韓國컴퓨터情報學會 論文誌
第15卷 第12號, 2010. 12.

2010-15-12-4-1

미등록어 거절 알고리즘에서 가우시안 모델 최적화를 이용한 신뢰도 정규화 향상

안 찬 식*, 오 상 엽**

In Out-of Vocabulary Rejection Algorithm by Measure of Normalized improvement using Optimization of Gaussian Model Confidence

Chan-Shik Ahn*, Sang-Yeob Oh**

요 약

어휘 인식에서는 인식 학습 시 나타나지 않는 미 출현 트라이 폰이 존재하며, 이들 시스템에서는 모델 파라미터들의 초기 추정치를 생성하지 못하고 음소 데이터에 대한 모델을 구성할 수 없는 단점으로 인하여 가우시안 모델의 정확성을 확보하지 못하게 된다. 이를 개선하기 위하여 확률 분포를 이용한 모델 파라미터의 가우시안 모델 최적화 방법을 제안한다. 확률 분포의 가우시안 모델을 최적화하여 가우시안 모델의 정확성을 제공하고, 음소 단위로 데이터의 탐색을 지원하여 신뢰도가 향상되었다. 제안된 방법의 성능 평가를 위하여 실제 다양한 미등록어가 관측될 수 있는 대상으로 실험을 수행하였으며 본 연구에서 제안한 정규화 신뢰도를 이용한 미등록어 거절 알고리즘이 기존의 방법들에 비하여 평균 1.7%의 성능향상을 나타내었다.

Abstract

In vocabulary recognition has unseen tri-phone appeared when recognition training. This system has not been created beginning estimation figure of model parameter. It's bad points could not be created that model for phoneme data. Therefore it's could not be secured accuracy of Gaussian model. To improve suggested Gaussian model to optimized method of model parameter using probability distribution. To improved of confidence that Gaussian model to optimized of probability distribution to offer by accuracy and to support searching of phoneme data. This paper suggested system performance comparison as a result of recognition improve represent 1.7% by out-of vocabulary rejection algorithm using normalization confidence.

• 제1저자 : 안찬식 교신저자 : 오상엽
• 투고일 : 2010. 08. 24, 심사일 : 2010. 09. 02, 게재확정일 : 2010. 09. 16.
* 광운대학교 컴퓨터공학과 박사과정 **경원대학교 IT대학 컴퓨터소프트웨어
※ 본 연구는 2010년도 경원대학교 지원에 의한 결과임.

▸ Keyword : 가우시안 모델(gaussian model), 모델 최적화(model optimization),
미등록어 거절(out-of vocabulary rejection), 신뢰도(confidence)

Ⅰ. 서 론

미등록어 거절 방식은 구현 방식에 따라 발화 검증 방식과 핵심어 검출 방식으로 구분된다. 핵심어 검출 방식은 문법을 설계할 때 핵심어만 고려하고 나머지 단어는 가비지(garbage) 모델을 사용하여 불필요한 단어를 제거하여 사용하는 방법이다 [1]. 발화 검증 방식은 인식 결과를 확인하는 과정이 추가되며 필러(filler) 모델을 이용한다. 필러 모델은 구성방식이 단어 기반이므로 가변 어휘 단어 인식 시스템을 위한 발화 검증 구현을 위해서는 매 음소단위의 검증기능이 있어야하며 반음소 모델을 사용하는 방식이 제안되고 있다[2].

어휘 음성 인식에서는 인식 학습 시 나타나지 않는 미등록어의 트라이 폰이 나타나므로 모델 파라미터들의 초기 추정치를 생성할 수 없으므로 음소 데이터에 대한 모델을 구성할 수 없다. 이러한 단점으로 인하여 가우시안 모델의 정확성이 떨어지게 되어 인식의 신뢰도가 저하된다. 이를 개선하기 위하여 본 논문에서는 확률 분포를 이용한 모델 파라미터의 가우시안 모델을 최적화한 방법을 제안하였다. 모델 파라미터들로부터 확률적 분포인 가우시안 모델을 이용하여 모델 파라미터를 최적화고 정확성을 제공한다. 또한 최적화한 모델 파라미터를 음소 단위로 지원하여 데이터 탐색의 효율성을 높이고 신뢰도를 향상시킨다.

이를 개선하기 위하여 주어진 표본 데이터 집합의 분포 밀도를 하나의 확률밀도함수로 모델링하는 방법을 개선한 밀도 추정 방법이며 복수 개의 가우시안 확률 밀도 함수로 데이터의 분포를 모델링하는 방법인 가우시안 모델을 사용하였으며 미리 정의 되지 않은 음소와 추가되어진 음소로부터 인식률이 저하되는 문제를 개선하기 위하여 확률 분포를 이용한 모델 파라미터의 가우시안 모델을 최적화한 방법을 제안하여 신뢰도를 향상을 시켰다.

기존의 어휘 인식에서는 일반적인 벡터 값을 데이터베이스를 이용하여 구하므로 탐색 중에 형성되는 음소를 처리하지 못하는 문제점을 제공하지만, 본 논문에서는 가우시안 확률 밀도 함수로 데이터의 분포를 모델링하는 방법인 가우시안 모델을 사용 음소를 관리 및 제어할 수 있도록 하였다. 또한, 기존에 연구된 어휘 인식 시스템들이 유용하지만 다른 시스템과의 인터페이스를 제공하지 않으며, 일부 단위로만 음소를 관리하므로 포괄적이지 못하여 제한적으로 사용되고 있는 문제점을 처리할 수 있도록 설계하였다.

연속 확률 분포의 공유로부터 가우시안 모델 최적화를 실험한 결과 향상된 신뢰도로 인해 높은 인식 성능을 확인하였으며 미등록어 거절 알고리즘이 기존의 방법들에 비하여 평균 1.7%의 성능 향상을 나타내었다.

제 2장에서는 CHMM 연속 확률분포와 가우시안 모델, 그리고 라이브러리에 대해 설명하고, 제 3장에서는 본 논문에서 제안한 시스템에 대하여 설명한다. 제 4장에서는 제안한 시스템의 실험결과에 대하여 설명하고 제 5장에서 결론을 기술한다.

Ⅱ. 기존 연구

2.1 미등록어

음성 인식 시스템은 인식 대상 단위에 따라 고립 단어 인식 시스템과 연속 음성 인식 시스템으로 구분된다. 고립 단어 인식 시스템은 사용자가 한 단어만을 말하거나 단어와 단어 사이에 구분을 둠으로써 시스템이 단어 단위로 인식을 하는 시스템이다. 연속 음성 인식 시스템은 여러 단어나 문장을 자연스럽게 말을 하고 시스템은 그 결과를 여러 단어나 문장 단위로 보여주는 것이다. 그러나 두 시스템 모두 미리 정해 놓은 특정 인식 대상 단어가 입력될 것이라는 가정 하에 음성 인식 기능을 수행하며 사용자의 실수 또는 고의로 인식 대상 단어 외의 말하면 인식 대상 단어중의 하나로 인식 결과를 보여주므로 다른 단어로 인식해 버리는 문제점이 있다[3].

고립 단어 인식 시스템의 경우는 그런 입력이 들어왔을 때 그 단어에 대해서만 오인식을 하게 될 것이므로 그리 큰 문제가 되지 않는다고 볼 수 있다. 그러나 연속음성인식 시스템의 경우에는 음향학적 처리기가 동작을 하게 되므로 더 큰 문제를 야기시킬 수 있다. 사용자가 발화한 문장 중 한 단어만 음성 인식 대상 단어가 아니라 하더라도 그 단어가 오인식 됨으로 인해서 그 뒤에 발화된 단어에까지 영향을 미쳐 오인식을 높이게 되므로 더 치명적인 결과를 초래하게 된다.

음소 필러 모델을 사용한 핵심어 검출 방식을 이용해서 미등록어를 거절시키는 방법이다. 핵심어 검출 방식은 핵심어 모델과 필러 모델을 사용하는 연결단어 인식 알고리즘을 기반으로 사용하고 있다. 필러 모델들은 핵심어에 해당하지 않는 음성 구간들인 비핵심어들과 비음성, 묵음 또는 배경 잡음 구간들을 표현하는데 사용된다[4].

기존 미등록어 거절 방법에는 가변 어휘 단어 시스템에서 비터비 탐색시 사용되는 네트워크 망을 그림 1과 같이 구성한다. 구성된 네트워크 망에서 인식된 결과는 등록어들과 음소들의 열로 나타나게 된다.

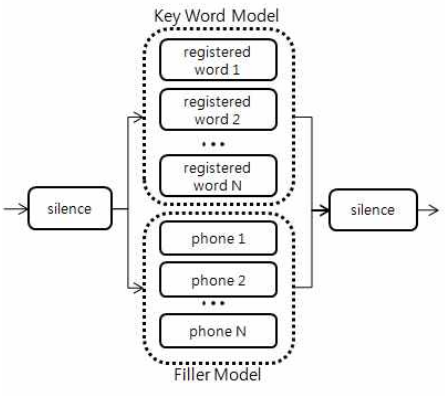

그림 1. 핵심어 모델과 필러 모델 분류 과정
Figure 1. Keyword model and filler model classification process.

즉, "묵음+(등록어 및 음소들의 열)+묵음"과 같은 형태가 된다. 단어 패널티를 잘 조정하면 입력된 음성이 등록어이면 인식된 결과는 "묵음+(등록어 및 약간의 음소들의 열)+묵음"으로 나타나게 되고, 미등록어이면 인식된 결과는 묵음+(등록어 및 다수의 음소들의 열)+묵음" 또는 "묵음+(다수의 음소들의 열)+묵음"으로 나타나게 된다. 이렇게 인식된 결과를 발화 검증 시스템으로 넘기게 되며 가변 어휘 단어 인식 시스템의 단어 패널티와 인식된 결과의 삽입된 음소들의 개수를 이용하여 미등록어를 거절시킬 수 있다. 삽입된 음소들은 필러 모델을 뜻하며 삽입된 음소가 많다는 것은 인식 결과에 핵심어가 없다는 의미이다. 즉 사용자가 미등록어를 발성하게 되면 필러 모델들로 인식됨을 알 수 있다. 또한 삽입된 음소가 미리 정해둔 임계값 이하라도 인식 결과에 등록어가 포함되어 있지 않거나 2개 이상이면 거절시킨다[5].

2.2 신뢰도 정규화

인식된 결과는 음소나 단어로부터 발화되었을 확률에 대한 상대값을 의미하며 신뢰도는 인식 결과에 대해 그 결과가 얼마나 믿을 만한 것인가를 나타내는 척도이다[6].

어떤 O를 관측 세그먼트라 하면 인식 과정에서 O가 입력되었을 때는 두 가지의 가정이 가능하다. O가 실제 세그먼트 k일 것이라는 가정이 가능하며 영가설이라 하고 H_0로 표시한다. 또한 O가 실제 세그먼트가 아닌 다른 유사 발화라 가정이 가능하며 대립가설이라 하고 H_1으로 표현한다. 주어진 테스트 세그먼트 O에 대해 발화 검증 과정은 영가설에 대한 확률과 대립가설에 대한 확률을 비교하여 영가설에 대한 확률이 크면 인식하고 아니면 잘 못된 인식으로 판단한다. 따라서 식 (1)가 만들어 진다[7].

$$P(O|H_0) > P(O|H_1) \quad \cdots\cdots (1)$$

베이시안의 정리는 불확실한 상황에서의 의사 결정 문제를 수리적으로 다룰 때 사용한다. 연역적 추론 방식인 확률을 사용하여 귀납적 추론을 만들어내는 방법이다. 전통적인 확률을 직접적인 확률(direct probability)이라 한다면 베이시안의 정리는 역의 확률(inverse probability)이라 하고 식 (2)과 같이 정리한다.

$$P(H_0|O) = \frac{P(O \cap H_0)}{P(O)} = \frac{P(O \cap H_0)P(H_0)}{\sum_{k=1}^{n} P(O \cap H_k)P(H_k)} \quad \cdots\cdots (2)$$

위의 식 (1)을 베이시안 룰(bayes rule)인 식 (2)에 의해 정리하면 식 (3)과 같이 표현되며

$$\frac{P(H_0|O)P(H_0)}{P(O)} > \frac{P(H_1|O)P(H_1)}{P(O)} \quad \cdots\cdots (3)$$

$P(H_0|O)$는 모델 λ_k에서 O가 관측될 확률이고, $P(H_1|O)$는 다른 모델에서 O가 관측될 확률이다. 식 (3)의 $P(O)$ 결정 규칙에 영향을 미치지 않는 동일한 값이므로 제거하여 정리하면 식 (4)와 같다.

$$\Lambda(O) = \frac{P(H_0|O)}{P(H_1|O)} > \frac{P(H_1)}{P(H_0)} \quad \cdots\cdots (4)$$

$P(H_0|O)$는 모델 λ_k에서 O가 관측될 확률이고, $P(H_1|O)$는 다른 모델에서 O가 관측될 확률이다. H_1을 모델링하기 위해서 각 음소마다 유사한 음소들을 구하여 파라미터로 훈련하고 훈련된 파라미터를 $\lambda_{\bar{k}}$로 표현하고 이를 안티모델이라 한다. 안티모델을 log을 취해서 우도(log-likelihood)로 사용하고 식(5)로 표현한다.

$$LLR_k(O, \lambda_k) = \log P(O|\lambda_k) - \log P(O|\lambda_{\bar{k}}) \quad \cdots\cdots (5)$$

우도 값이 너무 큰 범위에서 나타나지 않도록 정화규하며 시그모이드 함수를 사용하고 최종적인 음소 신뢰도는 식 (6)에 의해서 계산된다[8].

$$f(LLR) = \log \frac{1}{1 + \exp^{(-\alpha \cdot LLR)}} \quad \cdots\cdots (6)$$

로그우도비를 시그모이드 함수를 사용하여 정규화한 식을 표현한다. 신뢰도를 측정하기 위해 식 (6)에 표현한 로그우도비를 사용하여 측정하고 측정된 로그우도비가 높을수록 신뢰도가 높으며 인식 시에 정확한 인식으로 표현된다.

III. 시스템 모델

3.1 가우시안 최적화 모델

가우시안 최적화 모델은 주어진 표본 데이터 집합의 분포 밀도를 단 하나의 확률밀도함수로 모델링하는 방법을 개선한 밀도 추정 방법으로 복수 개의 가우시안 확률밀도함수로 데이터의 분포를 모델링하는 방법이다. 단일한 가우시안으로는 모델링 할 수 없는 복수개의 중심점을 가지는 1차원 데이터와 2차원 환형 데이터에 대하여 견고하게 모델링된다[9].

확률밀도함수는 가우시안 분포뿐 아니라 다른 분포가 될 수도 있다. 가우시안 최적화 밀도는 단지 확률밀도함수를 가우시안 분포로 가정하는 경우이다. 결국 최종적인 전체 확률밀도함수는 M개의 가우시안 확률밀도함수의 선형 결합으로 식(7)과 같이 표현된다.

$$p(x|\theta)=\sum_{i=1}^{M} p(x|\omega_i,\theta_i)P(\omega_i) \quad \cdots\cdots (7)$$

여기서 $p(x|\omega_i,\theta_i)$는 데이터 x에 대하여 ω_i번째 성분 파라미터 θ_i로 이루어진 확률밀도함수를 의미하며, $P(\omega_i)$는 혼합 가중치로 각 확률밀도함수의 상대적인 중요도를 의미한다. 혼합 가중치를 사전확률과 같은 형태로 α_i라고 하면 식(8)과 같은 제약조건이 따른다[10].

$$0 \le \alpha_i \le 1, \text{ and } \sum_{i=1}^{M}\alpha_i = 1 \quad \cdots\cdots (8)$$

확률밀도함수가 가우시안 분포를 따를 경우 θ_i는 식(9)와 같은 파라미터 집합이 된다.

$$\theta_i = (\mu_1, \mu_2, \ldots, \mu_M, \sigma_1, \sigma_2, \ldots, \sigma_M, \alpha_1, \alpha_2, \ldots, \alpha_M,) \quad \cdots\cdots (9)$$

전체 모델을 이루는 각 가우시안 성분은 완전대각 또는 정방형 공분산 행렬의 형태를 가질 수 있다. 또한 혼합 성분의 개수는 학습 데이터 집합의 크기에 따라 조절 가능하다.

가우시안 최적화 모델로 데이터의 분포를 모델링할 경우에 혼합 성분의 개수가 충분히 주어지고 적절한 파리미터 값들만 주어진다면 이론적으로는 어떠한 연속적인 분포도 완벽하게 추정하여 모델링한다[11].

단일 가우시안 출력 확률 밀도 함수를 갖는 3상태 단 음소 모델 초기 집합을 생성하고 훈련한다. 단 음소의 상태 출력 분포는 재 추정하여 훈련된 트라이폰 모델 집합을 초기화하기 위해 복사한다. 동일한 각 음소로부터 유도된 트라이폰들의 각 집합에 대해 대응되는 상태들을 군집화한다. 각 결과 군집에서 대표적인 상태가 선택되고 모든 군집 내의 상태들은 대표 상태로 묶이게 된다. 각 상태의 혼합 요소의 수를 증가시켜 재추정하여 모델의 정밀도를 향상시킨다.

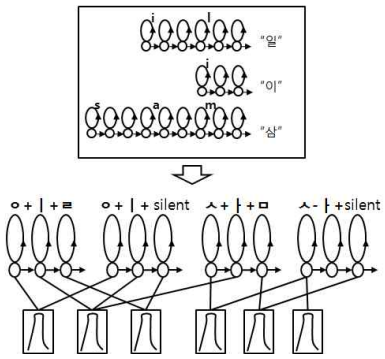

그림 2. 가우시안 최적화 모델
Figure 2. Gaussian optimization model

음성신호로부터 특징벡터들을 추출하여 가우시안 값을 추정하고 가우시안 모델을 만드는 과정을 그림 2에 나타내었다.

3.2 개선된 미등록 거절 알고리즘

음성 인식 시스템은 인식 대상 단위에 따라 고립 단어 인식 시스템과 연속 음성 인식 시스템으로 구분된다. 고립 단어 인식 시스템은 사용자가 한 단어만을 말하거나 단어와 단어 사이에 구분을 둠으로써 시스템이 단어 단위로 인식을 하는 시스템이다. 연속 음성 인식 시스템은 여러 단어나 문장을 자연스럽게 말을 하고 시스템은 그 결과를 여러 단어나 문장 단위로 보여주는 것이다. 그러나 두 시스템 모두 미리 정해 놓은 특정 인식 대상 단어가 입력될 것이라는 가정 하에 음성 인식 기능을 수행하며 사용자의 실수 또는 고의로 인식 대상 단어 외의 말하면 인식 대상 단어중의 하나로 인식 결과를 보여주므로 다른 단어로 인식해 버리는 문제점이 있다[12].

음성인식 기능과 검증기능이 동시에 검색되도록 구성되어진 One-pass 시스템과 인식기의 후처리 방식으로 검증 기능을 구현하는 Two-pass 방식으로 구성된다. Two-pass 방식은 기존 시스템을 그대로 적용하고 검증 과정을 추가한 것으로 구현이 쉽다는 장점을 가지고 있다. 발화 검증 시스템을 설계할 때 미등록어와 잘못 인식된 단어를 잘 선별할 수 있는 검증 모델에 기반한 적정한 신뢰도(confidence measure)를 정의해야 하고, 훈련 데이터에서 검증 오류를 최소화할 수 있도록 검증 모델을 적응시키는 훈련과정을 선택해야 하며 유사도의 변화와 검증 문턱치의 변화, 훈련과 테스트 상태의 변화

에 강해야 한다.

그림 3은 인식과 검증으로 구성된 2단계 발화 검증 시스템의 기본 구조를 나타낸다.

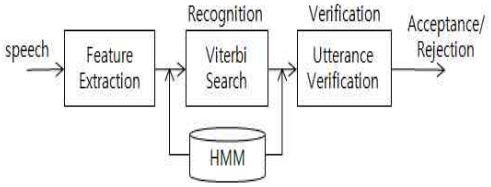

그림 3. 2단계 발화 검증 시스템
Figure 3. 2-step utterance verification system

1단계에서 인식 모델을 사용하여 비터비(viterbi)탐색 알고리즘에 의한 인식과정을 수행한다. 음소 모델들은 ML(Maximum Likelihood)를 이용하여 HMM(Hidden Markov Model)의 파라미터를 최적화한다. 인식 과정을 거치는 동안 각 단어의 발화는 음소 가설로 분할되며, 그 결과를 발화 검증 시스템으로 전달한다. 두 번째 단계인 발화 검증 과정은 인식된 후보 단어의 음소열에 대해 반음소 모델과의 신뢰도를 구하여 그 단어의 신뢰도 값을 결정한다. 이 신뢰도 값이 미리 정해둔 문턱치보다 크면 등록단어로 인식이 되고 아니면 거절된다.

3.3 신뢰도 정규화 처리

음소를 기본 단위로 한 음성 인식 시스템의 성능 향상을 위해서는 각 음소를 다양한 주변 음소 환경 하에서 정확히 모델링하는 것이 중요하며 각 음소를 달리 모델링하는 음소 모델링 기법을 사용한다[13].

신뢰도는 거절기능의 척도로 사용하며 음성 인식기의 출력인 음소 확률과 인식된 음소의 반 음소 확률과의 비로 정의된다. 음소 단위 신뢰도를 구하고 가중 평균을 취하여 단어 단위 신뢰도를 계산하여 사용한다[14].

음소 단위 신뢰도는 출력 음소 확률과 인식된 음소의 반 음소 확률과의 비로 정의 된다.

반 음소 모델의 평균 로그 확률은 $\log p_a$로 식 (10)과 같이 나타낸다. p_a는 반 음소 모델을 나타내고 M은 반 음소 모델의 수를 나타낸다.

$$\log p_a = \frac{1}{M}\sum_{i=0}^{M-1}\log p_{a_i} \quad \cdots\cdots (10)$$

단어 구성 음소 모델의 평균 로그 확률은 $\log p_w$로 식 (11)과 같이 나타낸다. p_w는 단어 구성 음소 모델을 나타내고 N은 단어 구성 음소 모델의 수를 나타낸다.

$$\log p_w = \frac{1}{N}\sum_{i=0}^{N-1}\log p_{w_i} \quad \cdots\cdots (11)$$

각 음소의 반 음소 모델과의 유사도 비는 Lm으로 식 (12)와 같이 나타낸다.

$$Lm = \frac{\log p_p - \log p_a}{|\log p_p|} \quad \cdots\cdots (12)$$

식 (12)에 음의 값을 갖는 가중치를 포함하여 음소 단위 신뢰도를 표현하면 식 (13)과 같이 나타낸다.

$$LM = \frac{1}{f_{Lm}} log\left(\frac{\sum_{p=o}^{n_p-1}\exp(f_{Lm} \cdot Lm)}{n_p}\right) \quad \cdots\cdots (13)$$

f_{cm}은 음의 값을 갖는 가중치를 나타내고, n_p는 단어 구성의 음소의 개수를 나타낸다. log-유사도 값의 크기로 정규화하여 프레임 길이의 정규화보다 일관적인 음소 단위의 검증 성능을 나타낸다.

일반적으로 반 음소는 전체 음소 집합에서 인식된 음소를 제외한 나머지 음소를 사용하게 되며 반 음소의 확률 계산을 위하여 음소 HMM(Hidden Markov Model)을 이용한다.

핵심어 검출 방식은 핵심어 모델과 필러 모델을 사용하는 연결단어 인식 알고리즘을 기반으로 사용하고 있다. 필러 모델들은 핵심어에 해당하지 않는 음성 구간들인 비핵심어들과 비음성, 묵음 또는 배경 잡음 구간들을 표현하는데 사용된다.

유효한 핵심어를 거부하는 오거부의 경우 각 음소 단위 신뢰도들의 통계적 특성의 불안정으로 인하여 발생한다. 이를 해결하기 위하여 각 음소 단위 신뢰도들의 평균과 표준편차를 이용하는 정규화를 사용한다.

통계적 분포가 불안정한 현상을 해결하기 위해서 정규화를 사용하게 되며 사전에 계산된 음소 단위 신뢰도의 평균과 표준편차를 이용하여 각 음소 단위 신뢰도들의 표준 정규분포를 정규화하여 식(14)와 같이 나타낸다.

$$nLm = \frac{Lm - Phone_q \cdot mean}{Phone_q \cdot sd} + \alpha \quad \cdots\cdots (14)$$

Lm은 기존의 음소 단위 신뢰도를 나타내고, nLm는 정규화된 음소 단위 로그 확률을 나타낸다. $Phone_q.mean$은 음소 단위의 신뢰도 평균을 나타내며 $Phone_q.sd$은 음소 단위의 신뢰도 표준편차를 나타낸다. α는 음소 단위 신뢰도의 정규화에 사용되는 가중치이며 각 음소 단위 신뢰도가 양의 영역에 속할 수 있도록 임계치로 사용한다. 식 (14)에 음의 값을 갖는 가중치인 f_{nLm}을 추가하여 계산하면 식 (15)와 같이 나타낸다.

$$NLM = \frac{1}{f_{nLm}} log\left(\frac{\sum_{p=0}^{n_p-1} \exp(f_{nLm} \cdot nLm)}{n_p}\right) \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (15)$$

각 음소 단위 신뢰도들의 표준편차의 차이가 큰 것을 임계치를 적용하여 신뢰도의 표준편차가 1의 값을 갖도록 유도하여 보다 안정된 신뢰도 특성을 나타내며 가중치 α에 의해 음소 단위 신뢰도를 항상 양의 값으로 안정시킬 수 있었다.

안정된 음소 단위 신뢰도는 단어 단위 신뢰도에 직접적인 영향을 주어 유효한 핵심어가 불안정한 신뢰도에 의해 거부되는 것을 막아준다.

IV. 실험결과 및 분석

본 논문에서 제안한 미등록어 거절 알고리즘에서 가우시안 모델 최적화를 이용한 신뢰도 정규화 시스템 모델을 구성하여 인식 실험을 수행하였다.

본 실험에서는 다양한 가우시안 모델을 포함하고 있는 PBW 445DB를 이용하였다. 어휘수가 총 445개로 구성되어 있으며 1명이 2회 발성한 것을 1개의 set로 구성하였다. 이러한 set이 남성음이 5set, 여성음이 5set으로 모두 10set으로 구성하였다[15].

실내 환경과 잡음 환경에서 이동기기에 내장되어 있는 내장형 마이크로폰을 사용하여 16kHz Mono로 녹음 하였고, 16bit PCM 양자화를 사용하였다.

녹음된 데이터는 인식기 학습을 위해 MFCC 특성 추출 방법을 사용하였고 인식기는 SITEC에서 개발한 ECHOS[16]를 이용하였다.

미등록어 거절의 성능은 다음과 같은 항목을 기준으로 평가하였다[17].

1) 등록어

① CA(Correctly Accepted for Keyword) 인식 대상 등록어를 제대로 accept한 경우의 확률

② FAI(False Accepted In-Grammar Word, Keyword) 인식 대상 등록어로 accept는 했지만 잘 못 인식한 경우의 확률

③ FR(False Rejected for Keyword) 인식 대상 등록어를 말했는데 reject한 경우의 확률

④ CA + FAI + FR = 100%

2) 미등록어

① CR(Correctly Rejected for OOV) 미등록어에 대해 reject한 경우의 확률

② FAO(False Accepted Out-of-Grammar Word, OOV) 미등록어인데 accept한 경우의 확률

③ CR + FAO = 100%

표 1과 표 2는 단어 패널티와 삽입된 음소들의 개수에 따른 성능을 보여준다.

표 1. 발화 검증의 성능(단어 패널티 1=-30.0일 때)
Table 1. Performance of existing utterance verification method(word penalty 1=-30.0)

데이터 종류 / 삽입된 음소개수	시험용 데이터				
	CA	CR	FAI	FAO	FR
3개이상 거절	82.18	79.62	3.15	25.15	12.48
2개이상 거절	81.54	85.61	2.67	19.72	14.52
1개이상 거절	69.72	96.10	1.68	9.58	29.21
데이터 종류 / 삽입된 음소개수	평가용 데이터				
	CA	CR	FAI	FAO	FR
3개이상 거절	89.74	73.42	2.67	26.13	8.50
2개이상 거절	87.62	83.82	2.64	18.19	9.64
1개이상 거절	79.51	90.25	1.35	9.56	20.12

표 2. 발화 검증의 성능(단어 패널티 1=-40.0일 때)
Table 2. Performance of existing utterance verification method(word penalty 1=-40.0)

데이터 종류 / 삽입된 음소개수	시험용 데이터				
	CA	CR	FAI	FAO	FR
3개이상 거절	91.03	56.78	3.91	47.92	6.81
2개이상 거절	87.46	68.12	3.61	36.41	7.15
1개이상 거절	79.78	79.65	2.68	18.67	18.12
데이터 종류 / 삽입된 음소개수	평가용 데이터				
	CA	CR	FAI	FAO	FR
3개이상 거절	95.04	55.91	3.01	49.31	3.15
2개이상 거절	90.35	68.35	2.21	34.25	3.94
1개이상 거절	83.10	70.54	1.95	18.64	9.11

표 1에서는 단어 패널티 1과 임계치 -30.0일 때의 발화 검증의 성능을 나타냈으며 표 2에서는 단어 패널티 1과 임계치 -40.0일 때의 발화 검증의 성능을 나타내었으며, 미등록어 거절 알고리즘 후처리 시스템에 대한 인식율과 거절율은 식 (13)과 식 (15)을 이용하여 표 3에 나타내었다.

표 3은 기존의 에러 패턴 학습을 이용한 방법[18][19]인 error pattern과 의미기반의 방법[20]인 semantic 그리고 본 논문의 제안 방법인 가우시안 모델 최적화의 결과를 나타내었다.

표 3. 오류 보정율 비교
Table 3. Comparison of error correction

오류 보정	인식율(%)	거절율(%)
error pattern	83.4	10.5
semantic	82.9	9.7
Gaussian optimization	84.1	11.8

에러 패턴 학습을 이용한 미등록어 거절의 경우 10.5%, 의미 기반의 미등록어 거절의 경우 9.7%의 미등록어 거절율을 보였으며, 본 논문에서 제안한 가우시안 최적화 모델을 이용한 미등록어 거절은 11.8%로 미등록어 거절 알고리즘이 기존의 방법들에 비하여 평균 1.7%의 성능 향상을 나타내었다.

V. 결론

본 논문에서는 미등록어 거절 알고리즘에서 가우시안 모델 최적화를 이용한 신뢰도 정규화 향상을 위한 시스템을 제안하였다.

가우시안 확률 밀도 함수로 데이터의 분포를 모델링하는 방법인 가우시안 최적화 모델을 사용하여 확률 분포를 이용한 모델 파라미터를 구성하고 미등록어 거절을 통해 신뢰도 정규화를 향상 시켰다.

기존의 어휘 인식에서는 데이터베이스를 벡터 값을 이용하여 패턴 패칭하므로 탐색 중에 형성되어진 음소를 인식하지 못하여 인식률 저하의 원인이 되었다. 하지만 본 논문에서는 가우시안 확률 밀도 함수로 데이터의 분포를 모델링하는 방법인 가우시안 최적화 모델을 사용하여 음소를 관리 및 제어할 수 있도록 하였으며 일부 단위로만 음소를 관리하므로 포괄적이지 못하고 제한적으로 사용되는 문제점을 해결할 수 있었다.

참고문헌

[1] 방기덕, 강철호, “가변 신뢰도 문턱치를 사용한 미등록어 거절 알고리즘에 대한 연구,” 한국멀티미디어학회논문지, 제11권, 제11호, 1471-1479쪽, 2008년 11월.

[2] 김우성, 구명완, “반음소 모델링을 이용한 거절기능에 관한 연구,” 한국음향학회지, 제1 8권, 제 3호, 3-9쪽, 1999년.

[3] 문광식, 김회린, 정재호, 이영직, “가변어휘 단어 인식에서의 미등록어 거절 알고리즘의 성능비교,” 신호처리합동학술대회논문집, 제12권, 제1호, 305-308쪽, 1999년 10월.

[4] 안찬식, 오상엽, “공유모델 인식 성능 향상을 위한 효율적인 연속 어휘 군집화 모델링,” 한국컴퓨터정보학회지, 제15권, 제 1호, 177-183쪽, 2010년 1월.

[5] 안찬식, 오상엽, “MLHF 모델을 적용한 어휘 인식 탐색 최적화 시스템”, 한국컴퓨터정보학회지, 제 14권, 제 10호, 217-223쪽, 2009년 10월.

[6] 김용현, 정민화, “에러패턴 학습과 후처리 모듈을 이용한 연속 음성 인식의 성능향상,” Proc. KISS Spring Semiannual Conf. 제 27권, 제 1호, 441-443쪽, 2000년 4월.

[7] A. S. Manos and V. W. Zue, “A study on out-of-vocabulary word modeling for a segment-based keyword spotting system,” Master Thesis, MIT, 1996.

[8] 김동주, 김한우, “문맥가중치가 반영된 문장 유사도 척도,” 전자공학회 논문지, 제43권, 제6호, 496-504쪽, 2006년.

[9] L. R. Bahl, P. V. deSouza, P. S. Gopalakrishnan, D. Nahamoo, and M. Picheny, “A Fast Match for Continuous Speech Recognition Using Allophonic Models,” InProc. IEEE ICASSP-92, Vol.1, pp.17-21, 1992.

[10] L. R. Rabiner, B. H. Juang, “Fundamentals of speech recognition,” Prentice Hall, 1993.

[11] T. Jitsuhiro, S. Takatoshi, and K. Aikawa, “Rejection of out-of-vocabulary words using phoneme confidence likelihood,” ICASSP, pp.217-220, 1998.

[12] 이경록, 김철, 김진영, 최승호, 최승호, “정규화 신뢰도를 이용한 핵심어 검출 성능향상,” 한국음향학회지, 제

21권, 제 4호, 380-386쪽, 2002년 5월
[13] 김동주, 김한우, “문맥가중치가 반영된 문장 유사도 척도,” 대한전자공학회논문지, 제 43권, 제 6호, 496-504쪽, 2006년.
[14] 김상운, 신성효, “ML/MMSE를 이용한 HMM-Net 분류기의 학습에 대한 실험적 고찰,” 대한전자공학회논문지C, 제 36C권, 제 6호, 44-51쪽, 1999년 6월.
[15] S. Young, D. Kershaw, J. Odell, D. Ollason, Valtcher, P. Woodland, “The HTK Book,” Cambridge University Engineering Department, 2002.
[16] 권석봉, 윤성락, 장규철, 김용래, 김봉완, 김회린, 유창동, 이용주, 권오욱, “한국어 음성인식 플랫폼(ECHOS)의 개선 및 평가,” 대한음성학회지:말소리, 제59호, 53-68쪽, 2006년 9월.
[17] 최승호, “정규화 신뢰도 기반 가변 어휘 고립 단어 인식기의 거절기능 성능 분석,” 한국음향학회지, 제25권, 제 2호, 96-100쪽, 2006년 2월.
[18] K. Demuynck, J. Duchateau, and D. Van Comp ernolle, “A static lexicon network representation for cross-word context dependent phones,” In Proc. EUROSPEECH, Vol.1, pp.143-146, 1997.
[19] 김기태, 문광식, 김회린, 이영직, 정재호, “가변어휘 단어 인식에서의 미등록어 거절 알고리즘 성능 비교,” 한국음향학회지, 제 20권, 제 2호, 27-34쪽, 2001년 2월.
[20] M. W. Jeong, B. C. Kim, and G. G. Lee, “Semantic-oriented error correction for spoken query processing,” Proc. IEEE Workshop on ASRU, pp.156-161, Nov, 2003.

저 자 소 개

안 찬 식

2002 : 광운대학교 컴퓨터공학과 공학석사.
2004 : 광운대학교 컴퓨터공학과 박사수료.
관심분야 : 음성인식, 분산처리, 음성/음향 신호처리

오 상 엽

1999 : 광운대학교 전자계산학과 이학박사.
현 재 : 경원대학교 IT대학 컴퓨터소프트웨어 교수
관심분야 : 소프트웨어공학, 버전관리, 소프트웨어재사용, 형상관리, 객체지향, 음성인식, 분산처리, 음성/음향 신호처리