한국음향학회지 제23권 제5호 pp. 410~419 (2004)

베이시안 신뢰도 융합을 이용한 신뢰도 측정

Bayesian Fusion of Confidence Measures for Confidence Scoring

김 태 윤*, 고 한 석*
(Tae-Yoon Kim, Hanseok Ko)

*고려대학교 전자컴퓨터공학과

(접수일자: 2004년 2월 25일; 수정일자: 2004년 5월 17일; 채택일자: 2004년 6월 30일)

본 논문에서는 베이시안에 기반한 신뢰도 융합 기법을 제안한다. 음성인식에서 신뢰도는 인식 결과에 대한 신뢰의 정도를 말하며, 인식 결과가 맞는 지의 여부를 판단할 수 있다. 개별 신뢰도 기법의 신뢰도 값을 융합하여 최종 판단을 내리는 집중형 융합 방식과 개별 신뢰도 기법의 판단 결과들을 융합하는 분산형 융합의 두 가지 방식에 대해 최적의 베이시안 융합규칙이 제시되었다. 고립단어 인식에서의 미등록어 거절 실험 결과 집중형 베이시안 신뢰도 융합 기법은 개별 신뢰도 기법에 비해 13% 이상의 상대적인 에러 감소 효과를 보였으나, 분산형 베이시안 융합은 성능의 향상을 보이지 못했다.

핵심용어: 음성인식, 신뢰도, 미등록어 거절, 베이시안 융합, 신뢰도 융합
투고분야: 음성처리 분야 (2.5)

In this paper, we propose a method of confidence measure fusion under Bayesian framework for speech recognition. Centralized and distributed schemes are considered for confidence measure fusion. Centralized fusion is feature level fusion which combines the values of individual confidence scores and makes a final decision. In contrast, distributed fusion is decision level fusion which combines the individual decision makings made by each individual confidence measuring method. Optimal Bayesian fusion rules for centralized and distributed cases are presented. In isolated word Out-of-Vocabulary (OOV) rejection experiments, centralized Bayesian fusion shows over 13% relative equal error rate (EER) reduction compared with the individual confidence measure methods. In contrast, the distributed Bayesian fusion shows no significant performance increase.

Keywords: Speech recognition, Confidence measure, OOV rejection, Bayesian fusion, CM fusion
ASK subject classification: Speech signal processing (2.5)

I. 서 론

음성인식에 있어서 신뢰도 (confidence measure)는 시스템의 인식 결과에 대한 신뢰의 정도를 나타내며, 일반적으로 그 인식된 결과를 받아 들일지 거절할 지를 판단하는데 사용된다. 실제 환경에서의 음성인식은 주변 잡음이나, 대화체 발음 등의 문제로 높은 에러율을 보이고 있다. 인식결과에 대한 신뢰도 측정은 이러한 에러를 검출하고, 처리하는 방법을 제공할 수 있다. 신뢰도를 이용할 경우 음성인식의 결과로 주어진 단어들에 대해 그것이 실제로 맞는 결과인지 또는 인식된 결과를 얼마나 믿을 수 있는지를 수치로 제시해 줄 수 있기 때문이다. 이것은 단어뿐만 아니라 인식된 문장에 대해서도 적용이 가능하다. 실제로 신뢰도는 음성인식의 과정에서 많은 부분에 응용되고 있다. 인식 단어가 등록 단어의 집합(vocabulary)에 속해 있는지를 판단하여 미등록어(out-of-vocabulary) 검출에 사용될 수 있으며, 입력 음성에서 특정 핵심어 (key word)의 검출에 사용될 수 있다. 또한 음성 대화 시스템에서 명확하지 못한 인식 결과에 대하여 재 질문에 사용되거나[1], 비교사(unsupervised) 화자적응에 응용된다[2].

신뢰도 측정은 결과적으로 음성인식의 결과를 맞는 것으로 볼 것인가, 아니면 틀린 것으로 볼 것인가의 판단(decision making)의 문제로 귀결될 수 있다. 즉, 두 개

책임저자 : 고 한 석 (hsko@korea.ac.kr)
136-701 서울시 성북구 안암동 5가
고려대학교 전자컴퓨터공학과
(전화: 02-3290-3239; 팩스: 02-921-0544)

의 클래스를 가지는 패턴 분류 문제 (two-class classification problem)로 생각해 볼 수 있는 것이다. 이 때 신뢰도는 패턴 분류를 위해 사용되는 특징 (feature) 벡터로 볼 수 있다. 이러한 관점에서 기존의 연구들은 크게 두 가지로 나눠 볼 수 있다. 첫째는 패턴 분류를 위한 좋은 신뢰도 측정 기법 혹은 특징을 선정하기 위한 작업이다. 좋은 신뢰도란 두 클래스에 대해 변별적인 정보를 제공하는 것이며, 이러한 특징은 음성인식 과정의 여러 곳에서 얻어질 수 있다. 이들은 크게 음향학적인 정보, 언어 모델의 정보, 디코딩 과정의 정보 등을 통해 다양하게 얻어진다[3-5]. 둘째는 측정된 신뢰도를 이용한 패턴 분류 기법에 대한 연구이다. 패턴 분류에 단일의 신뢰도 값을 사용 할 수 있으나, 다양한 신뢰도를 융합 (fusion)하여 패턴 분류를 할 수도 있다. 이러한 접근으로는 LDA (linear discriminative analysis) [7], SVM (support vector machine)[3], 결정트리[3], 신경망[8], 부스팅 (boosting)[9] 등이 있다.

본 논문에서는 베이시안 기법에 기반한 신뢰도 융합을 제안한다. 이 방법은 두 개 클래스의 확률밀도함수를 추정하고 이것을 이용하여 베이시안 기준에서 최적인 판단경계 (decision boundary)를 찾는 과정으로 구성된다. 개별 신뢰도 기법의 신뢰도 값을 특징벡터로 융합하여 최종 판단을 내리는 집중형 융합 방식과 개별 신뢰도 기법의 판단 결과들을 융합하는 분산형 융합의 두 가지 방식에 대해 최적의 베이시안 융합규칙이 제시되었다. 기존의 신뢰도 융합 연구들은 LDA의 경우에는 차원 감소를 통해서, SVM은 마진을 크게 하는 판단경계를 적용하여 두 클래스간의 변별력을 크게 함으로써 좋은 분류 성능을 얻는데 초점을 맞추었다. 그러나 제안한 방법은 신뢰도 융합에 사용된 특징벡터의 확률밀도함수의 추정에 초점을 맞춘다. 이 방법은 확률밀도함수의 모델링을 이용하는 다양한 통계적 처리의 기반을 제공할 수 있을 것으로 예상한다. 베이시안 신뢰도 융합은 베이즈 리스크 (Bayes risk) 측면에서 최적인 신뢰도 융합을 보장한다는 장점을 가지고 있다.

II. 기존 신뢰도 측정 기법들의 비교

음성인식 시스템의 인식 결과로 주어진 단어 w_0에 대해 신뢰도를 측정한다고 하자. 신뢰도는 인식 결과가 맞을 가능성을 수치화 한 것이므로 인식 단어에 대한 신뢰도는 사후 확률 $P(w_0|\vec{O})$로 적절히 모델링 될 수 있다. 여기서 $\vec{O}$는 입력 음성을 나타낸다. 사후 확률이 주어졌을 경우 신뢰도를 이용한 판단은 임계치 비교로 할 수 있다.

$$P(w_0|\vec{O}) \underset{\omega_0}{\overset{\omega_1}{\gtrless}} \eta \tag{1}$$

인식 단어의 사후 확률은 베이즈 정리를 이용하면 다음과 같이 나타낼 수 있다.

$$P(w_0|\vec{O}) = \frac{p(\vec{O}|w_0)P(w_0)}{p(\vec{O})} = \frac{p(\vec{O}|w_0)P(w_0)}{\sum_{w_i} p(\vec{O}|w_i)P(w_i)} \approx \frac{p(\vec{O}|w_0)}{\sum_{w_i} p(\vec{O}|w_i)} \tag{2}$$

본 논문에서는 고립단어 인식에서의 신뢰도 측정 문제만을 언급하려 한다. 이 경우 언어 모델링에 해당하는 $P(w_i)$는 적절히 무시할 수 있다.

사후 확률의 계산은 신뢰도 측정의 정확한 방법을 제공하지만, 실제 이것의 계산은 매우 어렵다. 식(2)에서 $p(\vec{O})$는 가능한 모든 단어 w_i의 확률 $p(\vec{O}|w_i)$의 합인데, 이것의 정확한 계산이 불가능하기 때문이다. 따라서, 분모의 계산을 적절히 근사할 수 있는 모델이 필요하다. 이를 위해 음소 필러 (phonetic filler) 모델과 가우시안 믹스쳐 (Gaussian mixture) 모델 등이 사용된다. 음소 필러는 병렬 음소 연결과 귀환 루프의 구조를 가진다 (그림 1). 이 구조는 가능한 모든 발음을 음소 수준에서 모델링한 것이며, 이 모델을 이용한 확률 계산은 사후 확률의 분모 부분을 계산한다고 볼 수 있다. 실제로 음소 필러 모델 λ_{filler}를 사용한 신뢰도 $CM_{filler}(\vec{O}, w_0)$는 다음과 같이 계산한다.

$$CM_{filler}(\vec{O}, w_0) = \frac{1}{T}\log\left[\frac{p(\vec{O}|w_0)}{p(\vec{O}|\lambda_{filler})}\right] \tag{3}$$

계산의 편의를 위해서 로그 확률을 이용하고 단어의 발음 길이에 따른 영향을 없애기 위해서 발음에 소요된 프레임 수 T로 정규화하여 사용된다.

음성을 가우시안 믹스쳐 확률밀도함수를 이용하여 모델링할 경우, 이것은 가능한 발음을 음향학적인 수준에서 모델링한 것으로 이것 역시 사후 확률의 분모 부분을 계산하는 것이다. 가우시안 믹스쳐 모델 λ_{Gmix} 을 이용한 신뢰도 $CM_{Gmix}(\vec{O}, w_0)$ 는 다음과 같이 계산한다.

$$CM_{Gmix}(\vec{O}, w_0) = \frac{1}{T}\log[\frac{p(\vec{O}\,|\,w_0)}{p(\vec{O}\,|\,\lambda_{Gmix})}] \tag{4}$$

즉, 이 두 가지 모델을 이용한 신뢰도 측정 기법은 $p(\vec{O})$ 를 근사적으로 계산함으로써 사후 확률을 측정하는 방법이라 할 수 있다.

사후 확률 계산에 의한 신뢰도 측정은 다른 형태의 신뢰도 계산 모델로 나타나기도 한다. 사후 확률은 다음과 같이 변형이 가능하다.

$$\begin{aligned} P(w_0\,|\,\vec{O}) &= \frac{p(\vec{O}\,|\,w_0)P(w_0)}{p(\vec{O})} \\ &= \frac{p(\vec{O}\,|\,w_0)P(w_0)}{p(\vec{O}\,|\,w_0)P(w_0) + p(\vec{O}\,|\,w_0^c)(1-P(w_0))} \\ &= \frac{1}{1+\frac{p(\vec{O}\,|\,w_0^c)}{p(\vec{O}\,|\,w_0)}\frac{1-P(w_0)}{P(w_0)}} \end{aligned} \tag{3}$$

여기서, w_0^c 는 인식단어 w_o 가 아닌 단어들로 이루어진 집합을 의미한다. 이 때 사후 확률을 이용하는 임계값 판정식은 다음과 같다.

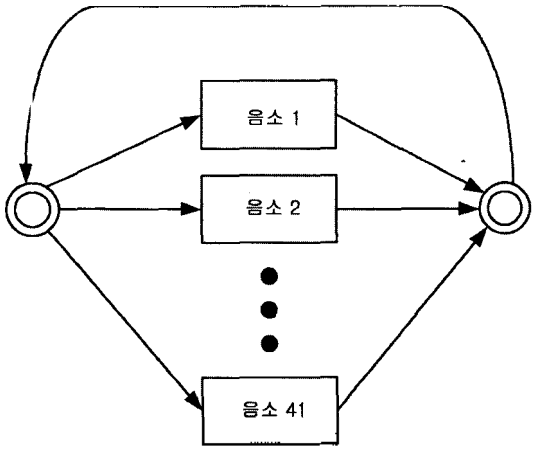


그림 1. 병렬 음소 연결 및 귀환 구조를 가지는 음소 필러 모델
Figure 1. Phonetic filler model of parallel phone connection and loop back path

$$\frac{1}{1+\frac{p(\vec{O}\,|\,w_0^c)}{p(\vec{O}\,|\,w_0)}\frac{1-P(w_0)}{P(w_0)}} \underset{\omega_0}{\overset{\omega_1}{\gtrless}} \eta \tag{4}$$

언어모델의 값을 임의 상수로 대치하고 정리하면,

$$\frac{p(\vec{O}\,|\,w_0)}{p(\vec{O}\,|\,w_0^c)} \underset{\omega_0}{\overset{\omega_1}{\gtrless}} \eta^{*} \tag{5}$$

즉, 우도비 (likelihood ratio) 계산을 통해서도 사후 확률 기반의 신뢰도 측정이 가능함을 알 수 있다.

신뢰도 측정 기법으로 많이 사용되는 서브워드 기반의 우도비 테스트 (sub-word likelihood ratio test)가 좋은 예이다. 비터비 알고리즘을 사용하여 인식 단어 w_0 를 구성하는 N 개 서브워드 q_i 의 세그멘테이션 (segmentation) $\vec{O} = (\vec{o}_{q_1}, \vec{o}_{q_2}, ..., \vec{o}_{q_N})$ 을 얻었다고 하자. 이 때 서브워드 q_i 의 로그 우도비는 다음과 같이 계산할 수 있다.

$$LLR(\vec{o}_{q_i}, q_i) = \frac{1}{T_{q_i}}\log[\frac{p(\vec{o}_{q_i}\,|\,\lambda_{q_i})}{p(\vec{o}_{q_i}\,|\,\bar{\lambda}_{q_i})}] \tag{6}$$

여기서 λ_{q_i} 는 서브워드 q_i 의 모델을, $\bar{\lambda}_{q_i}$ 는 q_i 의 반모델 (anti-model)을 나타낸다. 반모델은 인식에러를 내는 음성데이터를 이용하여 변별적인 훈련과정을 통해 얻는 방법이나[5], 가우시안 믹스쳐 모델 λ_{Gmix} 을 이용하는 방법, 자신을 제외한 서브워드 모델 $q(k)$ 의 확률을 다음처럼 평균하는 방법이 있다 [6].

$$p(\vec{o}_{q_i}\,|\,\bar{\lambda}_{q_i}) = \left[\frac{1}{K-1}\sum_{\substack{k=1 \\ q(k)\neq q_i}}^{K} p(\vec{o}_{q_i}\,|\,\lambda_{q(k)})^{\nu}\right]^{1/\nu} \tag{7}$$

여기서 K는 서브워드 모델의 개수, ν 는 평균에 사용되는 지수 가중치이다. 서브워드모델의 확률을 평균하는 방법은 좋은 성능을 보이지만 계산량이 많다는 단점이 있다. 단어 단위의 신뢰도는 $LLR(\vec{o}_{q_i}, q_i)$ 들을 평균하여 얻는데 이 값들은 변화의 폭이 매우 커서 국소적인 불일치가 단어 단위 신뢰도 값을 크게 좌우하는 문제점이 발생한다. 따라서 로그 우도비 값의 변화폭을 [0,1] 로 제한

하기 위해 다음과 같이 시그모이드 (sigmoid) 함수를 추가로 사용한 서브워드의 신뢰도 값을 사용한다.

$$S(\bar{o}_{q_i}, q_i) = \frac{1}{1+\exp[-\gamma(LLR(\bar{o}_{q_i}, q_i)-\tau)]} \tag{8}$$

여기서 $\gamma > 0$는 시그모이드 함수의 기울기, τ는 중심값을 결정하는 상수이다. 최종적으로 단어단위 신뢰도는 $S(\bar{o}_{q_i}, q_i)$의 산술평균이나 기하평균을 취하여 다음과 같이 얻어진다.

$$CM_{sub_arth}(\bar{O}, w_0) = \frac{1}{N}\sum_{i=1}^{N} S(\bar{o}_{q_i}, q_i) \tag{9}$$

$$CM_{sub_geo}(\bar{O}, w_0) = \exp[\frac{1}{N}\sum_{i=1}^{N} \log S(\bar{o}_{q_i}, q_i)] \tag{10}$$

따라서 서브워드 기반의 우도비 테스트는 서브워드 단위의 사후 확률 $LLR(\bar{o}_{q_i}, q_i)$을 시그모이드 함수를 이용해 임계값 판정하고 그 결과들을 평균한 것으로 볼 수 있다.

N-best 기법 또한 우도비 계산을 통한 신뢰도 측정의 한 방법이다. 이 기법은 인식결과 단어에 연관된 N-best 결과들을 이용하는 방법으로 신뢰도 $CM_{N-Best}(\bar{O}, w_0)$는 다음과 같이 계산된다.

$$CM_{N-Best}(\bar{O}, w_0) = \frac{1}{T}[\ln p(\bar{O}\,|\,w_0) - \ln \sum_{\substack{w_i \in N-best \\ w_i \neq w_0}} p(\bar{O}\,|\,w_i)] \tag{11}$$

이 기법은 N-best 단어들의 확률 (likelihood)을 합하여, $p(\bar{O}\,|\,w_0^c)$를 모델링하며, 로그를 정리하여 지수함수를 곱할 경우 우도비 테스트의 형태를 가짐을 알 수 있다.

위에서 살펴본 바와 같이 신뢰도 측정은 사후 확률 계산을 기본으로 하고 있으나, 적용에 있어 다른 가정과 모델링을 통하여 다양한 형태의 측정 기법이 있음을 알 수 있다.

III. 베이시안 신뢰도 융합 기법

3.1. 집중형 베이시안 신뢰도 융합

신뢰도 측정을 위해 많은 기법들이 사용되고 있는데, 이는 신뢰도라는 개념 자체가 복합적인 요인에 의해서 영향을 받기 때문이라고 생각된다. 복합적인 요인들을 반영하기 위해서는 기존 신뢰도 측정 기법의 융합이 적절하다. 신뢰도 융합은 두 종류의 접근 방법을 가질 수 있다 (그림 2). 첫 번째는 개별 기법에서 측정된 신뢰도를 융합하여 판단을 내리는 것이고, 두 번째는 개별 기법들이 신뢰도를 기반으로 판단을 내리면, 이 판단을 융합하여 최종의 판단을 내리는 것이다. 이러한 두 가지 접근 방식은 센서 융합의 문제에서 각각 집중형 (centralized) 센서 융합과 분산형 (distributed) 센서 융합으로 불리는 기법들과 동일하다[10-11].

집중형 융합은 N 개의 신뢰도 기법들로 구성된 특징벡터 $\vec{X}_N = (x_1, \ldots, x_{N-1}, x_N)$를 이용한 2 개 클래스의 패턴 분류 문제로 다음의 베이즈 리스크 $R_{\vec{X}_N}$를 최소화 함으로써 얻어진다.

$$\begin{aligned} R_{\vec{X}_N} &= \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij} P(\omega_j, r_N(\vec{X}_N) = \omega_i) \\ &= C_{10}P(\omega_0)\int_{\Omega_1} p(\vec{X}_N\,|\,\omega_0)d\vec{X}_N + C_{01}P(\omega_1)\int_{\Omega_0} p(\vec{X}_N\,|\,\omega_1)d\vec{X}_N, \\ \Omega_j &= \{\vec{X}_N\,|\,r_N(\vec{X}_N) = \omega_j\} \end{aligned} \tag{12}$$

여기서 $P(\omega_j)$는 클래스 ω_i, $j \in \{0,1\}$의 사전 확률, C_{ij}는 ω_j 클래스에서 특징벡터가 주어졌을 때 판단을 ω_i 클래스로 내리는 경우에 대한 코스트로 $C_{00} = C_{11} = 0$을 가

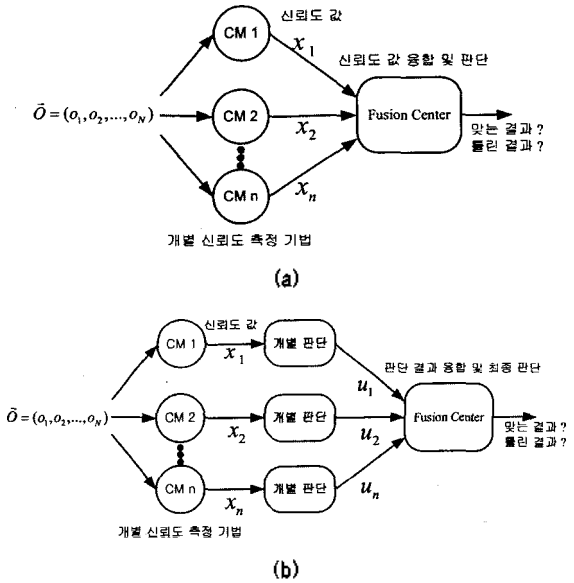


그림 2. 신뢰도 융합의 두 가지 방법
(a) 집중형 신뢰도 융합 기법
(b) 분산형 신뢰도 융합 기법

Figure 2. Two types of confidence measure fusion
(a) Centralized confidence measure fusion
(b) Distributed confidence measure fusion

정하였다. $r_N(\bar{X}_N)$은 판단함수 (decision function)로 신뢰도 융합의 관점에서는 융합규칙으로 볼 수 있다. 최소의 베이즈 리스크 $R^{\min}_{\bar{X}_N}$을 갖는 최적의 베이시안 융합규칙 $r^{opt}_N(\bar{X}_N)$은 다음과 같이 계산된다.

$$r^{opt}_N(\bar{X}_N)=\begin{cases}\omega_1 & p(\bar{X}_N|\omega_1)/p(\bar{X}_N|\omega_0)>C_{10}P(\omega_0)/C_{01}P(\omega_1)=\eta_0\\ \omega_0 & otherwise\end{cases} \tag{13}$$

이때 집중형 베이시안 융합에 의한 신뢰도는 다음과 같다.

$$CM_{Bayesian}(\bar{X}_N,w_0)=\log[\frac{p(\bar{X}_N|\omega_1)}{p(\bar{X}_N|\omega_0)}] \tag{14}$$

(14) 식은 신뢰도 융합의 결과가 특징벡터의 우도비임을 보이고 있다. 우도비의 형태를 갖는 신뢰도는 사후확률로 변환할 수 있으므로 집중형 베이시안 융합에 의한 신뢰도 역시 사후 확률 계산에 기반한 신뢰도 측정법이다.

만약 클래스의 사전 확률과 코스트가 알려져 있다면 베이시안 융합규칙은 이 값을 이용해 임계값 η_0를 정함으로써 얻을 수 있으며 $C_{00}=C_{11}=0, C_{01}=C_{10}=1$의 코스트를 사용하는 경우에는 최적 융합규칙은 최소의 분류 에러를 가짐이 보장된다. 실제로 음성인식에 적용하는 경우 코스트 값이 명시적으로 주어지지 않는 것이 일반적이다. 이러한 경우에는 C_{10}와 C_{01} 값의 상대적인 비율을 고려하여 η_0를 결정할 수 있을 것이다. 결론적으로 집중형 베이시안 신뢰도 융합은 신뢰도 x_i의 결합 확률밀도함수 (joint probability density function)를 이용한 우도비 테스트의 형태를 가지게 되며, 이 때 이 확률분포의 추정이 가장 중요한 문제가 된다.

결합 확률밀도함수의 추정을 위해서는 다변수의 가우시안 또는 가우시안 믹스쳐 모델을 사용한 파라미터 추정 (parameter estimation) 방법을 사용할 수 있다. 히스토그램 분석을 통해 기존의 신뢰도 측정 방법들 중 상당수가 가우시안 형태의 확률분포를 가짐을 알 수 있었다. 그러나 서브워드 기반의 우도비 테스트 기법에서 얻어지는 신뢰도는 그 분포가 가우시안에서 벗어남을 관찰하였다. 이는 서브워드 단위 신뢰도에 적용되는 시그모이드 함수때문인 것으로 생각된다. 실험적으로 관찰해 본 결과, 시그모이드 함수의 역함수를 얻어진 신뢰도 값에 적용할 경우 그 분포를 가우시안 형태로 만들 수 있었다. 실제로 가우시안 분포를 전혀 따르지 않는 신뢰도 측정 기법들도 있으므로, 일반적인 형태의 확률밀도함수 모델링을 위해서는 가우시안 믹스쳐 모델을 사용하는 것이 바람직하다.

베이시안 신뢰도 융합에 있어서 신뢰도 기법들을 융합했을 경우 성능 향상을 가져 올 수 있는가 역시 중요한 문제다. 이상적으로 베이시안 신뢰도 융합은 어떠한 개별 신뢰도를 선택하든지 융합 후 성능은 향상되거나 최소한 유지된다. 그 증명은 다음과 같다.

N 개의 개별 신뢰도로 구성된 특징벡터 $\bar{X}_N$에 대한 최적 융합규칙 $r^{opt}_N(\bar{X}_N)$이 가지는 베이즈 리스크를 $R^{\min}_{\bar{X}_N}$이라고 하자. N 번째 신뢰도 x_N가 제외된 특징벡터 $\bar{X}_{N-1}$을 사용할 경우 베이즈 리스크 $R^{\min}_{\bar{X}_{N-1}}$를 가지는 최적 융합 규칙 $r^{opt}_{N-1}(\bar{X}_{N-1})$은 주변분포 (marginal distribution)의 특징을 이용하면 다음과 같이 쓸 수 있다.

$$\begin{aligned}r^{opt}_{N-1}(\bar{X}_{N-1})&=\begin{cases}\omega_1 & p(\bar{X}_{N-1}|\omega_1)/p(\bar{X}_{N-1}|\omega_0)>\eta_0\\ \omega_0 & otherwise\end{cases}\\ &=\begin{cases}\omega_1 & \int_{-\infty}^{\infty}p(\bar{X}_N|\omega_1)dx_N/\int_{-\infty}^{\infty}p(\bar{X}_N|\omega_0)dx_N>\eta_0\\ \omega_0 & otherwise\end{cases}\\ &=r_N(\bar{X}_N)\end{aligned} \tag{15}$$

(15) 식으로부터 $\bar{X}_{N-1}$을 사용한 최적 융합규칙 $r^{opt}_{N-1}(\bar{X}_{N-1})$은 $\bar{X}_N$을 이용하는 일반적인 융합규칙 $r_N(\bar{X}_N)$로 볼 수 있다. 그러나 $\bar{X}_N$을 사용한 최적 융합규칙은 $r^{opt}_N(\bar{X}_N)$ 이므로, $R^{\min}_{\bar{X}_{N-1}}$은 $R^{\min}_{\bar{X}_N}$ 보다 작을 수 없다. 즉 i 개 신뢰도의 베이시안 융합의 경우 베이즈 리스크 $R^{\min}_{\bar{X}_i}$는 다음과 같은 관계를 만족한다.

$$R^{\min}_{\bar{X}_1}\geq\ldots\geq R^{\min}_{\bar{X}_i}\geq R^{\min}_{\bar{X}_{i+1}}\geq\ldots \tag{16}$$

이로부터 우리는 두 가지 사실을 알 수 있다. 첫째, 신뢰도 융합은 어떠한 경우라도 개별 기법의 신뢰도를 단독으로 사용하는 것과 동일하거나 그 이상의 성능을 가진다. 둘째, 계속적으로 융합하는 신뢰도의 수를 늘려감에 따라서 그 성능은 계속 향상 되거나 유지된다. 그러나 이것은 확률밀도함수가 정확하게 추정되었다는 가정하에서 성립하는 것이다. 실제로 추가되는 신뢰도 기법

의 수가 늘어날수록 결합 확률밀도함수를 정확하게 추정하는 것은 어려워지며 훈련 데이터의 양이 제한된 경우에는 그 정도가 심해질 수 있다.

3.2. 분산형 베이시안 신뢰도 융합

분산형 베이시안 신뢰도 융합은 각 개별 신뢰도 기법의 판단의 결과 값들을 묶어서 특징벡터로 사용하게 된다. 이것은 판단 융합 (decision fusion)으로 불릴 수 있다. 베이시안 기준하의 최적 융합규칙의 유도는 집중형 융합의 경우와 유사하다. $\bar{U}_N = (u_1, \ldots, u_{N-1}, u_N)$ 을 각 개별 신뢰도 측정 기법의 판단 값을 모아 놓은 특징벡터라고 하자. 이 때 각 개별 신뢰도에 의한 판단을 바이너리 (binary)라고 가정하면, u_i =j는 i 번째 개별 신뢰도에 의한 판단 결과가 클래스 ω_j 를 선택했음을 나타내며, 이것을 특징벡터로 하는 최적 베이시안 융합규칙 $r_N^{opt}(\bar{U}_N)$ 은 다음과 같다.

$$r_N^{opt}(\bar{U}_N) = \begin{cases} \omega_1 & p(\bar{U}_N \mid \omega_1)/p(\bar{U}_N \mid \omega_0) > C_{10}P(\omega_0)/C_{01}P(\omega_1) = \eta_0 \\ \omega_0 & otherwise \end{cases} \tag{17}$$

개별 신뢰도에 의한 판단들은 서로 독립적으로 이루어지므로, 융합규칙의 우도비 항은 다음과 같이 다시 쓸 수 있다.

$$\begin{aligned} \frac{P(u_1, u_2, \ldots u_N \mid \omega_1)}{P(u_1, u_2, \ldots u_N \mid \omega_0)} &= \prod_{i=1}^{N} \frac{P(u_i \mid \omega_1)}{P(u_i \mid \omega_0)} \\ &= \prod_{s1} \frac{P(u_i = 1 \mid \omega_1)}{P(u_i = 1 \mid \omega_0)} \prod_{s0} \frac{P(u_i = 0 \mid \omega_1)}{P(u_i = 0 \mid \omega_0)} \\ &= \prod_{s1} \frac{1 - P_{M_i}}{P_{F_i}} \prod_{s0} \frac{P_{M_i}}{1 - P_{F_i}} \end{aligned} \tag{18}$$

여기서 P_{M_i} 와 P_{F_i} 는 ω_1 클래스에 대한 미탐지 확률 (probability of miss)과 오탐지 확률 (probability of false alarm)로 다음과 같이 정의된다.

$$P_{M_i} = P(u_i = 0 \mid \omega_1) \tag{19}$$

$$P_{F_i} = P(u_i = 1 \mid \omega_0) \tag{20}$$

그리고 S_i 는 판단값이 i 를 갖는 개별 판단들의 집합을 나타낸다.

수식 (18)를 (17) 식에 대입하고 로그를 취하면 최적 융합규칙은 다음과 같이 정리된다 [12].

$$r_N^{opt}(\bar{U}_N) = \begin{cases} \omega_1 & \sum_{i=1}^{N} u_i \log[(1 - P_{M_i})/P_{F_i}] + (1 - u_i)\log[P_{M_i}/(1 - P_{F_i})] > \log \eta_0 \\ \omega_0 & otherwise \end{cases} \tag{21}$$

이때 분산형 베이시안 융합에 의한 신뢰도는 다음과 같다.

$$CM_{BayesD}(\bar{U}_N, w_0) = \sum_{i=1}^{N} u_i \log \frac{1 - P_{M_i}}{P_{F_i}} + (1 - u_i)\log \frac{P_{M_i}}{1 - P_{F_i}} \tag{22}$$

분산형 베이시안 융합에 의한 신뢰도 역시 사후 확률에 기반한 신뢰도 측정방법이다. (22)식은 (17), (18)에서 주어진 우도비 항을 정리한 것이기 때문이다.

2 클래스 분류 문제에 있어서 베이즈 리스크 R 은 다음과 같이 쓸 수 있다.

$$R = C_F P_F - C_D P_D + C \tag{23}$$

여기서 $C_F = C_{10}P(\omega_0)$, $C_D = C_{01}P(\omega_1)$, $C = C_{01}P(\omega_1)$ 이다. 분산형 신뢰도 융합의 경우 P_F 와 P_D 는 다음과 같다.

$$P_F = \sum_{\bar{U}_N} P(u_0 = 1 \mid \bar{U}_N) P(\bar{U}_N \mid \omega_0) \tag{24}$$

$$P_D = \sum_{\bar{U}_N} P(u_0 = 1 \mid \bar{U}_N) P(\bar{U}_N \mid \omega_1) \tag{25}$$

여기서 $u_0 = 1$ 은 융합규칙에 의해 선택된 클래스가 ω_1 임을 의미한다.

(24), (25)을 (23)에 대입하고 (18), (19), (20) 식을 참고하여 정리하면, 분산형 융합의 최적화 대상인 베이즈 리스크는 다음과 같다.

$$\begin{aligned} R &= C_F \sum P(u_0 = 1 \mid \bar{U}_N) \prod^{N} P(u_i \mid \omega_0) - C_D \sum P(u_0 = 1 \mid \bar{U}_N) \prod^{N} P(u_i \mid \omega_1) + C \\ &= C_F \sum P(u_0 = 1 \mid \bar{U}_N) \prod_{S_0} P_{F_i} \prod_{S_1} (1 - P_{F_i}) - C_D \sum P(u_0 = 1 \mid \bar{U}_N) \prod_{S_0} (1 - P_{M_i}) \prod_{S_1} P_{M_i} + C \end{aligned} \tag{26}$$

(21) 식은 P_{M_i} 와 P_{F_i} 이 고정되어 있을 경우의 최적

융합규칙이지만, (26)에서 보듯이 최적화된 신뢰도 융합을 위해서는 P_{M_i}와 P_{F_i}에 대하여 R의 최소화 과정이 필요하다. 이 최소화 과정은 복잡하기 때문에 N 개의 (P_{M_i}, P_{F_i}) 쌍에 대해 순차적으로 최적화 과정을 거치고 이 과정을 반복적으로 수행하는 person-by-person optimization (PBPO) 방식의 최적화를 이용한다 [12, 13]. 실제로 PBPO 최적화 역시 복잡한데, 신뢰도들이 서로 통계적으로 독립이라면 최적화 과정을 간소화 할 수 있다. 그러나 신뢰도들은 통계적으로 높은 상관성을 가지는 경우가 많다. 이 때 신뢰도들이 가우시안을 따르면 두 클래스의 공분산 (covariance) 행렬을 동시 대각화 (simultaneous diagonalization)해서 통계적으로 독립인 성분을 가지는 확률밀도함수를 얻을 수 있다 [14]. 가우시안 분포는 상관성 제거 (de-correlation)를 통해서 통계적 독립의 특성을 얻을 수 있기 때문이다.

분산형 신뢰도 융합의 성능 변화는 집중형 신뢰도 융합에서와 마찬가지로 융합하는 개별 신뢰도의 개수가 증가함에 따라서 베이즈 리스크가 감소하거나, 최소한 유지되는 특징을 지니고 있다. 이것에 대한 증명은 집중형 신뢰도 융합에 대한 증명에서 특징벡터가 개별 신뢰도 기법의 판단 값을 모아 놓은 $\vec{U}_N = (u_1, \ldots, u_{N-1}, u_N)$로 바뀌는 것을 제외하고는 동일한 과정을 따른다.

집중형 신뢰도 융합과 분산형 신뢰도 융합의 성능을 서로 비교해 보자. 집중형 신뢰융합은 개별 신뢰도 기법의 값들을 직접 이용하는데 반해, 분산형 융합은 각 신뢰도 기법에 의한 판단 결과들을 융합하게 된다. 따라서 신뢰도 값을 바이너리의 판단 값으로 변환하여 융합할 때 정보의 손실이 생기게 되어 분산형 신뢰도의 성능은 집중형 신뢰도보다 낮을 것으로 예상 할 수 있다.

IV. 실험 및 분석

신뢰도 융합 기법을 평가하기 위하여, 고립단어 인식에서 미등록어 제거 실험을 해 보았다. 훈련 데이터는 표준어를 사용하는 350명 정도의 남녀가 발성한 국내 주요 지명 6,000 개의 고립 단어 및 연결 단어로 구성되어 있으며, 조용한 사무실 환경에서 헤드셋 마이크를 이용하여 16 kHz 샘플링으로 녹음되었다. 음성 특징벡터는 MFCC와 1 차, 2 차 미분값의 39 차 벡터를 사용했다. 음향 모델은 트라이폰 단위로 결정트리 기반의 클러스터링을 통해 2641 개의 상태를 공유하는 HMM을 얻었으며, 각 상태는 8 개의 가우시안 성분을 가지는 믹스쳐로 모델링하였다. 미등록어 검출 테스트에는 공통음성 데이터베이스 PBW452를 사용하였다. 452 단어 중 200 단어를 등록어 (IV: In-Vocabulary)로 하고 나머지 252 단어를 미등록어로 하여, 남녀 각 15 명에 대하여 452 단어의 처음 발음을 이용하였다. 이 중 20 명의 약 9000 발화는 개발 집합 (development set)에, 나머지 10 명의 약 4500 발화는 테스트 집합으로 사용되었다. 본 실험은 미등록어 검출 능력을 평가하기 위한 목적을 가지고 있으므로, 등록어에 대한 오인식 결과는 평가에 사용되지 않았다.

신뢰도 측정 기법들 중에서 2 절에서 언급한 서브워드

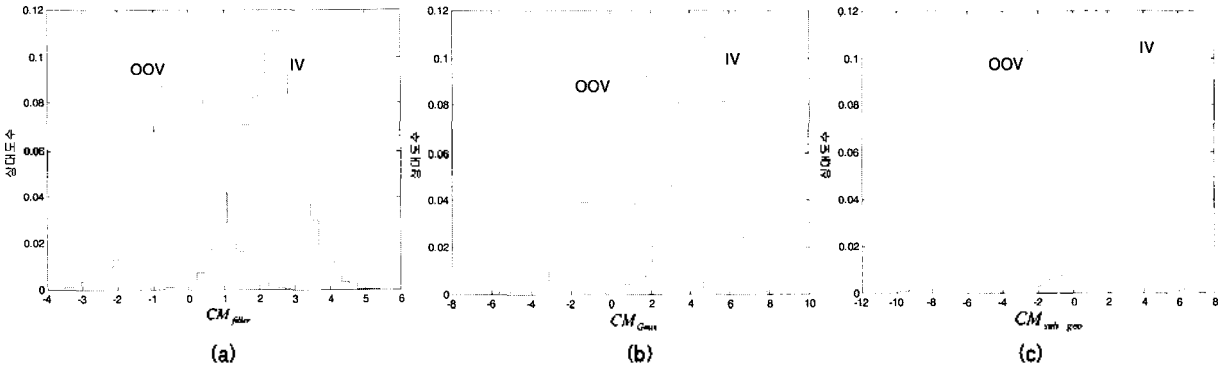


그림 3. 등록어와 미등록어에 대한 신뢰도 값의 정규화 히스토그램
(a) 음소 필러 모델
(b) 가우시안 믹스쳐 모델
(c) 서브워드 우도비 테스트

Figure 3. Normalized histograms of confidence value in IV and OOV cases
(a) Phonetic filler model
(b) Gaussian mixture model
(c) Sub-word likelihood ratio test

표 1. 가우시안 믹스처를 이용한 집중형 신뢰도 융합과 LDA 융합의 EER (%)
Table 1. EER of centralized Bayesian fusion and LDA fusion with various Gaussian mixture models in OOV rejection test (%)

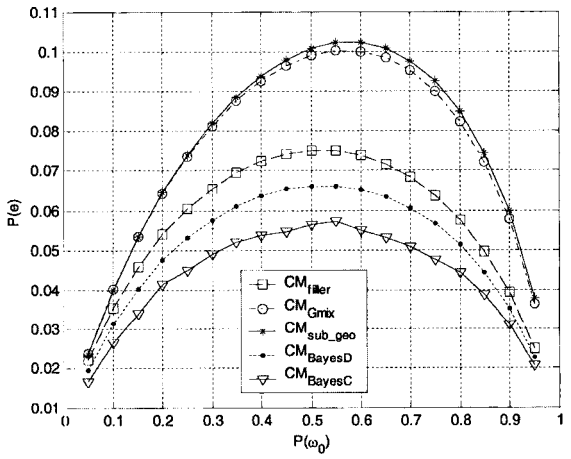
융합 신뢰도	평가 집합	1 가우시안	3 가우시안	7 가우시안	12 가우시안	LDA
3 신뢰도	개발 집합	6.5	6.2	6.2	6.2	6.4
	테스트 집합	6.4	6.2	6.2	6.2	6.4
7 신뢰도	개발 집합	7.6	6.4	6.3	6.1	6.6
	테스트 집합	7.6	6.9	6.6	6.6	6.6

단위 우도비 테스트를 이용한 $CM_{sub_geo}(\bar{O},w_0)$, 음소 필러 모델 기반의 $CM_{filler}(\bar{O},w_0)$, 가우시안 믹스처 모델 기반의 $CM_{Gmix}(\bar{O},w_0)$을 개별적인 신뢰도 기법으로 실험하였다. $CM_{sub_geo}(\bar{O},w_0)$의 계산에 있어서 반음소 모델의 확률은 (7) 식을 이용했으며, 지수 가중치는 $\nu=2$를 사용했다. (8)의 시그모이드 함수의 파라미터는 $\gamma=0.5, \tau=1$ 이었다. 이 값들은 실험적으로 결정하였다. $CM_{Gmix}(\bar{O},w_0)$의 경우에는 256 개의 가우시안 성분을 사용했다. 신뢰도에 대한 확률적인 모델링을 하기 위해 우선 히스토그램을 살펴보았다(그림 3). $CM_{filler}(\bar{O},w_0)$와 가우시안 믹스처 모델을 사용한 $CM_{Gmix}(\bar{O},w_0)$의 경우 신뢰도 값은 등록어와 미등록어에 대해 가우시안에 가까운 형태의 확률 분포를 가진다. $CM_{sub_geo}(\bar{O},w_0)$ 역시 시그모이드 함수의 역함수를 이용한 후처리 후에는 가우시안의 형태를 가진다. 따라서 3가지의 신뢰도 측정 기법에 대하여 가우시안 모델링을 이용한 파라미터 추정은 적절할 것으로 예상된다.

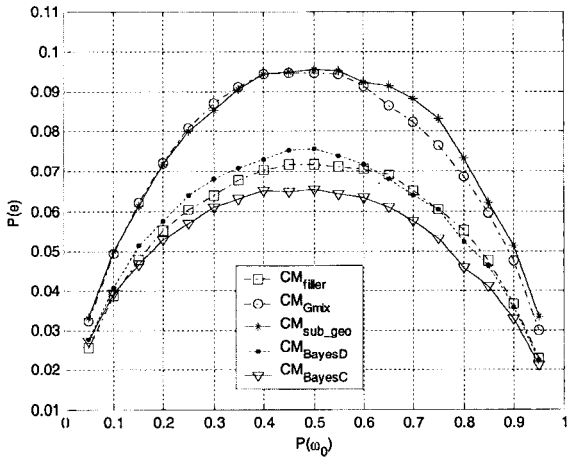
우선 신뢰도의 가우시안 모델링 및 이를 이용한 융합의 효용성을 살펴보았다. 가우시안 분포 모델링에서 신뢰도 값들 간의 높은 상관성 때문에 공분산 행렬은 전체 요소를 사용하였다 (full covariance). 분산형 융합은 결합 확률밀도함수를 동시 대각화하여 서로 독립의 3 가지 성분을 얻고 이를 이용하여 PBPO 과정의 최적화를 수행하였다[13]. PBPO 최적화는 안정적으로 수렴하였다. 성능 평가는 미등록어 발생확률 $P(\omega_0)$에 따른 분류에러 $P(e)$를 계산했다. $P(e)=(1-P(\omega_0))(1-P_D)+P(\omega_0)P_F$로 얻어진다. 신뢰도들이 가우시안을 따른다면 그 성능은 랜덤 데이터를 이용한 시뮬레이션이나, 수학적 계산을 통해서 예측할 수 있다. 그림 4.(a)는 가우시안 가정하에 시뮬레이션을 통해 얻은 각 개별 기법 및 융합 기법의 에러율을, 그림 4.(b)는 동일한 모델에 대해 실제 데이터를 이용한 성능을 나타낸다. 추정에 관련된 문제를 제외하기 위해 개발 집합과 테스트 집합의 데이터를 모두 이용하여 가우시안 모델 파라미터 및 융합 규칙을 추정하고 동일 데이터를 평가에 이용했다. 가우시안 파라미터들은 ML 추정을 하였다. 그림에서 보면 개별 기법과 융합 기법의 에러율은 예측 값과 약간의 차이를 보이고 있음을 알 수 있다. 이는 신뢰도들이 정확한 가우시안 분포를 따르지 않기 때문이다. 개별 기법들은 예측에 비해 대체로 에러율이 약간 감소하고 있으나, 융합 기법들은 다소 증가하는 경향을 보이고 있다. 등록어와 미등록어 발생 확률이 동일한 조건(equal prior condition)에서 $CM_{filler}(\bar{O},w_0)$ 신뢰도에 비해 집중형 융합은 약 25% 정도, 분산형의 경우는 약 13%의 상대적인 에러율 감소를 예측하였다. 그러나 실제로는 집중형 융합의 경우 10%의 성능 향상을 얻었으며, 분산형의 경우에는 에러가 증가하였다. 따라서 이후 실험은 집중형 융합의 경우만을 고려하였다.

가우시안과 가우시안 믹스처를 이용한 집중형 융합에 대해 추정 문제를 고려한 실제 성능을 평가해 보고 LDA를 이용한[7]의 방법과 비교해 보았다 (표 1). 가우시안 모델 파라미터의 추정은 개발 집합을 이용하였고, 성능 평가는 개발 집합과 테스트 집합에 대해 각각 해보았다. 가우시안 믹스처 모델을 사용한 융합에서 파라미터 추정은 EM 알고리즘을 사용하였다. 성능은 단일 수치로 비교 평가 할 수 있도록 P_D와 P_F에 대한 EER을 사용했으며, 좀 더 정확한 성능평가를 위하여 30 명의 화자 중 개발 집합 10 명과 테스트 집합 20 명의 화자를 확정하지 않고 50 회에 걸쳐 두 집합을 임의로 선정하여 파라미터 추정 및 평가를 반복하여 얻어진 EER을 평균하였다. 개별 기법 중에서는 $CM_{filler}(\bar{O},w_0)$가 가장 작은 7.1%의 EER을 보였다. 3 개 신뢰도의 융합에 있어서 3 개의 가우시안 믹스처를 사용한 경우 에러율이 약간 감소하지만, 7 개와 12 개로 늘렸을 경우에 추가적인 에러율 감소는 없다. 이는 3 개의 신뢰도가 가우시안을 적절히 따르기 때문에 많은 수의 가우시안 믹스처가 필요하지 않기 때문이다. 또한 개발 집합과 테스트 집합을 평가에 사용했을 때 에러율의 차이가 없는 것으로 보아 추정은 정확하게 이루어지고 있음을 알 수 있다. 단일 가우시안 모델링에서 융합에 사용되는 신뢰도를 7 개로 늘렸을 경우에는 성능의 저하가 나타났다. 추가된 신뢰도들은

$N=5$를 사용한 $CM_{N-Best}(\vec{O},w_0)$, $CM_{sub_geo}(\vec{O},w_0)$에서 반모델로 가우시안 믹스쳐 λ_{Gmix}를 사용하는 기법, 서브워드 우도비를 산술평균하는 $CM_{sub_arth}(\vec{O},w_0)$, 디코딩 과정에서 활성화된 경로의 평균 개수를 세는 기법 등이다. 이들 신뢰도들은 대부분 그 분포가 가우시안에 많이 벗어나기 때문에 단일 가우시안을 이용한 융합은 에러율이 증가한다. 또한 믹스쳐 모델을 사용하는 경우에는 가우시안의 개수를 계속적으로 증가시키면 개발집합에서의 에러율은 지속적으로 감소하지만 테스트 집합에서의 에러율은 7 개 이상의 가우시안에서 더 이상 감소하지 않는다. 즉, 집중형 가우시안 융합의 융합 신뢰도의 개수나, 믹스쳐 모델의 가우시안 개수를 증가시키는 것이 직접적인 성능 향상으로 이어지지 않음을 알 수 있다. 이는 추정 과정이 정확하지 못했기 때문이다. 이것은 오버피팅 (overfitting) 문제로 불리며 훈련 샘플의 개수에 비해 파라미터의 개수가 많은 경우에 개발 집합에서는 좋은 추정 값을 주는 것으로 보이나, 실제 테스트 집합을 이용한 평가 결과에서 개발 집합에서의 예측과 달라지는 현상을 말한다. 실제로 많은 수의 특징을 사용한 복잡한 확률밀도함수 모델링보다 중요한 특징의 선택과 효율적인 모델링이 성능 향상에 도움을 준다는 것을 알 수 있다. 전체적으로 집중형 융합의 성능은 3 개의 가우시안을 가지는 믹스쳐 모델을 이용하여 3 개 신뢰도를 융합하는 경우가 가장 적절했으며, 이는 개별 기법에 비해 약 13% 이상의 EER 감소가 있었다.



(a)



(b)

그림 4. 개별 기법과 집중형 및 분산형 신뢰도 융합의 에러율
(a) 가우시안 랜덤 데이터를 이용한 시뮬레이션
(b) 테스트 데이터를 이용한 실제 성능

Figure 4. [illegible] of individual CM's, centralized and distributed Bayesian fusions in OOV rejection test
(a) Simulation results with Gaussian random data
(b) Results with real speech data

LDA 융합 기법과 비교해 보았을 때 가우시안 믹스쳐 모델을 사용하는 융합의 성능은 비슷했으며, 융합 신뢰도의 선택과 믹스쳐 모델의 가우시안 개수가 적절히 선택되었을 경우에는 약간의 성능 향상이 있었다. 실험 결과에서 보듯이 베이시안 융합은 기존의 방법에 비해 월등한 성능 향상을 보여 주지는 못하였다. 분산형 융합의 경우에는 개별 기법보다 성능이 저하되기도 했다. 하지만 신뢰도의 확률밀도 함수 추정에 기반한 베이시안 융합은 신뢰도 측정의 확률적 모델링을 제공하고, 베이시안 최적화가 가능하다는 장점을 지닌다. 신뢰도에 대한 이 확률적 모델은 음성인식에서 고려되었던 다양한 통계적 처리에 사용될 수 있을 것이다. 그 한가지 예로써 신뢰도 모델에 대한 화자적응 기법을 들 수 있다. 이 외에도 응용분야는 더 있을 것으로 예상한다.

V. 결 론

본 논문에서는 음성인식을 위한 베이시안 신뢰도 융합 기법을 제안하였다. 기존의 신뢰도 측정 기법들은 사후확률의 계산을 위한 다양한 모델링의 측면에서 이해가 가능했으며, 이들 기법에 대한 집중형과 분산형의 베이시안 융합에 대해 최적 융합규칙을 얻을 수 있었다. 고립단어 미등록어 제거 실험에서 집중형 베이시안 융합은 개별 기법에 비해 단일 가우시안 모델은 8%, 가우시안 믹스쳐 모델은 13% 이상의 상대적인 EER 감소를 관찰할 수 있었다. 그러나 이 수치는 가우시안 가정의 부정확성으로 인해 예측된 것 보다는 작은 에러 감소율이었다. 가우시안 믹스쳐를 사용하는 경우에는 융합하는 신뢰도의 개수

와 가우시안 성분의 개수를 조절해 오버피팅 문제가 발생하지 않도록 해야 한다. 분산형 융합에서는 가우시안 가정의 문제로 인해 성능 향상을 얻지 못했다. LDA 융합 기법과 비교했을 때 베이시안 융합은 의미있는 성능 향상을 가져오지는 못했지만, 신뢰도 측정의 확률적 모델을 제공한다는 점에서 많은 응용이 가능하리라 생각한다.

감사의 글

본 연구는 산업자원부 공업기술개발 과제 (No.10011362)에 의해 지원되었습니다.

참 고 문 헌

1. Paul Carpenter, Chun Jin, et al., "Is This Conversation on Track?" in proc. EUROSPEECH , 2121 - 2124, 2001
2. Micheal Pitz, Frank Wessel, et al., "Improved MLLR Speaker Adaptation Using Confidence Measures for Conversation Speech Recognition," in proc. ICSLP, 2000.
3. Rong Zhang, Alexander Rudnicky, "Word Level Confidence Annotation using Combinations of Features," in proc. EUROSPEECH, 2105 - 2108, 2001.
4. Frank Wessel, Schluter, et al., "Confidence Measures for Large Vocabulary Continuous Speech Recognition," IEEE Trans. on Speech and Audio Processing, 9(3), 288-298, 2001.
5. Eduardo Lleida, Richard C. Rose, "Utterance Verification in Continuous Speech Recognition: Decoding and Training Procedures," IEEE Trans. on Speech and Audio Processing, 8(2), 126-139, 2000.
6. Rafid A. Sukkar, Chin-Hui Lee, "Vocabulary Independent Discriminative Utterance Verification for Nonkeyword Rejection in Subword Based Speech Recognition," " IEEE Trans. on Speech and Audio Processing, 4(6), 420-429, 1996.
7. Simo Kamppari, Timothy Hazen, "Word and Phone Level Acoustic Confidence Scoring," in proc. ICASSP, 3, 5-9, 2000.
8. Mitch Weintraub, Francoise Beaufays, et al., "Neural -Network Based Measures of Confidence for Word Recognition," in proc. ICASSP, 887-890, 1997.
9. Pedro J. Moreno, Beth Logan, et al., "A Boosting Approach for Confidence Scoring," in proc. EUROSPEECH, 2109 - 2112, 2001.
10. David L. Hall, James Llians, "An Introduction to Multisensor Data Fusion," Proceedings of the IEEE, 85(1), 6-23, 1997.
11. Ramanarayanan Viswanathan, Pramod Varshney, "Distributed Detection with Multiple Sensors: Part 1 - Fundamentals," Proceedings of the IEEE, 85(1), pp. 54-63, 1997.
12. Pramod Varshney, Distributed Detection and Data Fusion, Springer-Verlag, New York, 1997.
13. Zhunag-Bo Tang, Krishna R. Pattipati, et al., " An Algorithm for Determining the Decision Thresholds in a Distributed Detection Problem," IEEE Trans. on Sys., Man, and Cybernetics, 21(1), pp. 231-237, 1991.
14. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd edition, Academic Press, 1990.

저자 약력

• 김 태 윤 (Tae-Yoon Kim)

1997년 2월: 고려대학교 전자공학과 (공학사)
1999년 8월: 고려대학교 전자공학과 (공학석사)
1999년 8월~현재: 고려대학교 전자 컴퓨터 공학과 박사과정 재학 중
*주관심 분야: 음성인식, 통계적 신호처리, 패턴인식

• 고 한 석 (Hanseok Ko)

1982년 5월: Carnegie-Mellon Univ. Electrical Engineering (공학사)
1985년 5월: Univ. of Maryland System Engineering (공학석사)
1988년 5월: Johns Hopkins Univ. Electrical Engineering (공학석사)
1992년 5월: Catholic Univ. of America Electrical Engineering (공학박사)
1983년 9월~1995년 2월: White Oak 연구소 Principal Engineer
1992년 9월~1995년 2월: Univ. of Maryland Part-time Asst. professor
1995년 3월~현재: 고려대학교 전자컴퓨터공학과 교수
*주관심 분야: 음성 신호처리, 이미지 데이터 융합, 표적 신호 탐지 및 추적