

정규화 신뢰도를 이용한 핵심어 검출 성능향상

Improvement of Keyword Spotting Performance Using Normalized Confidence Measure

김 철*, 이 경 록**, 김 진 영**, 최 승 호***, 최 승 호****
(Cheol Kim*, Kyong Rok Lee**, Jin Young Kim**, Seung Ho Choi***, Seung Ho Choi****)

*KT, **전남대학교 전자공학과, RRC HECS,

삼성종합기술연구원, *동신대학교 정보통신공학과

(접수일자: 2001년 12월 20일; 채택일자: 2002년 4월 26일)

Rahim의 논문 (M. G. Rahim, et al., PROC. of ICASSP96, 1996)과 같은 기존의 후처리 방법은 음소 모델과 반모델 (anti-model)의 유사도를 이용하여 음소 단위 신뢰도를 계산하고, 이들의 평균을 단어 단위 신뢰도로 정의한다. 그런데 음소단위의 신뢰도가 동일한 확률밀도함수를 갖는 것이 아니기 때문에 특정단어의 경우 계산된 신뢰도는 대체로 낮은 값을 갖는다. 이를 극복하기 위한 방법으로서, 본 논문에서는 기존의 신뢰도를 통계적으로 정규화한 신뢰도를 제안한다. 즉 음소단위의 신뢰도가 가우시안 분포를 갖는다고 가정된 후 트라이 폰 (tri-phone) 단위로 정규화하여 동일한 정규분포를 갖도록 한다. 본 논문에서는 제안된 방법의 검증을 위하여 문맥중속 핵심어 모델과 문맥독립 필터 모델을 이용한 일반적인 핵심어 검출기를 사용하였다. 실험결과 제안된 정규화 신뢰도 (NCM: Normalized Confidence Measure)가 불검출율 (MDR: Missed Detection Rate) 8% 정도에서 오검출율 (FAR: false alarm rate)을 0.44에서 0.33 FA/KW/HR (false alarm/keyword/hour)로 저하시켰다. 이것은 오검출율에서 성능이 25% 향상된 것이다.

핵심용어: 정규화 신뢰도, 후처리, 핵심어 검출

투고분야: 음성처리 분야 (2.5)

Conventional post-processing as like confidence measure (CM) proposed by Rahim calculates phones' CM using the likelihood between phoneme model and anti-model, and then word's CM is obtained by averaging phone-level CMs[1]. In conventional method, CMs of some specific keywords are very low and they are usually rejected. The reason is that statistics of phone-level CMs are not consistent. In other words, phone-level CMs have different probability density functions (pdf) for each phone, especially tri-phone. To overcome this problem, in this paper, we propose normalized confidence measure. Our approach is to transform CM pdf of each tri-phone to the same pdf under the assumption that CM pdfs are Gaussian. For evaluating our method we use common keyword spotting system. In that system context-dependent HMM models are used for modeling keyword utterance and context-independent HMM models are applied to non-keyword utterance. The experiment results show that the proposed NCM reduced FAR (false alarm rate) from 0.44 to 0.33 FA/KW/HR (false alarm/keyword/hour) when MDR is about 8%. It achieves 25% improvement of FAR.

Keywords: Normalized confidence measure, Postprocessing, Keyword spotting

ASK subject classification: Speech signal processing (2.5)

I. 서론

요즘 기본적인 의사전달 수단인 음성을 MMI (man machine interface)로 사용하기 위한 음성 처리 기술에 대한 연구가 활발히 진행되고 있다. 이러한 음성 처리 기술 중 하나인 핵심어 검출은 연속적인 입력음성을 분석하여 중요도가 높은 핵심단어만을 검출한다[2].

본 연구의 목적은 핵심어 검출기의 후처리로서 사용되는 기존의 신뢰도 기반 발화검증의 문제점을 분석하고 이를 개선하는 것이다. 본 연구에서 사용된 핵심어 검출기는 대학 교내 전화번호 안내 서비스 구현을 위하여 개발된 것으로서, 핵심어 모델과 필러 모델을 토큰 패싱 알고리즘 (token passing algorithm)으로 결합시키고 이를 이용하여 핵심어를 검출한다. 핵심어 검출은 연속입력에서 핵심어가 발생되는 구간만을 검출하는 시스템의 특성상 비핵심어 구간에서 핵심어를 검출하거나, 핵심어를 오인식하는 오검출 (FA: false alarm)이 발생한다. 이를 제어하는 것을 후처리라 하며 핵심어 검출기의 성능에 큰 영향을 미친다.

본 논문에서는 오검출을 처리하는 기존의 후처리 방법 중 반음소 모델을 이용한 신뢰도 기반 후처리를 사용하였다. 이때 핵심어의 단어 단위 신뢰도는 음소 단위의 신뢰도를 확장하여 정의되고 이것이 문턱치보다 낮으면 거부된다[3]. 그런데 바르게 검출된 핵심어 중 신뢰도가 낮은 값을 가져서 문턱치를 만족시키지 못하고 거절되는 경우가 발생하였다. 이것은 단어를 구성하는 음소 단위 신뢰도가 서로 다른 pdf특성을 갖기 때문이다.

본 논문에서는 이를 보완하기 위해서 통계적 방법을 이용하여 각 음소별로 신뢰도의 분포를 분석하고, 이들의 신뢰도 평균과 표준편차를 이용하여 기존의 신뢰도를 정규화한 정규화 신뢰도를 제안한다.

II. 핵심어 검출 시스템

핵심어 검출 시스템에는 대어휘 연속음성 인식 시스템을 이용한 방법, 핵심어 모델과 필러 모델의 결합 시스템을 이용한 방법, 가상 필러 모델을 이용한 방법 등이 있다. 실험에서는 입력음성의 핵심어부와 비핵심어부를 독립적으로 구축하여 변별력을 높이는 핵심어 모델과 필러 모델의 결합 시스템을 채택하였다[2, 4].

핵심어 검출 시스템은 전처리부, 핵심어 검출부, 후처

리부로 구성되어 있다. 전처리부는 입력 음향으로부터 음성구간을 검출하고 이로부터 특징 파라미터를 추출한다. 핵심어 검출부는 핵심어 모델과 필러 모델로 구성된 핵심어 검출 네트워크를 사용하여 핵심어를 검출한다. 후처리부는 신뢰도를 기반으로 하여 핵심어의 발화여부를 검증한다.

2.1. 데이터베이스

데이터베이스는 핵심어 검출기를 최적화하기 위한 훈련용과 핵심어 검출기의 성능을 평가하기 위한 평가용으로 구분된다. 훈련용 데이터베이스는 음성부 훈련 데이터베이스와 비음성부 훈련 데이터베이스로 구성된다.

음성부 훈련 데이터베이스는 남성화자 30명의 1301문장 발성 데이터로 구축되었고, 비음성부 훈련용 데이터베이스는 남성화자 8명의 53문장 발성 데이터로 구축되었다. 평가용 데이터베이스는 남성화자 3명의 496문장 발성 데이터를 이용하였다. 데이터베이스 녹음환경은 조용한 사무실 환경이고 8 kHz, 16 bit로 A/D변환하였다.

2.1.1. 핵심어/비핵심어 선정

본 논문의 핵심어 검출 시스템에 사용된 핵심어와 비핵심어 정보는 표 1과 같다. 실험에 사용된 핵심어는 전국대학명 및 대학교내 부서명을 기반으로 하여 3단계 552개를 선정하였다. 비핵심어는 실제 전화번호 문의패턴을 반영하기 위해서 대학교내 교환실에서 수집한 전화문의 내용을 분석하여 출현위치에 따라 문두, 문중, 문미의 3단계 56개로 구분하였다.

2.1.2. 훈련용 데이터베이스

음성부 훈련 데이터베이스는 핵심어 모델, 모노폰 모델, 비음성부 중에서 묵음 (silence)과 sp (short pause)를 훈련하는데 사용된다. 훈련 데이터베이스는 트라이폰 분석을 통하여 선정한 1,301문장의 발성 데이터를 이용하여 구축하였다. 비음성부 훈련 데이터베이스는 비핵심어 중 문두와 문미의 끝에 입술소리와 숨소리를 첨부한 53문장의 발성 데이터를 사용하였다.

표 1. 텍스트 코퍼스 구축에 사용된 핵심어와 비핵심어 정보
Table 1. Keyword and non-keyword information that is used in text corpus construction.

구분	핵심어				비핵심어			
	상위	중위	하위	소계	문두	문중	문미	소계
수	366	20	166	552	18	3	35	56

표 2. 평가용 녹음문장의 핵심어 출현분포

Table 2. Keyword appearance distribution of recording sentence for test.

구분	0 회	1 회	2 회	3 회	계
문장수	50	170	170	106	496

2.1.3. 평가용 데이터베이스

평가용 훈련 데이터베이스는 핵심어 검출기의 성능평가를 위해서 문장 당 최대 3개까지의 핵심어가 출현하도록 허용하였다. 평가용 훈련 데이터베이스의 구성은 표 2와 같다. 그리고 객관적인 성능평가를 위해서 핵심어의 출현빈도를 조정하여 각 핵심어가 최소 1회 이상 출현하도록 하였다. 핵심어는 496문장에서 총 828회 출현하였다.

2.2. 핵심어 검출기

본 논문의 핵심어 검출 시스템은 핵심어 모델과 필터 모델로 구성된 핵심어 검출 네트워크를 이용한다.

2.2.1. 전처리부

전처리부는 입력 음향에 대해서 에너지와 영교차율을 이용하여 음성의 시작점/끝점을 검출하고 검출된 음성구간으로부터 핵심어 검출에 사용될 특징 파라미터를 추출한다. 핵심어 검출을 위해서는 사람의 청각특성을 반영한 12차의 멜 캡스트럼과 1차의 정규화 로그 에너지, 이들 파라미터의 델타를 포함하여 총 26차의 특징 파라미터를 사용하였다.

2.2.2. 핵심어 검출부

핵심어 검출부의 핵심부분인 핵심어 검출 네트워크는 핵심어 모델들과 필터 모델들을 토른 패싱 알고리즘을 이용하여 결합한 것이다[5]. 핵심어 검출 네트워크의 핵심어 모델들과 필터 모델들은 가우시안 혼합밀도함수 모델(GMM: Gaussian mixture model)을 이용하여 모델링되었다. 그림 1은 핵심어 검출기의 개략도를 나타낸 것이다.

핵심어 모델 네트워크는 552개의 핵심어 모델들로 구성되어 있다. 필터 모델 네트워크는 모노폰 모델들과 비음성부 모델들로 구성된다. 필터 모델 네트워크는 입력 음성 중 비핵심어부를 검출하고 이들을 필터 모델들의 시퀀스 형태로 출력한다. 필터 모델 구축에 사용된 모노폰 모델 네트워크는 총 44개의 모노폰 모델로 구성된다. 비음성부 모델 네트워크는 입력 음성 중 신체잡음 및 환경잡음 등의 비음성부를 처리하는 역할을 한다. 비음성부 모델로는 숨소리, 입술소리, 목음, sp의 4가지를 사용하였다.

노드는 핵심어 모델 네트워크와 필터 모델 네트워크 간의 입출력을 담당한다.

2.2.3. 후처리부

후처리부는 핵심어의 오검출을 방지하기 위해 핵심어가 실제 발생 여부를 검증한다. 후처리 방법으로는 신뢰도 기반 방법과 혼동 행렬(confusion matrix)을 이용하는 방법, 음소 인식기를 이용하여 검증하는 방법 등이 있다.

본 논문에서는 신뢰도 기반 후처리 방법을 사용하였으며, 신뢰도를 측정하기 위해서 핵심어를 구성하는 각 음소 모델과 반음소 모델간의 유사도를 이용하였다. 이에 대해서는 3장에서 자세히 설명할 것이다.

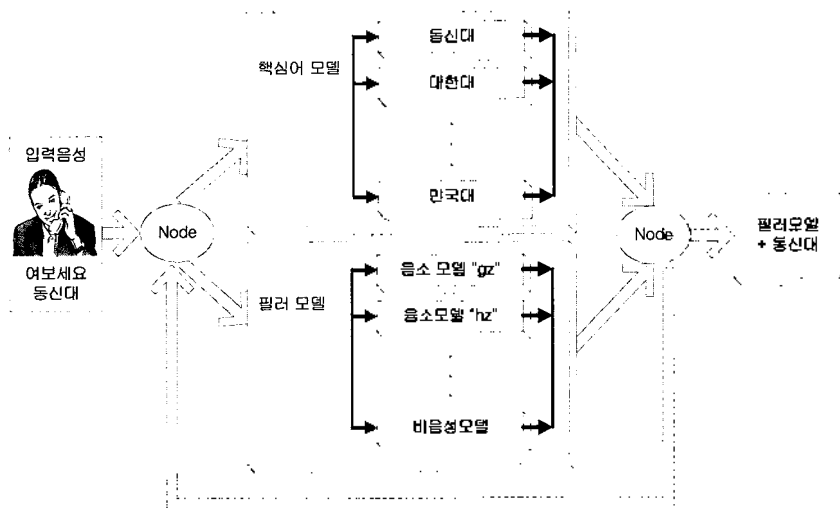


그림 1. 핵심어 검출부
Fig. 1. Keyword spotting model.

표 3. 최적화된 핵심어 검출부의 사양
Table 3. Specification of optimized keyword spotting model.

구분	내용
n-gram	tri-gram, bi-gram
핵심어 모델	3 state 3 mixture
필러 모델	3 state 9 mixture

2.3. 핵심어 검출 실험결과

핵심어 검출 단계에서는 불검출율을 최소화하는 것을 실험목표로 하였다. 이는 FA/KW/HR은 후처리를 통해서 보상이 가능하지만, 불검출율은 차후 보상이 불가능하기 때문이다.

핵심어 검출기의 성능을 평가하기 위해서 다음과 같은 기준을 정의하였다. 불검출율은 발생된 핵심어를 제대로 검출하지 못한 경우이고, FA/KW/HR은 각 핵심어당 오검출 출현횟수를 평가 시간으로 정규화한 것이다. 식 (1)에서 핵심어 총 출현횟수는 828회이다. 식 (2)에서 총 핵심어수는 552개이며, 평가용 DB의 전체 녹음시간은 1.6 시간이다.

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

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

핵심어 검출기를 최적화하기 위해서 다음과 같이 실험하였다. 첫째, 음소 단위 n-gram을 사용하여 필러 모델의 토큰 분배에 문법 정보를 적용하였고 둘째, 핵심어 모델과 필러 모델의 변별력을 향상시키기 위해서 각 모델의 mixture를 조절하여 실험하였으며 셋째, 일반인들의 핵심어 발화 패턴을 참고하여 핵심어 모델을 수정하였다. 이러한 실험을 통하여 최적화된 핵심어 검출부의 사양은 표 3과 같으며, 이 때의 성능은 불검출율 6.76%, 1.06 FA/KW/HR이다.

III. 후처리 시스템

후처리 시스템은 핵심어 검출부에 의해 검출된 핵심어의 실제 발화여부를 검증하여 오검출을 처리한다. 후처리의 목적은 불검출율의 열화를 최소화하면서 오검출에 대한 변별력을 최대로 하는 것이다. 본 논문에서는 신뢰도 기반의 발화검증을 후처리로 사용하였다.

3.1. 기존의 후처리

기존의 후처리 시스템은 음소 모델과 반모델과의 유사도를 이용하여 각 음소별 신뢰도를 계산하고, 음소 단위

신뢰도들의 평균을 단어 단위 신뢰도로 확장한다. 검출된 핵심어의 단어 단위 신뢰도가 문턱치보다 클 경우에만 유효한 핵심어 검출로 인정한다.

본 논문에서는 반모델로 반음소 모델 (anti-phoneme)을 사용하였다. 반음소 모델을 선정한 이유는 필러 모델로 사용하기 위해서 훈련된 음소 모델들을 재사용할 수 있어 별도의 훈련이 필요하지 않기 때문이다.

반음소 모델은 검출된 핵심어의 음소와 대립되는 개념의 모델로서 필러 모델 네트워크를 구성하는 44개의 모노폰 모델 중 자신을 제외한 나머지 43개를 사용하였다. 이러한 반음소 모델은 많이 정의할수록 반모델로서의 역할을 충실히 수행할 수 있지만, 정의된 수에 비례하는 대용량의 훈련 데이터를 필요로 하기 때문에 규모에 대한 고려가 필요하다[3].

3.1.1. 기존의 신뢰도 (CM)

참고논문 [1]에서 Rahim, Lee, Juang에 의해서 제안된 RLJ-CM은 핵심어 구성 음소 모델과 반음소 모델간의 유사도를 이용하여 신뢰도를 계산한다. 반음소 모델의 확률은 43개의 반음소 모델들의 평균 로그 확률을 사용하였다. 계산된 음소 단위 신뢰도들은 식 (5)를 이용하여 단어 단위 신뢰도로 확장된다. 다음은 RLJ-CM의 계산식을 나타낸 것이다[1,3].

$$\log pr_a = \frac{1}{M} \sum_{m=0}^{M-1} \log pr_{a,m} \quad (3)$$

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

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

위의 수식에서 $\log pr_a$ 는 반음소 모델의 평균 로그 확률이고, M은 반음소 모델의 수이다. cm_p 는 음소 단위 신뢰도이고, $\log pr_p$ 는 핵심어 구성 음소 모델의 로그 확률이다. n_p 는 핵심어의 구성 음소의 수이고, f_{cm} 은 가중치로서 -12로 정의되었다.

후처리부에서 RLJ-CM을 계산하는 과정은 다음과 같다. 핵심어 검출 시스템으로부터 핵심어의 발성구간 정보를 인계 받아 비더비 탐색과 후향검색을 통해 최적의 스테이트 시퀀스를 계산한다. 계산된 최적의 스테이트 시퀀스를 음소 단위로 확장하여 핵심어 구간에서의 음소 시퀀스의 구간 정보를 얻는다. 분석된 음소 구간 정보를 바탕으로 하여 핵심어 구성 음소 모델과 반음소 모델들의 확률을 계산하고 이를 이용하여 음소 단위 신뢰도를 계산한다.

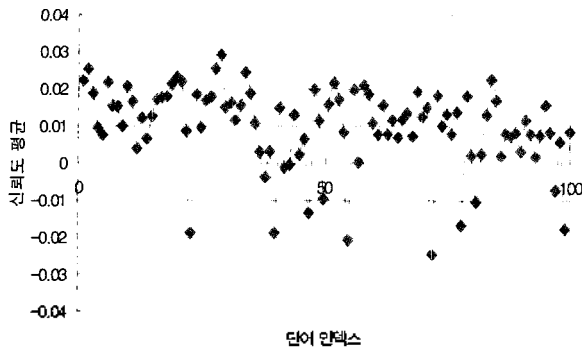


그림 2. 단어 단위 신뢰도 분포
Fig. 2. Confidence distribution of word.

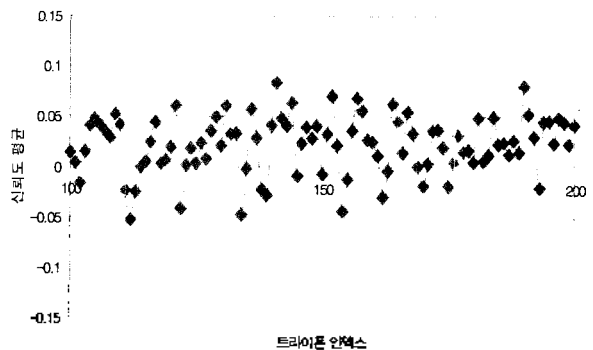


그림 3. 트라이폰의 신뢰도 분포
Fig. 3. Confidence distribution of tri-phoneme.

각 음소 단위 신뢰도들을 이용하여 단어 단위 RLJ-CM을 계산하고 이를 문턱치와 비교한다.

3.1.2. 기존 신뢰도의 문제점

기존 신뢰도를 적용한 후처리 결과를 분석해 보면 특정 단어에서 FR (false rejection)이 다수 발생한다. FR로 잘못 처리된 단어들의 신뢰도를 분석한 결과 일반적으로 낮은 신뢰도를 갖는 것으로 나타났다. 그래서 핵심어 검출부에서 바르게 검출한 핵심어들의 신뢰도 분포를 조사하였다. 그림 2는 이러한 핵심어들의 신뢰도 분포 중 일부를 나타낸 것이다. 그림 2에서 보는 바와 같이 음의 신뢰도를 갖는 단어들이 관찰되었다. 핵심어 모델의 신뢰도는 음소 단위 신뢰도를 기반으로 하여 계산되기 때문에 이러한 문제의 원인은 음소 단위 신뢰도에 있다.

이를 확인하기 위해서 핵심어들의 기본 음소 모델인 트라이폰들의 신뢰도 분포를 조사하였다. 그림 3은 바르게 검출된 핵심어들을 구성하는 트라이폰들의 신뢰도 분포 중 일부를 나타낸 것이다. 예상한 바와 같이 음의 신뢰도를 갖는 트라이폰들이 관찰되었다. 이렇게 음의 신뢰도를 갖는 트라이폰이 전체 트라이폰의 약 16%를 차지하였다. 그러므로 핵심어의 신뢰도 열화는 음소 단위 신뢰도의 불안정한 분포에 원인이 있다.

3.2. 제안한 정규화 신뢰도 (NCM: normalized CM)

기존의 신뢰도는 음소 단위 신뢰도 중 음의 값을 갖는 것이 발생하여 단어 단위 신뢰도가 열화되는 경향이 발생하여 불필요한 FR이 발생하게 된다. 전체 단어 중 약 12%가 신뢰도 열화로 인해서 FR되는 경향을 보인다. 이러한 단어 단위 신뢰도의 열화를 막기 위해서는 음소 단위 신뢰도들을 정규화하였다.

본 논문에서는 신뢰도들이 정규분포를 가진다고 가정하고 각 트라이폰 신뢰도들의 평균과 표준편차를 계산한 다음, 이를 이용하여 표준 정규분포로 정규화하였다.

본 논문에서 제안한 정규화 신뢰도의 계산식은 다음과 같다.

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

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

cm_p 는 기존의 CM에서 계산된 음소 단위 신뢰도이다. ncm_p 는 정규화된 음소의 로그확률이고, f_{ncm} 은 NCM가중치로서 -0.5로 정의되었다. $TriP_q.mean$, $TriP_q.sd$ 은 해당 트라이폰의 신뢰도 평균과 표준편차를 말한다. α 는 음소 단위 신뢰도의 정규화에 사용되는 가중치로서 실험에서는 2로 정의되었다. n_p 는 핵심어를 구성하는 음소의 수이다.

정규화 신뢰도는 검출된 핵심어를 구성하는 각 트라이폰의 신뢰도 평균과 표준편차를 이용하여 신뢰도의 음소 단위 신뢰도 cm_p 를 정규화한 ncm_p 를 얻고 이를 식 (7)에 대입하여 정규화 신뢰도를 계산한다.

IV. 실험 결과

4.1. 기존의 신뢰도 실험결과

기존의 신뢰도를 사용한 실험결과는 표 4와 같다. 실험 결과를 살펴보면, 문턱치가 증가함에 따라 불검출율이 완만히 증가하는데 반해서 FA/KW/HR은 급격히 감소하는 것을 볼 수 있다. 이는 신뢰도를 이용한 검중에 의해서

표 4. 기존의 신뢰도를 사용한 결과

Table 4. The result of Conventional CM experiment.

문턱치	MDR(%)	FA/KW/HR
적용전	6.76	1.06
-0.025	7.20	0.64
-0.010	8.57	0.44
-0.005	10.14	0.34
0	12.88	0.26
0.005	18.07	0.19

표 5. 제안된 정규화 신뢰도를 사용한 결과

Table 5. The result of proposed NCM experiment.

문턱치	MDR(%)	FA/KW/HR
적용전	6.76	1.06
0.75	7.00	0.61
1.25	8.57	0.33
1.35	9.66	0.29
1.40	10.22	0.28
1.50	12.35	0.23

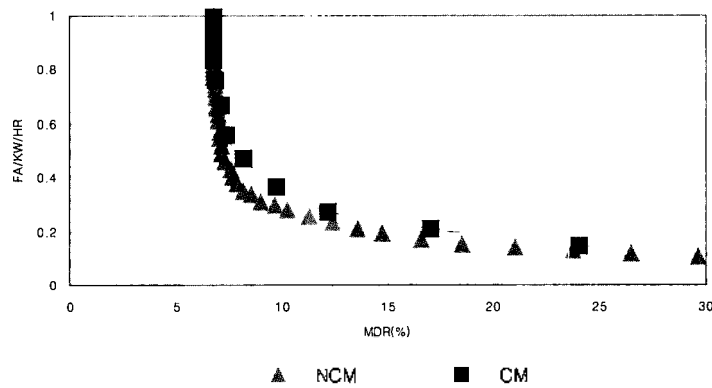


그림 4. 제안된 정규화 신뢰도 실험결과

Fig. 4. The result of proposed NCM experiment.

핵심어 검출 결과 중 신뢰도가 낮은 핵심어와 오검출이 효과적으로 제거된 것이다.

4.2. 제안한 정규화 신뢰도 실험결과

기존의 신뢰도를 정규화한 정규화 신뢰도의 실험결과는 표 5와 같다. 그림 4를 살펴보면 기존의 신뢰도에 비해서 제안한 정규화 신뢰도가 문턱치 증가에 따른 불검출율의 감소가 적고 FA/KW/HR에서도 보다 큰 격차로 감소한다. 이는 본 논문에서 제안한 방법이 각 음소 단위 신뢰도들을 효과적으로 정규화하여 각 단어 단위 신뢰도가 안정적으로 분포된 결과이다.

불검출율 8%대를 불검출율과 FA/KW/HR의 트레이드 오프 (trade-off) 지점이라고 정하면 기존의 신뢰도는 0.44 FA/KW/HR, 제안한 정규화 신뢰도는 0.33 FA/KW/HR로서 성능이 25% 향상되었다.

모델과 필러 모델들로 구성된 핵심어 검출 네트워크를 이용한다. 핵심어 검출부의 성능은 불검출율 6.76%, 1.06 FA/KW/HR 이다.

핵심어 검출과정에서 발생된 오검출은 신뢰도 기반 후처리를 이용하여 발화검증한다. 기존의 신뢰도는 반응소 모델과의 유사도를 이용해서 음소 단위 신뢰도를 계산하고 이를 단어 단위로 확장하여 사용하였다. 그러나 음소 단위 신뢰도 중 음의 값을 가지는 것들에 의해서 단어 단위 신뢰도가 열화되는 경향이 발생하였다. 이를 처리하기 위해서 각 음소 단위 신뢰도의 평균과 표준편차를 이용하여 정규화한 정규화 신뢰도를 사용하였다.

실험결과 불검출율 8% 부근에서 기존의 신뢰도는 0.44 FA/KW/HR, 제안한 정규화 신뢰도는 0.33 FA/KW/HR로서 제안한 정규화 신뢰도를 적용하였을 때가 기존의 신뢰도를 적용하였을 때에 비해 성능이 25% 향상되었다.

V. 결론

본 논문에서는 정규화 신뢰도를 이용한 핵심어 검출 성능향상에 대하여 연구하였다. 핵심어 검출기는 핵심어

참고문헌

1. M. G. Rahim, C. H. Lee, B. H. Juang and W. Chou, "Discriminative utterance verification using minimum string verification error

(MSVE) training," *Proc. of ICASSP 96*, 3585-3588, 1996.

2. P. Jeanrenaud, K. Ng, J. R. Siu, M. Rohlicek, H. Gish, "Phonetic-based word spotter : Various configurations and application to event spotting," *Proc. of Eurospeech 93*, 2, 1057-1060, 1993.
3. 김기태, 문광식, 김회린, 이영직, 정재호. "가변어휘 단어 인식에서의 미등록어 거절 알고리즘 성능 비교," *한국음향학회지*, 20 (2), 27-34, 2001.
4. G. Guo, Y. W. Z. Lin, B. S. Yaun, Q. Zhao, and J. A. Liu, "Keyword spotting in auto-attendant system," *Proc. of ICSLP 2000*, 2, 1050-1052, 2000.
5. S. J. Young, N. H. Russell, and J. H. S. Thornton, "Token Passing : a simple conceptual model for connected speech recognition systems," Technical report of Cambridge university engineering department, TR 38, 1989.
6. J. Junkawitsch, G. Ruske and H. Hoge, "Efficient methods for detecting keywords in continuous speech," *Proc. of Eurospeech 97*, 1, 259-262, 1997.

저자 약력

● 김 철 (Cheol Kim)



1970년 2월: 조선대학교 전기공학과 (공학사)
 1982년 2월: 한양대학교 산업대학원 (공학석사)
 1998년 2월: 동신대학교 정보통신공학과 (공학박사)
 1998년 2월~현재: KT 순천지사장
 ※ 주관심분야: 음성인식 및 신호처리, 통신망 운용/관리

● 이 경 록 (Kyong Rok Lee)

1997년 3월: 호남대학교 전자공학과 (공학사)
 2001년 8월: 전남대학교 정보통신공학부 (공학석사)
 2001년 8월~현재: 전남대학교 정보통신공학부 박사과정 재학 중
 ※ 주관심분야: 음성인식, 멀티미디어 인덱싱

● 김 진 영 (Jin Young Kim)

1986년 3월: 서울대학교 전자공학과 (공학사)
 1988년 3월: 서울대학교 전자공학과 (공학석사)
 1994년 3월: 서울대학교 전자공학과 (공학박사)
 1994년~1995년: 한국통신 소프트웨어 연구소
 1995년~현재: 전남대학교 정보통신공학부 부교수
 ※ 주관심분야: 음성인식, 음성합성, 멀티모달 MMI

● 최 승 호 (Seung Ho Choi)



1991년: 한양대학교 공과대학 전자공학과 학사과정 졸업
 1993년: 한국과학기술원 전기및전자공학과 석사과정 졸업
 1999년: 한국과학기술원 전기및전자공학과 박사과정 졸업
 1996년~현재: 삼성종합기술원 재직
 ※ 주관심분야: 음성인식, 음성코딩, 음성합성, 디지털 신호처리, 디지털통신

● 최 승 호 (Seung Ho Choi)

1981년 2월: 전북대학교 물리학과 (이화사)
 1984년 8월: 영지대학교 전자공학과 (공학석사)
 1992년 2월: 영지대학교 전자공학과 (공학박사)
 1992년 3월~현재: 동신대학교 정보통신공학과 교수
 ※ 주관심분야: 음성인식, 멀티미디어통신, 멀티모달 MMI