLL	32.85	32.83	32.88
전체 혼합 밀도 방식	산술	12.56	12.42	12.70
	기하	12.75	12.68	12.81
	조화	13.31	13.47	13.15

5.1. 이웃 모델의 결합

첫 번째 실험은 기존 LRT의 반모델 방식을 변경하여, 사전 확률 정보를 이용하지 않고 4장에서 설명한 두 가지 방법으로 반모델을 구축하여 실험한 경우이며 실험 결과는 <표 6>과 <표 7>에 각각 정리되어 있다.

위 실험 결과 균일 결합 방식이 KLD 값에 의한 가중 결합 방식에 비하여 EER

<표 6> 이웃 모델의 균일 결합 (EER:%)

이웃 수	산술평균	기하평균	조화평균
20	15.56	15.29	16.57
40	14.66	14.58	16.30
80	13.82	13.69	15.68
120	13.95	13.91	15.15
160	14.21	14.05	15.59
200	13.59	13.61	15.37
240	13.70	13.78	14.97
280	13.59	13.74	15.16
320	13.63	13.66	14.97
640	13.78	13.91	16.27
1280	15.60	15.49	16.94
2560	17.18	16.87	18.01
Inf	17.40	16.67	18.09

<표 7> KLD 값에 의한 가중 결합 (EER:%)

이웃 수	산술평균	기하평균	조화평균
20	15.52	15.32	17.04
40	14.51	14.55	16.73
80	14.24	14.24	16.17
120	14.13	14.40	15.27
160	14.24	14.29	15.60
200	14.15	14.05	15.79
240	14.28	14.25	15.59
280	14.29	14.37	15.78
320	14.02	14.16	16.22
640	14.61	14.43	16.55
1280	17.03	17.10	17.49
2560	19.35	19.18	18.61
Inf	19.35	19.20	18.41

값이 작은 것을 알 수 있다. 그것은 이웃 모델을 이용하여 반모델을 구축할 때 통념상으로 KLD 값을 적용한 가중치가 균등하게 적용한 방식보다는 좋은 성능을 보여야 하나, 실제 학습된 인식 모델의 경우 여러 가지 환경적 요인이나 화자의 정보에 의하여 KLD 값의 영향이 부정적인 방향으로 작용하였기 때문으로 해석된다.

5.2. 사전 확률의 적용

전 절에서는 이웃 모델을 이용하여 반모델을 구축할 때, 사전 확률의 적용 없이

단순히 이웃 모델의 우도 값을 산술 평균(균일 결합)하거나 가중 평균(가중 결합)하여 반모델의 우도 값을 계산하였다. 베이스 인수 방식은 각 클래스에 해당하는 특징 벡터의 우도 값과 모델의 사전 확률을 적용하여 가설 검증을 하기 때문에 정확한 사전 확률의 추정이 중요하다. 그러나 본 연구에서는 사전 확률 추정에 필요한 학습 데이터를 사용하지 못하는 제약조건으로 인하여 직접 사전 확률을 추정하지 않고, 간접적으로 근사화하여 사용하였다. 모델의 사전 확률은 학습에 사용된 학습 데이터에 의한 사후 확률 값을 이용하여 근사화될 수 있으나, 학습 데이터가 없는 경우 사용할 수 없다는 단점이 있다. 이에 모델의 사전 확률을 학습에 사용된 데이터에 대응되는 텍스트 정보를 이용하여 근사화하는 방안을 고려하였으나, 이것 역시 정보가 없어 어휘 사전에 등록된 단어 정보만을 이용하여 근사화하였다. 각각의 차이를 <표 8>에 정리하였다.

이들 방법 중에서 어휘, 클래스 정보만을 이용하여 사전 확률을 추정한 후, 각각의 결합 방법에 대해 적용하였다. 각각의 실험 결과는 <표 9>와 <표 10>에 정리하였다. 실험 결과, 사전 확률을 적용하였을 때 성능의 변이가 줄었으나, 두 방법 모두 성능의 변화는 미비하였다. 이것은 사전 확률이 정확하게 추정되지 않아 그 영향이 적었을 것으로 판단된다.

<표 8> 사전 확률 추정 방법 및 정확도 비교

근사화 방법	이용 정보	적용 방법	정확도
직접 근사화	음성 특징, 클래스 정보	점진적 학습 방법	상
간접 근사화	텍스트, 클래스 정보	Unigram	중
간접 근사화	어휘, 클래스 정보	Unigram	하

<표 9> 균일 결합 방식에 사전 확률을 적용한 경우 (EER:%)

이웃 수	산술평균	기하평균	조화평균
20	15.54	15.17	16.81
40	14.70	14.72	16.60
80	13.78	13.77	15.95
120	13.78	13.74	15.22
160	13.78	13.61	15.51
200	13.69	13.69	15.49
240	13.66	13.77	14.73
280	13.66	13.70	14.88
320	13.56	13.67	14.93
640	13.91	13.84	15.91
1280	15.35	15.38	16.96
2560	16.59	16.48	17.90
Inf	16.90	16.95	17.63

<표 10> 가중 결합 방식에 사전 확률을 적용한 경우 (EER:%)

이웃 수	산술평균	기하평균	조화평균
20	15.95	15.69	16.93
40	14.86	14.50	16.70
80	14.22	14.26	15.83
120	14.24	14.40	15.48
160	14.39	14.59	15.72
200	14.26	14.18	15.89
240	14.44	14.42	15.47
280	14.45	14.34	15.52
320	14.23	14.11	15.72
640	14.54	14.37	16.22
1280	17.05	16.80	17.31
2560	18.80	18.42	18.44
Inf	19.02	18.73	18.10

5.3. 이웃 모델 결합에 의한 반모델의 구축

이웃 모델 결합에 의한 반모델을 이용하여 LRT 방식의 발화검증 실험을 한 결과, 평가 기준 시스템이나 기존 균일 방식, 가중 결합 방식, 사전 확률을 적용한 방식보다 성능이 현저하게 향상됨을 알 수 있었다(<표 11> 참조).

<표 11> 이웃 모델들을 반모델로 구축한 경우의 성능 변화

이웃 수	산술평균	기하평균	조화평균
평가기준	12.56	12.75	13.31
20	14.36	14.11	16.20
40	13.26	13.17	15.82
80	12.60	12.34	14.55
120	11.88	11.74	14.52
160	11.63	11.74	13.67
200	11.52	11.52	13.45
240	11.44	11.43	13.31
280	11.63	11.64	13.41
320	11.74	11.74	13.24
640	11.60	11.62	12.70
1280	11.38	11.13	12.41
2560	10.68	10.74	12.56
Inf	-	-	-

특히 산술 평균을 이용한 경우, 최대 17%의 에러율 감소(error rate reduction)를 보인다. 그러나 제안된 방법은 모든 인식 모델에 대해 각각의 반모델을 구축하고 전체를 메모리로 적재하기 때문에, 구현상의 문제로 인해 일부 환경에서는 결과를

얻지 못했다.³⁾

5.4. 분석

발화 검증에 널리 사용되는 우도비 검증 방법의 한계를 파악하고 성능을 향상시키기 위한 방안으로서 베이스 인수 방법에서 사용되는 이웃 모델 방식과 사전 확률의 적용, 이웃 모델 방식의 반모델에 의한 가설 검증 방법을 제안하고, 실험을 실시하였다. 실험은 이웃 모델을 결합하는 방식에 따라 산술 평균을 적용하는 균일 결합 방식과 KLD 값에 의한 가중치를 적용하는 가중 결합 방식, 베이스 인수 방식과 유사한 방식을 적용하기 위하여 간접적인 방식에 의한 사전 확률 추정 방법, 이웃 모델을 이용하여 단일의 반모델 구축과 그 성능을 비교하였다.

이웃 모델을 이용하여 반모델을 구축하는 두 가지 방법에 대한 실험 결과, 균일 결합 방식이 가중 결합 방식에 비하여 근소하게나마 성능이 향상됨을 보였으나, 평가 기준 시스템에 비하여 여전히 성능이 저하됨을 알 수 있었다. 그 원인으로서는 전체 혼합 밀도를 이용하는 반모델의 경우, 각 상태를 구성하는 혼합 밀도는 상태 천이나 상태에서의 관측 확률에 제약사항이 없으나, 이웃 모델을 이용하는 경우 혼합 밀도는 모델간의 경계에 의하여 제약사항이 존재하기 때문으로 파악된다. 또한 가중 결합의 경우 KLD 값에 의하여 모델의 기여도를 판단하였으나, 모델의 기여도에 의한 가중치 설정이 부정확하였을 가능성도 배제할 수 없다.

사전 확률을 이웃 모델의 결합 방식에 적용하여 반모델을 구축한 실험의 경우, 사전 확률을 적용하기 전후의 성능 변화가 거의 없음을 알 수 있다. 다만, 이웃의 수에 따른 성능 변이의 차가 줄어들어 사전 확률의 정확한 추정이 이뤄진다면 환경 변화에 강인한 발화 검증 시스템을 기대할 수 있을 것이다.

마지막으로 이웃 모델을 이용하여 단일의 반모델을 구축하고, LRT방식의 가설 검증 방법을 적용한 결과 기존의 제안한 방법에 비하여 높은 성능을 보이고 있으나, 모든 이웃 모델의 혼합 밀도를 고려하기 때문에 시간적, 공간적 제약사항이 따른다는 문제점이 나타났다.

일반적으로 발화 검증에서 가설 검증 방식을 적용할 경우 실제 모델의 우도 값과 반모델의 우도 값의 차이가 중요한 역할을 하기 때문에 실제 모델의 정확한 모델 구축과 이웃 모델의 정확한 선정이 중요하다. 제안된 방식들의 성능 변화 원인을 파악하기 위하여 현재 구축된 인식 모델의 경우와 전체 혼합 밀도를 이용한 경우의 반모델의 우도 값, 이웃 모델을 이용한 경우 등의 우도 값을 비교하였다.

비교 결과, <그림 8>-<그림 11>에서 살펴 본 바와 같이 실제 모델의 우도 값이 양의 값을 갖는 경우 전체 혼합 밀도에 의한 경우에는 반모델의 우도 값이 실제

3) 이웃 수가 ∞ 인 경우, 즉, 자기 자신을 제외한 거의 모든 인식모델을 이용하여 반모델을 구축하는 경우

모델의 우도 값보다 높은 값을 보이는 경우(역전현상)가 드물지만, 이웃 모델을 이용한 경우에는 간혹 발생함을 알 수 있다. 특히 균일 결합 방식보다 가중 결합 방식을 이용한 경우 그 차이가 더 큼을 알 수 있어, 이웃 모델을 이용한 경우의 성능 하락 요인을 설명할 수 있을 것이다. 이웃 모델을 이용하는 경우 역전 현상을 제외하고 전체 혼합 밀도 방식과 비교해서 인식 음소 값과 반모델의 우도 값의 차이가 작음을 알 수 있다. 결국 발화 검증의 성능 저하는 역전현상과 대비되는 쌍들의 우도 값의 차이로 설명될 수 있다.

```

"SpeechDB/ETRI_1000/IV/fkjs00/UH100FKJS0000.MFCC"
0 3800000 sil 757.166504 sil
3800000 5000000 g-e -85.683655 anti_g -148.939010
5000000 5500000 g-e+m 7.026062 anti_e -3.622233
5500000 6500000 e-m+ju 8.452774 anti_e -48.812683
6500000 7100000 m-ju+M 35.542419 anti_jv -31.925032
7100000 8100000 ju-M+t -36.232727 anti_M -68.240571
8100000 8900000 N-t+u 9.024109 anti_t -62.510475
8900000 9700000 t-u+z 15.986877 anti_u -51.783588
9700000 10700000 u-z+a -15.839478 anti_z -80.337944
10700000 11500000 z-a+b -2.977661 anti_a -53.806931
11500000 12200000 a-b+a -58.161255 anti_b -102.804655
12200000 12800000 b-a+M 11.088684 anti_a -21.875876
12800000 13800000 a-M+s -3.479187 anti_M -74.834923
13800000 14800000 M-s+o -17.945198 anti_s -94.234383
14800000 15700000 s-o+M 23.263550 anti_o -10.339923
15700000 18300000 o-M+G 338.847908 anti_M 232.066883
18300000 22000000 sil 819.554138 sil

```

<그림 8> 전체 혼합 모델을 이용한 경우의 반모델 우도 값 분석

```

"SpeechDB/ETRI_1000/IV/fkjs00/UH100FKJS0000.MFCC"
0 3800000 sil 757.166504 sil
3800000 5000000 g-e -85.683655 anti_g -78.481230
5000000 5500000 g-e+m 7.026062 anti_e 19.816912
5500000 6500000 e-m+ju 8.452774 anti_e -1.2745719
6500000 7100000 m-ju+M 35.542419 anti_jv -10.558993
7100000 8100000 ju-M+t -36.232727 anti_M -45.972374
8100000 8900000 N-t+u 9.024109 anti_t -48.283504
8900000 9700000 t-u+z 15.986877 anti_u 9.842479
9700000 10700000 u-z+a -15.839478 anti_z -61.595449
10700000 11500000 z-a+b -2.977661 anti_a -31.188932
11500000 12200000 a-b+a -58.161255 anti_b -84.749237
12200000 12800000 b-a+M 11.088684 anti_a -5.317257
12800000 13800000 a-M+s -3.479187 anti_M -50.312885
13800000 14800000 M-s+o -17.945198 anti_s -77.365776
14800000 15700000 s-o+M 23.263550 anti_o -6.113953
15700000 18300000 o-M+G 338.847908 anti_o-M 355.971689
18300000 22000000 sil 819.554138 sil

```

<그림 9> 이웃 모델의 균일 결합에 의해 계산된 대안 가설 모델의 우도 값 분석

```

"SpeechDB/ETRI_1000/IV/fkjs00/UH100FKJS0000.MFCC"
0 3800000 sil 757.166504 sil
3800000 5000000 g-e -85.683655 anti_g -73.107513
5000000 5500000 g-e+m 7.026062 anti_e 25.710785
5500000 6500000 e-m+ju 8.452774 anti_e 6.510197
6500000 7100000 m-ju+M 35.542419 anti_jv -5.265594
7100000 8100000 ju-M+t -36.232727 anti_M -48.677181
8100000 8900000 N-t+u 9.024109 anti_t -42.989056
8900000 9700000 t-u+z 15.986877 anti_u 15.136986
9700000 10700000 u-z+a -15.839478 anti_z -56.305061
10700000 11500000 z-a+b -2.977661 anti_a -25.895691
11500000 12200000 a-b+a -58.161255 anti_b -79.454765
12200000 12800000 b-a+M 11.088684 anti_a -8.02197
12800000 13800000 a-M+s -3.479187 anti_M -45.818311
13800000 14800000 M-s+o -17.945198 anti_s -72.871358
14800000 15700000 s-o+M 23.263550 anti_o -8.820420
15700000 18300000 o-M+G 338.847908 anti_M 361.265128
18300000 22000000 sil 819.554138 sil

```

<그림 10> 이웃 모델의 가중 결합에 의해 계산된 대안 가설 모델의 우도 값 분석

```

"SpeechDB/ETRI_1000/IV/fkjs00/UH100FKJS0000.MFCC"
0 3800000 sil 757.166443 sil
3800000 5000000 g-e -85.683655 anti_g -87.224503
5000000 5500000 g-e+m 7.026062 anti_g-e+m 17.540944
5500000 6500000 e-m+ju 8.452774 anti_e-m+ju -7.778711
6500000 7100000 m-ju+M 35.542419 anti_m-ju+M -9.942322
7100000 8100000 ju-M+t -36.232727 anti_jv-M+t -37.235275
8100000 8900000 N-t+u 9.024109 anti_N-t+u -28.016985
8900000 9700000 t-u+z 15.986877 anti_t-u+z -1.445853
9700000 10700000 u-z+a -15.839478 anti_u-z+a -43.749481
10700000 11500000 z-a+b -2.977661 anti_z-a+b -23.843470
11500000 12200000 a-b+a -58.161255 anti_a-b+a -81.397171
12200000 12800000 b-a+M 11.088684 anti_b-a+M -16.120956
12800000 13800000 a-M+s -3.479187 anti_a-M+s -41.831661
13800000 14800000 M-s+o -17.945198 anti_M-s+o -64.598240
14800000 15700000 s-o+M 23.263550 anti_s-o+M -10.326693
15700000 18300000 o-M+G 338.847839 anti_o-M+G 280.177523
18300000 22000000 sil 819.554138 sil

```

<그림 11> 이웃 모델을 이용하여 반모델을 구축한 경우의 우도 값 분석

이와 같이 인식 결과와 경계 정보의 대칭 쌍에 대한 우도 값을 비교하였을 때 역전현상이 발생하는 것은 평가용 입력 벡터와 언어 정보에 의하여 인식과정에서 선택된 인식 모델의 평균 벡터와의 차이가 많이 나기 때문으로 보이며, 우도 값이 양의 값을 갖는 것은 특정 벡터의 평균값의 차에 비해 분산 값이 굉장히 작다는 것을 의미한다. 즉, 인식 모델이 학습 데이터에 대해서 과 조율(over tuning)되었거나 학습 데이터의 양이 충분치 않은 경우에 발생하는 현상으로 해석될 수 있다.

그러나 학습 과정이나 인식 과정에서 우도 값이 양의 큰 값을 갖는 경우를 제외 또는 제한한 경우 음성 인식 성능의 저하를 가져와 좀 더 깊은 분석이 필요하다고 본다.

각각의 실험 결과의 분석 및 검토 의견을 정리하면, <표 12>와 같이 요약될 수 있다.

<표 12> 실험 결과 분석 및 검토 요약

사전 확률 적용 여부	이웃 모델 구축 방법	분석 및 검토	공통 의견
미적용	균일 결합	- 각 모델의 기여도는 균일하게 고려되거나 실제 모델 간의 KLD 값에 의하여 모델이 선정됨	- 기존 전체 혼합 밀도 방식에 비하여 각 상태에서의 혼합 밀도의 차이가 모델 내에 제약됨으로 성능 저하 발생 (균일 결합, 가중 결합 방식) - 인식 모델의 정확한 구축과 이웃 모델의 선택이 성능 평가에 영향을 미침
	가중 결합	- 각 모델의 기여도는 실제 모델간의 KLD 값에 의한 가중치로 적용됨 - KLD 값에 의한 모델간의 기여도 정의에 대해 이론적, 실험적 검증 필요	
	반모델 구축	- 각 모델을 구성하는 혼합 밀도를 상태별로 결합하여 단일의 반모델 구축 - 단순 결합 방식보다는 역전 현상이나, 비교되는 모델간의 우도 값의 차이가 줄어듦	
적용	균일 결합	- 성능 저하의 폭을 줄여주는 효과를 가져옴	
	가중 결합		

6. 결론

본 논문에서는 신뢰도 척도나 발화 검증에 많이 사용되는 우도비 검사를 베이스 이론에 입각하여 재정의한 베이스 인수 방법의 측면에서 살펴보았다. 베이스 인수 방법을 직·간접으로 구현하는 과정은 대안 가설에 대응되는 반모델과 사전 확률의 영향을 많이 받는다. 따라서 본 논문에서는 발화 검증에 널리 사용되는 우도비 검증 방법의 한계를 파악하고 성능을 향상시키기 위한 방안으로서 베이스 인수 방법에서 사용되는 이웃 모델 방식과 사전 확률을 적용한 가설 검증 방법을 제안하고, 실험을 실시하였다. 실험에서는 이웃 모델을 결합하는 방식에 따라 산술 평균을 적용하는 균일 결합 방식과 KLD 값에 의한 가중치를 적용하는 가중 결합 방식, 간접적인 방식에 의한 사전 확률 추정 방법과 그 적용에 따른 성능 비교를 하였다. 또한 결합 방식에 의한 대안 가설 모델의 구축은 각 모델에서의 상

태간 천이 정보의 제약사항이 그대로 존재하기 때문에 이웃 모델의 혼합 밀도를 상태별로 재결합한 반모델 구축 방법을 제안하였다.

실험 결과, 이웃 모델의 결합방식에 의한 대안 가설 모델의 우도 값을 결정하는 방법에서는 성능 향상이 없었지만 이웃 모델의 혼합 밀도를 상태별로 결합하여 단일의 반모델로 구축한 방법에서는 최대 17%의 오류율 감소를 보였다.

제안한 방식은 이웃 모델에서의 모든 혼합 밀도를 보존하고, 계산에 포함되기 때문에 공간적·시간적 제약이 따른다. 따라서, 추후 연구로는 일반적인 반모델의 구축에 사용되는 혼합 밀도의 수만큼 이웃 모델에서 혼합 밀도를 선택하는 연구가 진행되어야 할 것이다.

감사의 글

음성 인식 플랫폼인 EchoS를 사용하도록 허락한 원광대학교 음성정보기술산업 지원센터에 깊은 감사를 드리며, EchoS 기반 위에서 발화검증 부분을 구현하여 본 연구에 많은 도움을 준 ICU의 권석봉씨에게 깊은 감사의 뜻을 전합니다.

참 고 문 헌

- [1] H. Jiang, C.-H. Lee, "A new approach to utterance verification based on neighborhood information in model space", *IEEE Transactions on Speech and Audio Processing*, Vol. 11, No. 5, pp. 425-434, 2003.
- [2] R. Vogt, S. Sridhar, "Bayes factor scoring of GMMs for speaker verification", *Proc. Odyssey: The speaker and Language Recognition Workshop*, pp. 173-178, 2004.
- [3] H. Jiang, L. Deng, "A Bayesian to the verification problem: Applications to speaker verification", *IEEE Transactions on Speech and Audio Processing*, Vol. 9, No. 8, pp. 874-884, 2001
- [4] 권석봉, 김희린, 강점자, 구명완, 류창선, "다양한 신뢰도 척도를 이용한 SVM 기반 발화 검증 연구", *말소리*, 제60호, pp. 165-180, 2006.
- [5] R. A. Sukkar, C.-H. Lee, "Vocabulary independent discriminative utterance verification for nonkeyword rejection in subword based speech recognition", *IEEE Transactions on Speech and Audio Processing*, Vol. 4, No. 6, pp. 420-429, 1996.

접수일자: 2008년 7월 30일

게재결정: 2008년 9월 19일

▶ 윤영선 (Young-Sun Yun) : 교신저자
주소: 306-791 대전광역시 대덕구 오정동 133번지
소속: 한남대학교 정보통신공학과
전화: 042) 629-7569
E-mail: ysyun@hnu.kr