韓國컴퓨터情報學會 論文誌
第15卷 第1號, 2010. 1.

2010-15-1-4-4

공유모델 인식 성능 향상을 위한 효율적인 연속 어휘 군집화 모델링

안 찬 식*, 오 상 엽**

Efficient Continuous Vocabulary Clustering Modeling for Tying Model Recognition Performance Improvement

Chan-Shik Ahn*, Sang-Yeob Oh**

요 약

연속 어휘 인식 시스템에서는 통계적 방법에 의한 어휘 인식을 수행하기 위하여 확률분포를 이용하며 이는 음소 단위의 클러스터링을 사용하여 모델링하여 샘플들을 기반으로 확률 파라미터를 추정한다. 어휘 검색 시 추정된 확률 파라미터로부터 인식 결과를 나타내는데 미리 정의되지 않은 음소와 추가되어진 음소로부터 인식률이 저하되는 문제점이 발생하며, 하나의 클러스터링으로 모델링하므로 가우시안 모델이 정확성을 확보하지 못한다는 단점이 있다. 이를 개선하기 위하여 확률 분포의 혼합 가우시안 모델을 최적화하여 유사도를 기반으로 Euclidean과 Bhattacharyya 거리 측정 방법을 혼합한 군집화 모델을 제안하고, 군집화된 모델에서 음소 단위로 확률 모델을 탐색할 수 있는 시스템을 모델링하였다. 본 논문에서 제안한 시스템을 적용한 결과 시스템 성능에서 어휘 종속 인식률은 98.63%, 어휘 독립 인식률은 97.91%의 인식률을 나타내었다.

Abstract

In continuous vocabulary recognition system by statistical method vocabulary recognition to be performed using probability distribution it also modeling using phoneme clustering for based sample probability parameter presume. When vocabulary search that low recognition rate problem happened in express vocabulary result from presumed probability parameter by not defined phoneme and insert phoneme and it has it's bad points of gaussian model the accuracy unsecure for one clustering modeling. To improve suggested probability distribution mixed gaussian model to optimized for based resemble Euclidean and Bhattacharyya distance measurement method mixed clustering modeling that system modeling for be searching phoneme probability model in clustered model. System performance as a result of represent vocabulary dependence recognition rate of 98.63%, vocabulary independence recognition rate of 97.91%.

▸ Keyword : 군집화 모델링(clustering modeling), 음소 모델링(phoneme modeling), 모델 공유(tying model), 가우시안 모델(gaussian model), 어휘 인식(vocabulary recognition)

• 제1저자 : 안찬식 교신저자 : 오상엽
• 투고일 : 2009. 11. 18, 심사일 : 2009. 11. 26, 게재확정일 : 2010. 01. 26.
* 광운대학교 컴퓨터공학과 박사과정 **경원대학교 IT대학 컴퓨터미디어학과 교수
※ 본 연구는 2010년도 경원대학교 지원에 의한 결과임.

Ⅰ. 서 론

연속 확률 분포 HMM(Hidden Markov Models)에 기반한 어휘 인식 시스템에서 높은 인식률을 얻기 위해서 문맥 종속 모델들과 혼합 가우시안 출력 확률 분포가 사용되고 있다.[1]

많은 모델 파라미터들로 인한 데이터 부족 문제가 수반되며 모델별 훈련 데이터의 수가 균일하지 않은 경우가 일반적이므로 CHMM(Continuous Hidden Markov Models) 기반 연속 어휘 인식 시스템을 위한 훈련용 데이터의 가용성과 모델의 복잡성 사이에 균형을 유지할 수 있는 방법으로 공유 방법이 사용된다.[2]

공유 방법은 모델들을 대상으로 하여 파라미터들을 공유할 수 있는 집합들을 결정하며 이러한 집합들을 사용함으로써 보다 강인한 모델들을 추정하며 군집화에 사용될 모델 파라미터들의 초기 추정치를 생성하기 위하여 각 문맥들에 대한 데이터가 반드시 필요하다.[3] 그러나 공유 모델은 검색 중에 나타나지만 훈련 중에는 나타나지 않는 데이터들에 대한 모델을 구성 할 수 없는 단점이 있다. 또한 미리 정의되지 않은 음소와 추가되어진 음소로부터 인식률이 저하되는 문제점이 발생하며, 하나의 클러스터링으로 모델링하므로 가우시안 모델이 정확성을 확보하지 못한다는 단점이 있다. 이러한 단점을 개선하기 위하여 효율적인 연속 어휘 인식 군집화 모델링을 제안하였다.

제안한 모델은 확률 분포의 혼합 가우시안 모델을 최적화하고 유사도를 기반으로 Euclidean과 Bhattacharyya거리 측정법을 결합한 혼합 거리 측정법을 사용하였다.

Euclidean 거리 측정은 연속 확률 밀도를 일반적인 벡터 값으로 구할 수 있는 장점이 있지만 가중치에 의한 벡터를 추정할 시 확률 값이 떨어지는 단점이 있다.

Bhattacharyya 거리 측정은 오류율 측정에 기반을 두고 비슷한 가중치를 갖는 가우시안의 조합이 쉽고 외부의 가우시안이 흡수되지 않는 장점이 있지만 오류율을 측정하여 계산이 복잡해진다는 단점이 있다.

본 논문에서는 두 거리 측정법을 혼합하여 연속 확률 밀도와 오류율을 측정하여 일반적인 벡터 값을 사용하고 비슷한 가중치를 갖는 가우시안의 조합이 쉽고 외부의 가우시안이 흡수되지 않는 방법을 사용하여 모델을 군집화하여 강인하면서 음성의 변화를 고려한 모델을 생성하였다.

본 논문에서 제안한 시스템을 적용한 결과 어휘 종속 인식률은 98.63%, 어휘 독립 인식률은 97.91%의 인식률을 나타내어 빠르고 좋은 인식성능을 확인하였다.

모델 공유로부터 탐색을 위한 알고리즘을 사용한 시스템을 실험한 결과 빠르고 좋은 인식성능을 확인하였다. 제2장에서는 모델 군집화의 기본구조를 설명하고, 제3장에서는 본 논문에서 제안한 시스템에 대하여 설명한다. 제4장에서는 제안한 시스템의 실험결과에 대하여 설명하고 제5장에서 결론을 기술한다.

Ⅱ. 기존 연구

2.1 HMM 어휘인식

HMM은 관측할 수 없는 "hidden" 과정와 음성 신호로부터 "hidden" 과정의 상태로 유도되는 음향학적 벡터를 연결하는 관측 과정으로 구성된다. HMM에서는 관측할 수 없는 음성의 통계적인 특성을 관측 가능한 벡터열을 통해 추정함으로서 음성의 통계적인 변이성을 반영한다.[4][5]

그림 1은 6-states HMM으로 모델링한 상태 천이가 가능한 음성모델을 나타낸다.

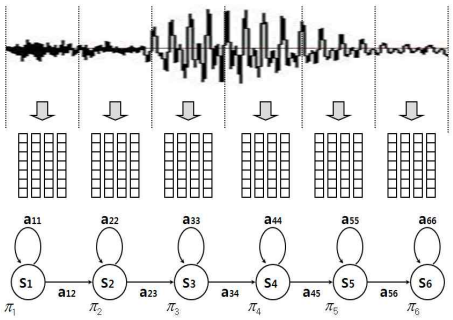

그림 1. 상태의 수가 6인 HMM
Fig. 1. HMM with 6 states

그림 1에서와 같이 HMM 모델은 음성 구간의 변이에 의한 상태천이 확률 A_i와 각 상태에서의 관측 심볼의 출력 확률 B_j, 초기 확률 π로 표현할 수 있으며 다음 식(1)과 같이 정리된다.

$$\lambda = (A, B, \pi) \quad \cdots\cdots\cdots\cdots \quad (1)$$

$$A = \{a_{ij}\}, \quad a_{ij} = p[q_{t+1} = j | q_t = i]$$

$$B = \{b_j(k)\}, \quad b_j(k) = p[o_t = v_k | q_t = j]$$

$$\pi = \{\pi_i\}, \quad \pi_i = p[q_1 = i]$$

HMM은 초기 $(t=0)$에 상태 i의 확률 $\pi_i = \Pr(s_0 = i)$, 상태 i에서 j로의 천이 확률 $a_{ij} = \Pr(s_t = i, s_{t+1} = j)$, 상태 j에

서 심볼 k를 관측할 확률 $b_j(k)=\Pr(x_t=k|s_t=j)$로 표현한다. 임의의 음성 특징벡터의 관측열 $O=(o_1,o_2,\ldots,o_T)$이 사실임을 가정할 때 주어진 N-states HMM 모델에서의 상태열이 $q=(q_1,q_2,\ldots,q_T)$라면 결국 관측열의 확률은 다음 식(2)와 같이 주어진다.[6]

$$P(O|q,\lambda)=\prod_{t=1}^{T}P(o_t|q_t,\lambda) \quad \cdots\cdots (2)$$
$$=\sum_{q\in Q}\pi_{q_1}b_{q_1}(o_1)a_{q_1q_2}b_{q_2}(o_2)\cdots a_{q_{T-1}q_T}b_{q_T}(o_T)$$
$$=\sum_{q\in Q}\prod_{t=1}^{T}a_{q_{t-1}q_t}b_{q_t}(o_t)$$

초기 상태 $t=1$에서 확률 $a_{s_0s_1}(\pi_{s_1})$로 천이가 시작되며, 관측 O_1는 출력 확률 $b_{s_1}(O_1)$로서 생성이 된다. 초기 상태 s_1에서 상태 s_2로의 천이는 천이 확률 $a_{s_1s_2}$로 이루어지며, 대응되는 상태 s_2에서의 관측 O_2를 생성될 확률은 $b_{b_2}(O_2)$가 된다. 이러한 과정은 상태 s_{T-1}에서 마지막 상태 s_T로 $a_{s_{T-1}s_T}$의 확률로 천이되어 기호 O_T를 출력 확률 $b_{s_T}(O_T)$로 생성할 때까지 계속된다. 이러한 과정을 정의에 의하여 직접 계산하면 모든 시간 $t=1,2,\cdots,T$에서는 진행 가능한 상태 수는 N개가 되어 계산의 복잡도는 $O(N^T)$이 된다.[7]

이러한 확률계산은 음성 구간에 따라 모델이 지수 함수적으로 증가하는 상태 열을 갖기 때문에 쉽게 계산할 수 없고 계산량이 지나치게 방대해지므로 전향, 후향 알고리즘을 이용하여 HMM 모델의 관측열의 확률을 추정한다.[8]

2.2 모델 공유

공유 구조를 이용하여 생성하는 주요 문제는 최소의 복잡성으로 데이터의 특성을 표현하는 일반적인 모델을 어떻게 얻느냐하는 것이다. 공유 구조를 사용함으로써 증가하는 파라미터를 줄일 수 있어 훈련의 효율성을 증대시킬 수 있고, 모델 파라미터의 수를 줄임으로써 인식에 필요한 계산량을 줄일 수 있다. 공유 수준에 따라 모델공유, 상태공유, 분포공유, 특징공유 수준으로 나눌 수 있다.[9]

결정트리 기반 상태공유 음향 모델링의 목적은 언어의 음향적 음성학적인 지식을 일정한 최대 확률 기법에 따라 모델에 효과적으로 포함시키는데 있다.[10]

그림 2는 상태를 공유하여 군집화하는 형태를 보여준다.

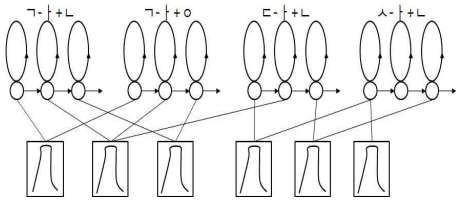

그림 2. 가우시안 상태 공유
Fig. 2. Gaussians Cluster tied- state

임의의 S를 HMM 상태들의 집합이라 하고 $L(S)$를 S의 log likelihood라 할때 집합 S에 있는 모든 상태들이 묶였다는 가정 하에 훈련 데이터의 프레임들의 집합인 F에서 생성된다고 하면, 공통 평균 $\mu(S)$과 분산 $\Sigma(S)$를 공유하게 된다. 또한 묶인 상태들의 상태 별 프레임의 정렬이 바뀌지 않는다고 가정하면 $L(S)$에 대해 다음 식(3)과 같은 근사식을 쓸 수 있다.[11]

$$L(S)=\sum_{f=F}\sum_{s=S}\log(\Pr(o_f;\mu(S)))^{*}G \quad \cdots\cdots (3)$$

(단, $G=\gamma_s(o_f)$)
G : 상태 s에 의해 생성되는 사후 확률
o_f : 관측 프레임

중심 음소로 갖는 HMM의 가우시안 분포들을 모아서 하나의 커다란 풀을 만들고 각 트라이폰의 관측확률은 공통의 풀에 속한 가우시안을 사용하고 가중치를 다르게 하여 트라이폰의 음향적인 특성을 반영한다. 공통의 가우시안을 구성하는 방법은 같은 중심음소를 갖는 음성 특징벡터를 모아서 군집화한다.[12]

2.3 Euclidean과 Bhattacharyya 거리 측정 방법

어휘 인식, 문자 또는 영상 인식과 같은 패턴인식에서 거리의 개념은 패턴들이 특징 공간상에서 서로 얼마나 떨어져 있는지를 통하여 패턴들 사이의 비슷한 정도를 측정하기 위한 기준으로 사용한다. 특징 공간상에서 매우 근접한 거리에 있는 두 패턴은 거의 동인한 특징을 가지므로 큰 유사도를 갖는다. 이러한 거리 측정 방법에는 Euclidean 거리, Bhattacharyya 거리 등이 사용되고 있다.[13]

Euclidean 거리 측정은 벡터를 추정하여 거리를 계산하는 측정 방법이며 Bhattacharyya 거리 측정은 오류율을 측정하여 거리를 계산하는 방법이다. 단순 거리 계산을 수행하는 방법이므로 실시간을 요구하는 인식과정에서는 일반적으로 동적 프로그램 기술인 Viterbi decoding 방법을 이용하여 상태경로의 변이와 최적의 모델을 추정하여 인식한다.[14]

전향, 후향 확률에 의한 연산을 이용하여 상태경로를 추적하는 경우 인식율이 다소 우수한 반면 주어진 모든 상태에서의 출력 심볼의 확률을 전부 추정하므로 계산량과 복잡도가 증가하게 된다. Viterbi decoding의 경우 전향, 후향 확률추정을 이용한 decoding에 비해 인식율이 다소 저하되나 연산에서의 부하를 월등히 감소시키므로 일반적인 인식과정에서 이용한다.[15]

그림 3은 주어진 HMM 모델에서 Viterbi 알고리듬에 의한 상태 천이도를 나타내고 있다.

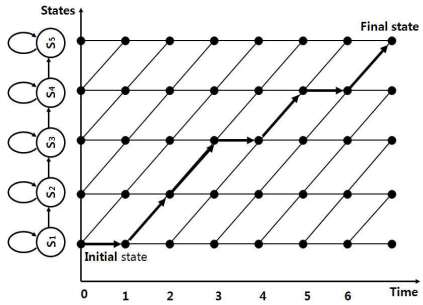

그림 3. 동적 프로그래밍으로서의 비터비 탐색도
Fig. 3. Viterbi beam searching as the dynamic programming

III. 효율적인 연속 어휘 군집화 모델링

3.1 시스템 모델

Euclidean 거리 측정은 가중치에 의한 벡터를 추정할 시 확률 값이 떨어지는 단점과 Bhattacharyya 거리 측정은 오류율을 측정하여 계산이 복잡해진다는 단점이 있다.

이러한 단점을 보완하기 위해 두 거리 측정법을 혼합하여 연속 확률 밀도와 오류율을 측정하여 일반적인 벡터 값을 사용하고 비슷한 가중치를 갖는 가우시안의 조합이 쉽고 외부의 가우시안이 흡수되지 않는 방법을 사용하여 모델을 군집화하여 강인하면서 음성의 변화를 고려한 모델을 그림 4와 같이 생성하였다.

군집하여 생성된 모델은 검색을 위한 음소단위 확률 모델을 지원하고 검색 시스템은 거리측정 방법을 사용하여 모델 유사도를 지원하여 결정 트리에 의해 검색 결과를 나타낸다. 사용자의 요구사항과 일치하는 검색에 실패했을 경우 다양한 검색 방법을 지원하여 검색 효율을 향상시켰으며 확률검색 모델을 이용하였다.

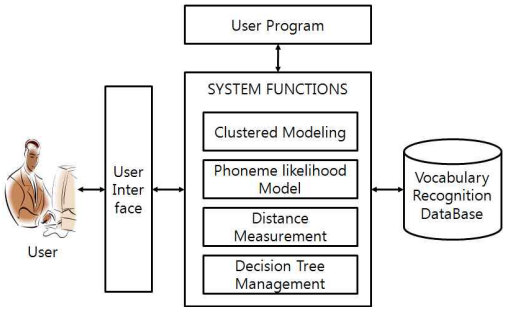

그림 4. 군집화 모델링 시스템 구성도
Fig. 4. Cluster modeling system architecture

3.2 군집화 모델링

단일 가우시안 출력 확률 밀도 함수를 갖는 3상태 단 음소 모델 초기 집합을 생성하고 훈련한다. 단 음소의 상태 출력 분포는 Baum-Welch 재 추정을 사용하여 훈련된 트라이폰 모델 집합을 초기화하기 위해 복사한다. 동일한 각 음소로부터 유도된 트라이폰들의 각 집합에 대해 대응되는 상태들을 군집화한다. 각 결과 군집에서 대표적인 상태가 선택되고 모든 군집 내의 상태들은 대표 상태로 묶이게 된다. 각 상태의 혼합 요소의 수를 증가시켜 재추정하여 모델의 정밀도를 향상시킨다.[16]

개선된 군집화를 위한 기본 알고리듬은 다음과 같다.

```
Begin
        Initialize the DistanceSearch
Job1 :  Find Nearest pair of gaussians
        If remain element then
        Get two element and calculate distance
        for if check list then
        add to list
        else next
        else goto job3

Job2 :  Nearest pair of gaussians
        Merge gaussians
        goto job3

Job3 :  If convergence then
         goto end
        else goto job1
end
```

가우시안들을 메모리에 로딩하고 가장 가까운 두 개의 가우시안을 찾는 과정을 반복한 후에 수렴하는지에 따라 계속할 것인지 종료할 것이지를 결정하게 된다. 가장 가까운 두 개의 가우시안을 찾는 과정은 가우시안 집합에 가우시안이 있는

지를 검사하고 남아 있으면 가장 가까운 두 개의 가우시안을 찾는다. 두 개의 가우시안을 찾으면 합성 목록에 추가한다.

3.3 거리 측정 알고리듬

거리 측정의 목적은 가장 비슷한 것을 찾아내는 것을 의미한다. 출력 확률 분포를 연속 확률 밀도로 갖는 가우시안의 경우, Euclidean 거리 측정법과 오류율 측정에 기반을 두고 있는 Bhattachayya 거리 측정법을 사용한다. 본 연구에서 사용한 거리 측정법은 Euclidean 거리 측정법과 Bhattachayya 거리 측정법을 혼합하여 사용하였으며 식(4)와 같다.

$$d(i,j)=\left[\frac{1}{n}\sum_{k=1}^{n}\frac{(\mu_{ik}-\mu_{jk})^2}{\sigma_{ik}-\sigma_{jk}}\right]^{\frac{1}{2}} \quad \cdots\cdots (4)$$

n은 데이터의 차수를 나타내고 μ_{ik}와 σ_{ik} 상태 s(i 혹은 j)의 가우시안 분포의 k번째 평균과 분산이다. 혼합 거리는 두 가우시안 사이의 겹치는 부분을 측정한다. 거리 범위는 0에서 ∞까지의 값을 가지며 각각 가우시안과 완전히 겹치거나 전혀 겹치지 않음을 나타낸다. 실제로 가우시안 간의 겹치는 부분이 조금이라도 있게 되면 거리 측정은 절대 ∞가 되지 않는다. 다음 식(5)는 혼합 거리 측정을 나타낸다.

$$B_{distance}=\frac{1}{8}(\vec{\mu_1}-\vec{\mu_2})^T\times\left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{-1}\times(\vec{\mu_1}-\vec{\mu_2})+\frac{1}{2}ln\frac{\left|\left(\frac{\Sigma_1+\Sigma_2}{2}\right)\right|}{\sqrt{|\Sigma_1||\Sigma_2|}} \quad \cdots\cdots (5)$$

μ_1과 μ_2은 각 분포의 평균이며 Σ_1과 Σ_2은 공분산이다. 혼합 거리는 비슷한 가중치를 갖는 가우시안들이 조합되기가 더 쉽기 때문에 축적된 형태로 구하게 된다.[17] 이것은 연속적으로 흡수하는 이웃 가우시안들로부터 공분산이 큰 단일 가우시안이 생겨나고, 외부의 것이 흡수되지 않는 것을 막아준다. 가중치 크기(weighting scalar) B_{scale}는 가우시안의 가중치 w_1과 w_2의 함수로 식(6)과 같이 나타낸다.

$$B_{scale}=\sqrt{\frac{w_1^2+w_2^2}{2w_1w_2}} \quad \cdots\cdots (6)$$

B_{scale}는 $w_1\rightarrow w_2$ 일때, $B_{scale}\rightarrow 1$이 된다. 반대로 $w_1\gg w_2$이거나 $w_1\ll w_2$일 때 $B_{scale}\rightarrow\infty$가 된다. 이것은 가우시안 간의 가중치 차이를 최소화한다.[18]

3.4 결정 트리 관리

음성학적 결정트리는 각 노드에 질의가 첨부된 이진 트리이다. 질의어들은 중심상태의 왼쪽이나 오른쪽으로 음성학적 문맥과 연관되어 있다. 이러한 트리는 훈련 데이터에 없는 트라이폰 모델을 합성해 낼 수 있고, 해당 트라이폰 모델의 문맥과 가장 비슷한 단말 트리 노드를 찾아 이와 연관된 공유상태들로 구성 할 수 있다.[19] 모든 질의는 왼쪽 혹은 오른쪽 음소가 집합 X의 원소인가?의 형태이다. 이때 집합 X는 비음(nasal), 마찰음(fricative), 모음(vowel) 등과 같은 광범위한 것에서 단일 집합 {l}, {m} 등과 같이 단순한 음성학적 분류로 이루어진다. 각 트리는 하향식으로 순차적 최적화 과정을 통해 생성된다.[20]

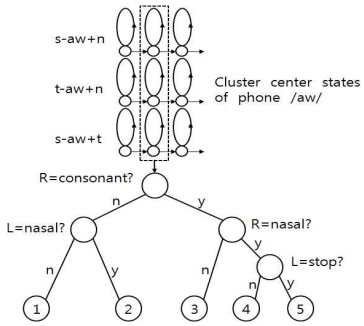

그림 5. 결정 트리 구성도
Fig. 5. Decision Tree architecture

처음에 군집화된 모든 상태들이 트리의 루트 노드로 대치되며, 그 노드내에서 모든 상태들이 공유되었다는 가정 하에 훈련 데이터의 로그 확률을 계산한다. 이 노드는 로그 확률을 최대로 증가시키는 부모 노드에서 상태들을 나누는 질의를 찾음으로써 두 개로 분할하게 된다.

IV. 실험결과 및 분석

본 논문에서 제안한 혼합 거리측정 방법의 객체 지향 검색 시스템의 성능 검증을 위하여 인식 실험을 수행하였다. 훈련 과정과 실험환경과의 불일치 문제를 해결하기 위해 잡음처리는 워너 필터를 사용하였다. 어휘 목록은 서울 시내의 지역명 200개, 지하철역명 100개로 구성하였다. 인식 실험에서는 실험에 참가한 화자가 어휘 목록을 5회 발음하여 총 1500단어를 대상으로 실험을 수행하였다. 어휘는 실내 환경과 잡음 환경에서 이동기기 등에 내장되어 있는 내장형 마이크로폰을 사용하여 16kHz Mono로 녹음 하였고, 16bit PCM 양자화를 사용하였다. 실험 어휘는 실내 10명, 실외 5명 등 총 15명의 성인 남성이 참가하였다. 실내 환경은 50~55dB이고, 실외 환경은

70~75dB의 소음환경 하에서 실험하였다.

표 1은 기존의 방식인 Euclidean, Bhattachayya를 이용한 인식 구조와 본 논문에서 제안한 혼합한 방법을 이용한 인식 구조 인식 속도에 관한 실험 결과이다.

표 1. 가우시안 수의 따른 인식률
Table 1. Gaussians Recognition Rated

가우시안 수	인식률(%)		
	Euclidean	Bhattachayya	Mixed
1만	98.23	98.75	98.97
2만	97.51	98.33	98.56
3만	96.93	97.51	97.89

표 2와 3은 기존의 방식인 Euclidean, Bhattachayya를 이용한 인식구조와 본 논문에서 제안한 혼합 거리측정 방법의 객체 지향 검색 시스템을 이용한 인식 구조를 실내 환경에서의 실험과 실외 환경에서의 실험을 나타낸다. 실내 환경은 50~55dB에서 실험 하였으며, 실외 환경은 70~75dB의 소음환경 하에서 실험하였다. 결과에서 보는 것과 같이 시스템 성능 평가 결과 어휘 종속 인식률은 98.63%, 어휘 독립 인식률은 97.91%의 인식률을 나타내었다.

표 2. 실내 환경 인식률
Table 2. Indoor Environment Recognition Rate

어휘	인식률(%)		
	Euclidean	Bhattachayya	Mixed
어휘 종속	97.51	93.31	98.31
어휘 독립	96.97	91.91	96.91

표 3. 실외 환경 인식률
Table 3. Outdoor Environment Recognition Rate

어휘	인식률(%)		
	Euclidean	Bhattachayya	Mixed
어휘 종속	91.01	89.21	91.11
어휘 독립	90.07	87.11	90.01

V. 결론

본 논문은 공유 모델링 방법인 결정트리 기반 상태공유 모델링을 기반으로 출력 확률 분포의 혼합 가우시안 수를 줄임으로써 모델을 최적화하여 유사도를 기반으로 Euclidean과 Bhattacharyya거리 측정 방법을 혼합하여 모델을 군집화하고 검색을 위한 음소 단위로 지원한다.

Euclidean 거리 측정은 연속 확률 밀도를 일반적인 벡터 값으로 구할 수 있는 장점과 Bhattacharyya 거리 측정의 오류율 측정에 기반을 두고 비슷한 가중치를 갖는 가우시안의 조합이 쉽고 외부의 가우시안이 흡수되지 않는 장점을 활용함으로써 모델의 군집화를 원활화 할 수 있었으며 벡터 값과 오류율을 조절함으로 원하는 결과를 얻을 수 있었다.

제안한 혼합 거리측정 방법의 검색 시스템으로 인하여 가우시안 상태의 모델을 최적화하여 혼합 모델을 군집화 함으로써 훈련 중에 나타나지 않는 모델에 대해 인식률을 향상 시킬 수 있는 장점을 확인하였으며 검색 시 속도와 인식률에서 기존 시스템보다 나은 결과를 얻을 수 있었다. 시스템 성능 평가 결과 어휘 종속 인식률은 98.63%, 어휘 독립 인식률은 97.91%의 인식률을 나타냈으며 검색 후 다양한 사용자 지원을 통하여 검색결과를 사용자에 위한 형태로 변형이 가능하였다. 따라서 혼합 거리측정 방법의 객체 지향 검색 시스템은 기존 인식 시스템의 높은 인식률을 그대로 적용할 수 있다.

참고문헌

[1] S. Young, D. Kershaw, J. Odell, D. Ollason, Valtcher, P. Woodland, "*The HTK Book*," Cambridge University Engineering Department, 2002.

[2] L. R. Rabiner, B. H. Juang, "*Fundamentals of speech recognition*," Prentice Hall, 1993.

[3] 안태옥, "*혼합 가우시안 군집화를 이용한 상태공유 음향모델 최적화*," 대한전자공학회 논문지, 제 42권, SP편 제 6호, 167-176쪽, 2005년 11월.

[4] D. Jurafsky and J. H. Martin, "*Speech and Language Processing*," Prentice-Hall, 2000.

[5] 우인성, 신좌철, 강홍순, 김석동, "다양한 연속밀도 함수를 갖는 HMM에 대한 우리말 음성인식에 관한 연구," 전기전자학회 논문지, 제11권, 제2호, 89-94쪽, 2007년 6월.

[6] 이호웅, 정희석, "지능형 홈네트워크 시스템을 위한 가변어휘 연속음성인식시스템에 관한 연구," 한국ITS학회 논문지, 제7권, 제2호, 37-42쪽, 2008년 4월.

[7] K. Demuynck, J. Duchateau, and D. Van Compernolle, "A static lexicon network repre- sentation for cross-word context dependent phones," In Proc. EUROSPEECH, Vol.1, pp.143-146, 1997.

[8] 조영수, 이기정, 김광태, 홍재근, “HMM을 이용한 한국어 음소인식,” 대한전자공학회 학술발표회 논문집, 제 16권, 제 1호, 81-84쪽, 1994년. 6월.

[9] M. F. Gales, "*Model-based techniques for nosie robust speech recognition*," Ph. D. dissertation, University of Cambridge, Sept, 1995.

[10] 안찬식, 오상엽, “MLHF 모델을 적용한 어휘 인식 탐색 최적화 시스템,” 한국컴퓨터정보논문지, 제 14권, 제 10호, 217-223쪽, 2009년 10월.

[11] A. S. Manos and V. W. Zue, "A study on out-of-vocabulary word modeling for a segment- based keyword spotting system," Master Thesis, MIT, 1996.

[12] 김광호, 임민규, 김지환, “지식베이스를 이용한 임베디드용 연속음성인식의 어휘 적용률 개선,” 대한음성학회지, 말소리, 제68호, 115-126쪽, 2008년 12월.

[13] 김동주, 김한우, "문맥가중치가 반영된 문장 유사도 척도," 전자공학회 논문지, 제 43권, 제 6호, 496-504쪽, 2006년. 3월

[14] 김기백, 최종호, “음성인식 기반 컨텐츠 네비게이션 시스템,” 한국컴퓨터정보학회지, 제 15권, 제 1호, 99-102쪽, 2007년 6월.

[15] S. Ortmanns, A. Eiden, H. Ney, and N. Coenen, "Look-ahead Techniques for Fast Beam Search," InProc. IEEE ICASSP-1997, pp. 1783-1786, 1997.

[16] Kris Demuynck, Tom Laureys, Dirk van Compernolle, and Hugo van Hamme, "FLaVor:a flexible architecture for LVCSR," In EUROSPEECH -2003, pp.1973-1976, 2003.

[17] Justin Zobel and Philip Dart, "Phonetic String Matching: Lessons from Information Retrival," SIGIR'96, pp.166-173, 1996.

[18] T. Jitsuhiro, S. Takatoshi, and K. Aikawa, "Rejection of out-of-vocabulary words using phoneme confidence likelihood," ICASSP, pp. 217-220, 1998.

[19] L. R. Bahl, P. V. deSouza, P. S. Gopalakrishnan, D. Nahamoo, and M. Picheny, "A Fast Match for Continuous Speech Recognition Using Allophonic Models," In Proc. IEEE ICASSP-92, Vol.1, pp.17-21, 1992.

[20] W. Daelemans, S. Buchholz, and J. Veenstra, "Memorybased shallow parsing," in Proc. CoNLL, pp.53-60, 1999.

저 자 소 개

안 찬 식

2002: 광운대학교 공학석사.
2002 - 현재: 광운대학교 박사과정
관심분야: 음성인식, 분산처리, 음성/음향 신호처리

오 상 엽

1999: 광운대학교 이학박사.
1998 - 현재: 경원대학교 IT대학 컴퓨터미디어학과 교수
관심분야: 소프트웨어공학, 버전관리, 소프트웨어재사용, 형상관리, 객체지향, 음성인식, 분산처리, 음성/음향 신호처리