

BMS 알고리즘을 이용한 거절기능 성능 향상

김 병 돈*, 이 경 록**, 김 진 영*, 최 승 호***

*전남대학교 전자공학과, **남부대학교 컴퓨터전자공학과, ***동신대학교 정보통신공학부

Improvement of Confidence Measure Performance using Background Model Set Algorithm

*ByoungDon Kim, **KyongRok Lee, *JinYoung Kim, ***SeungHo Choi

*School of Electronics, Chonnam National University,

**School of Computer and Electronic Engineering, Nambu University,

***School of Electronics Computer and Information Engineering, Dongshin University

E-mail : doni96@dsp.chonnam.ac.kr

Abstract

In this paper, we proposed Background Model Set algorithm for the speaker verification to improve the shortcoming of calculating process in conventional confidence measure(CM). CM is to display relative likelihood between recognized models and unrecognized models. Unrecognized models is known as antiphone models. Calculate probability and standard deviation using all phonemes at process that compose antiphone model. At this process, antiphone CM brought bad result. Also, recognition time increases. In order problem, we studied about method to reconstitute average and standard deviation taking BMS algorithm using antiphoneme that near phoneme of CM calculation.

I. 서론

연속으로 발음된 음성 중에서 핵심어 출현을 인식하는 핵심어 검출기는 연속음성인식기에 비해서 검출부의 구조가 간단하여 연산비용이 저렴하다. 이 때문에 연속 음성인식기에 비해서 우월한 특성을 갖는다.

후처리 성능향상에 사용된 핵심어 검출기는 입력음성의 핵심어부와 비핵심어부를 독립적으로 구축하여 변별력을 높일수 있는 핵심어 모델과 필러모델을 사용

하는 일반적인 핵심어 검출 시스템을 이용하였다[1].

신뢰도 측정은 오검출을 처리하는 후처리 방법 중 음소 모델과 반음소 모델의 유사도를 이용하는 신뢰도를 측정한다. 이 과정에서 사용되는 반음소 모델은 자기 음소를 제외한 유사음소 집합을 말한다. 이러한 유사음소의 집합은 정의된 유사음소가 많을수록 반음소 모델로서의 역할을 충실히 수행하지만 유사음소 수의 증가는 그에 상응하는 훈련 데이터베이스의 증가를 요구하기 때문에 모든 음소의 계산이 필수적이다. 이에 따라서 음소수의 증가는 검출 시스템의 계산 속도를 증가시켜 전체적인 시스템의 메모리의 과다한 소요와 인식시간의 증대를 가져오게 되었다.

위의 단점을 해결하기 위해서 본 논문에서는 인식 향상과 계산시간 단축을 위해서 화자 인증에서 사용되는 BMS(Background Model Set) 알고리즘을 이용한 새로운 신뢰도를 제안한다.

II. 인식 시스템

2.1. 핵심어 검출기

핵심어 검출 시스템은 크게 전처리부와 핵심어 검출부, 후처리부로 구성되어 있다. 전처리부에서는 음성구간을 검출하여 특징 파라미터를 추출하고 이를 바탕으로 핵심어 검출부에서는 핵심어 모델과 필러 모델로 구성된 핵심어 검출 네트워크를 사용하여 핵심어를 검출한다. 핵심어 검출 네트워크의 각 모델은 GMM

(Gaussian Mixture Model)을 이용하여 모델링되었다. 후처리부는 음소모델과 반음소 모델의 상대적 유사도를 통해서 발화여부를 검증한다[2].

2.2 데이터베이스 구성

훈련 데이터베이스는 인식 모델과 필러모델을 위한 음성부 훈련 데이터베이스와 비음성부 훈련 데이터베이스로 구성된다. 핵심어 검출기는 트라이폰을 분석한 결과 1,031 문장으로 구성하였다. 평가용 데이터베이스는 핵심어 검출기의 성능평가를 위해 남성화자 6명의 496문장에서 최대 3개까지의 핵심어를 출현하도록 허용하여 이를 평가용 데이터베이스로 사용하였다. 데이터베이스 녹음 환경은 조용한 사무실 환경이고 8kHz, 16bit로 A/D 변환하였다.

2.3 검출 실험 결과 평가 기준

인식 성능 평가를 위하여 다음과 같은 평가기준을 선정하였다. 먼저 미검출율은 테스트 문장에서 출현한 핵심어를 검출기가 제대로 검출하지 못한 경우이고, FAR로는 각 핵심어당 FA의 출현횟수를 평가시간으로 정규화한 FA/KW/HR을 사용하였다.

다음의 식은 실험에 사용된 미검출율과 FA/KW/HR을 식으로 정의한 것이다.

$$MDR = \frac{\text{missed detection 수}}{\text{핵심어 총 출현횟수}} \quad (1)$$

$$FA/KW/HR = \frac{\text{false alarm 수}}{\text{총 핵심어 수} \times \text{평가 DB 녹음시간}} \quad (2)$$

III. 후처리 시스템

3.1 기존의 신뢰도 기반 후처리(RLJ-CM, NCM)

본 연구에서는 참고논문 [3]에서 제안된 RLJ-CM과 참고논문 [4]에서 제안된 NCM(normalized CM)을 이용하여 신뢰도를 측정하였다. RLJ-CM의 경우 각 음소 단위의 신뢰도를 음소모델과 반음소 모델의 유사도를 이용하여 계산하고 이를 기준으로 단어단위로 확장하여 검출어의 신뢰도를 계산하였다[3].

$$\log pr_a = \frac{1}{M} \sum_{i=0}^{M-1} \log pr_{a_i} \quad (3)$$

$$cm_p = \frac{\log pr_p - \log pr_a}{|\log pr_p|} \quad (4)$$

$$CM = \frac{1}{f_{cm}} \log \left(\frac{\sum_{p=0}^{n_p-1} \exp(f_{cm} \cdot cm_p)}{n_p} \right) \quad (5)$$

위의 수식에서 $\log pr_a$ 는 반음소 모델의 평균 로그 확률이고, $\log pr_p$ 는 핵심어 구성 음소모델의 로그확률이다. n_p 는 핵심어의 구성 음소수이고, M 은 반음소 모델의 수이다. cm_p 는 음소단위의 신뢰도이고 f_{cm} 은 음의 가중치이다.

정규화 신뢰도(normalized CM)은 RLJ-CM의 통계적 분포가 불안정한 현상을 해결하기 위해서 정규화라는 방법을 선택하였다. 먼저 각 음소 단위 신뢰도들이 정규분포를 가진다고 가정하고 평균과 표준편차를 조사하여 사전에 계산된 음소 단위 신뢰도의 평균과 표준편차를 이용하여 각 음소 단위 신뢰도들을 표준 정규분포로 정규화하였다[4].

$$ncm_p = \frac{cm_p - TriP_q \cdot mean}{TriP_q \cdot sd} + a \quad (6)$$

$$NCM = \frac{1}{f_{ncm}} \log \left(\frac{\sum_{p=0}^{n_p-1} \exp(f_{ncm} \cdot ncm_p)}{n_p} \right) \quad (7)$$

f_{ncm} 은 음소 구성 신뢰도들의 가중치이다. cm_p 는 기존의 음소단위 신뢰도이고, ncm_p 는 정규화된 음소단위 로그확률이다. a 는 음소단위 신뢰도의 정규화에 사용되는 가중치이다.

3.2 기존의 신뢰도 기반 반음소 모델 선정

신뢰도란 인식된 모델과 인식되지 않은 모델간의 상대적인 유사도이므로 각 모델에 대한 가장 혼돈되기 쉬운 유사한 것을 찾아 그것에 따른 모델을 구성하는 것을 반음소 모델이라고 한다.

핵심어 검출기에서 반음소 모델을 구성하기 위해서 사용된 방법은 bhattacharyya 거리 측정을 통하여 음소간의 거리를 계산하였다. 이 방법은 GMM 기반 음소모델에 대해 평균과 공분산 매트릭스만을 가지고 음소들간의 거리를 계산하는 방법으로 별도의 사전계산과 준비과정을 필요로 하지 않는다. 이 측정 방법은 계산이 매우 단순하고, 정확히 거리를 계산하는 것보다 에러에 대한 경계를 제공하여 유연성을 갖는다.

식(8)은 Bhattacharyya 거리를 구하는 공식이며 M 은 음소의 평균, Σ 는 음소의 공분산 매트릭스를 나타낸다 [5].

$$D_{bhat} = \frac{1}{8} (M_1 - M_2)^T \left[\frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (M_2 - M_1) + \frac{1}{2} \ln \frac{|\Sigma_1 - \Sigma_2|}{\sqrt{|\Sigma_1| |\Sigma_2|}} \quad (8)$$

3.3 Background Model Set

BMS에서의 평균 로그 유사도는 Background Model에 의해 계산되어진다. 식(9)는 평균 로그 유사도 계산식을 나타낸 것이다.

$$\iota(X|\lambda_c) = \log \left[\frac{1}{N_B} \sum_{b=1}^{N_B} \exp \iota(X|\lambda_b) \right] \quad (9)$$

위의 식에서 N_B 는 mixture 개수이다.

Background model Set의 각각의 클라이언트들은 그 클라이언트 모델 셋으로부터 선택되어진다. 트레이닝 데이터를 사용함으로써 클라이언트 모델들 사이의 거리를 계산할 수 있게 된다. 모델 λ_D 와 λ_E 는 모델 구성시에 사용되었던 특징 벡터 셋인 X_D 와 X_E 에 대응된다. 셋의 거리를 구하는 식은 식(10)과 같다.

$$d(\lambda_D, \lambda_E) = [\iota(X_D|\lambda_D) - \iota(X_D|\lambda_E)] + [\iota(X_E|\lambda_E) - \iota(X_E|\lambda_D)] \quad (10)$$

식(10)에서 거리 계산식은 모델 λ_D 와 λ_E 둘간의 유사함 또는 둘간의 접근 상태를 계산할 수 있다. BMS는 각 모델들이 클라이언트 모델과 얼마나 근접해 있는지 혹은 얼마나 떨어져 있는가의 정보를 가지고 있다.

이 정보를 통해서 우리는 가장 클라이언트로부터 근접해 있는 모델들을 사용하고 클라이언트로부터 가장 멀리 떨어져 있는, 즉 취약함을 보이는 모델을 제거할 수 있다.

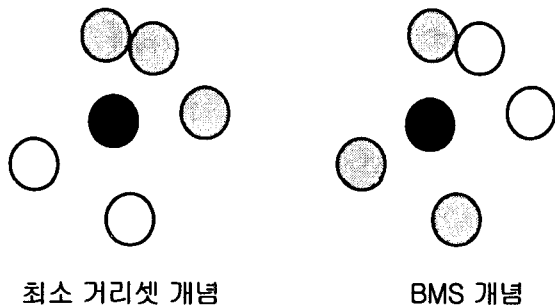


그림 1. 최소거리셋과 BMS 비교

클라이언트 모델이 λ_c 로 주어지고, 모델 셋 B 에서 가장 근접해 있는 모델을 λ_b 라고 하면 BMS 수식은 아래와 같다.

$$\lambda_i = \arg \max_{\lambda_i} \left[\frac{1}{N_{B_{class}}} \sum_{\lambda_j \in B_{class}} \frac{d(\lambda_b, \lambda_j)}{d(\lambda_c, \lambda_j)} \right] \quad (11)$$

λ_i 는 모델셋 B 에서 λ_c 에는 가장 가깝고 λ_b 와는 가장 멀리 떨어진 client model argument 이고 $d(\lambda_c, \lambda_j)$ 는 λ_c 와 λ_j 의 거리를 나타낸다.

BMS argument를 계산하기 위해서는 3단계의 과정이 필요하다.

첫째, 모델셋 B 에서 가장 가까운 곳에 있는 음소모델의 거리를 측정한다.

둘째, 식(11)을 이용하여 argument λ_i 를 계산한다.

셋째, 위의 두 과정을 $N_{B_{class}} = N_B / 2$ 일때까지 반복한다[6].

이와 같은 방법을 통해서 반응소 셋을 결정한다. 이 과정에서 BMS 알고리즘에서 필요한 거리함수를 계산하기 위하여 Bhattacharyya 거리를 사용한다.

3.4 BMS를 이용한 반응소 모델 선정

BMS 알고리즘을 이용한 반응소 모델 선정 과정에서 BMS 알고리즘 적용을 두가지 방법을 통하여 실험을 실시하였다. 음소 모델간의 거리의 임계값을 사용하여 반응소 모델을 선정하는 방법과 반응소 개수에 따른 반응소 모델을 선정하였다. 각각의 방법을 통해서 전체의 반응소셋을 사용하였을 때보다 BMS 알고리즘을 이용하여 반응소셋의 개수를 줄였을 때와의 신뢰도의 성능 향상을 비교하였다.

IV. 실험 결과

실험 결과는 6명의 결과의 평균한 값을 데이터로 나타낸 것이다. 그림 2는 핵심어 검출기에서 기존의 CM과 BMS-CM과의 비교를 나타낸 것이다. MDR 22% 부근에서 BMS 알고리즘을 적용했을 때 RLJ-CM은 0.322 FA/KW/HR(false alarm/keyword/hour)로서 BMS 적용전의 반응소셋에 비해서 성능이 29% 향상됨을 알 수 있다.

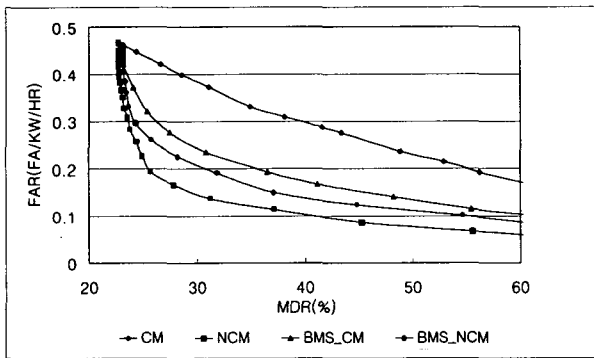


그림 2. 핵심어 검출기에서 실험 결과

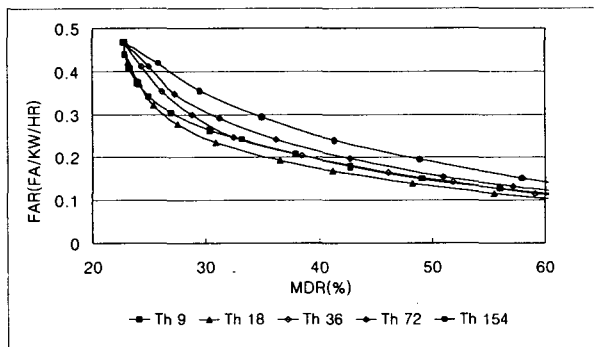


그림 3. 임계값 변화에 따른 실험 결과

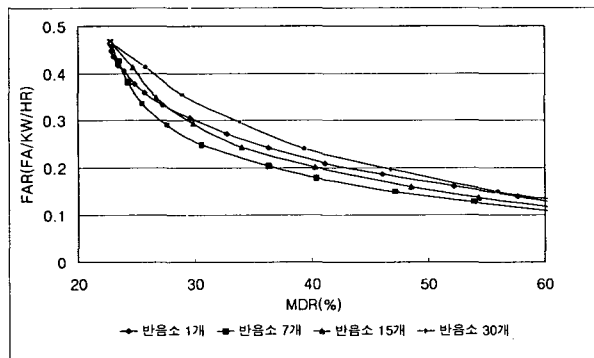


그림 4. 반응소 개수에 따른 실험 결과

하지만 NCM에서는 MDR 25%에서 0.385 FA/KW/HR (false alarm/keyword/hour)로서 BMS 적용시에 기존의 NCM보다 뒤떨어지는 결과를 보인다.

그림 3과 4에서 보는 바와 같이 값이 커질수록 인식 성능이 향상되지만 특정 범위를 넘게 되면 인식 성능이 저하됨을 알 수 있었다. 실험 결과 임계값 18을 주었을 때와 반응소 개수가 7개일 때 인식 성능이 가장 좋음을 알 수 있다. 표 1은 임계값에 따라 사용된 음소의 평균 개수를 나타낸 것이다.

표 1. 임계값에 따라 사용된 반응소 평균 개수 비교

임계값	3	5	9	18	36	72	154
반응소 평균 개수	1.43	2.13	3.88	7.04	12.12	20.37	35.24

표를 통해서 임계값 18을 사용했을 때 모든 음소가 사용하는 반응소의 평균 개수가 7.04 임을 알 수 있다. 모든 반응소셋을 사용했을 때보다 전체 반응소셋의 14%만을 사용했을 때 가장 좋은 성능을 보임을 알 수 있었다. 이 결과를 통해서 우리는 인식 음소를 제외한 모든 음소를 반응소셋으로 사용했을 때의 인식 성능에 비해 BMS 알고리즘을 이용하여 반응소셋의 생성에 제한을 두었을 때의 신뢰도 성능이 향상됨을 알 수 있었다.

V. 결론

본 논문에서는 기존에 핵심어 검출기 후처리부로 사용되고 있던 RLJ-CM에 BMS 모델을 이용한 신뢰도 개선 방안을 제안하였다. 반응소 모델셋에 따라서 신뢰도가 변하였고, 기존의 RLJ-CM에서의 취약점인 반응소 모델 선정 과정에서 음소모델의 개수의 증가를 BMS 알고리즘을 이용하여 대상 음소와 가까운 음소만을 선택함으로써 FA 발생시에 기존의 CM에 비해 보다 견인한 성능을 보고 또한 음소의 개수를 제한하게 되어 인식 속도의 증가를 가져올 수 있다.

참고문헌

- [1] P. Jeanreanud, K. Ng, M. Siu, etc. "Phonetic-based word spotter : Various configuration and application to event spotting", Proc. ESCA eurospeech, 1993.
- [2] S. J. Young, N. H. Russell, J. H. S. Thornton, "Token Passing : a simple conceptual model for connected speech recognition systems", Cambridge University Engineering Department, 1989.
- [3] M. G. Rahim, C. -H. Juang and W. Chou, "Discriminative utterance verification using minimum string verification error (MSVE) training", Proc. of ICASSP 96, 3585-3588, 1996.
- [4] 이경록, 서현철, 최승호, 최승호, 김진영, "하이브리드 신뢰도를 이용한 제한영역 핵심어 검출 성능향상", 한국음향학회지 제 21권 제 7호, pp.632-640, 22002.
- [5] Brain Mak, Etienne Barnard, "Phone clustering using the bhattacharyya distance", 2000
- [6] Conrad Sanderson and Kuldip K. Paliwal, "Likelihood normalization for face authentication in variable recording conditions", Proc. ICIP 2002.