Optimal Decision Tree를 이용한 Unseen Model 추정방법*

김성탁(ICU), 김회린(ICU)

<차 례>

<Abstract>

Unseen Model Prediction using an Optimal Decision Tree

Sungtak Kim, Hoi-Rin Kim

Decision tree-based state tying has been proposed in recent years as the most popular approach for clustering the states of context-dependent hidden Markov model-based speech recognition. The aims of state tying is to reduce the number of free parameters and predict state probability distributions of unseen models. But, when doing state tying, the size of a decision tree is very important for word independent recognition. In this paper, we try to construct optimized decision tree based on the average of feature vectors in state pool and the number of seen modes. We observed that the proposed optimal decision tree is effective in predicting the state probability distribution of unseen models.

* Keywords: decision tree, unseen model, optimal tree

* 본 연구는 한국정보통신대학교 디지털미디어연구소의 정보통신연구개발사업의 연구비 지원에 의하여 수행되었음.

1. 서 론

HMM (Hidden Markov Model)을 적용하는 음성인식에서 파라미터 공유의 중요성은 자주 언급되어왔다[1]. 파라미터 공유의 목적은 음성인식 시스템의 파라미터 수를 줄여서, 제한된 훈련 데이터로도 강인한 모델 파라미터 추정을 가능토록 하는데 있다. 이러한 파라미터 공유에서 유의해야 할 점은 공유된 모델의 정밀성(precision)과 훈련성(trainability)에 적절한 균형을 유지하는 것이다[2]. 즉, 파라미터의 수를 줄이되 최고의 성능을 보장해 줘야 한다는 것이다. 특히 어휘독립음성인식(Vocabulary Independent Speech Recognition)을 수행할 경우 그 중요성은 더욱 커진다. 어휘독립음성인식을 위한 시스템 구축시 인식 단위를 triphone model로 사용할 경우 훈련데이터에 나타나지 않은 triphone model이 인식 환경에서 나타나는 경우가 많이 있다. 이런 경우 관측모델들의 상태 확률값을 기반으로 만들어진 decision tree[3]를 이용해 같은 leaf node로 분류되는 비관측모델들을 기존 관측모델의 상태 확률값을 이용해 추정을 한다[4]. 어휘독립인식에선 비관측모델 추정의 정확성에 따라 성능에 큰 영향을 받게 된다. 이때 decision tree의 크기에 따라 예측에 사용되어지는 관측모델들의 상태 확률값들의 강인성이 결정되므로 decision tree의 크기가 어휘독립인식 시스템의 성능을 결정하는 중요한 요인이 된다. 하지만 decision tree의 크기를 결정하기 위해서 필요한 정보를 사용자가 직접 결정하기에는 값의 범위를 정하기가 어려울 뿐만 아니라 정해야 할 값의 수가 너무 많은 어려움이 있다.

본 논문에서는 이러한 어려움을 극복하고 추가적인 성능 향상을 기하기 위한 효과적인 방법을 제안한다. 어휘독립인식 시스템에서 binary decision tree를 이용해서 비관측모델의 상태 확률값을 예측할 때, 성능에 많은 영향을 주는 decision tree의 크기를 관측모델들의 분포와 수에 따라 최적으로 구성하여 비관측모델들의 상태 확률값 예측 성능을 향상시켰다. 즉, 사용자가 직접 크기를 결정하지 않고 decision tree의 크기를 관측벡터들의 평균 log likelihood 값과 관측모델들의 수를 이용하여 decision tree의 크기를 결정함으로써 어휘독립인식에서 최적의 성능을 보장하는 optimal decision tree의 구성을 가능하게 하였다.

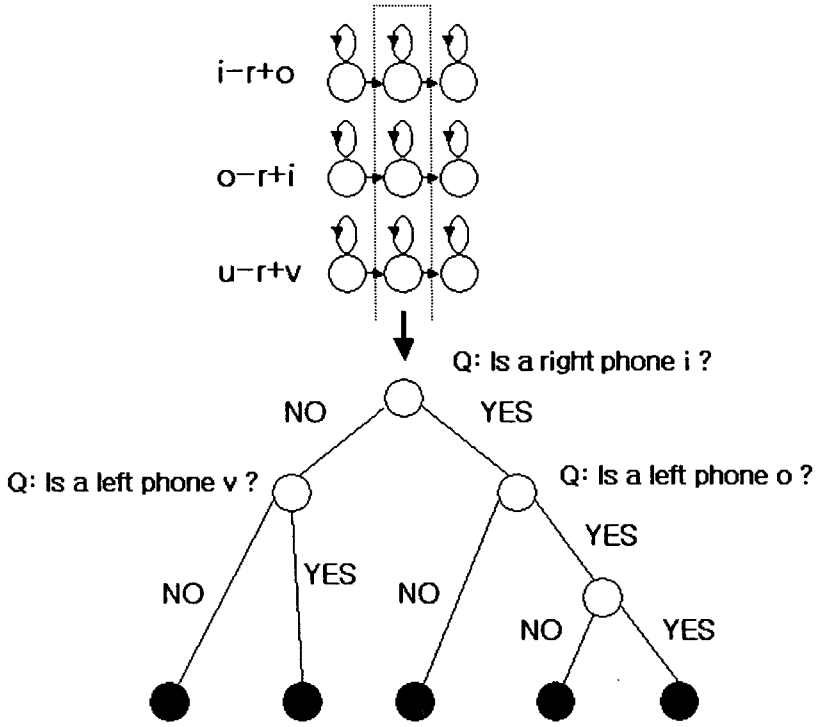
본 논문은 다음과 같은 순서로 기술된다. 2장에서는 binary decision tree의 간단한 설명과 decision tree에 사용되는 threshold 값인 log likelihood gain을 구하는 방법에 대해 기술하고, 3장에서는 optimal tree 구성을 위한 알고리즘을 기술한다. 4장에서는 실험 결과에 대한 소개를 하고, 마지막으로 5장에서 제안된 방법의 실효성 및 추후 연구 방향을 제시한다.

2. Binary decision tree

State tying 방법에는 크게 data-driven method와 binary decision tree method 가 있다. 그러나 data-driven method로 모델들의 state를 공유하면 관측모델의 파라미터수는 줄일 수 있지만, 비관측모델의 상태 확률값 예측이 불가능한 단점이 있다. 그러나 binary decision tree를 이용하면 관측모델들의 파라미터 수를 줄일 수 있을 뿐만 아니라, 비관측모델들의 상태 확률값을 관측모델들의 상태 확률값을 이용해 추정이 가능하다. 그래서 보다 우수한 어휘독립인식 시스템 구현이 가능하게 된다.

2.1. Binary decision tree 구조

일반적으로 binary decision tree를 사용하여 state tying을 수행할 때 triphone model의 state를 각 node에서 분류하기 위해 question set을 사용한다. 그리고 이 question set 중에서 정해진 척도(criterion)에 따라 가장 큰 값을 가지는 question을 선정하여 state를 분류한다. state tying을 위해 state pool을 구성하게 되는데 사용하는 state는 triphone model들 중에 중간 음소가 같은 state들을 사용한다. 정해진 척도에 따라 구성된 binary decision tree에서 각 leaf node에 있는 state들을 공유한다. 비관측모델의 상태 확률값은 관측모델들을 이용해서 결정된 binary decision tree의 각 node의 question에 따라 분류하여 leaf node에 있는 훈련모델들의 state들과 같은 값으로 예측한다. 여기서 question을 문맥상관 question을 사용하면 비관측모델들이 관측모델의 문맥에 따른 분류로 예측이 되어진다. decision tree를 구축시 구성을 중지하는 조건은 주어진 question set으로 더 이상 분류가 불가능할 때와 분류를 했을 때 주어진 threshold 값보다 작을 경우에 decision tree의 증가를 중지한다. <그림 1>은 binary decision tree 구조를 나타낸다.



<그림 1> Binary decision tree 구성 예

2.2. State tying 과정

Binary decision tree를 이용해 state 공유를 하는 목적은 파라미터의 수를 줄이는 데 있으므로 가장 비슷한 확률분포를 갖는 state들을 공유해야 한다. 그러므로 인식에서도 사용되어지는 log likelihood를 이용한 log likelihood gain을 이용한다. 각 node에서 log likelihood gain을 구하는 식은 식 (1)과 같고, 이때 LL은 log likelihood를 나타낸다[2].

(1) $$G(A,B) = (LL(A) + LL(B)) - LL(AB)$$
$$= \frac{1}{2}\left(n_A \sum_{d=1}^{p} \log\left[\frac{\sigma_{d,AB}}{\sigma_{d,A}}\right]^2 + n_B \sum_{d=1}^{p} \log\left[\frac{\sigma_{d,AB}}{\sigma_{d,B}}\right]^2 \right)$$

여기에서 AB, n_X 그리고 $\sigma_{d,X}$는 binary decision tree의 root node, node X에서

훈련벡터의 수, 그리고 node X에서 훈련벡터 d 차원에서의 분산을 나타낸다. 그리고 A와 B는 root node AB에서의 child node이다.

Log likelihood gain을 구하기 위한 식 (1)은 쉽게 state들이 가지고 있는 평균과 분산 그리고 state에 할당된 훈련벡터의 수를 가지고 다시 재구성이 가능하다. 각 node에서의 전체 훈련벡터들의 평균과 분산은 아래 식 (2)와 식 (3)으로 구할 수 있다.

(2) $$\tilde{\mu}_{d,X}=\frac{1}{n_X}\sum_{s\in X} n_s\mu_{s,d}$$

(3) $$\sigma^2_{d,X}=\frac{1}{n_X}\left(\sum_{s\in X} n_s\sigma^2_{s,d}+\sum_{s\in X} n_s\mu^2_{s,d}\right)-\left(\tilde{\mu}_{d,X}\right)^2$$

식 (2)에서 s는 state를 나타낸다. $\tilde{\mu}_{d,X}$, n_s 그리고 $\mu_{d,s}$는 각각 node X에서의 훈련벡터 d차원에서의 평균값, state에 할당된 특징벡터의 수 그리고 state s가 가지고 있는 훈련벡터 d차원에서의 평균벡터를 나타낸다. 식 (3)에서 $\sigma^2_{d,X}$ 그리고 $\sigma^2_{s,d}$는 node X에서의 분산과 state가 가지고 있는 분산을 나타낸다. 위의 식 (2)와 식 (3)을 이용함으로써 decision tree를 구성할 때 훈련데이터를 반복해서 사용하는 것을 피할 수 있다.

3. Unseen model 대처를 위한 optimal binary decision tree

Binary decision tree의 장점인 비관측모델의 상태 확률값 예측시, 문제가 되는 것은 관측모델의 수가 적을 때 비관측모델의 상태 확률값 예측에 이용되는 관측모델의 상태 확률값의 수가 작아져서 비관측모델의 대부분이 정확하게 예측되지 않는 경우가 발생한다. 그리고 시스템의 파라미터수를 줄이기 위해 decision tree의 크기를 제한하면 비관측모델의 정확도가 저하된다. 이런 문제를 해결하기 위해 본 논문에서는 비관측모델의 상태 확률값 예측을 위한 optimal binary decision tree 구성을 위한 방법을 제안한다. Optimal binary decision tree를 구성함으로써 파라미터의 수를 줄일 때, decision tree의 크기와 비관측모델의 상태 확률값 예측시 사용되어지는 관측모델의 상태 확률값의 수를 적절하게 사용할 수 있게 한다.

Unseen model의 적절한 예측을 위해선 관측모델의 state 확률값을 최대한 많이 이용하고, 또 관측벡터들의 분포를 참조하여 decision tree의 크기를 결정해야 한다. 그리고 어휘독립인식을 위한 비관측모델의 상태 확률값 예측시 관측모델의 상태

확률값을 100% 완전히 이용하게 되면 오히려 성능 저하가 일어나므로 decision tree의 크기가 최적화되어야 한다. 기존의 binary decision tree를 사용하여 state tying을 수행시 decision tree의 크기를 결정하는 값은 log likelihood gain을 threshold로 한다. 이때 훈련자가 직접 값을 결정해야 되는데 어려움이 많고 정확한 기준이 없는 것이 사실이다. 예를 들면 monophone의 종류가 47종이고 state 수가 3인 HMM을 사용할 경우, binary decision tree를 구성하기 위해 141개의 threshold 값를 결정해야 한다. 그래서 decision tree의 크기를 결정하는 threshold 값을 결정할 때 초기 decision tree의 root node에서 관측벡터들의 평균 log likelihood 값과 관측모델의 수를 이용하여 적절한 threshold 값을 정한다. 비관측모델의 적절한 예측을 위한 threshold 값을 구하는 식은 식 (4)와 같다

$$Threshold = LL_{average} \times N_{seen} \times \eta \tag{4}$$

$LL_{average}$ 와 N_{seen}은 state pool에서 관측 벡터들의 평균 log likelihood 값과 관측모델의 수를 나타낸다. 여기서 관측 벡터들의 평균 log likelihood 값을 사용함으로써 훈련데이터의 구성에 따른 민감도를 최소화할 수 있다. η는 weighting 값이다. 식 (4)를 보면 관측벡터들의 평균 log likelihood 값과 관측모델의 수와 threshold 값이 비례함을 볼 수 있다. 이것은 시스템의 파라미터 값을 줄이는 목표를 log likelihood 값이 크고 관측모델의 수가 많은 state pool에서는 시스템 파라미터의 수를 많이 줄이고, 그 반대의 경우는 적게 줄여서 decision tree의 크기를 최적화 할 수 있게 한다.

4. 실험 및 결과

4.1. 실험 데이터

실험을 위해 70명(남자: 38명, 여자: 32명)의 화자가 어휘내용이 다른 452 균일 음소 분포 단어(Phonetically Balanced Words, PBW)를 2회씩 발성한 데이터베이스를 훈련데이터로 하고, 2명(남자: 1명, 여자: 1명)의 화자가 1회씩 발성한 고빈도 2,000어절 중 PBW의 단어 452를 제외한 1,548개의 단어를 포함하는 데이터베이스를 테스트데이터로 사용하여 어휘독립인식실험을 수행하였다. 음성신호는 16kHz로 샘플링 되어있고 16bit로 양자화 되어있다.

음성신호는 10ms 단위의 프레임마다 총 39차 특징벡터로 표현하였다. 시스템에서의 39차 특징벡터는 12차 MFCC (Mel Frequency Cepstral Coefficients)와 로그에너

지, delta 및 delta-delta로 구성되어있다. 본 실험에서의 모델은 triphone을 사용하였고, 3 state의 left-to-right 방식의 연속밀도 HMM (Hidden Markov Model) 기반으로 하였다.

4.2. 실험 결과

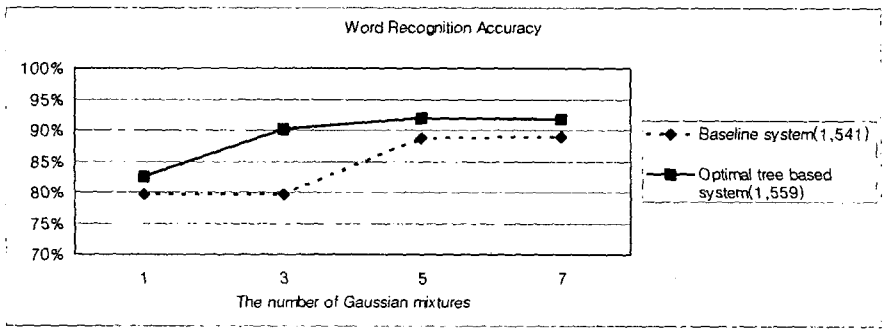
제안된 Optimal tree의 성능을 비교하기 위해 baseline system에서는 총 141개의 threshold 값인 log likelihood gain을 5,300, 5,600, 6,000 그리고 6,300으로 하여 state 수가 1,541, 1,469, 1,357 그리고 1,308개인 시스템을 구성하였다. 제안된 Optimal tree는 η의 값을 17, 19, 21 그리고 22로 하여 baseline system의 state 수와 상응하는 1,559, 1,433, 1,349 그리고 1,303개의 state를 가지는 decision tree을 구성하여 실험을 수행하였다. 그리고 사용되어진 question set은 47개의 monophone으로 구성되어진 94개의 question을 사용하여 실험을 수행하였다. State tying을 수행하기 전, 전체 state의 수는 6,408개이다. Baseline 어휘독립인식기에서 state의 수를 줄일수록 성능이 향상되다가 전체의 약 21%정도를 사용했을 때가 성능이 가장 높았다. 실험 결과는 <표 1>과 같다.

<표 1> Baseline system과 optimal tree based system의 state 수에 따른 성능

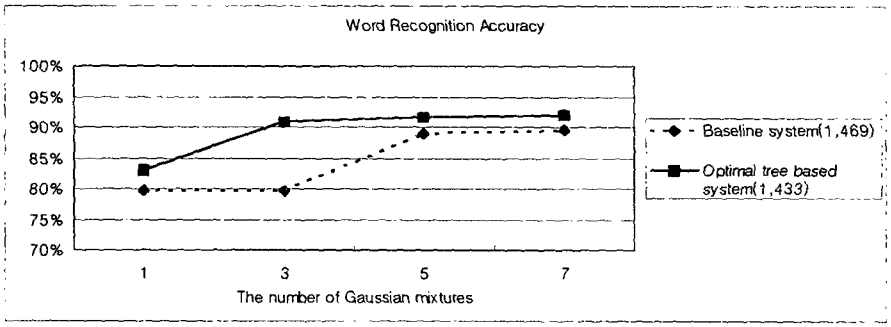
Tied state 수	Baseline		Tied state 수	Optimal tree	
	Mixture 수	인식률(%)		Mixture 수	인식률(%)
1,541	1	79.59	1,559	1	82.49
	3	79.59		3	90.25
	5	88.70		5	91.93
	7	88.89		7	91.67
1,469	1	79.59	1,433	1	83.07
	3	79.59		3	90.83
	5	89.02		5	91.73
	7	89.47		7	92.05
1,357	1	79.33	1,349	1	82.30
	3	79.33		3	90.57
	5	89.08		5	91.73
	7	89.47		7	91.73
1,308	1	79.07	1,303	1	82.30
	3	79.07		3	90.70
	5	88.50		5	91.47
	7	88.76		7	91.67

<표 1>을 보면 제안한 optimal tree를 사용할 경우 Gaussian mixture 수에 관계없

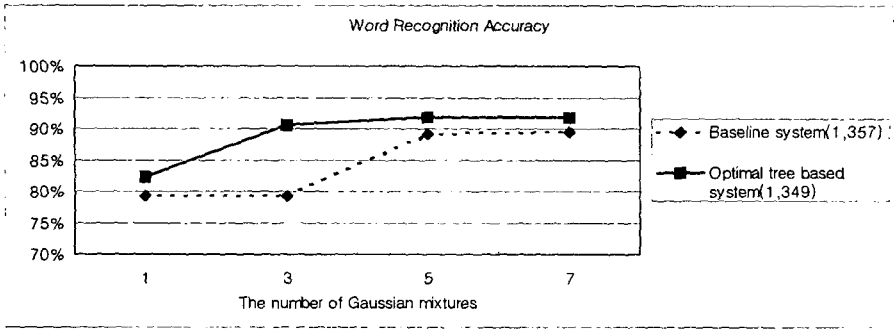
이 성능이 우수함을 보여준다. ERR (Error Reduction Rate)는 약 24%이다. 아래 그림들을 보면 제안된 optimal tree를 이용할 경우 성능이 향상되고, 또 Gaussian mixture가 3일 때부터 안정된 성능을 보장하는 반면 baseline system에서는 Gaussian mixture가 5일 때부터 안정된 성능을 보장한다.



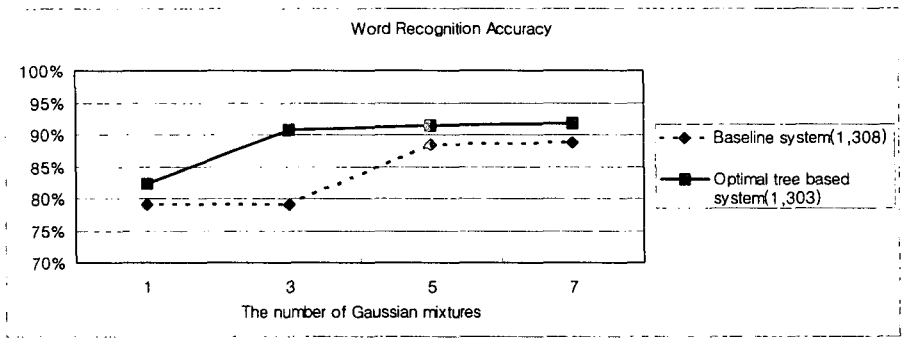
<그림 2> Baseline(state: 1,541)과 optimal tree(state: 1,559)의 결과



<그림 3> Baseline(state: 1,469)과 optimal tree(state: 1,433)의 결과



<그림 4> Baseline(state: 1,357)과 optimal tree(state: 1,349)의 결과



<그림 5> Baseline(state: 1,308)과 optimal tree(state: 1,303)의 결과

5. 결 론

본 논문에서는 어휘독립인식기에서 비관측모델의 효과적인 예측을 위해 주어진 관측모델들의 정보를 최대한, 그리고 효과적으로 이용하기 위해 optimal tree를 구축하는 방법을 제안하였다. 그 결과, 음소 47개를 이용하는 시스템을 구축할 때 사용되어지는 threshold 값 141개를 시스템 구축자가 직접 결정하지 않고, state pool에 있는 관측벡터들의 평균 log likelihood 값과 관측모델의 수를 이용하여 비관측모델들의 상태 확률값 예측이 쉽고 효과적인 optimal tree를 구축할 수 있었다. 이렇게 optimal tree를 구축함으로써 어휘독립음성인식의 성능 향상에 기여하였다.

향후에는 optimal tree 구축을 위해 사용한 weight 값인 η의 값을 조정하여 좀 더 관측모델에 적합한 tree를 구축하는 연구를 하고, 더 나아가 optimal tree를 구축하더라도 정확하게 예측이 되지 않는 비관측모델들의 상태 확률값들을 예측할 수 있는 알고리즘의 연구를 할 것이다.

참 고 문 헌

[1] S. Young, *The HTK BOOK (for HTK version 3.0)*, 2000.

[2] K. Beulen, H. Ney, "Automatic question generation for decision tree based state tying", in *proc. ICASSP*, pp.12-15, 1998.

[3] R. Reddy, A. Acero et al., *Spoken Language Processing: A guide to theory, algorithm & system development*. 2001.

[4] M-Y. Hwang, X. Huang, F. A. Alleva, "Predicting Unseen Triphones with Senones", *IEEE Trans. on Speech and Audio Processing*, Vol. 4, No. 6, pp.412-419, 1996.

접수일자: 2003년 2월 13일
수정일자: 2003년 3월 7일
게재결정: 2003년 3월 8일

▶ 김성탁(Sungtak Kim)
주소: 305-732 대전광역시 유성구 화암동 58-4번지 한국정보통신대학교 공학부
소속: 한국정보통신대학교 공학부 음성인식연구실
전화: 042) 866-6207
FAX: 042) 866-6245
E-mail: stkim@icu.ac.kr

▶ 김회린(Hoi-Rin Kim)
주소: 305-732 대전광역시 유성구 화암동 58-4번지 한국정보통신대학교 공학부
소속: 한국정보통신대학교 공학부 음성인식연구실
전화: 042) 866-6139
FAX: 042) 866-6245
E-mail: hrkim@icu.ac.kr