

# 상태 공유와 결정트리 방법을 이용한 효율적인 문맥 종속 프로세스 모델링

안찬식<sup>†</sup>, 오상엽<sup>\*\*</sup>

## 요 약

HMM(Hidden Markov Model)을 사용하는 어휘 인식 시스템에서 인식 시 훈련 중에 나타나지 않는 모델들로 인해 인식률의 저하를 가져오며 인식 대상 어휘가 변경되거나 추가되면 데이터베이스의 수집과 훈련 과정을 수행하여 모델을 재생성해야 하고 그에 따른 시간과 추가 비용이 초래된다. 본 논문에서는 결정 트리 방법과 모델 공유 방법을 사용하여 효율적인 문맥 종속 프로세스 모델링 방법을 제안하였다. 제안한 방법은 생성된 모델들로부터 모델 공유 방법을 이용하여 모델의 재생성 과정을 줄이고 강인하고 정확한 문맥 종속 음향 모델링을 제공한다. 또한, 모델의 수를 줄이고 훈련 중에 나타나지 않는 모델들에 대해 문맥 종속 유사 음소 모델을 제공하여 훈련 중에 나타나지 않는 모델의 문제점을 해결하고 훈련성을 확보하였다. 제안된 방법으로 6종류의 음성 데이터베이스를 이용하여 어휘 종속 인식과 어휘 독립 인식 실험을 수행한 결과 어휘 종속 인식 실험에서는 98.01%의 성능을 보였고, 어휘 독립 인식 실험에서 97.38%의 성능을 보였다.

## Efficient context dependent process modeling using state tying and decision tree-based method

Chan Shik Ahn<sup>†</sup>, Sang Yeob Oh<sup>\*\*</sup>

## ABSTRACT

In vocabulary recognition systems based on HMM(Hidden Markov Model)s, training process unseen model bring on show a low recognition rate. If recognition vocabulary modify and make an addition then recreated modeling of executed database collected and training sequence on account of bring on additional expenses and take more time. This study suggest efficient context dependent process modeling method using decision tree-based state tying. On study suggest method is reduce recreated of model and it's offered that robustness and accuracy of context dependent acoustic modeling. Also reduce amount of model and offered training process unseen model as concerns context dependent a likely phoneme model has been used unseen model solve the matter. System performance as a result of represent vocabulary dependence recognition rate of 98.01%, vocabulary independence recognition rate of 97.38%.

**Key words:** vocabulary recognition(어휘인식), speech database(음성 데이터베이스), decision tree(결정 트리), independence vocabulary(가변 어휘)

※ 교신저자(Corresponding Author): 오상엽, 주소: 경기도 성남시 수정구 복정동(461-702), 전화: 031)750-5798, FAX: 02)426-9159, E-mail: syoh@kyungwon.ac.kr  
접수일: 2009년 11월 16일, 수정일: 2010년 1월 13일  
완료일: 2010년 3월 11일

<sup>†</sup> 정회원, 광운대학교 컴퓨터공학과  
(E-mail: csan1004@paran.com)

<sup>\*\*</sup> 정회원, 경원대학교 IT대학 컴퓨터미디어 교수

※ 본 연구는 2010년도 경원대학교 지원에 의한 결과임.

## 1. 서 론

기계에게 인간의 음성을 인식시키기 위한 연구는 언어학, 음성학, 음운학, 생리학, 해부학 등 다양한 학문적인 배경 하에서 진행되어져 왔으며, 최근 컴퓨터 하드웨어의 발전과 신호 처리 기술 및 알고리즘의 발전에 힘입어 단순한 실험적인 결과가 아닌 실용적인 측면에서 인식 시스템의 결과들을 활용하는 연구가 전 세계적으로 활발하게 진행되고 있다[1,2].

어휘 인식 시스템은 인식 대상 어휘가 고정적인 시스템으로써 인식 대상 어휘가 정해지면 인식 어휘에 대한 많은 사람의 발성을 수집하여 음성 데이터베이스를 구축하게 되고 구축된 음성 데이터베이스를 바탕으로 학습을 통하여 인식할 대상 어휘에 대한 음소와 유사 음소의 단위 모델을 생성하게 된다. 어휘 인식 시스템은 단위 모델을 인식하며 이미 생성된 단위 모델에 대해서는 인식 성능이 우수하나 인식 시 훈련 중에 나타나지 않는 모델들로 인해 인식률이 저하된다. 이러한 인식률 저하는 어휘 인식 시스템의 성능으로 연결되어지므로 모델들로부터 새로운 모델을 생성하기 위하여 모델 공유 방법을 사용하여 훈련 중에 나타나지 않는 모델들에 적용한다. 생성된 모델들로부터 모델 공유 방법을 이용하여 모델의 재 생성 과정을 줄이고 강인하고 정확한 문맥 종속 음향 모델링을 제공한다[3,4].

인식 대상 어휘가 변경되거나 추가되면 새로운 인식 어휘에 대해서 별도의 음성 데이터베이스의 수집과 훈련 과정을 반복적으로 수행하고 모델을 생성하여야 하므로 많은 시간과 추가 비용이 초래되는 문제점을 지니고 있다[5-7].

이러한 문제점을 극복하기 위한 인식 시스템은 인식 대상 어휘에 종속되지 않는 결정 트리 방법이 사용되며, 이것은 음운 현상을 충분히 반영한 음향 모델을 학습을 통하여 단위 모델로 학습하고 인식 어휘가 추가 및 변경되어도 인식한다[8].

본 논문에서는 많은 화자로부터 음운 현상이 충분히 반영된 음성 데이터베이스를 가지고 시스템의 훈련성을 확보하기 위해 상대 수준 공유 방법인 결정 트리 방법과 모델 기반 공유 방법을 사용하였다. 결정 트리 방법을 적용하여 음운 현상을 반영한 음향 모델을 단위 모델로 학습하고, 인식 어휘의 추가 및 변경작업을 용이하게 처리한다. 또한 스키마를 효율

적으로 적용하여 하나의 문맥 종속적인 음소가 주어졌을 때 루트노드에서부터 시작하여 현재의 문맥에 관한 질문의 결과에 따라 다음 노드를 선택하는 방식의 트리 순회를 통하여 하나의 모델을 선택하게 하는 공유 방법을 사용함으로써 모델 공유 시 나타나지 않는 모델로 인하여 인식 성능이 저하되는 훈련 중에 나타나지 않는 모델의 문제점 해결하고 문맥 종속적인 유사 음소 모델의 공유로 인하여 훈련성을 확보할 수 있다. 제안된 방법은 트리 및 의미론적 문맥 종속 프로세스 관리를 통해 결정 트리 방법을 보다 효율적으로 지원한다.

본 연구에서는 6종류의 음성 데이터베이스를 이용하여 어휘 종속 인식과 어휘 독립 인식 실험을 수행한 결과 어휘 종속 실험에서는 98.01%의 인식 성능을 보였고, 어휘 독립 실험에서 97.38%의 인식 성능을 보였다. 어휘 독립 인식실험 시 성능은 어휘 종속 인식실험에 비해 약간 떨어졌지만 스키마를 이용하여 가변 어휘를 사용할 수 있다는 큰 장점을 지니고 있다. 제 2장에서는 어휘 인식 시스템에 대해 간략히 소개하고, 제 3장에서는 본 논문에서 제안한 문맥 종속 프로세스 모델링에 대하여 설명한다. 제 4장에서는 제안한 시스템의 실험 결과에 대하여 설명하고 제 5장에서 결론을 맺는다.

## 2. 관련연구

### 2.1 어휘 인식

훈련 데이터베이스로부터 음소 단위의 HMM (Hidden Markov Model)을 학습하고 인식 단계에서는 발음 사전을 참조하여 음소 모델을 이용하고 인식하는 단어의 모델을 조립한다. “일”이라는 단어의 HMM은 “i”모델과 “l”모델을 연결함으로써 언어진다. 단어모델이 구성되면 입력된 특징 벡터 열에 대한 관측 확률을 계산하여 가장 확률이 높은 모델을 찾는다. 어휘 인식은 음향 모델로부터 발음 사전을 참조하여 단어 모델을 생성한다. 언어 모델에 의하여 주어지는 단어 네트워크의 각 단어를 단어 HMM으로 치환하고 단어 네트워크, 단어 모델, 음소 단위 HMM이 하나의 커다란 네트워크를 구성한다. 탐색 알고리즘은 입력된 특징 벡터 열에 대하여 가장 확률이 높은 단어 열을 찾음으로써 인식 결과를 구한다 [9,10].

탐색 알고리즘은 입력 음성 벡터의 확률을 최대로 하는 단어 열을 찾는 과정으로서 다음과 같은 수식으로 표현된다.

$$\hat{w} = \arg \max_W F(W|X) = \arg \max_W \frac{F(W)F(W|X)}{F(X)} \quad (1)$$

$F(W|X)$ 는 사후 확률이고  $F(W)$ 는 단어 열의 확률이며 언어모델이다.  $F(W|X)$ 는 단어 열이 주어질 때 음향 모델의 확률이고  $F(X)$ 는 관측 벡터 열  $X$ 의 확률로서 인식 과정에서는 단어 열에 영향을 미치지 않으므로 다음과 같이 나타낼 수 있다.

$$\hat{w} = \arg \max_W F(W)F(W|X) \quad (2)$$

문맥 독립 어휘 모델들로 구성되어진 인식 시스템에서는 문장을 구성하고 있는 어휘들의 대한 데이터 베이스가 어휘 모델별로 하나씩 존재하게 되어 인식 시 변형된 어휘에 대한 인식률이 떨어지게 된다. 하지만 문맥 종속 어휘 모델을 사용하게 되면 음성 발생에서 나타나는 규칙적인 변화를 인식 시스템에 반영할 수 있고 독립되어진 어휘 모델들을 종속시켜 공유함으로써 모델 데이터베이스의 수를 늘릴 수 있어 변형된 어휘에 대한 인식률을 향상시킬 수 있다.

### 2.2 문맥 종속 어휘 모델

질의어 집합은 음성이 혀의 위치나, 성대의 떨림, 입의 개폐의 정도, 그리고 구강 또는 비강을 거치느냐에 따라 구분하여, 음소로 나누어 구분한다. 질의어 집합을 어떻게 구분하느냐에 따라 어휘 인식에 매우 큰 영향을 미치게 되며, 질의어 집합의 질의 순서는 각 자소가 지니는 음가의 우선순위로 구분한다. 조음의 위치에 따라 비강이나 구강의 어느 쪽으로 공기가 흐르는가에 따라, 비음, 마찰음, 파찰음, 유음, 경음, 격음 등으로 구분하며, 성대의 떨림과 연계되는 발생 방법에 따라 유성음과 무성음을 구분한다[11].

문맥 종속의 정도에 따라 모델의 수는 급격하게 늘어나며 시스템 구현 시 triphone 음소 모델을 사용한다. triphone을 표기할 때 "a-b+c"와 같은 표현 방법을 사용하며 이것은 바로 앞과 뒤의 문맥이 a와 c인 음소 b를 의미한다.

문맥 종속 어휘 모델을 사용함으로써 음성 발생에서 나타나는 규칙적인 변화를 인식 시스템에 반영한다. 문맥 종속의 정도에 따라 monophone, diphone,

triphone, quenphone 음소 문맥 등이 사용된다. monophone 음소 문맥에서는 하나의 모델이 하나의 음소를 나타낸다. 45개의 음소 집합을 가질 경우 모두 45개의 모델들이 필요하게 된다. diphone 음소 문맥에서는 왼쪽 혹은 오른쪽 문맥과 함께 하나의 음소를 모델링하여, 모두 2,071개의 모델들이 있고, triphone 음소 문맥에는 한 음소를 모델링할 때 왼쪽과 오른쪽에 있는 특정 문맥을 함께 고려하여, 모두 95,221개의 모델들이 존재한다. 한 음소에 quenphone 음소 문맥은 한 음소에 대한 왼쪽 및 오른쪽 2개씩의 문맥을 함께 모델링한다[12].

### 2.3 문맥 종속 처리 스키마

스키마는 시스템내의 데이터가 조직되는 논리적인 구조, 계획 또는 방법으로서 기본적인 데이터 요소 또는 속성을 모델화한 것이다. 데이터 모델에는 모든 데이터 요소를 정의하여 어떤 정보를 필드로 나타낼 것인지, 필드는 얼마나 많은 문자를 지닐 수 있는지, 필드의 속성은 무엇으로 할 것인가를 정확히 기술한다. 문맥 종속 처리를 위한 데이터 모델을 구성할 때에는 데이터 모델이 무엇을 위해 사용될 것인가를 구체화하여야 하며, 조작과 사용의 요구에 대한 평가를 해야 한다. 어떤 스키마는 특정지역의 데이터 처리를 위해 구성되어 특정 지역 사용자만이 접근할 수 있도록 하는 반면, 여러 곳에서 데이터가 유지, 보수될 수 있도록 분산 데이터베이스를 제공하는 스키마도 있다. 한 회사 내에 원거리와 근거리 사용자를 모두 갖는 스키마라면 데이터베이스 관리자는 어떤 정보에 대해서는 모든 사용자가 접근할 수 있도록 구성해야 한다. 시스템 내의 모든 사용자에게 접근 가능한 스키마라고 한다면 이것은 총체적 데이터 요소(global data elements)로서 정의되며, 선택된 단말기에 의해 시스템내의 일부 사용자만이 접근하도록 제한된다면 이 스키마는 제한적 국지 데이터 요소(restricted local elements)로 정의된다. 4가지의 일반적인 데이터베이스 스키마는 리스트 구조, 트리 또는 계층구조, 네트워크 구조, 그리고 관계형 구조 등이 있다[13].

최근 문맥 종속 처리를 위한 스키마 기술의 발전으로 스키마의 속성 타임 스탬핑을 이용하는 분산(disjoint)시간 간격의 유한 집합인 시간 원소와 이력 릴레이션을 제한하며 각 속성 값은 시간 원소를 속성

의 정의역에 대응시키는 함수로 표현하고 있다. [14]에서는 처음으로 속성 수준에서 시간차원을 도입하였으며, 이 모델은 릴레이션 스키마의 각 속성과 릴레이션의 각 튜플에 대한 시간을 할당하는 방법을 사용한다. 스키마의 동적 변경은 단일 스키마 변경과 스키마 버저닝(schema versioning)이 있다. [15]의 연구에서는 스키마 변경에 대한 유형을 정의하고 각각에 대한 의미를 가지는 단일 스키마 변경 모델을 제시하였다. 스키마 버저닝은 단일 논리 스키마에 버전을 만드는 것으로 서로 다른 스키마 하에서 사용자들이 데이터베이스의 서로 다른 뷰(view)를 볼 수 있도록 해주며, 각 스키마 버전에 대한 스키마 변경을 독립적으로 적용해도 서로 영향을 주지 않는 방법이다.

### 3. 어휘인식 문맥 종속 프로세스 모델링

#### 3.1 시스템 모델

사용자의 음성 신호를 라이브러리에 등록된 다음 두 개의 triphone이 몇 개의 상태를 공유함으로써 모델 수준에서의 공유보다는 높은 융통성을 가지게 된다. 더 정교한 음향 모델링을 위하여 한 음성 모델에 대해 더 많은 숫자의 마코프(Markov) 상태들을 자유롭게 사용한다.

상태 숫자의 증가는 자유 파라미터의 개수를 증가시키게 되지만 구분되어야 할 상태들은 그대로 유지하면서 서로 중복되는 상태들은 동일한 출력 분포를 공유함으로써 하나의 상태로 줄인다. 상태를 줄임과 동시에 새로운 상태를 만들고 만들어진 상태들을 관리하며 인식 시에 부족한 상태들과 mis-average 문제를 보완하기 위해 스키마를 이용한 프로세스 관리를 통해 해결하였다.

본 연구에서 제안한 시스템 모델은 그림 1과 같으며 문맥 독립 어휘 모델을 상태 기반 공유를 수행하여 문맥 종속 어휘 모델로 구성한다. 상태 기반 공유를 수행하기 위해 클러스터링 과정을 거치게 되고 결정 트리를 생성하여 관리하게 되며 문맥 종속 처리 지원을 위한 스키마 변경 관리를 수행한다.

#### 3.2 문맥 종속 상태 공유

파라미터 공유는 공유 수준에 따라 모델 기반 공유와 상태 기반 공유로 나누어진다. 모델 기반 공유

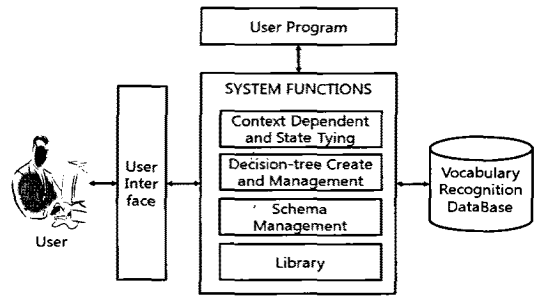


그림 1. 시스템 모델

는 모델이 제대로 훈련되지 않았거나 서로 닮은 경우에 비슷한 triphone들을 병합시켜 모델들을 더욱 많은 데이터로 사용하므로 시스템의 전체 크기는 줄이면서 더욱 정확한 모델을 생성한다. 모델 기반 공유는 왼쪽과 오른쪽의 문맥들을 독립적으로 처리할 수 없기 때문에 좌우의 문맥이 서로 다른 triphone들도 하나로 병합될 수 있는 mis-average 문제가 발생한다. 이러한 문제를 모델 수준의 공유보다 한 단계 낮은 수준인 상태 수준의 공유를 사용함으로써 해결한다. 모델별 상태의 개수와 공유되는 출력 분포들의 전체 개수를 훈련 데이터와 고려하여 적절한 수준으로 결정하여 스키마에서 관리할 수 있는 정교하고 강인한 모델을 생성한다[16].

상태 기반 공유는 상태 합병(merging)으로 이루어지는데 훈련 데이터베이스로부터 얻을 수 있는 모든 문맥 종속적인 모델을 생성하고 반복을 줄이기 위해 상태를 합병한다. 독립적으로 훈련된 모델들은 공유된 구조를 만들기 위해 비슷한 분포를 갖는 모델들로 클러스터링 된다.

상태 공유를 위해 표 1과 같이 HMM 상태 수만큼 합병을 수행하여 공유 블록을 만든다.

#### 3.3 결정 트리 생성 관리

공유 블록을 만든 후 결정 트리를 구성하며 그림 2와 같이 문맥 종속적인 모델에 대한 상태들의 결정을 통해 나온 최종 단 노드에 질의어에 대한 상태 공유 결과의 모델로 구성된다[8].

HMM 모델에 기반한 음성 인식 시스템에서 높은 인식률을 얻기 위해서 문맥 종속 모델들이 주로 사용된다. 많은 숫자의 모델 파라미터들로 인한 데이터 부족 문제가 필수적으로 수반되고, 모델별 훈련 데이터의 수가 균일하지 않은 경우가 일반적이다. 데이터

표 1. 음소 공유 블록(HMM 상태 5)

TB 350.0 "as_s2" ((as, *-as, *-as+*, as+*).state[2])
TB 350.0 "ae_s2" ((ae, *-ae, *-ae+*, ae+*).state[2])
TB 350.0 "ss_s2" ((ss, *-ss, *-ss+*, ss+*).state[2])
TB 350.0 "as_s4" ((as, *-as, *-as+*, as+*).state[4])
TB 350.0 "ae_s4" ((ae, *-ae, *-ae+*, ae+*).state[4])
TB 350.0 "ss_s4" ((ss, *-ss, *-ss+*, ss+*).state[4])

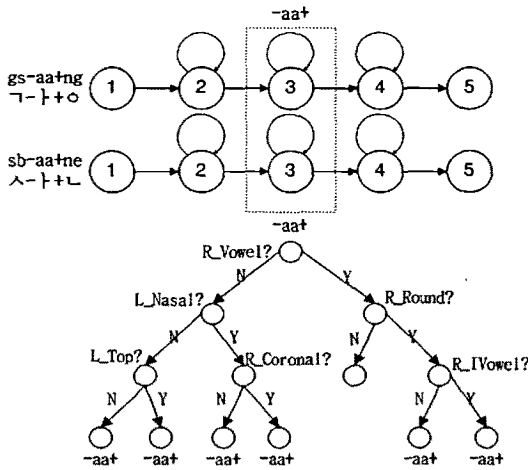


그림 2. Tri-phone /t/에 관한 결정트리

부족 문제는 단어 간 문맥 종속 모델이 사용될 때 더욱 심각해진다. 많은 숫자의 단어 간 triphone들로부터 추정해야할 모델의 숫자가 많아지며, 그 가운데 상당수의 triphone들은 훈련데이터의 발생 횟수가 매우 적거나 전혀 발생하지 않는 현상이 일어나게 된다. 특별한 응용을 위해 필요한 triphone의 전체 개수는 음소 집합과 사전에 따라 결정하게 된다.

### 3.4 문맥 종속 처리 지원을 위한 스키마 변경 관리

스키마는 시스템내의 데이터가 조직되는 논리적인 구조, 계획 또는 방법으로서, 데이터베이스 시스템에서 데이터 타입의 변경과 스키마의 일관성을 유지하기 위한 방법이 스키마 변경이다.

스키마 변경 관리의 어떤 시스템에 포함된 데이터베이스의 일부분을 변경할 필요가 생겼을 때 전체 시스템에 미치는 영향을 줄여야 한다. 이를 위해 프로그램, 설명 문서, 모델 베이스 등이 포함된 소프트웨어 시스템의 개발 관리 문제의 하나로 스키마 변경 관리가 다루어져 왔다. 스키마 변경 관리는 수작업에 의해 비효율을 줄이고 일관성 있는 변경 처리를 기하기 위해 자동화된 시스템을 필요로 한다. 대형 소프트웨어 시스템은 특히 많은 숫자의 개체 및 관계성들로 구성되고, 여러 명의 설계자와 프로그래머가 가진 지식과 기술이 집약되며, 변경 처리 자체가 일상적이면서 상당한 시간을 요하는 작업이 되기 때문에 효과적인 변경 관리 시스템을 필요로 한다.

객체지향 데이터베이스 시스템인 Orion, GemStone, Ode[16]등은 데이터베이스 스키마의 변경관리 기능을 제공하고 있다. 변경 관리 시스템의 기반 모형으로는 관계형(relational) 데이터 모형, E-R모형을 포함한 의미론적 데이터 모형, 객체지향 데이터 모형, 프레임(frame) 기반 지식 표현 모형 등이 사용되었다.

표 2는 기존 스키마 변경 관리 시스템들을 기반 모형의 종류에 따라 유형별로 분류한 것이다[17-20]. 본 연구는 표 2의 유형 A를 대상으로 적용하였다.

스키마 변경사항이 발생하였을 때는 언제든지 추출 작업을 수행하며, 스키마의 새로운 테이블의 삽입이 있을 시는 자동으로 버전을 부여하여 관리할 수 있도록 하였다. 또한, 시간 개념의 타임 스탬프에 의해 자동으로 추출 시간과 발생 시점의 버전을 관리한다. 이에 대한 삽입 알고리즘은 다음과 같다.

기존 스키마의 테이블이 삭제되었을 때는 추출 작업을 수행하여 스키마를 현재의 스키마에서 제거하고, 삭제된 스키마는 시간 개념을 사용하여 후에 다시 사용할 수 있도록 기존 BackUp\_Table에 저장한다.

표 2. 스키마 변경관리 시스템 유형

유형	변경대상 객체의 정의	변경처리 규칙의 정의	변경관리 시스템의 예
A	트리 및 의미론적 DB 스키마	시스템/응용 프로그램	Inscape. Version Server
B	관계형 DB 스키마	시스템/응용 프로그램	Gabdalf.HCMS
C	객체지향 DB 스키마	객체지향 DB 스키마, 시스템 프로그램	Orion.GemStone. Ode
D	프레임/규칙기반 지식표현 스키마	시스템/응용 프로그램	Orion.GemStone. Ode. RUBRIC

```

sub Insert_Schema()
  Set Current_Table = prodb.OpenTable("tbmain")
  Current_Table.Seek "=", Read_Table_Name

  If Current_Table.NoMatch Then
    BeginTrans
    Current_Table.AddNew
    Current_Table.tbsta = Current_Date
    Current_Table.tbver = Version + 1
    Current_Table.Update
    CommitTrans
  End If
  
```

다. 스키마 검색은 현재 존재하는 시스템에서 스키마 내용을 검색 처리하여 사용자에게 업무의 효율과 생산성 향상을 가져오게 한다. 다음의 스키마 검색 알고리즘은 인식 대상 어휘로부터 트리 및 의미론적 문맥 종속 프로세스 관리를 통해 결정 트리 방법을 보다 효율적으로 지원한다.

```

sub Display_Schema()
  Q = "SELETE * FROM tbmain"
  Q = "WHERE dbname = " Input_Db_Name
  Set Current_Table = prodb.CreateDynaSet(Q)
  Do Until Current_Table.EOF
    likelihood like(j,t)=like(i,t-1)+log aij+log b(j,xt)
    for all leaf nodes
    Consider transitions from leaf nodes to root node.
    like(j,t)=like(end,t)
    Record the previous node, entering time,
    likelihood for backtracking.
  end
  Display data
  Loop
  
```

3.5 시스템 프레임워크

인식 시스템을 구현하기 위하여 그림 3과 같이 음운 현상의 최대화를 통한 6종류의 단어 어휘 데이터베이스를 가지고 문맥 종속 음향 모델을 구축하였고, 최적의 HMM의 구조를 나타낼 수 있는 상태수와 Gaussian Mixture의 개수를 결정하였으며, Left-to-Right의 상태를 갖는 3가지의 상태의 유사 음소 단위의 문맥 종속적인 triphone을 바탕으로 인식 시스템을 구성하였으며 특징 파라미터로 12차의

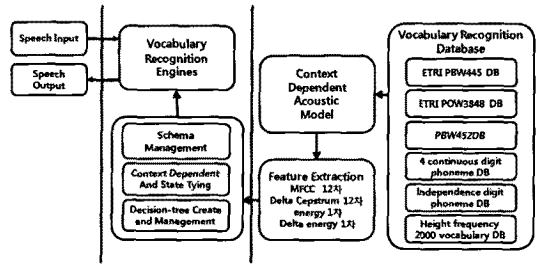


그림 3. 어휘 인식 데이터베이스 프레임워크

표 3. 인식 실험에 사용한 음성 DB

DB 종류	어휘수	어휘당 반복 횟수
PBW	445	40회
POW	3,848	10회
PBW	452	70회
4연속 숫자음	41	40회
단독 숫자음	35	70회
고빈도 2000DB	2,000	2회

MFCC와 각 캡스트럼의 미분형태인 12차의 Delta Cepstrum과 1차의 에너지, 1차의 Delta 에너지를 사용하여 총 26차의 특징을 추출하였다.

음소 모델 훈련을 위하여 표 3과 같이 총 6종류의 ETRI의 PBW445 DB, POW3848 DB, 국어공학연구소의 PBW452DB, 4연숫자음, 단독숫자음, 고빈도 2000어휘 DB를 사용하였다.

4. 실험결과 및 분석

본 논문에서 제안한 결정 트리 방법을 적용하여 어휘 인식 성능 실험을 수행하였다. 훈련 과정과 실험 환경과의 불일치 문제를 해결하기 위해 잡음처리는 위너(wiener) 필터를 사용하였다. 인식 실험은 단말기에서 사용되는 단어 20개를 선정하여 표 4와 같

표 4. 어휘 데이터

No.	단어	No.	단어	No.	단어	No.	단어
1	위	2	아래	3	시작	4	다음
5	종료	6	상위	7	하위	8	정지
9	크게	10	작게	11	확대	12	축소
13	취소	14	예	15	찾기	16	도착지
17	출발지	18	경로	19	확인	20	아니오

이 선별하였고 100단어씩 발음하여 총 2,000단어를 데이터로 구축하여 수행하였다.

음성은 실내 환경과 잡음 환경에서 이동기기 등에 내장되어 있는 내장형 마이크로폰을 사용하여 16kHz Mono로 녹음 하였고, 16bit PCM 양자화를 사용하였다. 실험 음성은 실내 10명, 실외 5명 등 총 15명의 성인 남성이 참가하였다. 실내 환경은 50~55dB이고, 실외 환경은 70~75dB의 소음 환경 하에서 실험하였다.

6종류의 단어 어휘 데이터베이스를 가지고 문맥 종속 음향 모델을 1,320개 구축하고 HMM의 구조를 나타내는 상태 3,960개를 구성하였으며 임계치(threshold)에 따라 상태 공유와 Gaussian Mixture의 개수를 결정하여 Left-to-Right의 상태를 갖는 3가지의 상태의 특징을 추출하여 캠브리지 대학에서 만든 HTK(Hidden Markov models toolKit)[21]를 사용하여 실험하였다.

표 5는 상태 수가 임계치에 따라 변화되는 결정트리 개수와 상태 공유의 결과를 보였다. 묵음을 포함한 음소 집합을 선정한 후 훈련용 데이터베이스에서 만들어진 Triphone을 가지고 실험한 결과 Triphone의 3,960개의 상태수를 가지고 임계치 10일 경우 상태 공유 수는 213개가 나타났고 결정트리 개수는 3,747로 나타났다. 인식률에 있어서는 94.39%로 가장 높게 나타났다. 임계치에 따라 상태 공유 개수와 인식률이 변화되는 형태를 보이고 있으며 임계치가 10일 때 가장 좋은 결과를 보이고 있고 임계치가 높아질수록 인식률이 저하되는 결과를 볼 수 있다.

표 6과 표 7은 제안한 알고리즘을 적용하여 어휘 종속 인식과 어휘 독립 인식 실험한 결과이다. 표 5에서 얻은 결과를 바탕으로 임계치를 50으로 적용하여 실험하였다. 어휘 종속 및 어휘 독립의 인식률을 살

표 5. 임계치에 따른 상태 공유

임계치	상태수 (개)	상태 공유	결정트리 (개)	인식률 (%)
0	3,960	0	3,960	93.12
10	3,960	213	3,747	94.39
50	3,960	754	3,206	94.35
100	3,960	2,031	1,929	94.31
150	3,960	2,980	980	94.04
200	3,960	3,585	375	94.13

표 6. 어휘 종속 인식 결과

DB종류	어휘수	인식된 어휘(개)	오인식된 어휘(개)	인식률 (%)
PBW	445	439	6	98.65
POW	3,848	3,808	40	98.96
PBW	452	443	9	98.53
4연속숫자음	41	35	6	93.03
단독숫자음	78	78	0	100.00
고빈도2000DB	299	295	4	98.89

표 7. 어휘 독립 인식 결과

DB종류	어휘수	인식된 어휘(개)	오인식된 어휘(개)	인식률 (%)
PBW	445	432	13	97.08
POW	3,848	3,800	48	98.75
PBW	452	437	15	96.68
고빈도2000DB	299	290	9	96.99

표 8. 어휘 인식률 비교

Euclidean		Bhattacharyya		제안 방법	
어휘	PBW 445	어휘	PBW 445	어휘	PBW 445
인식 어휘	421 개	인식 어휘	428 개	인식 어휘	432 개
오인식 어휘	24 개	오인식 어휘	17 개	오인식 어휘	13 개
인식률	94.61 %	인식률	96.18 %	인식률	97.08 %

펴보면 어휘 종속 실험에서 각 데이터베이스의 평균 98.01%의 인식률을 보이고 어휘 독립 실험에서는 평균 97.38% 인식 성능을 보였다. 표 8에서는 기존의 어휘 인식 방법인 Euclidean, Bhattacharyya 방법과 비교하여 실험한 결과 인식률은 각각 94.61%와 96.18%를 나타냈다. 제안한 방법이 Bhattacharyya 방법보다 0.9% 정도 낮은 성능을 보였다.

실험 결과 어휘 종속 환경에서의 인식률과 근사함을 확인할 수 있었다. 어휘 독립 환경에서도 어휘 종속 환경에 근접하는 인식 성능을 보임으로써 인식 대상 어휘가 제한되는 현재의 시스템에서 인식 대상 어휘를 가변적으로 사용할 수 있는 인식 시스템의 상용화에 효과적으로 대응할 수 있다는 것을 보여주고 있다.

## 5. 결 론

본 논문에서는 데이터베이스를 가지고 시스템의 훈련성을 확보하기 위해 상태 수준 공유 방법인 결정 트리 방법과 모델 기반 공유 방법을 사용하였고 트리 및 의미론적 문맥 종속 프로세스 관리를 통해 결정 트리 방법을 보다 효율적으로 지원하는 방법을 제안하여 임계치에 대한 상태 공유 수의 변화와 어휘 독립 실험 및 어휘 종속 실험을 수행하였다.

모델 공유 방법을 사용함으로써 훈련 중에 나타나지 않는 모델의 문제점 해결하고 문맥 종속적인 음소 음소 모델의 공유로 인하여 훈련성을 확보하였다. 결정 트리 방법을 사용하여 하나의 문맥 종속적인 음소가 주어졌을 때 루트 노드에서부터 시작하여 트리 순회를 통하여 하나의 모델을 선택할 수 있었으며 문맥 종속 프로세스 관리를 통해 결정 트리 방법을 보다 효율적으로 사용할 수 있었다.

6종류의 음성 데이터베이스를 가지고 어휘 종속 인식과 어휘 독립 인식 실험을 수행한 결과 어휘 종속 실험에서는 98.01%의 인식 성능을 보였고, 어휘 독립 실험에서 97.38%의 인식 성능을 보였다. 어휘 독립 인식 실험 시 성능은 어휘 종속 인식 실험에 비해 약간 떨어졌지만 스키마를 이용하여 가변 어휘를 사용할 수 있다는 큰 장점을 지니고 있다. 음소 집합의 Triphone 개수를 3,960개에서 결정 트리 기반 상태 공유를 통해 3,206개로 줄일 수 있었다. 임계치를 조절하여 실험할 경우 여러 결과를 얻을 수 있으며 더 좋은 성능의 결과를 나타낼 수 있다.

따라서 트리 및 의미론적 문맥 종속 프로세스 관리를 통하여 가변 어휘 인식 시스템에서 모델을 재수집하여 모델링하는 시간과 비용을 줄일 수 있으며 재 생성하는 과정을 모델 공유와 효과적인 결정 트리 방법으로 해결할 수 있었다.

## 참 고 문 헌

- [1] L. Rabiner, B. H. Juang, "Fundamentals of speech recognition," Prentice Hall, 1993.
- [2] L. R. Bahl, P. V. deSouza, P. S. Gopalarishnan, D. Nahamoo, and M. Picheny, "A Fast Match for Continuous Speech Recognition Using Allophonic Models," InProc. IEEE ICASSP-92, Vol.1, pp. 17-21, 1992.
- [3] 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.
- [4] 안태옥, "혼합 가우시안 군집화를 이용한 상태 공유 음향모델 최적화," 전자공학회논문지, 제 42권 SP편 제6호, pp. 167-176, 2005. 11.
- [5] 신광호, 정호열, 정현열, "음성 개선 기반의 모델 보상 기법을 이용한 강인한 잡음 음성 인식," 음향학회논문지, Vol.27, No.4, pp. 191-199, 2008. 5.
- [6] 방기덕, 강철호, "가변 신뢰도 문턱치를 사용한 미등록어 거절 알고리즘에 대한 연구," 멀티미디어학회논문지, Vol.11, No.11, pp. 1471-1479, 2008. 11.
- [7] W. Daelemans, S. Buchholz, and J. Veenstra, "Memorybased shallow parsing," in Proc. CoNLL, pp. 53-60, 1999.
- [8] W. Reichl and W. Chou, "Decision Tree State Tying Based on Segmental Clustering for Acoustic Modeling," Proc ICASSP, Seattle, pp. 801-804, 1998.
- [9] M. F. Gales, "Model-based techniques for noise robust speech recognition," Ph. D. dissertation, University of Cambridge, Sept, 1995.
- [10] D. Jurafsky and J. H. Martin, "Speech and Language Processing," Prentice-Hall, 2000.
- [11] 김동주, 김한우, "문맥가중치가 반영된 문장 유사도 척도," 전자공학회논문지, 제 43권 6호, pp. 496-504, 2006.
- [12] T. Jitsuhiro, S. Takatoshi, and K. Aikawa, "Rejection of out-of-vocabulary words using phoneme confidence likelihood," ICASSP, pp. 217-220, 1998.
- [13] Herry F. Korth, Abraham Silberschat., Database system concepts, 1991.
- [14] Tichy, W. F., "Rcs-A System for Version Control," Software Practice & Experience, Vol.15 No.7, pp. 637-654, 1985.
- [15] Hyoungjoo Kim, and Henery F. Korth. Schema Version and View a in Object-ori-



ented Database. Proc. INFO JAPAN. 1990.

[16] Agrawal, R. Buroff, S., Gehani, N. and Shasha, D. "Object Versioning in code," processing of 7th International Conference on Data Engineering, pp. 446-455, 1991.

[17] 김덕현, 박성주, "확장된 객체지향데이터 모형을 이용한 소프트웨어 변경관리 시스템," 정보과학회논문지, Vol.22, No.2. pp. 249-260, 1995.

[18] Keith E. Gorlen. "An Object-Oriented Class Library for C++ Program," Software-Practice and Experience. Vol.17, No.12. pp. 899-922. 1987.

[19] 박서영, 김갑수, 명선영, 신영길, 우치수, "객체지향 패러다임에서의 소프트웨어 컴포넌트 분류에 관한 연구," 정보과학회논문지, Vol.20, No.2, pp. 879-882. 1993.

[20] R. Helm, Y. S. Maarek, "Integrating Information Retrieval and Domain Specific Approaches for Browsing and Retrieval in Object Oriented Class Libraries," Proceeding of OOPSLA'91, pp. 47-61, 1991.

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



안 찬 식

2002년 광운대학교 컴퓨터공학과 석사  
 2004년 광운대학교 컴퓨터공학과 박사수료  
 관심분야 : 음성인식, 분산처리, 음성/음향 신호처리



오 상 엽

1991년 광운대학교 전자계산학과 석사  
 1999년 광운대학교 전자계산학과 박사  
 1993년~현재경원대학교 IT대학 컴퓨터미디어 교수

관심분야 : 소프트웨어공학, 버전관리, 소프트웨어 재사용, 형상관리, 객체지향, 음성인식, 분산처리, 음성/음향 신호처리