

고립 단어 인식 결과의 비유사 후보 단어 제외 성능을 개선하기 위한 다양한 접근 방법 연구¹⁾

Various Approaches to Improve Exclusion Performance of Non-similar Candidates from N-best Recognition Results on Isolated Word Recognition

윤 영 선²⁾

Yun, Young-Sun

ABSTRACT

Many isolated word recognition systems may generate non-similar words for recognition candidates because they use only acoustic information. The previous study [1,2] investigated several techniques which can exclude non-similar words from N-best candidate words by applying Levenstein distance measure. This paper discusses the various improving techniques of removing non-similar recognition results. The mentioned methods include comparison penalties or weights, phone accuracy based on confusion information, weights candidates by ranking order and partial comparisons. Through experimental results, it is found that some proposed method keeps more accurate recognition results than the previous method's results.

Keywords: isolated word recognition, N-best candidate selection, positional accuracy, word similarity

1. 서론

음성 인식은 사람과 기계와의 가장 직관적이며 편리한 입력 수단으로 여겨져 많은 연구가 진행되어 오고 있다. 음성 인식 기술은 대상 어휘와 발성 방법, 잡음 환경에 따라 다양하게 분류된다. 일반적으로 널리 알려진 음성 인식 방법으로는 소용량 고립단어 인식과 대용량 연속 음성 인식이 있다. 소용량 고립단어 인식은 인식 대상 어휘가 작고 단어 단위의 발성을 인식하기 때문에 PDA나 휴대폰, 텔레매틱스 단말기와 같은 소형 정보 단말기에 적재되어 사용된다. 반면 대용량 연속 음성 인식은 인식하려는 어휘의 수가 많고 연속적으로 발성되기 때문에 단말기는 음성을 수집하는 입력 장치 또는 특징 추출 장치로만 사용되고 인식 과정은 고성능 서버에서 수행한다.

최근에는 반도체 기술과 통신 기술이 발전하면서 정보 단말기에서 대용량 어휘를 이용한 음성 인식 기술을 채용한 제품들이 사용되고 있다. 그러나 정보 단말기 등에서는 장치의 제약으로 인하여 여러 후보를 출력하기 어렵거나, 일부 제한된 후보들만 제시하는 것을 요구받기도 한다.

특히 키패드 입력에 의하여 목적지나 상호를 입력하는 휴대폰이나 자동차 운행 보조 단말기(네비게이터) 등에서 음성인식 기능을 많이 채용하고 있으며, 입력 방식에 따라 단일 후보만을 요구하기도 한다. 그러나 목적지나 상호를 입력하는 경우 인식 대상 어휘는 많고, 단어 단위의 발성이 요구되기 때문에, 문맥 정보나 언어 정보를 사용하기 어려운 경우도 있다.

선행 연구 [1], [2]에서는 이와 같이 음성 인식 결과로 여러 후보 열이 제시되는 경우 다른 후보들에 비해 연관성이 적은 단어들을 인식 결과 목록에서 제외하여 사용자가 느끼는 심리적인 인식 성능의 향상과 신뢰도를 높이는 방법을 제안하였다. 제안된 방법은 기존 인식 후보들 중에서 정답(정확하게 인식된 결과)을 최대한 보존하면서 후보 단어들 중에서 관련성이 가장 적은 (다른 후보 단어들과 이질적인) 후보를 제거하는 것이다. 본 연구에서는 정답 인식률을 향상시키고 관련성이 적은 단어

1) 이 논문은 2010학년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음.

2) 한남대학교 정보통신공학과, ysyun@hnu.kr, 교신저자

접수일자: 2010년 10월 8일

수정일자: 2010년 12월17일

게재결정: 2010년 12월18일

의 제거율을 높이기 위하여 다양한 방법을 적용하고 그 결과를 살펴본다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구에서 사용하는 음성 인식 시스템의 구성을 간단하게 설명하고, 3장에서는 선행 연구에서 제안하는 후보 단어들의 비교 척도를 요약, 정리한다. 4장에서는 선행 연구의 비교 척도에 기반하여 성능을 향상시키기 위한 다양한 방법들을 제안하며, 5장에서는 제안된 방법의 유효성을 확인하기 위한 실험 및 결과를 제시하고, 마지막으로 결론을 맺는다.

2. 음성인식시스템

본 연구에서 사용한 음성 인식 시스템은 한국전자통신연구원에서 개발된 것으로서 텔레매틱스 시스템에 26만단어의 행선지 입력을 목적으로 개발되었다[3]. 텔레매틱스에 채용된 시스템의 한계로 인한 계산 량의 제약을 극복하기 위하여 2단계 인식 시스템을 채용하였다. 1단계 인식 과정에서는 연속 밀도 HMM(continuous density Hidden Markov Model)의 음향 모델을 변환한 모노폰 반연속 밀도 HMM(monophone semi-continuous HMM)을 사용하였으며, 1단계의 음성 인식 결과를 kNN(k-Nearest Neighbor) 알고리즘을 적용하여 2단계 인식 시스템에서 적용할 대상 어휘를 선정한다. 2단계 인식 시스템에서는 트라이폰(triphone) 반연속 밀도 HMM의 음향모델을 이용하여 축소된 대상 어휘(약 1000여개)에서 최대 10개의 후보 단어를 생성한다. 기존 시스템의 성능은 10-best에 대해서 약 90%의 성능을 보이며, 환경 적응을 고려할 경우 약 93%의 성능을 보인다.

3. 문자열 정확도 척도

2장에서 설명한 인식 시스템의 결과로 최대 10개의 후보 단어가 출력되는데, 이 후보 열은 각 음향 모델의 우도(likelihood) 값으로 결정되기 때문에 서로 관련이 없는 후보들이 선택될 수 있다. 본 장에서는 각각의 후보열들의 비교 척도를 요약한다[1], [2].

3.1 레벤스타인 거리 척도

레벤스타인 거리 척도는 두 개의 문자열이 주어질 때, 한 문자열을 다른 문자열로 변환하는 데 필요한 비용을 산출한다.

$$D_{(i,j)} = \min \begin{cases} D(i-1, j-1), & s_i = t_i \\ D(i-1, j-1) + \alpha, & s_i \neq t_i \\ D(i-1, j) + \beta, & \text{insertion} \\ D(i, j-1) + \beta, & \text{deletion} \end{cases} \quad (1)$$

(1)식에서 s_i 와 t_i 는 비교 문자열의 위치를 나타내며, α , β 는

치환오류와 삽입, 삭제오류 시 추가되는 비용을 의미한다.

레벤스타인 거리 방법은 정보 이론, 기계 번역 등에서 널리 사용되는 방법으로 그 변형도 많이 존재한다. 레벤스타인 거리에 의해 정렬된 두 문자열은 각 문자의 비교에 따라 삽입, 삭제, 치환 오류로 정의된다. 삽입(insertion) 오류는 원본 문자열에는 없으나 비교 문자열에 존재하는 경우이며, 삭제(deletion) 오류는 원본 문자열에는 존재하나 비교 문자열에 존재하지 않는 경우, 치환(substitution) 오류는 원본 문자열과 비교 문자열의 문자가 다른 경우이다.

3.2 음소 및 음절 거리를 이용한 단어 거리 정의

레벤스타인 거리 척도를 후보 단어의 비교에 이용하는 방법은 후보 단어를 구성하는 음소들을 문자열로 간주하여 레벤스타인 거리를 이용하는 것과, 후보 단어를 음절로 분리하고, 음절간의 거리를 더하여 단어 거리로 사용하는 방법이 있다.

음소 문자열을 그대로 후보 단어의 거리 계산에 사용하는 방법은 구현하기 쉬우며, 음절 간의 경계를 고려하지 않아도 되는 장점이 있다. 반면에 음절간의 경계를 고려하지 않기 때문에 음절 단위로 출력되는 인식 단어와 약간의 차이를 보이게 된다. 선행 연구에서는 정확하지 않은 음절 경계로 인하여 음절 거리 방법보다 음소 거리에 의한 정확도가 높은 것으로 보고되어, 본 연구에서는 음소 거리 방법을 이용하였다.

3.3 비교 위치에 따른 유사도 정의

레벤스타인 거리를 이용하여 두 문자열을 비교한 경우, 비교 위치에 따라 비교 문자열에 대하여 삽입(I), 정위치(C), 삭제(D)의 위치 속성을 표기한다. 정위치 표시는 비교 결과 맞음(correct)과 치환 오류가 발생할 수 있으나 비교 위치는 정확하다는 것을 의미한다.

원본 문자열과 비교 문자열의 비교 시, 비교 경로에 따라 삽입과 삭제, 정위치의 특성이 달라질 수 있다(문자열의 정렬이 달라진다) [4]. 이 경우 각 오류 특성에 따른 추가되는 비용을 다르게 한다면, 좀 더 정확한 문자열 정렬을 얻을 수 있을 것이다[5].

<그림 1>은 원본 문자열과 비교 문자열이 주어지면 비교 문자열의 각 문자에 대하여 위치 속성을 표기한 예를 보인다.

원본 문자열 :	-	S	O	R	R	Y
비교 문자열 :	O	S	O	R	-	I
누적 거리 :	1	1	1	1	2	3
문자 속성 :	I	C	C	C	D	S

그림 1. 레벤스타인 거리에 의한 문자열 정렬
Figure 1. Alignment of strings by Levenstein distance

문자열을 비교하기 위해서는 원본과 비교 문자열이 필요하다. 음성 인식의 경우 원본을 정확히 알 수 없다는 제약이 따른다. 따라서 본 연구에서는 10-Best 후보 열중에서 임의의 단어를 원본 문자열로 지정하고 나머지 후보들을 비교 문자열로 선정한다. 비교 문자열은 원본 문자열이 달라지면 각 오류에 따라 비교되는 위치 (이하 기준 위치)가 달라지기 때문에, 기준 위치를 원본 문자열의 위치(치환 오류 고려)와 원본 문자열에 대한 삽입 위치(원본 문자열 사이를 비교 문자열의 삽입 가능한 위치로 지정, 삽입 오류 고려)로 선정한다. 이 과정을 모든 후보 단어에 대하여 적용한 후, 그 평균을 각 후보 단어의 유사도로 정의한다.

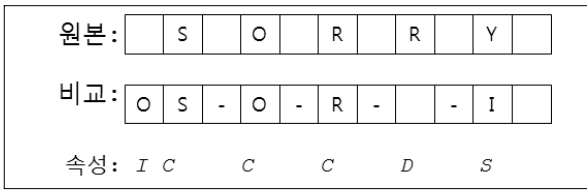


그림 2. 문자열 비교와 문자별 위치 속성

Figure 2. Comparison of strings and positional tag per character

3.3.1 원본 문자열의 정확도 계산

원본 문자열의 비교 시 기준 위치가 자기 자신이기 때문에 모든 문자의 비교 위치는 같다고 정의한다(문자의 삽입과 삭제가 발생하지 않는다.). 이 경우, 동일한 문자 위치에서의 문자열의 빈도수를 정위치(원본의 문자위치와 같다는 의미)에서의 확률이라 정의하고 원본 문자열의 정확도로 정의한다.

3.3.2 비교 문자열의 정확도 계산

비교 문자열은 원본 문자열과 비교하여 정위치(원본 문자 위치와 동일)와 삽입 위치(원본 문자 위치와 비교되지 않고, 추가되었다고 판단)로 구분되기 때문에 정위치에서의 정확도와 삽입 위치에서의 정확도가 필요하다. 정위치에서의 정확도는 원본 문자열의 정확도 계산과 동일하며, 삽입 위치에서의 정확도는 원본 문자열에 대한 삽입 위치에서의 출현 확률을 정확도로 정의한다.

위와 같이 각 문자의 특성에 따라 정위치 정확도와 삽입 위치 정확도가 정의되면, 후보 단어의 유사도를 정의하여야 한다. 후보 단어의 유사도는 삽입 위치와 정위치를 어떻게 고려하느냐에 따라 다음 두 가지로 구분할 수 있다.

첫 번째 경우, 후보 단어의 유사도는 원본 문자열과 비교 문자열의 정렬 후, 위치별 정확도를 이용하여 구한다. 모든 후보 단어를 원본 문자열로 지정하여 지정된 후보 단어와 비교한 후, 계산된 유사도의 평균을 현재 후보 단어의 유사도로 정의한다. 즉, 후보 단어 i 에 대하여 원본 문자열을 j 로 간주하고, 후보 단어 (i, j) 쌍에 대한 유사도 $P_{acc}^{(i, j)}$ 를 계산한다. 이 과정을 모든 후

보 문자열에 대하여 적용한 후 확률 평균을 구하면 $P_{acc}^{(i)}$ 를 계산할 수 있다.

$$P_{acc}^{(i, j)} = \left\{ \prod_{k \in \text{correct}}^{N_j} P_{correct}^{(i, j, k)} \right\}^{1/N_{correct}^{(j)}} \cdot \left\{ \prod_{k \in \text{insert}}^{N_j} P_{insert}^{(i, j, k)} \right\}^{1/N_{insert}^{(j)}} \quad (2)$$

$$P_{acc}^{(i)} = \left\{ \prod_{j=1, j \neq i}^{N_{cand}} P_{acc}^{(i, j)} \right\}^{1/N_{cand}}$$

이 때 $P_{correct}$ 와 P_{insert} 는 정위치 정확도와 삽입위치 정확도를 의미한다. N_j 는 원본 문자열 j 에 대한 기준 위치의 수를 나타내며, $N_{correct}^{(j)}$ 와 $N_{insert}^{(j)}$ 는 원본 문자열 j 에 대한 정위치와 삽입위치의 수를 나타낸다. 또한 N_{cand} 는 후보 단어의 수를 의미한다. 이 방법을 개별 유사도로 명명한다.

두 번째 경우에는 단일 후보 단어의 유사도를 구하는 과정은 비슷하나, 첫 번째 경우와 달리 위치별 정확도를 먼저 구한 후, 최종적으로 N_{cand} 의 후보에 대하여 유사도 계산이 완료될 때, 후보 단어 i 에서 발생하는 정위치와 삽입위치 정보를 이용하여 평균을 구한다. 이 방법은 식 (2)의 방법과 달리 전체 후보 단어의 위치별 정보를 이용하여 평균 유사도를 계산하였기 때문에 통합 유사도라 한다.

$$P_{acc}^{(i)} = \left\{ \prod_{j=1, j \neq i}^{N_{cand}} \prod_{k \in \text{correct}}^{N_j} P_{correct}^{(i, j, k)} \cdot \prod_{k \in \text{insert}}^{N_j} P_{insert}^{(i, j, k)} \right\}^{\frac{1}{N_{correct} + N_{insert}}} \quad (3)$$

식 (3)에서 $N_{correct}$ 와 N_{insert} 는 i 번째 후보 단어를 비교 문자열로 가정한 후, 각 후보 단어가 원본 문자열이라고 간주하고 순차적으로 비교하였을 경우에 관측되는 정위치와 삽입위치의 총 수를 나타낸다.

선행 연구의 결과에서 미세하나마 개별 유사도 방법이 통합 유사도 방법에 비하여 정확도가 높기 때문에, 본 연구에서는 개별 유사도 방법을 적용하였다.

4. 비유사 단어의 제외 개선

본 장에서는 선행 연구의 성능을 높이기 위한 다양한 방법들을 제시하고, 그 특징들을 살펴본다.

4.1 비교 가중치의 변화

식 (1)에서는 α 와 β 의 값을 동일한 값으로 적용하였다. α 와 β 의 값이 동일하다는 것은 치환 오류와 삽입·삭제 오류의 가중치를 동일하게 적용하여 두 문자열의 정렬과정에서 치환오류 또는 삽입·삭제 오류를 동일하게 인정하겠다는 것이다. 만약 비교 가중치를 다르게 적용한다면, 치환 오류와 삽입, 삭제 오류의 중요도를 다르게 함으로써 후보로 선택되는 문자열의 특

성을 선택할 수 있다. 예를 들어 $\alpha=10, \beta=4$ 라고 한다면 치환 오류가 발생한 경우의 거리 값이 삽입 오류나 삭제 오류가 발생한 경우의 거리 값보다 크기 때문에, 삽입 오류나 삭제 오류의 발생이 증가하게 된다.

4.2 혼동 정보를 이용한 음소 정확도(phone accuracy using confusion information)의 이용

음소 정확도는 구현된 시스템의 인식 과정에서 발생하는 혼동 오류 정보를 이용하여 계산된다. 기본적인 아이디어는 음소 정확도가 높은 경우에는 치환오류에 대한 가중치를 높여, 치환 오류 발생을 억제하는 것이며(상대적으로 삽입이나 삭제 오류를 선택하도록 함), 음소 정확도가 낮은 경우에는 치환 오류가 발생하더라도 비교 정확도가 떨어지기 때문에 삽입이나 삭제 오류보다는 치환오류를 선호하도록 하는 것이다. 예를 들어, 10개의 대상 음소 a에서 8개를 제대로 맞추었다면 음소 a의 정확도는 0.8이 된다. 이와 같이 계산된 음소 정확도 정보를 이용하여 식 (1)의 레벤스타인 비교 척도에서 치환 오류의 가중치인 α 의 값을 변경하는 것이다. 즉, $\alpha' = w \cdot \alpha$ 로 표현되며, w 는 음소 정확도 정보를 나타내며, 식 (1)에서 α 대신 α' 을 적용한다. 일반적으로 α 와 β 의 값이 동일한 경우에는 음소 정확도 정보를 이용하는 경우 (음소 정확도는 ≤ 1.0 이다), α 의 값은 항상 β 의 값보다 작게 되기 때문에 효과가 적으나, 4.1절의 비교 가중치를 적용하는 경우 음소 정확도 정보를 이용하는 것은 효과가 있을 것으로 판단된다. 즉, 음소 정확도가 높은 경우에는 치환 오류의 가중치가 증가하기 때문에, 삽입이나 삭제 오류의 발생 빈도가 증가하며, 반대로 음소 정확도가 낮은 경우에는 치환 오류에 대한 가중치가 낮아져 치환 오류를 선호하게 된다. 그러나 단어를 구성하는 음소의 개수가 작을 경우에는 음소 정확도 정보에 의하여 치환오류가 삽입 또는 삭제 오류로 변환되거나 반대로 오류가 전환될 확률이 작아, 단어를 구성하는 음소열이 길이가 긴 경우에 효과가 있을 것으로 판단된다.

4.3 후보 단어들에 순위 정보를 이용한 가중치 적용

각 후보들은 교차 방식에 의하여 정위치 정확도와 삽입위치 정확도를 계산한 후 개별 유사도 결합 방식에 의하여 전체 유사도를 계산하며, 일정 값 이상인 경우에는 비유사 후보로 결정한다. 이 방법을 이용한 경우, 인식 결과의 순위 정보를 고려하지 않아 최상위 후보 즉, 정답일 확률이 가장 높은 후보 단어를 제외할 수도 있다. 이런 점을 보완하기 위하여 선형 가중치(<그림 3> 참조)를 적용하여 상위 후보들의 거리 또는 유사도 값은 감소시키고 하위 후보들의 거리 또는 유사도 값은 증가시켰다.

비교에 사용되는 값의 성격에 따라 거리 척도인 경우에는 가중치를 곱하고, 유사도 척도의 경우에는 승을 이용하여 계산하였다.

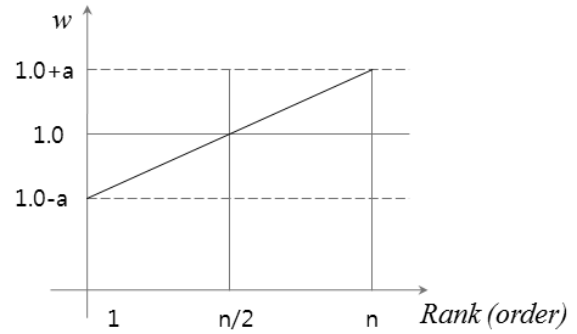


그림 3 순위 정보를 이용한 선형 가중치
Figure 3. Linear weight by ranking order

$$\begin{cases} D_{(n)}' = D_{(n)} \cdot w_n, & \text{거리 척도인 경우} \\ L_{(n)}' = L_{(n)}^{w_n}, & \text{유사도 척도인 경우} \end{cases} \quad (4)$$

4.4 상위 후보의 부분 정합에 의한 가중치 적용

각 후보단어들의 순위에 따른 가중치의 적용은 순위 값에 영향을 받기 때문에, 아주 작은 값에 의하여 순위가 결정된 경우에는 성능의 저하를 가져올 수 있다. 일반적인 음성 인식 시스템의 경우 3-4 정도까지의 순위에 정답이 존재할 확률이 95% 이상이기 때문에 상위 m 순위까지의 경우에만 일정정도 가중치를 적용하고 하위 $m+1$ 순위 이하의 경우에는 가중치를 적용하지 않는 방법을 고려하였다. 상위 m 개의 후보 단어에 대한 가중치를 결정하기 위하여 부분 정합 방식을 이용하여 결정하였다. 전체 교차 비교 방식은 n 후보 단어들에 대하여 $(n-1)^2$ 의 비교를 수행하나 (자기 자신 제외), 부분 정합의 경우에는 $m(n-1)$ 의 비교를 수행하도록 반복 알고리즘을 수정하였다 (<표 1> 상위 m 개의 후보에 대한 부분 정합 알고리즘 참조).

표 1. 상위 m 개의 후보에 대한 부분 정합 알고리즘
Table 1. Partial matching algorithm by top m candidates.

```

for each i in top-m candidates
  for each j in all candidates
    if two candidates are not same,
      calculate distance between two candidates
      and accumulate it to make dist[] to be symmetric
      as follows:
        dist[i] += distance;
        dist[j] += distance;
    end if
  end for
end for
for each i in top-m candidates
  divide its distance by m, s.t., dist[i] /= m;
end for
    
```

거리 값의 가중치 적용은 상위 m 순위까지는 m 으로 나누고, 하위 $m+1$ 이하의 후보 단어는 적용하지 않았다.

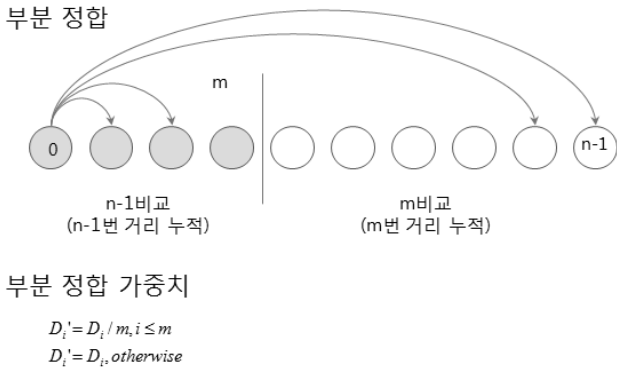


그림 4. 상위 m 후보 정합에 의한 가중치 적용
Figure 4. Weighting by partial top m candidates matching

부분 정합 가중치를 적용하는 경우, 기준이 되는 단어는 자기 자신을 제외하기 때문에 n-1번의 비교를 통한 거리 값이 누적되고, 비교 대상 단어는 기준 단어가 m개이기 때문에 m번의 비교 값이 누적된다. 부분 정합 가중치를 적용한다면 기준 단어의 누적 값이 D_i 인 경우 새로운 가중치를 적용한 거리 값은 D_i/m 이 된다. 따라서 다음의 조건이 만족되는 경우에는 부분 정합에 의한 가중치 방법이 효과적으로 판단할 수 있다.

$$\frac{n-1}{m} < m \text{ 또는 } n-1 < m^2 \quad (5)$$

그렇지 않은 경우 상대적으로 기준 단어의 누적 거리 값을 비교 단어의 누적 거리 값보다 키워, 기준 단어의 정확도를 감소시키는 효과가 있다. (기준 단어를 비유사 단어로 선정할 가능성이 높아진다.)

5. 실험 및 결과

표 2. 평가 데이터의 기본 인식률
Table 2. Base recognition rate of evaluation data

전체 문장수	Best	10-Best	오인식
3,675	2,975	3,299	376
500	405	449	51
인식률	80.9%	89.8%	10.3%

선행 연구와 동일한 환경에서 제안된 방식을 이용하여 비유사 단어를 검출하여 지정된 비유사 단어와 비교하였다. 전체 인식 결과는 IV (In-Vocabulary) 3,675 문장과 OOV (Out-Of-Vocabulary) 1,525문장으로 구성되었다. 이 중 비유사 단어 제거 실험을 위하여 IV 3,675 문장에서 Best, 10-best, 오인식의 비율에 따라 선정된 500문장을 사용하였다. 전체 3,675문장을 사용하는 것이 바람직하나, 비유사단어를 자동으로 판단하기 어렵기 때문에 선정된 500문장에 대하여 전문가가 수작업으로 비유사단어를 평가하였다. 원래 인식 결과와 비유사 단어 제거를 위

해 선택된 500 문장의 기본 인식률은 <표 2>와 같다.

5.1 실험 자료 및 평가 기준

사람의 경우 정답의 유무에 따라 비유사 단어의 선택이 달라지기 때문에, 전체 평가 데이터에서 추출된 500문장에 대해 정답을 알고 있는 경우와 모르는 경우에 대해서 비유사 단어를 선택하도록 하였다. 각각의 경우에 대하여 비유사 단어를 제외하고 남은 후보 단어들을 대상으로 인식률을 다시 계산하면 <표 3>과 같다.

표 3. 정답 제시 유무에 따른 평가 데이터 정보
Table 3. Evaluation data information with or without reference text

정답제시	인식률		총 비유사 단어 수
	Best	10-best	
제시	81.0%	89.8%	2,343
미제시	38.4%	44.2%	2,648

평가 방법은 전문가에 의하여 선택된 비유사 단어의 검출 여부(재현율, recall)로 판단하지 않고, 제안된 방법에 의하여 판단된 비유사 단어가 전문가에 의해 선택된 비유사 단어인지(정확도, precision)를 검사하였다. 일반적으로 재현율은 정보 검색에서 전체 존재하는 관련 데이터에서 검색엔진이 찾은 참(true) 데이터의 비율을 말하며, 정확도는 검색엔진에 의하여 찾은 데이터 중에서 참 데이터의 비율을 나타낸다.

5.2 실험 결과

5.1절에서 설명한 500문장의 평가 데이터를 이용하여 선행 연구에서 제안한 음소 거리와 개별 유사도 방식을 기준 시스템으로 설정하고, 4장에서 제안한 방법을 적용하여 실험을 하였다. 평가 대상이 500문장이기 때문에 비유사 후보단어는 개략적으로 각 인식 문장에서 1개의 후보 단어를 제외하는 경우와 2개의 후보 단어를 제외하는 경우로 나누어 실험하였다. 실험 결과 비교는 제안한 방식의 실험 결과와 기준 시스템의 성능을 비교하여 각 방식에 따른 성능의 변화를 살펴보았다.

<표 4>와 <표 5>는 선행 연구의 결과로서 전체 평가 데이터 중에서 500개와 1000개의 비유사 단어를 제거한 결과이다. 즉, 평균적으로 인식 결과에서 1개 또는 2개의 비유사 단어를 선택하고 제거한 결과이다.

실험 결과의 기준시스템1은 전문가들이 인식 결과 중에서 정답을 알고 있는 경우이기 때문에 Best와 10-best의 인식률은 이 상적인 경우이며, 비유사 단어의 선택은 정답과 유사도 차이가 많은 단어들을 선택하게 된다. 반면 기준시스템2와 정확도2 척도는 전문가가 정답을 모른 상태에서 비유사 단어를 선택하였기 때문에 기존의 정답까지 제거하며, 정확도2의 값은 후보 단

어 들 중 에서 유 사 성 이 낮 은 단 어 들 을 얼 마 나 많 이 선 택 하 였 는 지 를 나 타 낸 다. 따 라 서 시 스템 에 의 한 비 유 사 단 어 제 거 의 성 능 은 기 준 시 스템 1의 Best와 10-best 인 식 륜 에 가 장 접 근 하 면 서, 기 준 시 스템 2의 정 확 도 2보 다 큰 값 을 보 일 수 록 성 능 이 우 수 하 다 고 말 할 수 있 다.

표 4. 500개의 비유사 단어를 지정한 경우 기준 시스템 정확도 (기준시스템1과 정확도1은 정답이 제시된 경우, 기준시스템2와 정확도2는 정답이 제시되지 않은 경우)

Table 4. Baseline precision for 500 non-similar words (Baseline1 and precision1 are evaluated with reference words, baseline2 and precision2 are without them)

시스템	Best	10-best	정확도1	정확도2
기준시스템1	81.0	89.8	82.6	-
기준시스템2	38.4	44.2	-	83.8
음소 거리	75.4	83.4	79.5	91.1
개별 유사도	75.8	84.2	76.9	86.4

표 5. 1,000개의 비유사 단어를 지정한 경우 기준 시스템 정확도 (기준시스템1과 정확도1은 정답이 제시된 경우, 기준시스템2와 정확도2는 정답이 제시되지 않은 경우)

Table 5. Baseline precision for 1,000 non-similar words (Baseline1 and precision1 are evaluated with reference words, baseline2 and precision2 are without them)

시스템	Best	10-best	정확도1	정확도2
기준시스템1	81.0	89.8	79.8	-
기준시스템2	38.4	44.2	-	81.5
음소 거리	69.4	76.8	78.0	88.8
개별 유사도	73.6	81.2	75.3	84.5

5.2.1 비교 가중치의 변화

표 6. 500개의 비유사 단어를 지정한 경우 비교 가중치 변화에 의한 정확도 변화

Table 6. Variation of system precision for 500 non-similar words along to error dependent loss value

시스템	Best	10-best	정확도1	정확도2
개별 유사도	73.6	81.2	75.3	84.5
10:3.5:3.5	75.8	83.8	73.6	88.2
10:7:7	76.4	84.6	77.4	87.0

표 7. 1,000개의 비유사 단어를 지정한 경우 비교 가중치 변화에 의한 정확도 변화

Table 7. Variation of system precision for 1,000 non-similar words along to error dependent loss value

시스템	Best	10-best	정확도1	정확도2
개별 유사도	73.6	81.2	75.3	84.5
10:3.5:3.5	70.0	77.6	74.2	85.4
10:7:7	71.4	79.0	74.9	85.4

치 환, 삽 입 및 삭 제 오 류 에 대 한 정 확 도 손 실 을 계 산 하 기 위 하 여 비 교 가 중 치 를 변 화 하 여 실 험 하 였 다. 음 소 거 리 방 식 과 같 은 일 반 적 인 동 적 정 합 방 식 에 서 는 오 류 정 보 에 의 한 손 실 값 의 변 화 는 전 체 성 능 에 큰 영 향 을 미 치 지 않 는 다. 그 러 나 본 연 구 에 서 제 안 한 위 치 정 보 정 확 도 (개 별 유 사 도 방 식) 같 이 위 치 에 기 반 한 정 보 를 기 반 으 로 정 확 도 를 계 산 하 는 경 우, 그 성 능 차 이 는 발 생 할 수 있 다. 기 준 시 스템 의 손 실 비 용 은 치 환:삽 입:삭 제 = 1:1:1 이 기 때 문 에, 10:3.5:3.5와 10:7:7에 대 하여 실 험 을 진 행 하 였 으 며, 그 결 과 는 <표 6>과 <표 7>에 제 시 된 다. 실 험 은 개 별 유 사 도 방 식 만 을 고 려 하 였 다.

실 험 결 과, 500단 어 의 비 유 사 도 실 험 의 경 우 에 서 는 가 중 치 변 화 에 의 한 성 능 향 상 이 보 였 으 며, 10:7:7 비 율 이 좋 은 성 능 을 보 였 다. 그 러 나 1000 단 어 의 비 유 사 도 실 험 에 서 는 정 답 이 비 유 사 단 어 로 선 정 되 어 전 체 인 식 륜 은 하 락 하 는 양 상 을 보 였 다. 다 만, 미 세 하 나 마 비 유 사 단 어 의 정 확 도 가 상 승 하 는 것 을 알 수 있 다. 이 것 은 1000 단 어 비 유 사 실 험 의 경 우 10개 의 후 보 단 어 들 중 에서 2개 의 비 유 사 단 어 를 선 정 하 기 때 문 에 상 대 적 으 로 낮 은 유 사 도 값 을 보 이 는 정 답 후 보 가 제 거 되 었 기 때 문 으 로 판 단 된 다.

5.2.2 음소 정확도 정보의 이용

표 8. 500개의 비유사 단어를 지정한 경우 음소 정확도 정보 적용에 의한 정확도 변화 (*는 음소 정확도 이용 결과)

Table 8. Variation of system precision for 500 non-similar words using confusion information (* indicates the system using phone accuracy based on confusion information)

시스템	Best	10-best	정확도1	정확도2
10:3.5:3.5	75.8	83.8	73.6	88.2
10:3.5:3.5*	75.2	83.6	73.6	87.2
10:7:7	76.4	84.6	77.4	87.0
10:7:7*	76.6	84.8	75.9	86.45

표 9. 1,000개의 비유사 단어를 지정한 경우 음소 정확도 정보 적용에 의한 정확도 변화 (*는 음소 정확도 이용 결과)

Table 9. Variation of system precision for 1,000 non-similar words using confusion information (* indicates the system using phone accuracy based on confusion information)

시스템	Best	10-best	정확도1	정확도2
10:3.5:3.5	70.0	77.6	74.2	85.4
10:3.5:3.5*	69.6	77.0	73.7	85.5
10:7:7	71.4	79.0	74.9	85.4
10:7:7*	71.8	79.4	74.6	85.1

음 소 정 확 도 정 보 의 경 우 치 환 오 류 에 대 해 서 만 0.0~1.0의 가 중 치 를 적 용 하 기 때 문 에 동 일 한 비 교 가 중 치 의 경 우 그 영 향 은 미 미 할 것 으 로 판 단 하 여, 5.2.1의 비 교 가 중 치 의 변 화 에 따 른

두 조건 (10:3.5:3.5와 10:7:7)에 대하여 실험을 실시하였다. 따라서, 비교 기준 시스템은 개별 유사도 방식이 아닌, 5.2.1의 실험 결과가 된다.

실험결과 음소 정확도 정보를 이용한 경우, 10:7:7의 경우에는 약간의 성능 향상이 있으나, 10:3.5:3.5의 경우에는 성능 하락이 관측되었다. 이는 음소 정확도 정보의 경우 인식 시스템의 성능에 따라 정확도가 결정되며, 정확도에 따라 치환 오류의 선택 여부가 결정되기 때문으로 판단된다. 예를 들어 10:3.5:3.5의 경우 음소 정확도가 0.35보다 낮은 경우에는 치환 오류를 선호하나, 10:7:7의 경우 음소 정확도 정보가 0.7보다 낮아야 치환 오류를 선호하기 때문이다. 즉, 일반적인 인식 시스템에서 혼동 정보를 이용한 음소 정확도의 경우 0.5 이하인 경우가 극히 드물다는 점을 감안하면, 10:3.5:3.5의 경우에는 거의 대부분 치환 오류보다는 삽입이나 삭제 오류를 선호하게 된다고 파악할 수 있다.

5.2.3 순위 정보에 의한 가중치 적용

순위 정보에 따라 계산된 음소 거리나, 개별 유사도 값에 대하여 가중치를 적용하는 방법에 대하여 실험을 하였다. <그림 3>의 경우와 같이 $a=0.2$ 를 적용하여 첫 번째 후보의 경우 0.8의 가중치를 적용하고, 마지막 후보의 경우 1.2의 가중치를 적용하였다. 적용된 가중치만의 성능 변화를 살펴보기 위하여 오류에 따른 비교 가중치는 별도로 지정하지 않았으며 (1:1:1 사용), 음소 거리와 개별 유사도 방식에 대하여 실험하였다.

표 10. 500개의 비유사 단어를 지정한 경우 순위 정보에 의한 정확도 변화 (*는 순위 정보 가중치 결과)

Table 10. Variation of system precision for 500 non-similar words using ranking order weights (* indicates the system using ranking order weights)

시스템	Best	10-best	정확도1	정확도2
음소 거리	75.4	83.4	79.5	91.1
음소 거리*	80.4	88.4	86.6	92.6
개별 유사도	75.8	84.2	76.9	86.4
개별 유사도*	80.0	88.2	85.7	92.1

표 11. 1,000개의 비유사 단어를 지정한 경우 순위 정보에 의한 정확도 변화 (*는 순위 정보 가중치 결과)

Table 11. Variation of system precision for 1000 non-similar words using ranking order weights (* indicates the system using ranking order weights)

시스템	Best	10-best	정확도1	정확도2
음소 거리	69.4	76.8	78.0	88.8
음소 거리*	79.2	86.6	85.6	90.5
개별 유사도	73.6	81.2	75.3	84.5
개별 유사도*	76.8	84.4	81.7	87.0

실험결과, 음소 거리 방식이나 개별 유사도 방식 모두 성능이 향상됨을 알 수 있다. 개별 유사도 방식에 비하여 음소 거리 방식의 성능 변화가 큰 이유로는 가중치 적용에 따른 값의 차이가 많이 발생하기 때문으로 파악된다. 즉, 음소 거리의 경우 각 후보 단어에 대한 거리 값은 1보다 크나 개별 유사도의 경우에는 1보다 작은 값을 갖는다. 따라서 동일한 가중치 w 를 적용할 경우 음소 거리 값의 값 차이가 많이 발생한다. ($D_1 \cdot 0.8$ 와 $D_2 \cdot 0.84$ 의 차이와, $L_1^{0.8}$ 과 $L_2^{0.84}$ 의 차이)

5.2.4 부분 정합 이용

마지막으로 전체 후보 단어를 모두 기준 단어로 설정하지 않고 상위 몇 단어만을 기준 단어로 선정할 후, 가중치를 적용한 경우이다. 이 방법은 선형 가중 방법인 순위 정보에 의한 가중치 적용과는 다르게, 기준 단어에 따라 상위와 하위 정보를 다르게 평가하는 방식이다. 본 연구에 사용된 후보 단어의 총수는 $n=10$ 이기 때문에, 식 (5)을 만족하는 $m=4$ 로 설정하였다. 이 경우 후보 1위부터 4위까지는 9개의 비교 단어 누적 거리를 4로 나눈 즉, 2.25 값이 가중치로 적용된 반면 후보 5부터 10위까지는 4의 가중치가 적용된다. 부분 정합에 의한 실험 결과는 다음 표와 같다.

표 12. 500개의 비유사 단어를 지정한 경우 부분 정합에 의한 정확도 변화 (*는 부분 정합 결과)

Table 12. Variation of system precision for 500 non-similar words based on partial match (* indicates the partial matching system)

시스템	Best	10-best	정확도1	정확도2
음소 거리	75.4	83.4	79.5	91.1
음소 거리*	81.0	88.8	90.0	91.0
개별 유사도	75.8	84.2	76.9	86.4
개별 유사도*	77.6	86.2	79.3	84.7

표 13. 1,000개의 비유사 단어를 지정한 경우 부분 정합에 의한 정확도 변화 (*는 부분 정합 결과)

Table 13. Variation of system precision for 1000 non-similar words based on partial match (* indicates the partial matching system)

시스템	Best	10-best	정확도1	정확도2
음소 거리	69.4	76.8	78.0	88.8
음소 거리*	80.6	88.2	86.9	87.7
개별 유사도	73.6	81.2	75.3	84.5
개별 유사도*	74.6	82.4	76.2	84.3

실험 결과 순위 정보에 의한 가중치 적용 방법과 유사하게 음소 거리 방법이 개별 유사도 방법보다 성능 변화가 큰 것을 알 수 있다. 이것 또한, 가중치 정보의 적용 방법의 차이로 파악된다.

5.2.5 제안된 방법의 통합

제안된 방법이 시스템의 성능을 향상시킨다고 해서, 모든 방법을 통합한 경우 전체 시스템의 성능이 향상된다고 보장할 수 없다. 제안된 방식의 통합 방식에 따른 성능 변화를 살펴보기 위하여 비교 가중치 적용 + 부분 정합, 부분정합 + 순위 정보에 의한 가중치 적용에 대하여 실험을 하였다. 모든 조건에 대하여 실험할 수 있으나, 음소 정확도 정보의 이용은 성능 변화가 미미하기 때문에 제외하였고, 순위 정보 가중치와 부분 정합 가중치 방식은 접근 방법이 비슷하기 때문에 성능 변화가 큰 부분 정합만을 고려하였다. 또한, 부분 정합과 순위정보 가중치를 같이 결합하는 경우에 대해서 추가적인 실험을 하였다. 실험 결과는 <표 14~17>에 정리하였다.

표 14. 500개의 비유사 단어를 지정한 경우 비교 가중치와 부분 정합에 의한 정확도 변화 (*는 부분 정합 결과)

Table 14. Variation of system precision for 500 non-similar words along to error dependent loss value on the partial match (* indicates the partial matching system)

시스템	Best	10-best	정확도1	정확도2
개별 유사도	73.6	81.2	75.3	84.5
10:3.5:3.5	75.8	83.8	73.6	88.2
10:3.5:3.5*	79.0	87.20	81.5	89.8
10:7:7	76.4	84.6	77.4	87.0
10:7:7*	78.4	87.0	80.9	88.7

표 15. 1,000개의 비유사 단어를 지정한 경우 비교 가중치와 부분 정합에 의한 정확도 변화 (*는 부분 정합 결과)

Table 15. Variation of system precision for 1,000 non-similar words along to error dependent loss value on the partial match (* indicates the partial matching system)

시스템	Best	10-best	정확도1	정확도2
개별 유사도	73.6	81.2	75.3	84.5
10:3.5:3.5	70.0	77.6	74.2	85.4
10:3.5:3.5*	74.4	82.2	81.5	86.4
10:7:7	71.4	79.0	74.9	85.4
10:7:7*	74.4	82.2	78.3	86.0

비교 가중치와 부분 정합을 동시에 적용한 경우, 비교가중치 (<표 6, 7>)나 부분 정합만을 이용한 경우(<표 12, 13>)보다 성능이 향상됨을 알 수 있다. 비교 가중치 만을 적용한 경우 기준 시스템보다 성능이 하락하는 경우가 있었으나, 비교 가중치나 부분 정합을 이용한 경우에는 기준 시스템보다 성능이 향상됨을 알 수 있다. 이는 부분 정합에 의한 가중치의 효과가 큰 것으로 파악된다.

다음으로는 가중치의 효과를 살펴보기 위하여 부분정합과 순위 정보에 의한 가중치 적용을 동시에 적용하여 실험하였다. 서로 다른 성격의 가중치 효과를 살펴보기 위하여 비교 가중치

정보는 변경하지 않았다.

표 16. 500개의 비유사 단어를 지정한 경우 부분 정합과 순위 정보 가중치에 의한 정확도 변화

Table 16. Variation of system precision for 500 non-similar words based on partial match and ranking order weights

시스템	Best	10-best	정확도1	정확도2
음소 거리	80.6	88.8	87.8	89.0
개별 유사도	80.0	88.4	83.5	91.6

표 17. 1,000개의 비유사 단어를 지정한 경우 부분 정합과 순위 정보 가중치에 의한 정확도 변화

Table 17. Variation of system precision for 1000 non-similar words based on partial match and ranking order weights

시스템	Best	10-best	정확도1	정확도2
음소 거리	80.6	88.2	81.5	86.2
개별 유사도	77.8	85.8	81.5	86.7

실험결과 서로 다른 가중치를 적용하였을 경우, 음소거리 적용 방법이 인식 성능이 약간 더 높았으나, 개별 유사도 방식이 정확도2 측면에서 좋은 성능을 보였다. 이는 음소 위치에 따른 정확도를 반영한 개별 유사도 방식이 전체 후보 단어 열에서 연관성이 적은 후보를 정확히 검출할 수 있다는 것을 의미하지만, 동시에 정답 또한 같이 제거한다는 것을 알 수 있다.

6. 결론

본 논문에서는 선행 연구에서 제안되었던 비유사 단어 제외 방법 (정보 단말기에 널리 사용되는 대용량 고립단어 인식 시스템의 인식 결과로 제시되는 다중 후보 단어들 중에서 유사성이 낮은 단어들을 제거)을 개선하는 방법을 제안하였다. 제안된 방법은 1) 각 후보단어를 비교하는 과정에서 단어를 구성하는 음소열의 치환과 삽입, 삭제에 대한 비교 오류 손실 값을 차별화하여 오류 형태를 변화시키는 방법, 2) 음성 인식 시스템의 혼동 정보에 기초한 음소 정확도의 이용, 3) 우선 순위 정보를 이용한 가중치 적용, 4) 전체 후보 단어를 기준 단어로 설정하지 않고 일부 후보 단어를 기준 단어로 지정하는 부분 정합 방법을 제안하였다. 제안된 방법의 유효성을 확인하기 위한 실험에서 제안된 방법은 기존의 선행 연구보다 우수한 성능을 보였다. 그러나, 제안된 방식은 음성 인식 시스템의 성능에 따른 정보의 이용이나, 인식 대상 언어의 특성에 따라 오류의 형태가 바뀔 수 있으며, 사용된 거리 척도에 따른 효과가 다르게 나타날 수 있음을 알 수 있었다. 따라서 앞으로는 인식 시스템의 특성을 반영할 수 있는 변수의 설정이나 비유사 후보 단어의 제외 연구가 필요하다고 생각된다.

참고문헌

- [1] Yun, Y. S. (2008), "Exclusion of non-similar words form N-best word list of isolated word recognition system", Proceedings of Conference of KSPS and KASS, pp. 183-186, November
(윤영선, (2008). "고립단어 인식 시스템의 후보 단어열에서 비유사 후보단어의 제외 연구", 2008 대한음성학회 · 한국음성학회 공동학술대회 발표논문집, pp. 183-186)
- [2] Yun, Y. S., Kang, J. J. (2009), "Exclusion of non-similar candidates using positional accuracy based on Levenstein distance from N-best recognition results of isolated word recognition", Phonetics and Speech Sciences, vol. 1, no. 3, pp. 111-117, September
(윤영선, 강점자, (2009). "레벤스타인 거리에 기초한 위치 정확도를 이용한 고립 단어 인식 결과의 비유사 후보 단어 제외", 말소리와 음성과학, vol. 1, no. 3, pp. 111-117)
- [3] Park, J. G., Chung, H., Lee, Y. (2007), "Development of the Point-of-Interest Input System based on Large-vocabulary Embedded Speech Recognition", Proceedings of Conference of KSPS, pp. 108-111, November
(박전규, 정훈, 이윤근, (2007). "내장형 대어휘 음성인식 기술에 기반하는 행선지 입력 시스템 개발", 2007 대한음성학회 가을학술대회 발표논문집, pp. 108-111)
- [4] Ueffing, N., Macherey, K., Ney, H. (2003), "Confidence measures for statistical machine translation", Proceedings of MT Summit IX
- [5] Wikipedia (2009), Sequence Alignment, http://en.wikipeida.org/wiki/Sequence_alignment

• **윤영선 (Yun, Young-Sun)**, 교신저자
 한남대학교 정보통신공학과
 대전광역시 대덕구 오정동 133번지
 Tel: 42-629-7569 Fax: 042-629-7843
 Email: ysyun@hnu.kr
 관심분야: 음성인식, 음성모델링, 발화검증
 2001~현재 정보통신공학과 교수