

엔트로피 거절을 활용한 음성인식 시스템의 성능 향상

송점동*

요 약

본 논문은 음성인식 시스템에서 정확도를 높이기 위해 후처리 단계에서 후보 단어들의 엔트로피 정보를 이용하였다. 기존의 우도비 검출방법은 음성 데이터에 따라 음성인식 시스템의 성능이 변하고 N개의 후보단어들의 우도값이 비슷하여 오인식 발생확률이 높았다. 그러나 본 논문에서는 각 후보 단어들의 엔트로피 값보다 인식대상 단어 외의 단어들의 엔트로피 값이 상대적으로 낮은 후보를 거절하는 후처리 방법을 사용하여 음성 데이터에 독립적이면서도 변별력을 높인 정확한 음성인식 시스템을 얻을 수 있었다. 실험 결과 본 논문에서 제안하는 엔트로피에 의한 후처리 방법은 우도비에 의한 방법보다 인식 시스템의 성능을 false alarm이 20%일 때 최대 3.6% 향상시킬 수 있었다.

I. 서론

음성인식 시스템의 궁극적인 목표는 화자 독립적인 자연스런 연속음을 빠른시간에 인식하는 것이다.

그러나 이는 어려운 일이므로 기존의 인식 시스템은 많은 제한적인 기능을 갖게 되었다. 반면 실제 사용자들은 시스템의 기능적 제한을 알지 못한 상황에서 간투사나 잘못된 음성으로 발성할지라도 시스템이 사용자가 의도한 대로 올바르게 작동하기를 바란다.

이런 경우 사용자의 잘못된 발성에 의해 시스템에서 동작의 오류를 일으키기 보다는 인식결과를 내지 않고 사용자에게 응답을 다시 요청하거나 사용자의 잘못된 입력 음성을 거절하는 것이 시스템의 신뢰성을 높인다[1]. 그러므로 음성

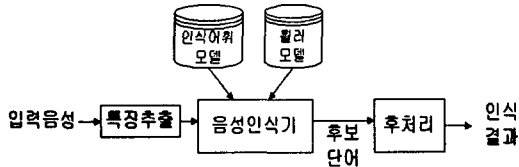
인식 시스템의 실용성을 위해서 사용자의 입력 중에서 시스템이 인식할 입력음성은 검출하되 시스템에서 원하지 않는 입력음성은 거절해야 한다.

본 논문에서는 후처리 단계에서 엔트로피를 이용한 거절 방법을 통해 적합하지 않은 입력음성을 거절하여 기존 시스템보다 정확도를 향상시키는 방법을 제안한다.

II. 인식대상 단어 구별

인식대상 단어 구별 기법은 입력 신호로부터 시스템이 필요로 하는 부분은 구별하고 적합하지 않은 부분에 대해서는 거절하도록 하는 음성인식의 응용분야를 말하며 (그림 1)과 같다.

* 경문대학 전산정보과 조교수



(그림 1) 인식대상 단어구별 시스템

여기서 필터 모델은 입력음성 중에서 적합하지 않은 부분을 모델링하는데 사용되며 비등락 어휘, 외부잡음, 배경 음성 구간들을 대상으로 모델링한다.

실제로 인식 시스템은 필터 모델이 인식대상 부분을 잠식하지 않으면서 비음성 부분을 효과적으로 표현해 줄 수 있어야 안정적인 성능을 나타낸다.

필터 모델을 구성하는 방법에는 인식대상어휘 이외의 데이터중에서 빈도가 높은 어휘를 독립적인 단어 모델로 구성하는 방법[2][3], 한 개 또는 여러 개의 대표 모델로 필터를 표현하는 방법[2], 기본 음소를 비롯한 부분 단어 단위로 필터를 모델링하는 방법[4], 인식과정중에 인식 대상 어휘의 모델링에 사용되는 모델을 이용하여 on-line으로 필터 모델을 찾는 방법 등이 사용되고 있다.

III. 우도비를 이용한 후처리

후처리 과정은 이미 구해진 인식 후보들의 신뢰도를 판단하여 잘못 검출된 후보를 효율적으로 제거하여 인식 시스템이 신뢰도를 높이는 데에 주안점을 두고 있다.

후처리 과정으로 대표적인 것은 인식 어휘 모

델 및 필터 모델에 우도비를 이용하는 방법이 있다[4]. 이는 후보단어에 대해 인식 어휘 모델에 대한 우도값과 필터 모델에 대한 우도값의 비를 사용하여 이 값이 임계치를 넘지 못하면 이 후보단어는 신뢰할 수 없는 것으로 판단하여 거절한다.

$$LR = \frac{P(O|A^{(k)})}{P(O|A^{(l)})} \quad (1)$$

식(1)에서 LR은 후보 단어에 대한 우도비에 의한 신뢰도를 나타내며, $A^{(k)}$ 은 인식 어휘 모델을 나타내고, $A^{(l)}$ 은 필터 모델을 나타낸다. 후보 단어에 대해 식(1)의 LR값이 임계값보다 적을 때 그 단어는 신뢰할 수 없으므로 거절한다.

그러나 이는 인식 대상 어휘 집합에 따라서 거절의 성능 차이가 심하다는 단점이 있다. 즉 입력음성과 비슷한 음향적 특징을 갖는 인식 대상 어휘에 대해서는 우도값이 비슷하여 후보단어를 기각하기 쉽다는 문제점이 있다.

IV. 엔트로피를 이용한 거절 시스템

통신이나 정보이론에서 사용되는 엔트로피는 통신채널에서 메시지가 갖는 평균 정보량을 나타내고 모호한 정도, 불확실한 정도를 의미한다.[5]

이를 음성인식 관점으로 나타내면 엔트로피는 음성특징벡터에 대한 정보량이 되며, 엔트로피

값이 높은 음성특징 벡터는 음성특징에서 모호한 정도가 높음을 의미하고 엔트로피 값이 낮은 음성특징벡터는 모호한 정도가 낮다는 것을 의미한다. 예를 들어 단어 W_i, W_j 대하여 각각의 엔트로피를 비교하여 $H(W_i) > H(W_j)$ 라고 할 경우 단어 W_j 가 더 모호하지 않은 정보를 나타낸다.

그러므로 본 논문에서는 후보 단어별로 엔트로피값을 구해 후보단어의 엔트로피값이 필터모델의 엔트로피보다 상대적으로 높다면 모호하다는 것을 의미하여 인식 어휘로 선택할 것을 거절한다.

본 논문에서 적용할 음성인식의 알고리즘인 DHMM으로 엔트로피 정보량을 나타내면 샘플 공간은 코드북에 해당하고 출력은 코드워드의 색인에 해당한다. 따라서 하나의 특징 파라미터를 위한 엔트로피는 식(2)와 같다[6].

$$H_s(p) = -1.0 \times \prod_{i=1}^{M_p} P_s(j) \log_2 P_s(i) \quad (2)$$

여기서 $H_s(p)$ 는 상태 s 에서 특징 파라미터 p 에 대한 엔트로피를 나타낸다. p 는 특징 파라미터의 색인을 나타내고, s 는 상태를 나타낸다. M_p 는 특징 파라미터 p 의 코드워드 총수를 나타내며, $P_s(i)$ 는 상태 s 에서 i 번째 코드워드의 출력 확률을 나타낸다.

이 때는 후보단어 W_k 의 m 번째 음소에 대한 엔트로피는 식(3)과 같다.

$$H(W_{k(m)}) = \sum_{v_s} H_s(P) \quad (3)$$

식(3)을 단어에 대한 엔트로피 값으로 정규화

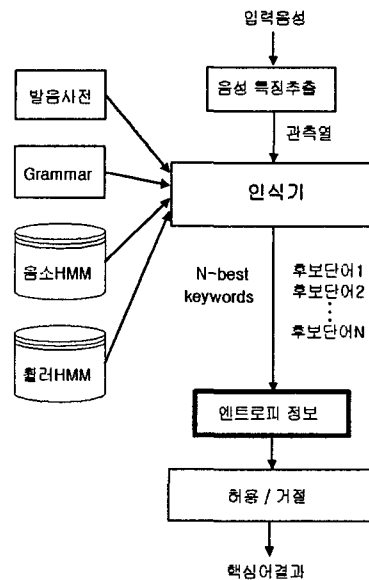
시키기 위해 식(4)와 같이 단어의 음소길이에 대해서 정규화시킨다. 여기서 M 은 후보 단어의 음소의 총 개수이다.

$$H(W_k) = \frac{1}{M} \sum_{m=1}^M H(W_{k(m)}) \quad (4)$$

위와 같이 후보 단어에 대해 정규화된 엔트로피를 구한후 식(5)와 같이 필터 모델의 엔트로피 값과 비교하여 임계값보다 높으면 이 후보단어는 필터 모델보다 모호한 정보로 판단하고 거절한다.

$$H(W_k) - H(W_i) > \tau \quad (5)$$

본 논문에서 제안하는 전체 시스템을 도식화하면 (그림 2)와 같다.



(그림 2) 엔트로피를 이용한 거절

V. 실험 및 평가

본 논문에서는 핵심어의 추가나 변경이 자유롭고 비교적 연산량이 적은 음소 단위를 핵심어 모델로 선택하였다.

인식 알고리즘으로는 다른 알고리즘에 비해 통계적인 문법이나 의미들과 쉽게 결합할 수 있고 적은 연산량으로 빠른 계산시간을 갖는 DHMM을 사용하였다[7].

음성 데이터는 ETRI에서 제공하는 PBW 음성 DB를 사용하였다. 이는 음소적으로 균형이 잡힌 452어절을 남자38명, 여자 32명이 발음하여 16KHz, 16bit로 녹음하였다. 본 논문에서는 이 음성을 DB를 이용하여 46개의 음소모델을 구하였다. 음성 특징을 나타내기 위해 사용된 특징 벡터는 16차의 MFCC계수이고 각 음소모델은 3개의 상태를 가지는 left-to-right HMM으로 표현했다. 인식률의 평가를 위해서는 남자 5명, 여자 5명이 2번씩 발음한 데이터를 평가에 사용하였다.

음성 DB에서 추출된 음소의 발음사전을 이용하여 인식 단어를 나타내었으며 필터 모델은 N개의 후보단어에 대해 관측확률 값의 평균값을 사용하였다.

실험에서 사용한 인식단어는 병원부서명을 사용하였으며 각 단어에 대한 엔트로피 값은 <표 1>과 같다.

거절에 대한 음성 인식 시스템의 성능지표로는 잘못 검출된 비율에 대한 검출비율로써 성능을 나타내며 검출 비율은 식(6)를 이용해 구한다.

<표 1> 인식단어와 엔트로피

| | | | |
|---------|----------|---------|----------|
| 방사선종양학과 | 16.23873 | 신경과 | 15.76531 |
| 진단방사선과 | 16.23588 | 마취과 | 15.77880 |
| 해부병리과 | 15.32797 | 피부과 | 15.53195 |
| 가정의학과 | 16.0913 | 소아과 | 15.7598 |
| 응급의학과 | 16.11642 | 안과 | 14.73357 |
| 산업의학과 | 15.13451 | 내과 | 14.76745 |
| 재활의학과 | 15.2298 | 치과 | 14.77454 |
| 의무기록과 | 15.44083 | 예방 | 14.75534 |
| 이비인후과 | 15.42397 | 진료의료센터 | 16.53441 |
| 영상병리과 | 15.5512 | 응급의료센터 | 16.41195 |
| 핵의학과 | 16.60472 | 건강증진센터 | 16.40312 |
| 비뇨기과 | 15.4689 | 순환기센터 | 16.34781 |
| 신경외과 | 15.55982 | 내시경센터 | 16.22314 |
| 정형외과 | 15.4689 | 신장센터 | 16.88991 |
| 성형외과 | 15.53441 | 암센터 | 15.86645 |
| 산부인과 | 16.61467 | 진료불편상담실 | 16.89012 |
| 흉부외과 | 16.73463 | 기능검사실 | 16.88091 |
| 일반외과 | 16.43091 | 입원실 | 15.38236 |
| 정신과 | 15.88236 | 영안실 | 15.48236 |

$$\text{검출비율} = 1 - \text{Type I 에러} \tag{6}$$

여기서 Type I 에러는 옳게 인식된 단어를 거절하는 경우를 말한다.

본 논문에서 제안한 실험 결과 잘못 검출된 비율에 대해서 우도비를 이용한 후처리와 엔트로피를 이용한 후처리의 검출 비율은 <표 2>와 같다.

<표 2> 실험 결과 (단위: %)

| | | | | | | | |
|----------------|-------|-------|-------|-------|-------|-------|-------|
| false alarm | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
| 우도비후처리 | 76.50 | 86.49 | 89.20 | 94.70 | 94.80 | 95.60 | 96.00 |
| 엔트로피후처리 | 80.10 | 86.52 | 91.00 | 94.00 | 95.89 | 97.60 | 97.30 |
| 엔트로피후처리-우도비후처리 | 3.60 | 0.03 | 1.80 | -0.70 | 1.09 | 2.00 | 1.30 |

실험 결과 우도비를 이용한 후처리 방법에 비해 엔트로피를 이용한 후처리방법이 잘못 검출된 비율이 20%일 때 최대 3.6%의 성능을 향상시켰다.

VI. 결론

본 논문에서는 DHMM 음성인식기를 사용하여 엔트로피를 이용한 거절을 제안하였다. 이는 후처리단계에서 잘못 검출된 인식 후보를 효율적으로 제거하여 오입력에 의한 인식 시스템의 오동작을 방지하는데 효과적이다.

실험 결과 기존의 우도비를 이용한 후처리 방법보다 본 논문에서 제안한 엔트로피를 이용한 후처리 방법을 통해 잘못 검출된 비율이 20%일 때 인식 성능이 최대 3.6%가 향상됨을 볼 수 있었다.

향후 연구로는 엔트로피를 이용한 거절 방법 이외에 엔트로피를 통한 핵심어와 비핵심어의 변별적인 모델링을 연구하고자 한다. 또한 인식 단어 집합을 확장하여 거절 성능을 비교 연구하고자 한다.

"Optimizing recognition and refection performance in wordspotting systems," in Proc. IEEE ICASSP, pp.373-376, 1994.

- [4] R.C. Rose and D.B. Paul, "A hidden markov model based keyword recognition system," in Proc. IEEE ICASSP, pp.129-132, 1990.
- [5] 강창언, 오용선, 이명호, 정보이론-코딩이론과의 접목, 생능사, pp.233-285, 1987.
- [6] 최환진, 오영환, "음성인식을 위한 은닉 마르코프모델에서 엔트로피에 기반한 상태별 특징 파라미터 가중," 정보과학회논문지(B)제 25권 제 2호, pp.294-301, 1998.
- [7] X.D. Huang, Y. Ariki, M.A. Jack, hidden Markov Models for speech recognition, Edinburgh Univ. Press, pp.186-188, 1990

참고문헌

- [1] Rafid A. Sukkar, Chin-Hui Lee, "Vocabulary independent discriminative utterance verification for nonkeyword rejection in subword based speech recognition," IEEE Transaction on speech and audio processing, Vol.4, No.6, pp.420-429, 1996.
- [2] J.G. Wilpon, L.R. Rabiner, C.H. Lee, E.R. Goldman, "Automatic hidden Markov models," IEEE Trans. Acoust., Speech, Signal Processing, Vol.38, No.11, pp.1870-1878, Nov.1990.
- [3] H. Bourlard, B. D'hoore and J.M. Boite,

Improvement of Speech Recognition System using Entropy Rejection

Jeom-Dong, Song*

Abstract

This thesis is a study on using of entropy information about the additional words in the after processing step to promote an accuracy in speech recognition system.

The existing ratio of Woodo detective method changes the efficiency of speech recognition system according to speech data and increases the probability of producing error recognition because of similarity of value of Woodo in the additional words.

But we could obtain the accurate speech recognition system which heightens discrimination becoming independent of speech data by using of after processing method refusing a candidate which entropy price is lower among words except words we could recognize than entropy price of each additional word.

As a result of this experiment when the false alarm is 20 percent, we could put out the maximum 3.6 percent efficiency of recognition system through this after processing method by entropy more than the method by ratio of Woods.

* Dept. of Computing and Information, Kyoung Moon College.