

# HMM을 기반으로 한 사전 확률의 문제점을 해결하기 위해 베이시안 기법 어휘 인식 모델에의 사후 확률을 융합한 잡음 제거

오상엽  
가천대학교 컴퓨터공학과

## Noise Removal using a Convergence of the posteriori probability of the Bayesian techniques vocabulary recognition model to solve the problems of the prior probability based on HMM

Sang-Yeob Oh

Dept. of Computer Media Convergence, Gachon University

**요 약** 사전 확률분포를 모델링하는 HMM을 사용하는 어휘 인식에서 인식 어휘의 모델들의 대한 인식 확률이 이산적인 분포를 나타내며 인식을 위한 계산량이 적은 장점이 있지만 인식률을 계산했을 때 상대적으로 낮은 단점이 있다. 이를 개선하기 위하여 베이시안 기법 어휘 인식 모델을 융합한 잡음 제거 인식을 향상할 방안을 제안한다. 본 논문은 베이시안 기법 어휘 인식을 위한 모델 구성을 베이시안 기법의 최적화한 인식 모델을 구성하였다. HMM을 기반으로 한 사전 확률 방법과 베이시안 기법인 사후확률을 융합하여 잡음을 제거하고 인식률을 향상시켰다. 본 논문에서 제안한 방법을 적용한 결과 어휘 인식률에서 98.1%의 인식률을 나타내었다.

**주제어** : 베이시안 기법, 어휘 인식, 인식 모델, 잡음 제거, 인식률

**Abstract** In vocabulary recognition using an HMM model which models the prior distribution for the observation of a discrete probability distribution indicates the advantages of low computational complexity, but relatively low recognition rate. The Bayesian techniques to improve vocabulary recognition model, it is proposed using a convergence of two methods to improve recognition noise-canceling recognition. In this paper, using a convergence of the prior probability method and techniques of Bayesian posterior probability based on HMM remove noise and improves the recognition rate. The result of applying the proposed method, the recognition rate of 97.9% in vocabulary recognition, respectively.

**Key Words** : Bayesian Method, Vocabulary Recognition, Recognition Model, Noise Removal, Recognition Rate

Received 20 June 2015, Revised 27 July 2015  
Accepted 20 August 2015  
Corresponding Author: SangYeob Oh  
(The University of Gachon)  
Email: syoh1234@gmail.com

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. 서론

신호처리 과정 중 예상하지 못한 잡음이 부가되거나 디지털 신호에서 기본적으로 양자화 잡음이 첨가되는 현상으로 인해 원 신호가 변경되거나 손상되어 인식을 저하되는 경우가 발생한다. 따라서 변경되거나 손상되어진 신호에서 잡음을 제거하여 원 신호로 복원하기 위한 연구가 진행되고 있으며, 잡음의 종류와 변경, 또는 손상과 첨가 되어진 경우의 다양한 형태로 인해 효과적인 잡음 제거 방법, 또한 다양하게 연구되고 있다[1].

잡음제거 방법은 공간 및 변화영역에서의 잡음 제거 방법으로 분류되며 잡음제거에 비교적 효과적인 비선형 추정 방법인 최대최소추정(minimax estimation)과 베이스 추정(Bayes estimation)을 사용한다[2].

베이지안 기법은 사전정보를 필요로 하며 베이스 정리를 적용하여 사용한다. 베이스 정리는 두 사상이 동시에 발생하는 결합 확률은 한 사상이 발생된 후 다른 사상이 발생할 조건부 확률과 주변 확률의 곱으로 사용된다[3]. 주변 확률은 사전 확률과 조건부 확률의 사후 확률로 표현된다. 일반적으로 신호처리 과정에서 사전 확률 분포는 알 수 없는 형태로 존재하므로 사전 확률을 모델링하여 베이스 정리에 의해 사후 확률 분포를 계산하여 원 신호를 추정한다[4,13,14,15].

사전 확률 분포를 모델링하기 위해 HMM(Hidden Markov Model) 기법이 널리 사용되어 지고 있다. HMM 기법이 널리 사용되어지는 이유는 음성과 문자 등에서 나타나는 많은 변형들을 흡수 할 수 있고, 시간에 따라 변해가는 특성을 지니는 데이터를 잘 모델링하며 파라미터의 수가 클수록 잘 모델링하기 때문이다. 하지만 HMM 기법은 인식 어휘에 대한 모델들을 구성하고 모델들의 관측 확률이 이산적인 분포를 나타내야 하며 구성된 모델만을 인식하여 인식이 상대적으로 낮은 단점이 있다. 이를 개선하기 위해 HMM을 기반으로 사후 확률을 이용한 베이지안 기법 모델을 이용한 잡음 제거 방법을 적용하여 인식을 향상을 위한 방법을 제안한다.

본 논문은 어휘 인식에서 베이지안 기법은 모수에 대한 정보와 모수에 대한 과거의 경험을 사전 분포로 표현하여 보다 정확한 모수의 불확실한 값에 대한 탐색을 통해 결과를 얻으며 이를 통해 잡음을 추정하여 인식을 향상시켰다. 어휘 구성을 위한 베이지안 기법 모델을 최

적화하여 인식 모델로 사용하였으며 어휘 인식 모델에 대한 음소 모델을 제안한 베이지안 기법 모델 방법을 실험한 결과 어휘 인식률에서 98.1%의 인식률을 나타내어 기존의 방법들에 비하여 인식 성능의 향상을 확인하였다.

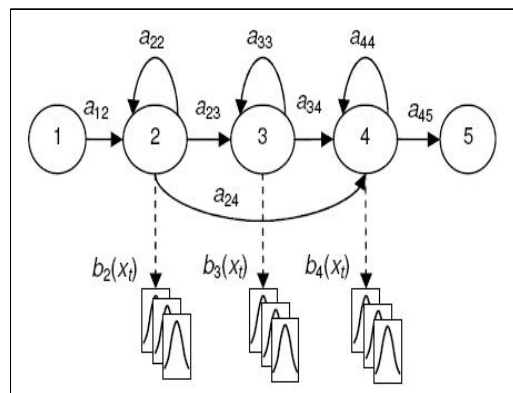
본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 언급하고 3장에서는 베이지안 기법 모델을 이용한 잡음 제거에 대해 설명하며, 4장에서는 성능 평가를 수행하고 마지막으로 5장에서 결론을 맺는다.

## 2. 관련 연구

### 2.1 HMM 모델

HMM 모델은 학습 과정에서 Markov 모델의 파라미터를 구하여 Markov 프로세스로 모델링한다. 관측이 불가능한 프로세스를 관측이 가능한 다른 프로세스를 통해 추정하는 이종 확률 처리 기법을 사용한다. 특히 인식 가능한 기준 패턴을 음소 단위로 음절 단위로 설정하여 인식하는 방법을 사용한다[5].

일련의 연속된 상태들로부터 이산 신호를 생성하는 확률 과정으로 표현되며 전이 확률에 따라 상태를 바꾼다. HMM 모델의 파라미터를 추정하기 위하여 각 모델을 위한 충분한 데이터가 있을 경우 강한 모델 구축이 가능하다. HMM 모델은 확률 밀도 함수를 갖는 상태의 모델을 생성하여 훈련된 모델로 사용된다. [Fig. 1]은 HMM 모델 생성을 나타낸다.



[Fig. 1] Hidden Markov Model

확률 밀도 함수는 n-차원을 가지는 특징 벡터  $x$ 를 확

를 변수로 확률 밀도 함수는 다음과 같이 나타낸다[6].

$$k(y) = \frac{1}{z^{-1}} \cdot \exp\left(-\frac{1}{2}(y - \mu_i, k) \sum_{i,k} (y - \mu_{i,k})\right) \quad (1)$$

$n$ 차원 확률 밀도 함수는  $n$ 차원 중심  $\mu$ 으로 표현되고,  $\mu$ 와  $z$ 는 파라미터로 표현된다[7].

### 2.2 잡음 제거

베이스 추정에는 사전 정보를 필요로 하며 베이스 정리를 적용하여 사용하며 두 사상이 동시에 발생하는 결합 확률은 한 사상이 발생된 후 다른 사상이 발생할 조건부 확률과 주변 확률의 곱으로 사용된다. 사전 확률을 모델링하여 베이스 정리에 의해 사후 확률 분포를 계산하여 원 신호를 추정하므로 잡음을 제거한다[8].

특정 확률 밀도 함수의 모수를 추정하기 위해서는 다양한 방법이 적용될 수 있지만 확률적인 접근론에 기반하기 때문에 모형의 모수는 임의계수로 취급된다. 따라서 고전 모형에서는 모형의 모수가 확률 분포로 추정되며 알려진 모델을 이용하여 알지 못하는 모델에 대한 학습을 반영하는 과정으로 수행된다. 확률 밀도 함수는 다음과 같이 표현된다[9].

$$p(\theta|x, x_{n+1}) \propto p(x_{n+1}|\theta) \cdot p(\theta|x) \quad (2)$$

$\theta$ 는 추정해야 하는 모수를 나타내고,  $x$ 는 알고 있는 모델을 나타낸다.  $p(\theta|x)$ 는 사후 분포(posterior distribution)이며,  $p(x|\theta)$ 는 주어진 모델 모수의 데이터의 함수인 우도 함수(likelihood function)이다.

모수의 추정은 모델로부터 얻은 모수에 대한 정보와 모수에 대한 과거의 경험 또는 주관을 사전 분포로 표현함으로써 보다 정확한 모수의 불확실성에 대한 추정을 수행한다.

## 3. 베이시안 기법 잡음 제거

### 3.1 베이시안 기법 인식 모델

베이시안 기법은 모수에 대한 정보와 모수에 대한 과거의 경험을 사전 분포로 표현하여 보다 정확한 모수의 불확실한 값에 대한 탐색을 통해 결과를 얻는다. 추정된

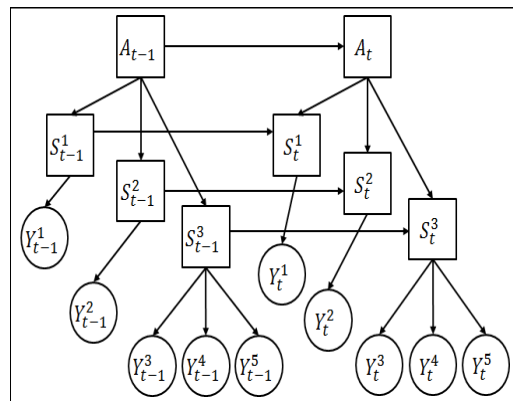
모수들의 평균과 같은 통계적 특성 값을 계산하여 구한다. 기존 통계학에서 사용되는 신뢰 구간(confidence interval)과 유사한 개념의 신용 구간(credible interval)을 구할 수 있고 추정된 모수 값으로 나타낼 수 있다. 모수의 개수  $n$ 과 유의수준  $\alpha$ 에서 베이시안은 다음과 같이 나타낸다.

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(x)} \quad (3)$$

$p(\theta|x)$ 는 사후 확률을 나타내며,  $p(\theta)$ 는 사전 확률나타낸다.  $p(x|\theta)$ 우도를 나타내며,  $p(x)$ 는 관찰값을 나타낸다. 베이시안의 정리를 확률 밀도 함수로 나타내면 다음과 같이 표현된다.

$$p(\theta|x_1, x_2, \dots, x_n) = \frac{1}{Z} p(\theta) \prod_{i=1}^n p(x_i|\theta) \quad (4)$$

$p(\theta|x_1, x_2, \dots, x_n)$ 는 사후 분포(posterior distribution)를 나타내고  $p(\theta)$ 는 사전 분포(prior distribution)를 나타낸다.  $\frac{1}{Z}$ 은 주변 분포(marginal distribution)를 나타내고,  $p(x_i|\theta)$ 는 발생 가능한 우도 함수(likelihood function)를 나타낸다. [Fig. 2]는 베이시안 기법 인식 모델 생성을 나타낸다.



[Fig. 2] Bayesian Method Model

베이시안 기법 인식 모델은 전이 확률에 따라 상태를 바꾸며 특정 상태는 출력 확률에 따라 하나의 관측을 만든다. 인식 모델의 파라미터를 추정하기 위하여 카테고리 정보가 있는 음성 데이터베이스를 사용하며 각 모델

을 위한 충분한 데이터가 있을 경우 실제 음성에 존재하는 다양성을 잘 표현할 수 있는 강인한 모델링이 가능하다.

### 3.2 잡음 편차 추정을 이용한 잡음 제거

가우시안 인식 모델은 주어진 표본 데이터 집합의 분포 밀도를 하나의 확률 밀도 함수로 모델링하는 방법을 사용한다. 하나의 가우시안으로는 모델링 할 수 없는 복수 개의 중심점을 가지는 1차원 데이터와 2차원 환형 데이터에 대하여 견고하게 모델링된다. 복수 개의 가우시안 확률 밀도 함수로 데이터의 분포를 모델링하므로 밀도 추정 방법이라 한다.

확률 밀도 함수는  $M$ 개의 가우시안 확률 밀도 함수의 선형 결합으로 다음과 같이 나타낸다.

$$p(x|\theta) = \sum_{i=1}^M p(x|\omega_i, \theta_i) P(\omega_i) \quad (5)$$

$p(x|\omega_i, \theta_i)$ 는 데이터  $x$ 에 대하여  $\omega_i$ 번째 성분 파라미터  $\theta_i$ 로 이루어진 확률 밀도 함수를 의미하며,  $P(\omega_i)$ 는 혼합 가중치를 의미한다. 잡음 추정 잡음 제거를 위해 잡음 추정을 위한 확률 분포를 사용한다. 모델의 추정을 위한 확률 분포를 다음과 같이 나타낸다.

$$P_x(x) = C \cdot \exp\left(-\left|\frac{x}{\alpha}\right|\right)^\beta \quad (6)$$

좌우 대칭 분포이며 평균이 0이고  $\alpha$ 와  $\beta$ 를 모수로 갖는 특징이 있다. 변환에 의한 잡음 편차 추정은 잡음 신호를 log-normal 분포로 가정하여 추정된 잡음 분산으로 다음과 같이 나타낸다.

$$\sigma_x^2 = \ln\left(1 + \frac{\sigma_y^2}{\mu_y^2}\right) \quad (7)$$

$\sigma_y^2$ 는 잡음 신호 대역의 분산의 제곱을 나타내고,  $\mu_y^2$ 은 잡음 신호 대역의 평균의 제곱을 나타낸다.

$$\sigma = \omega \sigma_x \quad (8)$$

잡음 신호 분포는 다양한 확률 분포를 나타내므로 추정된 잡음 편차를 보정하기 위해 가중 편차를 사용한다.

## 4. 실험 결과

본 연구에서는 잡음 음성 인식 실험을 위하여 Aurora 2 데이터베이스를 사용하였다[10]. 인식 실험에는 잡음인호(Subway, Babble, Car, Exhibition)들로부터 생성된 잡음 음성으로 구성하였다. 각각의 모델들로부터 특징 벡터를 추출하여 39차의 특징 벡터를 사용하였다. 음향 모델의 훈련과 인식을 위해서는 HTK(Hidden Markov Toolkit)를 사용하였다[11].

본 논문에서 제안한 베이시안 기법 모델 최적화를 이용한 잡음 제거 인식을 항상 방법의 성능 검증을 위하여 어휘 인식 실험을 수행하였다. 훈련 과정과 실험 환경과의 불일치 문제를 해결하기 위해 잡음 처리는 워너 필터를 사용하였으며 음성 인식 목록은 서울 시내 지역명 30개, 지하철역명 30개로 구성하였다. 인식 실험에서는 실험에 참가한 화자가 음성 인식 목록을 3회 발음하여 총 300단어를 대상으로 실험을 수행하였다[12].

제안한 시스템의 성능 평가를 위하여 기존 방식과 비교 실험을 하였으며 유클리디안 알고리즘, DTW 알고리즘[13], 제안 방법에 대해 인식률을 측정한 결과 제안 방법에서 인식률이 97.9%의 성능을 보였다.

<Table 1>은 기존 방식인 HMM 알고리즘, CHMM 알고리즘, 제안 방법을 실내 환경에서의 실험을 나타낸다. 표 1에서 보는 것과 같이 잡음이 거의 없는 실내 환경에서 실험한 결과 HMM 알고리즘을 이용한 음성 인식을 평균 96.9%로 나타냈으며 CHMM 알고리즘을 이용한 음성 인식을 평균 96.2%의 인식을 나타내었고 제안방법의 인식률 평균 98.1%를 나타내었다.

<Table 1> Non-Noise Environment Recognition Rate

Speech	Recognition Rate (%)		
	HMM	CHMM	Proposed Method
Speech Dependent	96.5	97.1	98.4
	95.9	96.3	97.6
	96.3	97.3	98.2
Average	96.2	96.9	98.1
Speech Independent	94.3	95.7	96.1
	93.8	94.6	95.2
	94.8	95.6	96.3
Average	94.3	95.3	95.9

<Table 2>는 기존 방식인 HMM 알고리즘, CHMM

알고리즘, 제한 방법을 실외 환경에서의 실험을 나타낸다.

<Table 2>에서는 잡음 환경에서 실험한 결과 HMM 알고리즘을 이용한 음성 인식을 평균 83.4%로 나타났으며 CHMM 알고리즘을 이용한 음성 인식을 평균 85.3%의 인식을 나타내었고 제안방법의 인식을 평균 86.0%를 나타내었다.

<Table 2> Noise Environment Recognition Rate

Speech	Recognition Rate (%)		
	HMM	CHMM	Proposed Method
Speech Dependent	83.6	85.3	86.1
	83.4	85.1	85.7
	83.1	85.6	86.2
Average	83.4	85.3	86.0
Speech Independent	82.5	83.4	85.2
	81.6	82.9	84.7
	82.4	83.2	85.3
Average	82.2	83.2	85.1

## 5. 결론

본 논문은 어휘 인식은 인식을 위한 모델을 구성하여 구성된 인식 모델들의 대한 관측 확률이 이산적인 분포를 나타내도록 한다. 이산적인 분포는 계산량이 적은 장점을 가지고 있지만 인식률은 상대적으로 낮은 단점을 가지고 있다. 이러한 단점을 보완하기 위해 베이시안 기법을 이용한 모델 최적화를 제안한다. 모델을 최적화한 인식 모델을 베이시안 기법인 사전확률과 사후확률을 계산하여 인식 모델을 추론하였다. 이를 확인하기 위해 HMM, CHMM 알고리즘과 비교하였다. 실내 환경에서 실험한 결과 HMM 알고리즘을 이용한 음성 인식을 평균 96.2%로 나타났으며 CHMM 알고리즘을 이용한 음성 인식을 평균 96.9%의 인식을 나타내었고 제안방법의 인식을 평균 98.1%를 나타내었다. 또한 실외 환경(잡음 환경)에서 실험한 결과 HMM 알고리즘을 이용한 음성 인식을 평균 83.4%로 나타났으며 CHMM 알고리즘을 이용한 음성 인식을 평균 85.3%의 인식을 나타내었고 제안방법의 인식을 평균 86.0%를 나타내었다.

## REFERENCES

- [1] Chan-Shik Ahn, Sang-Yeob Oh. Gaussian Model Optimization using Configuration Thread Control In CHMM Vocabulary Recognition. The Journal of Digital Policy and Management. Vol. 10, No. 7, pp. 167-172, 2012.
- [2] Chan-Shik Ahn, Sang-Yeob Oh. Echo Noise Robust HMM Learning Model using Average Estimator LMS Algorithm. The Journal of Digital Policy and Management. Vol. 10, No. 10, pp. 277-282, 2012.
- [3] Chan-Shik Ahn, Sang-Yeob Oh. Efficient Continuous Vocabulary Clustering Modeling for Tying Model Recognition Performance Improvement. Journal of the Korea Society of Computer and Information. Vol. 15, No. 1, pp. 177-183, 2010.
- [4] Chan-Shik Ahn, Sang-Yeob Oh. CHMM Modeling using LMS Algorithm for Continuous Speech Recognition Improvement. The Journal of digital policy and management. Vol. 10, No. 11, pp. 377-382, 2012.
- [5] Chan-Shik Ahn, Sang-Yeob Oh. Vocabulary Recognition Retrieval Optimized System using MLHF Model . Journal of the Korea Society of Computer and Information. Vol. 14, No. 10, pp. 217-223, 2009.
- [6] A. Srinivasan, Speech Recognition Using Hidden Markov Model, Applied Mathematical Sciences, vol. 5, no. 79, pp. 3943-3948, 2011.
- [7] S. M. Naqvi, M. Yu, J. A. Chamber. A Multimodal Approach to Blind Source Separation of Moving Sources. IEEE Trans. Signal Processing. Vol. 4, No. 5, pp. 895-910, 2010.
- [8] Beaufays, F., Vanhoucke, V., & Strope, B. Unsupervised discovery and training of maximally dissimilar cluster models. Proc. Interspeech, pp. 66-69, 2010.
- [9] Sang-Yeob Oh. Improving Phoneme Recognition based on Gaussian Model using Bhattacharyya Distance Measurement Method. Journal of Korea

- Multimedia Society. Vol. 14, No. 1, pp. 85-93, 2011.
- [10] Hirsch, H. G. & Pearce, D. The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions, in Proc. ICSLP. pp. 18-20. 2000.
- [11] Young, S. HTK: Hidden Markov Model Toolkit V3.4.1. Cambridge University, Engineering Department, Speech Group. 1993.
- [12] Jong-Young Ahn, Sang-Bum Kim, Su-Hoon Kim, Kang-In Hur. A study on Voice Recognition using Model Adaptation HMM for Mobile Environment. The Journal of the Institute of Webcasting, Internet and Telecommunication. Vol. 11, No. 3, pp. 175-179, 2011.
- [13] Sang-Yeob Oh. Selective Speech Feature Extraction using Channel Similarity in CHMM Vocabulary Recognition. The Journal of digital policy and management. Vol. 11, No. 7, pp. 453-458, 2013.
- [14] Sang-Yeob Oh. Bayesian Method Improve Recognition Rates using HMM Vocabulary Recognition Model Optimization. The Journal of digital policy and management. Vol. 12, No. 7, pp. 273-278, 2014.
- [15] Sang-Yeob Oh. Decision Tree State Tying Modeling Using Parameter Estimation of Bayesian Method The Journal of Digital Policy and Management. Vol. 13, No. 1, pp. 1243-248, 2015.

오 상 엽(Oh, Sang Yeob)



- 1991년 2월 : 광운대학교 대학원 전자계산학과 (이학석사)
- 1999년 2월 : 광운대학교 대학원 전자계산학과 (이학박사)
- 2007년 2월 ~ 현재 : 가천대학교 IT대학 컴퓨터공학과 교수
- 관심분야 : 버전관리, 형상관리, 음성/음향 신호 처리, 차량 통신
- E-Mail : syoh1234@gmail.com