

# HMM 어휘 인식 모델 최적화를 이용한 베이지안 기법 인식률 향상

오상엽  
가천대학교 컴퓨터미디어융합학과

## Bayesian Method Recognition Rates Improvement using HMM Vocabulary Recognition Model Optimization

Sang Yeon Oh

Dept. of Computer Media Convergence, Gachon University

**요 약** HMM(Hidden Markov Model)을 이용한 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 HMM(Hidden Markov Model) 모델 최적화를 이용한 베이지안 기법 인식률 향상을 제안한다. 본 논문은 HMM 어휘 인식에서 인식을 위한 모델 구성을 가우시안 믹스처 모델로 최적화한 인식 모델을 생성하였으며 베이지안 기법인 사전확률과 사후확률을 이용한 인식률을 향상시켰다. 본 논문에서 제안한 방법을 적용한 결과 어휘 인식률에서 97.9%의 인식률을 나타내었다.

**주제어** : HMM, 어휘 인식, 모델 최적화, 베이지안, 인식률

**Abstract** In vocabulary recognition using HMM(Hidden Markov Model) by model for the observation of a discrete probability distribution indicates the advantages of low computational complexity, but relatively low recognition rate. Improve them with a HMM model is proposed for the optimization of the Bayesian methods. In this paper is posterior distribution and prior distribution in recognition Gaussian mixtures model provides a model to optimize of the Bayesian methods vocabulary recognition. The result of applying the proposed method, the recognition rate of 97.9% in vocabulary recognition, respectively.

**Key Words** : HMM, Vocabulary Recognition, Model Optimize, Bayesian, Recognition Rate

### 1. 서론

HMM이 널리 사용되어지는 이유는 음성과 문자 등에서 발견되는 많은 변형들을 흡수 할 수 있고, 시간에 따라 변해가는 특성을 지니는 데이터를 잘 모델링하며 파라미터의 수가 클수록 잘 모델링하기 때문이다. HMM

구조의 최적화를 위한 연구는 다양한 방법으로 진행되어지고 있다. 높은 점유를 갖는 상태에서 순차적으로 분할해서 점차 상태수를 증가시키는 방법이나 사전 확률에 기반한 사후 확률을 사용하는 복잡한 구조로부터 점차 구조를 감소시켜나가는 방법이 연구되었다[1].

구조화된 모델들을 대상으로 파라미터의 집합들을 사

Received 10 May 2014, Revised 7 June 2014

Accepted 20 July 2014

Corresponding Author: Sang Yeon Oh (Dept. of Computer Media Convergence, Gachon University)

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.

용함으로써 보다 강인한 모델들을 추정하며 군집화에 사용될 모델 파라미터들의 초기 추정치를 생성하기 위하여 각 문맥들에 대한 데이터가 필요하다[2].

HMM을 이용하는 어휘 인식 시스템은 인식 어휘에 대한 모델들을 구성하고 그에 대한 모델들의 관측 확률이 이산적인 분포를 나타내며 인식과 모델링 구성에 대한 계산량이 적은 장점이 있지만 구성된 모델만을 인식하며 인식률이 상대적으로 낮은 단점이 있다. 이를 개선하기 위해 HMM(Hidden Markov Model) 모델 최적화를 이용한 베이지안 기법 인식을 향상을 제안한다.

본 논문은 어휘 인식에서 HMM 가우시안 믹스처 모델을 최적화한 인식 모델을 베이지안 기법인 사전확률과 사후확률을 이용하여 인식률을 향상시켰다. 어휘 구성을 위한 HMM 가우시안 믹스처 모델을 최적화하여 인식 모델로 사용하였으며 어휘 인식 모델에 대한 음소 모델을 제안한 HMM 가우시안 믹스처 연속 확률 밀도를 이용한 최적화 모델 방법을 실험한 결과 어휘 인식률에서 97.9%의 인식률을 나타내어 기존의 방법들에 비하여 인식 성능의 향상을 확인하였다.

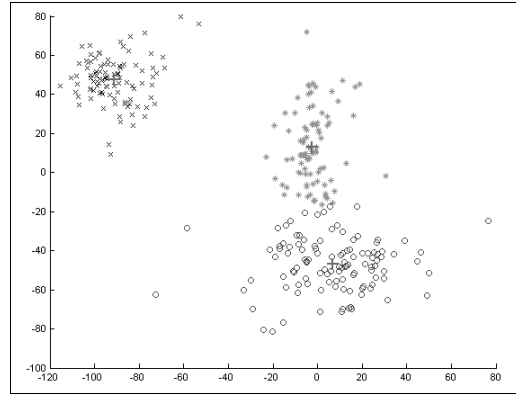
본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 언급하고 3장에서는 HMM 가우시안 믹스처 모델을 최적화한 인식 모델을 베이지안 기법인 사전확률과 사후확률을 이용하여 인식률을 향상 방법에 대해 설명하며, 4장에서는 성능 평가를 수행하고 마지막으로 5장에서 결론을 맺는다.

## 2. 관련 연구

### 2.1 k-means 기법

클러스터링의 대상이 되는 객체(Object)들은 각 객체의 특성을 나타내는 속성을 가지고 있다. 객체들은 클러스터링을 통해서 특정 군집에 속하게 되며, 각 군집은 소속 객체들의 속성 정보를 소유한다[3]. 객체에 대한 클러스터링 결과를 분석하면 각 군집에 분포된 객체들의 분포도에 대한 정보를 얻을 수 있다. 가장 많이 사용되는 클러스터링 기법으로 k-means 기법이 있다. [Fig. 1]에서 기본 적인 k-means 기법을 이용한 클러스터링을 나타내었다. 가장 가까운 중심점을 갖는 군집에 각 항목을 할당하는 과정을 반복하여 k개의 군집으로 항목들을 나

누는 것이다. 거리 기반 클러스터링 방법으로 고객의 선호도를 다차원 공간상의 점으로 표시하고, 거리를 계산함으로써 전체 고객들의 집합을 k개의 군집으로 나눈다[4].



[Fig. 1] K-means implementation

모델  $a$ 와  $k$ 사이의 거리는 다음 식으로 계산되고 식에서  $a_j$ 는 모델  $a$ 의 속성  $i$ 에 대한 선호도 값을 의미한다.

$$d_{a,k} = \sqrt{\sum_t (a_t - k_t)^2} \quad (1)$$

변수로 구성된 모델 클러스터링(Clustering)을 위해 k-means 기법을 적용한다. 적용된 k-means 기법은 계산 속도가 빠르고 대량의 자료에서 군집을 발견하는데 상당히 효과적이다.

### 2.2 HMM

HMM 알고리즘은 음성이 Markov process로 모델링하여 음성의 학습과정에서 Markov 모델의 파라미터를 구한다. 기존 Markov 모델을 만들고 입력된 음성과 저장된 기존 Markov 모델을 비교하여 유사도가 가장 높은 기존의 Markov 모델을 인식된 어휘로 결정한다[5].

HMM(Hidden Markov Model) 알고리즘은 관측이 불가능한 프로세스를 관측이 가능한 다른 프로세스를 통해 추정하는 이중 확률 처리 기법을 사용하여 모델링한다. 특히 인식 가능한 기준 패턴을 음소 단위와 음절 단위로 설정하여 모델을 구성하고 인식하는 방법을 사용한다[6].

기존 HMM 방법보다 확률의 분포 함수를 사용하기

위한 방법으로 가우시안 확률 밀도 함수(Gaussians Hidden Markov Model)를 사용한다. 가우시안 확률 밀도 함수는 기존 확률에 비해 평균과 표준편차를 이용하여 확률의 분포를 이용한다. 평균  $\mu$  과 표준편차  $\sigma$  를 구하기 위하여 가우시안 확률 밀도 함수로 다음과 같이 나타낸다.

$$b_i(y) = f(y; \theta_i) \tag{2}$$

가우시안 확률 밀도 함수는 2차 지수 함수의 형태를 가지므로 2차원 이상의 다차원 식으로 표현이 가능하며, n-차원을 가지는 특정 벡터  $x$  를 확률 변수로 한 가우시안 확률 밀도 함수는 다음과 같이 나타낸다[7].

$$g_i, k(y) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_{i,k}|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}(y - \mu_i, k)^T \Sigma_{i,k}^{-1} (y - \mu_i, k)\right) \tag{3}$$

식 (2)에서 n차원 가우시안 확률 밀도 함수는 n차원 공간에서의 한 점인 중심  $\mu$  으로 표현되고,  $\Sigma$  는  $n \times n$  의 가역적인 양의 정부호 대칭 행렬로 표현되므로  $X^T \Sigma^{-1} X$  의 형태를 가질 수 있다. 여기서,  $\mu$  와  $\Sigma$  는 가우시안 분포의 핵심 파라미터로 표현된다[8].

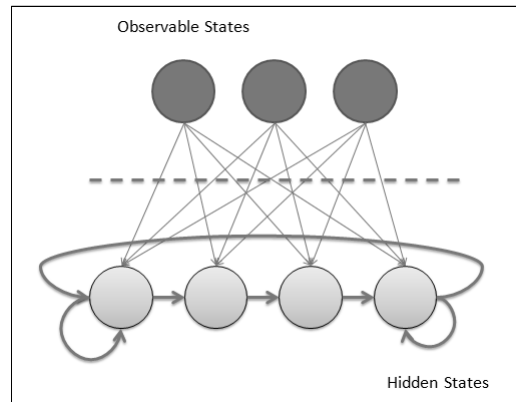
### 3. HMM 모델을 이용한 베이시안 기법

HMM을 이용하는 어휘 인식은 인식을 위한 모델이 구성되며 구성된 인식 모델들의 대한 관측 확률이 이산적인 분포를 나타낸다. 이산적인 분포는 계산량이 적은 장점을 가지고 있지만 인식률은 상대적으로 낮은 단점을 가지고 있다. 이러한 단점을 보완하기 위해 베이시안 기법을 이용한 모델 최적화를 제안한다. HMM 가우시안 믹스처 모델을 최적화한 인식 모델을 베이시안 기법인 사전확률과 사후확률을 계산하여 인식 모델을 추론하고 인식률을 향상시킨다.

#### 3.1 가우시안 HMM 모델

HMM은 일련의 연속된 상태들로부터 이산 신호를 생

성하는 확률 과정 모델로 표현된다. 모델은 전이 확률에 따라 상태를 바꾸며 특정 상태는 그 상태의 출력 확률에 따라 하나의 관측을 발생시킨다. 모델의 파라미터를 추정하기 위하여 카테고리 정보가 있는 음성 데이터베이스를 사용하며 각 모델을 위한 충분한 데이터가 있을 경우 실제 음성에 존재하는 다양성을 잘 표현할 수 있는 강인한 모델링이 가능하다. 가우시안 HMM 모델은 확률 밀도 함수를 갖는 3상태의 음소 모델을 생성하여 훈련된 모델을 사용한다. [Fig. 2]는 가우시안 HMM 모델의 음소 모델 생성을 나타낸다[9].



[Fig. 2] Hidden Markov Model

가우시안 HMM 모델은 주어진 표본 데이터 집합의 분포밀도를 하나의 확률밀도함수로 모델링하는 방법을 개선하여 밀도 추정 방법으로 복수 개의 가우시안 확률 밀도함수로 데이터의 분포를 모델링하는 방법이다. 하나의 가우시안으로는 모델링 할 수 없는 복수개의 중심점을 가지는 1차원 데이터와 2차원 환형 데이터에 대하여 견고하게 모델링된다. 확률밀도함수는 가우시안 분포뿐 아니라 다른 분포가 될 수도 있다. 가우시안 혼합 밀도는 단지 확률밀도함수를 가우시안 분포로 가정하는 경우이다. 확률밀도함수는 M개의 가우시안 확률밀도함수의 선형 결합으로 다음과 같이 나타낸다.

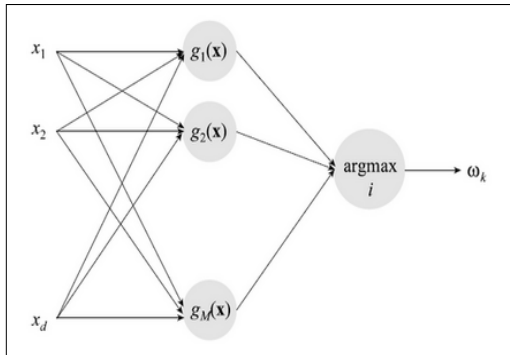
$$p(x|\theta) = \sum_{i=1}^M p(x|\omega_i, \theta_i) P(\omega_i) \tag{4}$$

$p(x|\omega_i, \theta_i)$  는 데이터  $x$  에 대하여  $\omega_i$  번째 성분 파라미

터  $\theta_i$ 로 이루어진 확률밀도함수를 의미하며,  $P(w_i)$ 는 혼합 가중치로 각 확률밀도함수의 상대적인 중요도를 의미한다.

### 3.2 베이시안 기법

Bayesian 기법을 이용한 모수의 추정은 자료로부터 얻은 모수에 대한 정보와 모수에 대한 과거의 경험 또는 주관을 사전분포로 표현함으로써 보다 정확한 모수의 불확실성에 대한 탐색에 목적이 있다. [Fig. 3]과 같이 베이시안 결정 기법을 사용하여 탐색하여 결정 값을 나타낸다.



[Fig. 3] Bayesian Decision Theory

Metropolis-Hastings 알고리즘을 수행하고 최종적으로 몬테카를로 적분을 적용하면 추정된 모수들의 평균과 같은 통계적 특성치를 근사적으로 얻을 수 있다. 또한 기존 통계학에서 사용되는 신뢰구간(confidence interval)과 유사한 개념의 신용구간(credible interval)을 얻을 수 있고 이를 이용하여 추정된 모수의 불확실성을 표현할 수 있다. 반복하여 추출된 모수의 개수  $n$ 과 유의수준  $\alpha$ 에서 Bayesian은 다음과 같이 나타낸다.

$$\tau = \frac{\sum_{i < j} \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j)}{\sqrt{\left(\frac{\sum t_i(t_i - 1)}{2}\right) \left(\frac{\sum u_i(u_i - 1)}{2}\right)}} \quad (5)$$

$n$ 은 변수의 관측치를 의미하며,  $t_i$ 는 주어진 순위에서 동일한  $X$ 의 관측치를 나타내며  $u_i$ 는 주어진 순위에서 동일한  $Y$ 의 관측치를 나타낸다.  $\text{sgn}$ 은  $\text{sgn}(z) = 1$  이

면  $z > 0$ 를 나타내고  $\text{sgn}(z) = 0$ 이면  $z = 0$ 를 나타내며  $\text{sgn}(z) = -1$ 이면  $z < 0$ 를 나타낸다. 계산된 계수는 정규화를 통해 0~1사이의 값을 갖는다.

특정 확률밀도함수의 모수를 추정하기 위해서는 다양한 방법이 적용될 수 있지만 본 연구에서는 Bayesian 기법을 사용하였다. Bayesian의 정리를 연속 확률밀도함수로 나타내면 다음과 같이 표현된다.

$$\pi(\theta|x_1, x_2, \dots, x_n) = \frac{f(x_1|\theta) \dots f(x_n|\theta)\pi(\theta)}{\int_{\theta} f(x_1|\theta) \dots f(x_n|\theta)\pi(\theta)d\theta} \quad (6)$$

좌변의  $\pi(\theta|x_1, x_2, \dots, x_n)$ 는 사후분포 (posterior distribution), 우변 분자의  $\pi(\theta)$ 는 사전분포 (prior distribution)라 한다. 우변의 분모는 상수로서 주변분포 (marginal distribution)이고, 우변의 분자의  $f(x_1|\theta) \dots f(x_n|\theta)$ 는 발생할 수 있는 모든 가능성을 고려한 우도함수(likelihood function)이다.

### 4. 실험 결과

본 논문에서 제안한 HMM(Hidden Markov Model) 모델 최적화를 이용한 베이시안 기법 인식을 향상 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[10].

제안한 시스템의 성능 평가를 위하여 기존 방식과 비교 실험을 하였으며 유클리디안 알고리즘, DTW 알고리즘[11], 제안 방법에 대해 인식률을 측정한 결과 제안 방법에서 인식률이 97.9%의 성능을 보였다.

<Table 1>은 기존 방식인 유클리디안 알고리즘, DTW 알고리즘, 제안 방법을 실제 환경에서의 실험을 나타낸다. <Table 1>에서 보는 것과 같이 잡음이 거의 없는 실내 환경에서 실험한 결과 유클리디안 알고리즘을 이용한 음성 인식을 평균 95.3%로 나타났으며 DTW 알

고리즘을 이용한 음성 인식을 평균 96.2%의 인식을 나타내었고 제안방법의 인식을 평균 97.9%를 나타내었다.

<Table 1> Non-Noise Environment Recognition Rate

Speech	Recognition Rate (%)		
	Euclidean	DTW	Proposed Method
Speech Dependent	95.2	96.9	98.2
	94.7	96.1	97.1
	95.6	96.4	98.3
Average	95.2	96.5	97.9
Speech Independent	95.6	96.3	98.1
	95.1	95.2	97.7
	95.8	96.5	98.2
Average	95.5	96.0	98.0

<Table 2>는 기존 방식인 유클리디안 알고리즘, DTW 알고리즘, 제한 방법을 실외 환경에서의 실험을 나타낸다.

<Table 2> Noise Environment Recognition Rate

Speech	Recognition Rate (%)		
	Euclidean	DTW	Proposed Method
Speech Dependent	82.5	83.9	84.9
	81.4	82.1	83.9
	82.1	83.9	84.7
Average	82.0	83.3	84.5
Speech Independent	79.8	82.7	84.1
	79.1	81.2	83.2
	80.1	83.2	83.8
Average	79.7	82.4	83.7

<Table 2>에서는 잡음 환경에서 실험한 결과 유클리디안 알고리즘을 이용한 음성 인식을 평균 80.8%로 나타났으며 DTW 알고리즘을 이용한 음성 인식을 평균 82.9%의 인식을 나타내었고 제안방법의 인식을 평균 84.1%를 나타내었다.

## 5. 결론

본 논문은 HMM을 이용하는 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 베이지안 기법을 이용한 모델 최적화를 제안한다. HMM 가우시안 믹스처 모델을 최적화한 인식 모델을 베이지안 기법인 사전확률과 사후확률을 계산하여 인식 모델을 추론하였다. 이를 확인하기 위해 유클리디안 알고리즘, DTW 알고리즘과 비교하였다. 실내 환경에서 실험한 결과 유클리디안 알고리즘을 이용한 음성 인식을 평균 95.3%로 나타났으며 DTW 알고리즘을 이용한 음성 인식을 평균 96.2%의 인식을 나타내었고 제안방법의 인식을 평균 97.9%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] Y. Shao, S. Srinivasan, Z. Jin, D. Wang. A Computational Auditory Scene Analysis System for

- Robust Speech Recognition. Computer Speech & Language. Vol. 24, No. 1, pp. 77-93, 2010.
- [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] S. Y. Cho, D. M. Sun, Z. D. Qiu. A Spearman correlation coefficient ranking for matching-score fusion on speaker recognition. Proc. TENCON Conf. pp. 736-741, 2011.
- [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] 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.
- [11] Sang-Yeob Oh. Selective Speech Feature Extraction using Channel Similarity in CHMM Vocabulary Recognition. The Journal of digital policy and management. Vol. 11, No. 10, pp. 453-458, 2013.

#### 오 상 엽(Oh, Sang Yeob)



- 1991년 2월 : 광운대학교 대학원 전자계산학과 (이학석사)
- 1999년 2월 : 광운대학교 대학원 전자계산학과 (이학박사)
- 2007년 2월 ~ 현재 : 가천대학교 IT대학 인터랙티브미디어학과 교수
- 관심분야 : 버전관리, 형상관리, 음성/음향 신호 처리, 차량 통신

· E-Mail : syoh1234@gmail.com