반복학습 음소모델을 이용한 핵심어 검출 시스템의 성능 향상

김주곤, 임수호, 이여송, 김범국*, 정현열
영남대학교 전자정보공학부
*대구과학대학 영상통신미디어과

Performance Enhancement of Keyword Spotting System Using Repeated Training of Phone-models

Joo-Gon Kim, Soo-Ho Lim, Young-Song Lee, Bum-Guk Kim*, Hyun-Yeol Chung
Dept. of Information and Communication Eng., Yeungnam University
*Dept. of Image Communication Media, Taegu Science College
E-mail : speech@yumail.ac.kr

요 약

본 논문에서는 반복학습으로 음소모델을 강건하게 하여 음소기반 핵심어 검출 시스템의 성능을 개선하고자 하였다. 가변어휘 핵심어 검출 시스템은 인식 대상 핵심어의 추가와 변경이 용이하도록 모노폰 단위로 핵심어 모델과 필러 모델을 구성하였다. 핵심어 모델과 필러 모델은 동일한 음소모델을 이용하므로 각각의 음소모델의 분별력 향상은 핵심어 검출 성능과 밀접한 관계에 있다. 따라서 본 논문에서는 음소 HMM(Hidden Markov Model)의 학습시에 반복 학습을 통하여 음소 모델을 강건하게 만든 후 핵심어 검출 실험을 수행하였다. 그 결과, 10회의 반복학습을 통하여 얻어진 음소 HMM을 이용한 핵심어 검출의 성능은 반복학습을 하지 않은 경우보다 핵심어 검출의 CA-CR 평균 성능이 4% 향상됨을 확인할 수 있었다.

1. 서 론

핵심어 검출이란 미리 특정한 핵심어를 정한 후 자유로운 음성입력에서 핵심어가 포함되어 있는지의 여부를 찾아내고 식별해 내는 작업을 의미한다. 일반적으로 핵심어 검출 시스템에 사용되는 방식은 HMM(Hidden Markov Model)을 이용하여 핵심어와 비핵심어들을 각각의 HMM으로 모델링하고 아무런 문법적 제한 없이 자연스럽게 발성된 음성을 HMM들이 연결된 것으로 표현한다[1][2][3]. 본 연구실에서는 핵심어를 효과적으로 검출하기 위하여 의사 N-gram 언어모델을 이용한 핵심어 검출 시스템을 구성하였다[4]. 이 시스템은 가변어휘 핵심어 검출 시스템으로 검출 대상으로 하는 핵심어가 바뀌어도 핵심어에 대한 음성 훈련과정을 새로 수행하지 않고 단지 발음 사전만을 교체하여 새로운 핵심어를 검출할 수 있는 장점이 있다. 핵심어 이외의 모든 비 핵심어에 대응하는 필러모델은 별도의 음향모델로 만들지 않고 핵심어를 구성하는 음소모델을 사용하였다. 하지만 이렇게 구성된 필러모델은 각각의 음소모델의 분별력이 높지 않으면 비 핵심어 부분을 잘 표현해주지 못하거나 핵심어 부분을 잠식하여 핵심어 검출 성능을 저하시키는 요인이 될 수 있다.

본 논문에서는 가변어휘 핵심어 검출 시스템에서 음소모델을 강건하게 하기위하여 반복학습을 통하여 음소 HMM을 생성한 후 핵심어 검출 실험을 실시하고, 검출 성능을 비교 검토하였다.

2. 핵심어 검출 시스템

핵심어 검출 시스템은 그림 1과 같이 전처리부와 핵심어 검출부, 후처리부로 구성되어 있다. 전처리부에서

는 입력된 음성의 끝점을 검출한 후 특징파라미터를 추출한다. 핵심어 검출부에서는 핵심어 모델과 필러 모델들을 의사 N-gram 언어모델로 핵심어 검출 네트워크를 구성하여 핵심어를 검출한다. 후처리 부에서는 반음소 모델을 이용하여 신뢰도를 계산하여 핵심어 여부를 판별한다.

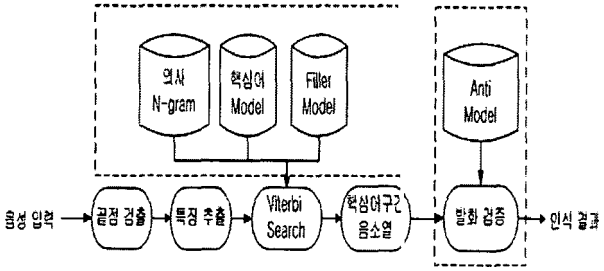


그림 1. 핵심어 검출 시스템의 구성도

의사 N-gram 언어 모델은 그림 2와 같이 구성된다. 연속 음성인식을 수행하기 위하여 사용되고 있는 N-gram 언어모델을 핵심어 검출 시스템에 적합한 형태로 구성하였다.

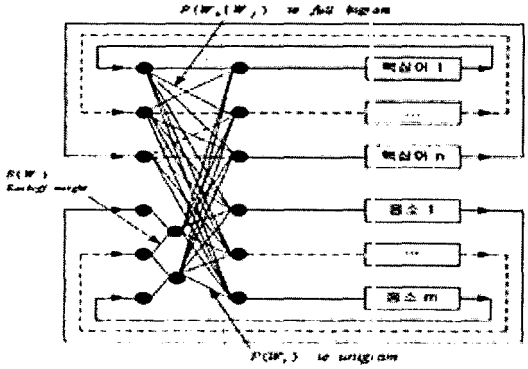


그림 2. 핵심어 검출 네트워크

핵심어와 필러모델의 출현확률 $P(W_i)$를 "1" (대수우도 0)로 설정하고, Bigram확률과 Backoff weight를 모두 "1"로 설정할 경우 연결 단어 형태의 핵심어 검출 네트워크를 구성할 수 있다. 효과적으로 핵심어를 검출하기 위하여 핵심어모델과 필러모델 사이에 임의로 적정한 비율의 언어제약을 부가하였다. 이렇게 구성한 핵심어 검출 시스템은 기존의 연속 음성인식 시스템에서 의사 N-gram 언어 모델의 수정만으로 연속 음성인식과 핵심어 인식을 병행하여 수행할 수 있는 장점을 가진다.

3. 반복학습을 통한 음소모델 생성

HMM을 이용한 음성인식은 학습과정과 인식과정으로 나눌 수 있다. 음소 HMM을 학습할 때 대량의 음성데이터베이스로부터 정확한 레이블링 정보에 의한 음소들의 분류를 통한 학습이 필요하지만 많은 시간과 비용 면에서 자동 레이블링으로 음소들을 분류하여 학습에 이용하고 있다. 자동레이블링 기술은 자동분할에 의한 음소 경계 위치중 80% 정도가 음성학 전문가에 의한 음소 경계 위치와 15ms 오차 이내에 들어왔다고 보고되고 있다[5]. 또한 Entropic 사는 영어에 대해 음성을 분할하는 Toolkit에서 자동분할 경계와 수동분할 경계의 차이가 16ms 이내에 들어오는 경우가 97%인 것으로 소개되고 있다[6].

본 연구에서는 음소 HMM을 학습시 HTK Tool을 이용하였다[7]. HTK에서 자동레이블링을 수행한 후 음소들에 대하여 HMM을 생성한다. 이때 HTK에서 제공하는 기본적인 학습 절차만으로는 강건한 음소 HMM을 생성하기가 어렵다. 따라서 본 논문에서는 그림 3과 같이 학습에 사용된 음성 데이터의 트레이닝률이 수렴할 때 까지 반복 학습을 통하여 강건한 음소 HMM을 생성하였다. 이렇게 얻어진 음소 HMM은 주어진 문장과 환경에 너무 의존할 가능성이 있어 적절한 반복학습을 수행할 필요가 있다.

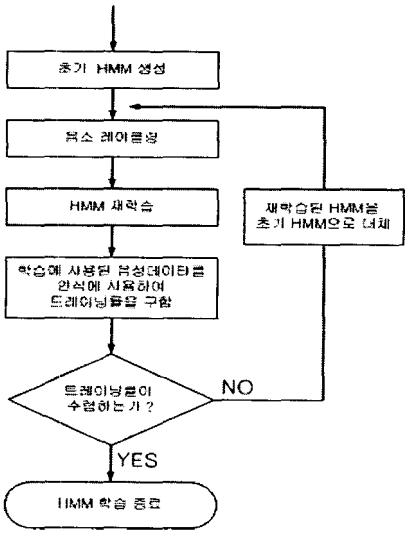


그림 3. 음소 HMM 모델의 반복학습

4. 실험 및 결과

음소기반 핵심어 검출 인식실험을 위하여 다양한 음소의 조합을 고려한 ETRI 445DB와 본 연구실에서 구축한 항공편 예약 200문장 DB를 사용하였다. 태스크 독립 실험을 위하여 KAIST 통신 연구실에서 구축한 무역상담용 연속 음성 데이터를 사용하였다.

특징파라미터는 사람의 청각특성을 반영한 12차의 멜 켑스트럼과 그것의 델타성분 그리고 1차의 정규화 로그 에너지를 CMN(Cepstral Mean Normalization)처리 과정을 거친 총 25차의 특징파라미터를 사용하였다. 핵심어 모델과 필러 모델은 모노폰만으로 구성하고, 실험을 위한 핵심어는 항공편 예약 음성DB에서 사용된 사람의 이름 "15단어"를 선택하였다.

4.1 반복 학습에 따른 트레이닝률

강건한 음소 HMM을 얻기 위해 3절에서 기술한 반복학습 방법으로 트레이닝률을 구하였다. 음성DB는 ETRI 445DB중 남성 15인과 항공편예약 음성DB 남성 10인을 학습에 사용하였다. 트레이닝률을 구하기 위하여 학습에 사용된 항공편예약 DB를 인식 실험에 사용한 후 그 결과를 표 1에 나타내었다.

표 1. 반복 학습에 따른 트레이닝률(%)

반복 회수	문장 인식률	단어 인식률
1	83.64	88.32
2	85.99	90.08
4	88.92	92.17
6	90.07	92.98
8	90.35	93.21
10	90.67	93.45

표 1에서 기본 1회보다 2회 반복한 경우 약 2% 이상의 문장 인식률과 단어 인식률이 향상됨을 알 수 있었다. 반복회수가 증가함에 따라 인식 성능이 향상이 되지만 계속적인 반복학습은 주어진 문장과 환경에 너무 의존할 가능성이 있어 적절한 반복학습을 수행할 필요가 있다.

4.2 태스크 독립에서 반복학습에 따른 음성인식

적절한 반복학습을 알아보기 위하여 문맥과 녹음 환경이 서로 다른 음성DB를 이용한 연속음성 인식 실험을 수행하였다.

사용된 음성 DB는 KAIST 통신 연구실에서 구축한 무역상담용 음성DB중 남성 90인을 학습에 사용하고 연속음성인식을 위하여 항공편 예약 음성DB중 3인을 사용하였다.

표 2. 반복 학습에 따른 태스크 독립 인식률(%)

반복 회수	문장 인식률	단어 인식률
1	58.55	87.66
2	65.89	90.52
4	67.72	91.53
6	69.65	92.56
8	69.45	92.48
10	70.16	92.76

표 2에서 기본 1회보다 2회를 반복한 경우 약 7%, 3%의 문장 인식률과 단어 인식률이 각각 향상됨을 알 수 있었다. 이는 대량의 학습데이타는 HTK에서 제공하는 기본 학습만으로는 음소 HMM의 학습이 부족하다는데서 기인한 것으로 추측된다. 태스크 독립 음성인식 실험에서도 반복회수가 증가함에 따라 음성인식 성능이 계속 증가함을 알 수 있다. 따라서 본 연구에서 제안한 반복학습의 필요성을 확인할 수 있었다.

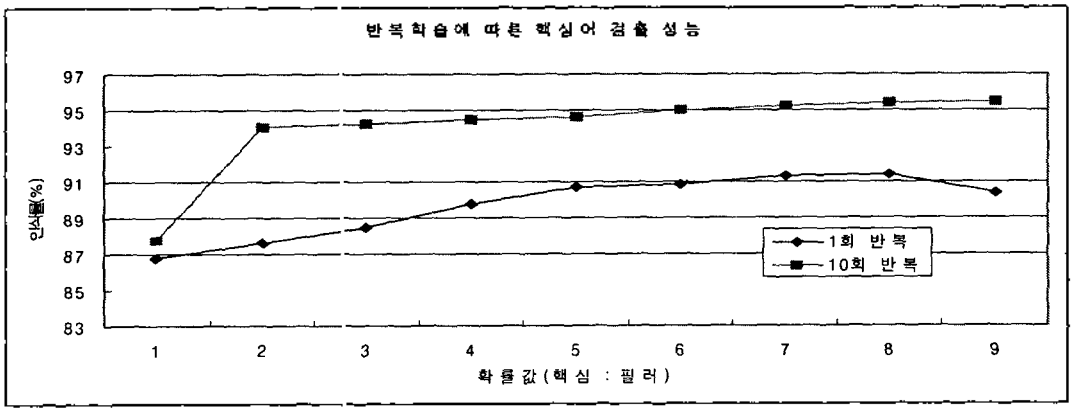
4.3 반복학습에 따른 핵심어 검출

음소기반 핵심어 검출 실험을 위한 음성DB는 4.1절에 기술된 음성DB를 학습에 사용하고 학습에 사용되지 않는 항공편 예약 DB중 3인을 사용하였다. 본 연구실에서 구축된 의사 N-gram을 이용한 핵심어 검출 시스템을 이용하여 핵심어모델과 필러모델의 확률을 가변적으로 적용하여 핵심어 검출 실험을 수행하였다. 핵심어의 검출 성능은 CA(Correctly Accept for Keyword: 핵심어를 제대로 인정한 경우)와 CR(Correctly Reject for Out of Vocabulary : 비핵심어에 대해 제대로 거절한 경우)의 평균 인식률을 기준으로 하였다.

표 3에서 기본 1회보다 10회 반복한 경우 약 4%의 CA-CR 평균인식률 향상을 나타내었다. 따라서 본 연구에서 제안한 반복학습을 통한 강건한 음소 HMM의 생성은 음성인식뿐만 아니라 음소기반의 핵심어 검출에서도 그 유효성을 확인할 수 있었다.

표 3. 반복학습에 따른 핵심어 검출 성능

언어 모델	1회 반복 (기존)			10회 반복		
핵심어 : 필러	CA	CR	CA-CR	CA	CR	CA-CR
0.9 : 0.1	93.3	80.2	86.75	97.9	83.6	90.75
0.8 : 0.2	91.1	84.1	87.60	95.6	92.6	94.10
0.7 : 0.3	91.1	85.8	88.45	95.6	92.8	94.20
0.6 : 0.4	91.1	88.5	89.80	95.6	93.3	94.45
0.5 : 0.5	91.1	90.3	90.70	95.6	93.7	94.65
0.4 : 0.6	91.1	90.6	90.85	95.6	94.4	95.00
0.3 : 0.7	91.1	91.5	91.30	95.6	94.9	95.25
0.2 : 0.8	91.1	91.7	91.40	95.6	95.1	95.35
0.1 : 0.9	88.9	91.9	90.40	95.6	95.3	95.45



5. 결 론

본 논문에서는 반복학습으로 음소모델을 강건하게 하여 음소기반 핵심어 검출 시스템의 성능을 개선하고자 하였다. 핵심어 검출 시스템은 핵심어 모델과 필러 모델들을 의사 N-gram 언어모델로 핵심어 검출 네트워크를 구성하고, 인식 대상 핵심어의 추가와 변경이 용이하도록 모노폰 단위로 핵심어 모델과 필러 모델을 구성하였다. 핵심어 모델과 필러 모델은 동일한 음소모델을 이용하므로 각각의 음소모델의 분별력 향상은 핵심어 검출 성능과 밀접한 관계가 있으므로 음소 HMM의 학습 시에 반복 학습을 통하여 음소 모델을 강건하게 만든 후 핵심어 검출 실험을 수행하였다. 그 결과, 10회의 반복 학습을 통하여 얻어진 음소 HMM을 이용한 핵심어 검출의 성능은 반복학습을 하지 않은 경우보다 핵심어 검출의 CA-CR 평균 성능이 약 4% 향상됨을 확인할 수 있었다.

참고문헌

1. Eng-Fong Huang, Hsiao-Chuan Wang, and Frank K. Soong, "A Fast Algorithm for Large Vocabulary Keyword Spotting Application," IEEE Tran. on S Speech and Audio Progression, VOL. 2, NO.3, pp. 449-452, 1994.
2. Mazin G Rahim, Chin-Hui Lee, Biing-Hwang Juang and Wu Chou, "Discriminative utterance Verification Using Minimum String Verification Error (MSVE) Training", ICASSP, 1996
3. A. Kenji, K. Kazushige, T. Kazunari, O. Sumio, F. Hiroya, "A New Method for Dialogue Management in an Intelligent System for Information Retrieval," ICSLP, pp. 118-121, 2000.
4. 김주곤 외 3명, "의사 N-gram 언어모델을 이용한 핵심어 검출 시스템의 성능향상," 음향학회 추계학술발표대회 논문집, pp. 89-90, 2003.
5. A. Ljolie, and M. D. Riley, "Automatic segmentation and labeling of speech," ICASSP, pp.473-476, 1991.
6. The Aligner : A system for automatic time alignment of English test and speech, Entropic Research Laboratory, Inc., 1994
7. Steve Young, " The HTK BOOK," 2000.