

모수 추정을 위한 베이시안 기법과 바타차랴 알고리즘을 융합한 어휘 인식 성능 향상

오상엽
가천대학교 컴퓨터공학과

Vocabulary Recognition Performance Improvement using a convergence of Bayesian Method for Parameter Estimation and Bhattacharyya Algorithm Model

Sang-Yeob Oh

Dept. of Computer Engineering, Gachon University

요약 어휘 인식 시스템은 학습 모델을 구성하여 인식하므로 구성되어진 모델에서 벗어난 어휘의 입력과 유사한 어휘의 입력은 인식하지 못하거나 유사한 어휘로 인식되어 인식을 저하가 나타난다. 이런 경우 인식 모델을 확장할 수 있도록 재구성하거나 인식 모델 구성 시 확장성을 반영하므로 해결할 수 있다. 본 논문에서는 모델 구성 시 확장성을 반영할 수 있는 모수 추정을 위한 베이시안 기법을 사용하여 바타차랴 알고리즘 음성 인식 학습 모델 구성 방법을 융합하여 제안하였다. 음소가 갖는 특징을 기반으로 학습 데이터의 음소에 모수 추정을 위한 베이시안 기법을 이용하였고 유사한 학습 모델은 바타차랴 알고리즘을 이용하여 정확한 학습 모델로 인식하도록 하였다. 바타차랴 알고리즘 인식 모델을 구성하여 인식 성능을 평가하였다. 본 논문에서 제안한 시스템을 적용한 결과 어휘 인식률에서 97.5%의 인식률과 1.2초의 학습 시간을 나타내었다.

주제어 : 베이시안 기법, 모수 추정, 바타차랴 알고리즘, 인식 모델, 어휘 인식

Abstract The Vocabulary Recognition System made by recognizing the standard vocabulary is seen as a decline of recognition when out of the standard or similar words. In this case, reconstructing the system in order to add or extend a range of vocabulary is a way to solve the problem. This paper propose configured Bhattacharyya algorithm standing by speech recognition learning model using the Bayesian methods which reflect parameter estimation upon the model configuration scalability. It is recognized corrected standard model based on a characteristic of the phoneme using the Bayesian methods for parameter estimation of the phoneme's data and Bhattacharyya algorithm for a similar model. By Bhattacharyya algorithm to configure recognition model evaluates a recognition performance. The result of applying the proposed method is showed a recognition rate of 97.3% and a learning curve of 1.2 seconds.

Key Words : Bayesian Method, Parameter Estimation, Bhattacharyya Algorithm, Recognition Model, Vocabulary Recognition

Received 9 August 2015, Revised 25 September 2015
Accepted 20 October 2015
Corresponding Author: SangYeobOh(Gachon University)
Email: syoh1234@gmail.com

© 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.

ISSN: 1738-1916

1. 서론

어휘 인식 시스템의 인식 알고리즘은 퍼지로지, Neural Network 및 HMM(Hidden Markov Model) 방법을 사용한다. 인식 알고리즘은 학습 모델을 구성하여 인식하므로 학습 모델의 구성이 인식의 영향을 미치게 되어 학습 모델 구성에 많은 노력을 한다. 학습 모델로 구성된 모델을 바탕으로 인식할 경우 구성되어진 모델에서 벗어난 어휘의 입력과 유사한 어휘의 입력은 인식하지 못하거나 근사한 어휘로 인식되어지는 경우가 발생하게 되고 이는 인식을 저하로 나타낸다. 이를 보완하기 위해 인식 오류 보정에 대한 연구가 진행되고 있다[1].

인식 오류 보정은 모델의 재구성과 모델의 확장을 위해 필요하게 되며 통계적 방법을 이용한다. 유클리디안 알고리즘, DTW(Dynamic Time Warping) 알고리즘, 바타차라(Bhattacharyya) 알고리즘을 사용한다. 모델의 인식을 위한 통계를 이용하여 패턴들을 비교하고 정합을 거쳐 학습 모델 중 유사도가 가장 높은 모델을 인식 결과로 채택한다[2].

패턴을 비교하고 통계적으로 계산하기 위해 사전 확률 분포를 이용한 모델링 방법으로 HMM(Hidden Markov Model) 기법이 사용된다[3]. HMM 기법은 음성과 문자 등에서 나타나는 유사 어휘로의 변형을 방지하여 모델을 구성할 수 있는 장점이 있다. 하지만 인식을 위해 입력되는 어휘가 모델에 포함되어 있지 않거나 유사한 어휘는 인식하지 못하거나 다른 어휘로 인식하는 단점이 있다. 이를 개선하기 위해 모수 추정 베이지안 기법과 바타차라 알고리즘을 이용한 인식 모델 구성 방법을 제안한다.

본 논문은 어휘 인식에서 베이지안 기법의 모수에 대한 정보와 모수에 대한 과거의 경험을 사전 분포로 표현한다. 따라서 정확한 모수의 불확실한 값에 대한 탐색이 가능하므로 어휘의 모델을 추정하여 인식률을 향상시켰다. 유사한 음소는 바타차라 알고리즘을 이용하여 정확한 음소로 인식할 수 있도록 유도하였고 유사 음소 인식과 오인식 오류를 최소화하여 인식률을 향상시켰다. 이를 확인하기 위해 시스템 성능 평가를 유클리디안 알고리즘, DTW 알고리즘과 비교하였다. 비교한 시스템 성능 평가 결과 98.7%의 인식률을 나타내었다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구

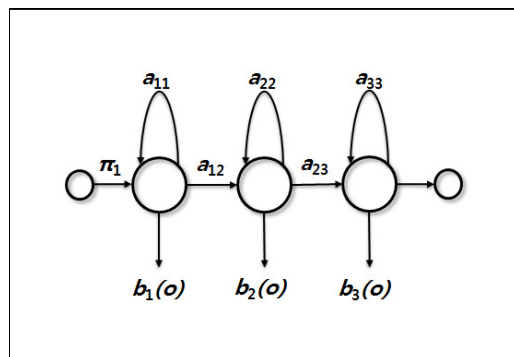
에 대해 언급하고 3장에서는 바타차라 알고리즘을 이용한 음성 인식 최적 학습 모델 구성 방법에 대해 설명하며 4장에서는 시스템 평가를 수행하고 마지막으로 5장에서 결론을 맺는다.

2. 관련연구

2.1 HMM 모델

HMM의 모델 학습과정은 관측열에 대하여 최대 확률을 가지는 적합한 모델을 선택하고, 이 모델의 내부 상태들 중에서 최적의 상태 순서를 결정한 후, 관측열 X에 대한 $P(X|\lambda)$ 를 최대화하는 모델 $\lambda = (\pi, A, B)$ 의 매개변수를 결정한다. Markov 매개변수는 사전 확률 (prior probability), 상태 천이 확률 (transition probabilities), 관찰 심볼 확률 (emission probabilities)의 세 가지 매개변수를 갖는다[4,14].

HMM 모델은 학습 과정에서 모델의 파라미터를 이용하여 Markov 프로세스로 모델링하여 연속된 상태들로부터 이산 신호를 생성하는 확률 과정으로 표현되며 전이 확률에 따라 상태를 바꾼다. 파라미터의 추정은 각 모델을 위한 충분한 데이터가 있는 경우 강인한 모델 구축이 가능하며 확률 밀도 함수를 갖는 상태의 모델을 생성하여 훈련된 모델로 사용하게 된다[5,15]. [Fig. 1]은 HMM 모델 생성을 나타낸다.



[Fig. 1] Hidden Markov Model

확률 밀도 함수는 특징 벡터 x 를 확률 변수로 다음과 같이 나타낸다. n 차원 확률 밀도 함수는 n 차원 중심 μ 으로 표현되고, μ 와 k 는 파라미터로 표현된다[6].

$$f(x) = \frac{1}{k} \left[1 - \sigma \frac{x - \mu}{k} \right]^{\frac{1}{\sigma} - 1} \exp \left[- \left[1 - \sigma \left(\frac{x - \mu}{k} \right) \right]^{\frac{1}{\sigma}} \right] \quad (1)$$

2.2 유클리디안 알고리즘

유클리디안 알고리즘(Euclidean Algorithm)은 모델들의 속성인 각 변수 값에 각각의 속성 값을 대입해서 거리를 구하는 방법이며 유사도 알고리즘과 달리 학습 조건이 있고 계산된 결과가 가장 작은 값을 가진 데이터를 추출하는 것이다. 학습 조건이 명확하지 않거나 다른 데이터가 추가되었을 경우에는 모든 데이터를 추출하기 때문에 추출 시간과 필터링이 시간이 오래 걸리는 단점을 가지고 있다. 하지만 명확한 학습 조건과 추가되지 않는 데이터 환경에서는 정확한 결과와 쉬운 알고리즘으로 인해 많이 사용되어진다[7].

유클리디안 알고리즘은 모델간의 거리를 구하기 위하여 사용되며 두 모델의 속성 값에 의한 두 모델 사이의 유사도를 구하여 확률이 높은 값을 인식한다. 관측 확률의 각 변수는 모델을 나타내는 벡터의 한 성분으로 관측 모델 사이의 거리를 구하기 위한 척도로 사용된다. 관측 모델의 거리 값이 작게 나타나면 대응하는 변수 값은 유사하다고 판단하여 두 모델간의 관계를 하나의 숫자로 표현할 수 있다[8].

n 개의 행벡터 w_i 와 w_j 가 각각 $w_i = \{w_{i1}, w_{i2}, \dots, w_{in}\}$ 과 $w_j = \{w_{j1}, w_{j2}, \dots, w_{jn}\}$ 으로 존재한다고 가정할 때, 다음과 같이 유클리디안 거리에 대해 다음 식과 같이 정의한다.

$$d(d_i, d_j) = \sqrt{\sum_{k=1}^n (w_{ik} - w_{jk})^2} \quad (2)$$

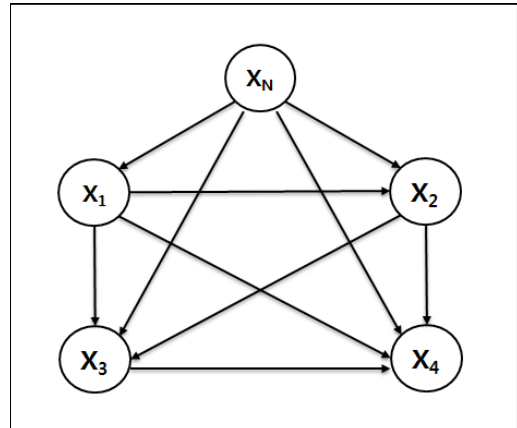
3. 베이시안 기법과 바타차라 알고리즘

3.1 베이시안 기법

베이시안 기법은 확률적인 접근론에 기반하여 모형의 모수는 임의의 계수를 사용하고 확률분포로 추정하며 어떠한 현상을 설명하는데 포함되는 불확실성을 표현한다. 베이시안 기법은 알려진 데이터를 이용하여 모르는 데이터에 대한 학습을 반영하는 과정이다[9]. 베이시안 기법의 조건부 확률식은 다음과 같다.

$$p(\theta|x) \propto p(x|\theta) \cdot p(\theta) \quad (3)$$

θ 는 추정할 모수를 나타내고, $p(\theta|x)$ 는 사후분포 (posterior density function), $p(\theta)$ 는 사전분포(prior density function)를 나타내며 $p(x|\theta)$ 는 주어진 모형 모수에 데이터의 함수인 우도함수(likelihood function)를 나타낸다. [Fig. 2]는 베이시안 기법의 모수 추정을 나타낸다.



[Fig. 2] Bayesian Method Model

사전분포는 추정해야 할 모수에 대해서 정보를 필요로 하지 않으므로 모수 θ 의 속성은 모델 구축 전에 반영하여 모델을 구축한다. 속성 값이 없는 상태에서 모델을 구축하여도 베이시안 기법은 추론을 이용하여 모델을 구축하므로 학습을 통해 강인한 모델을 만들어 가며 이러한 베이시안 기법은 추론에 의한 학습이다. 관측 값 $x = (x_1, \dots, x_n)$ 일 때 사후분포는 다음과 같다.

$$p(\theta|x, x_{n+1}) \propto p(x_{n+1}|\theta) \cdot p(\theta|x) \quad (4)$$

사전분포는 과거 실험 결과의 사후분포를 학습하여 추론한다.

3.2 바타차라 알고리즘

바타차라 알고리즘은 거리 측정에 사용되며 불연속 또는 연속적인 두 확률 분포의 유사성을 측정하는 통계 학습 기반으로 하며 확률 계산을 바탕으로 오류율을 측정하여 거리를 계산한다[10]. 단순 거리 계산을 수행하는 방법으로 실시간을 요구하는 인식 과정에서 동적 프로그

램 기술인 비터비 코딩 방법을 적용하여 상태 경로의 변이와 최적의 모델을 확률적인 모수 추정 방식으로 사용하였다.

비터비 확률 방법인 전향, 후향 확률에 의한 연산을 이용하여 상태 경로를 확률적인 모수 추적 방식으로 인식을 향상시킬 수 있다. 반면 주어진 모든 상태에서의 출력 심볼의 확률을 전부 추정하므로 계산량과 복잡도가 증가하게 된다. 전향, 후향 확률 추정을 이용한 방식에 비해 인식이 다소 저하되나 연산에서의 계산량과 복잡도를 감소시키므로 인식 과정에서 사용하였다[11].

오류율 측정에 기반을 두고 있는 바타차라 거리 측정법을 사용하여 거리 측정을 하였으며 다음과 같이 정의한다.

$$d^2(i, j) = \frac{1}{2} \sum_{k=1}^n \frac{(\mu_{ik} - \mu_{jk})^2}{\sigma_{ik} + \sigma_{jk}} \quad (5)$$

n 은 차수를 나타내고 μ_{ik} 는 i 에 대한 k 번째 평균을 나타내며 σ_{ik} 는 i 에 대한 k 번째 분산을 나타낸다. μ_{jk} 는 j 에 대한 k 번째 평균을 나타내며 σ_{jk} 는 j 에 대한 k 번째 분산을 나타낸다. 두 모델 사이의 거리 계산을 위해 모델의 속성을 측정하며 거리 범위는 0에서 ∞ 까지의 값을 가지므로 각각 모델 속성과 완전히 일치하거나 무한대로 떨어져 있는 거리 측정에 대해 다음과 같이 정의한다.

$$BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)} \quad (6)$$

$p(x)$ 는 p 번째 분포의 분산을 나타내며 $q(x)$ 는 q 번째 분포의 분산을 나타내고 p, q 는 서로 다른 두 분포를 나타낸다.

4. 실험결과

본 연구에서는 잡음 음성 인식 실험을 위하여 Aurora 2 데이터베이스[11]를 사용하였고 각각의 모델들로부터 특징 벡터를 추출하여 39차의 특징 벡터를 사용하였다. 음향 모델의 훈련과 인식을 위해서는 HTK(Hidden Markov Toolkit)를 사용하였다[12].

본 논문에서 제안한 모수 추정을 위한 베이지안 기법을 사용하여 바타차라 알고리즘 음성 인식 학습 모델 구

성 모델을 이용한 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 50개, 지하철역명 50개로 구성하였다. 인식 실험에서는 실험에 참가한 화자가 음성 인식 목록을 3회 발음하여 총 300단어를 대상으로 실험을 수행하였다[13].

제안한 시스템의 성능 평가를 위하여 기존 방식과 비교 실험을 하였으며 유클리디안(Euclidean) 알고리즘, DTW 알고리즘[13], 제안 방법에 대해 인식을 측정하고 제안 방법에서 97.5%의 평균 인식 성능과 1.2초의 평균 학습 시간을 보였다.

<Table 1>은 기존 방식인 유클리디안 알고리즘, DTW 알고리즘, 제안 방법에 대한 실내 환경에서의 실험 결과를 나타낸다.

<Table 1> Non-Noise Environment Recognition Rate

Speech	Euclidean		DTW		Proposed Method	
	Recognition		Recognition		Recognition	
	Rate (%)	Time (sec)	Rate (%)	Time (sec)	Rate (%)	Time (sec)
Speech Dependent	96.7	1.5	97.1	1.4	97.8	1.2
	96.9	1.4	97.5	1.4	97.9	1.2
	97.2	1.3	98.2	1.1	98.1	1.1
Speech Independent	95.2	1.7	96.5	1.6	96.8	1.4
	95.6	1.6	96.1	1.6	97.1	1.3
	95.9	1.4	97.3	1.4	97.1	1.2

<Table 1>에서 보는 것과 같이 잡음이 거의 없는 실내 환경에서 인식을 실험한 결과 유클리디안 알고리즘을 이용한 음성 인식을 평균 96.3%로 나타났으며 DTW 알고리즘을 이용한 음성 인식을 평균 97.1%의 인식을 나타내었고 제안방법의 인식을 평균 97.5%를 나타내었다. 인식률에 따른 학습 시간을 측정하고 유클리디안 알고리즘을 이용한 학습 시간은 평균 1.5초로 나타났으며 DTW 알고리즘을 이용한 학습 시간은 평균 1.4초의 인식을 나타내었고 제안방법의 학습 시간은 평균 1.2초를 나타내었다.

<Table 2>에서는 잡음 환경에서 실험한 결과 유클리디안 알고리즘을 이용한 음성 인식을 평균 83.4%로 나타났으며 DTW 알고리즘을 이용한 음성 인식을 평균 84.4%의 인식을 나타내었고 제안방법의 인식을 평균

86.4%를 나타내었다. 인식률에 따른 학습 시간을 측정 한 결과 유클리디안 알고리즘을 이용한 학습 시간은 평균 1.9초로 나타났으며 DTW 알고리즘을 이용한 학습 시간은 평균 1.8초의 인식률 나타내었고 제안방법의 학습 시간은 평균 1.6초를 나타내었다.

안방법 86.4%의 인식률 평균을 나타내었으며 인식률에 따른 학습 시간을 측정 한 결과 유클리디안 알고리즘 1.9 초, DTW 알고리즘 1.8초, 제안방법 1.6초의 학습 시간을 나타내었다.

<Table 2> Noise Environment Recognition Rate

Speech	Euclidean		DTW		Proposed Method	
	Recognition		Recognition		Recognition	
	Rate (%)	Time (sec)	Rate (%)	Time (sec)	Rate (%)	Time (sec)
Speech Dependent	83.1	1.8	84.2	1.6	86.8	1.5
	84.1	1.7	85.1	1.7	87.3	1.6
	83.9	1.6	84.6	1.6	87.4	1.5
Speech Independent	83.2	1.9	84.1	1.9	85.2	1.7
	83.1	2.0	84.3	1.8	85.7	1.6
	83.2	2.1	84.2	1.9	85.8	1.7

5. 결론

본 논문은 모델 구성 시 확장성을 반영할 수 있는 모 수 추정을 위한 베이시안 기법을 사용하여 바타차라 알고리즘 음성 인식 학습 모델 구성 방법을 제안하였다. 모델의 속성을 특징 기반으로 학습 데이터의 음소에 모 수 추정을 위한 베이시안 기법을 이용하였고 유사한 학습 모델은 바타차라 알고리즘을 이용하여 정확한 학습 모델로 인식하도록 하였다.

어휘 인식 시스템은 학습 모델을 구성하여 인식하므로 구성되었던 모델에서 벗어난 어휘의 입력과 유사한 어휘의 입력은 인식하지 못하거나 유사한 어휘로 인식되어 인식률이 낮아진다. 따라서 모델 구성 시 확장성을 반영할 수 있는 모 수 추정을 위한 베이시안 기법을 사용하여 바타차라 알고리즘 음성 인식 학습 모델 구성 방법으로 인식률을 향상시켰다.

이를 확인하기 위해 유클리디안 알고리즘, DTW 알고리즘과 비교하였다. 실제 환경에서 실험한 결과 유클리디안 알고리즘 96.3%, DTW 알고리즘 97.1%, 제안방법 97.5%의 인식률 평균을 나타내었으며 인식률에 따른 학습 시간을 측정한 결과 유클리디안 알고리즘 1.5초, DTW 알고리즘 1.4초, 제안방법 1.2초의 학습 시간을 나타내었다. 또한 실외 환경(잡음 환경)에서 실험한 결과 유클리디안 알고리즘 83.4%, DTW 알고리즘 84.4%, 제

REFERENCES

- [1] SangYeob Oh. Decision Tree State Tying Modeling Using Parameter Estimation of Bayesian Method. Journal of digital convergence v.13 no.1, pp.243-248, 2015.
- [2] SangYeob Oh. Bayesian Method Recognition Rates Improvement using HMM Vocabulary Recognition Model Optimization. Journal of digital convergence v.12 no.7, pp.273-278, 2014.
- [3] Jong-Sub Lee, Sang-Yeob Oh. Vocabulary Retrieve System using Improve Levenshtein Distance algorithm. The Journal of digital policy & management v.11 no.11, pp.367-372, 2013.
- [4] Sang-Yeob Oh. Decision Tree for Likely phoneme model schema support. The Journal of digital policy & management v.11 no.10, pp.367-372, 2013.
- [5] Sang-Yeob Oh. Selective Speech Feature Extraction using Channel Similarity in CHMM Vocabulary Recognition. The Journal of digital policy & management v.11 no.10, pp.453-458, 2013.
- [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] Chan-Shik Ahn, Sang-Yeob Oh. CHMM Modeling using LMS Algorithm for Continuous Speech Recognition Improvement. The Journal of digital policy & management v.10 no.11, pp.377 - 382, 2012.
- [9] Chan-Shik Ahn, Sang-Yeob Oh. Gaussian Model Optimization using Configuration Thread Control In CHMM Vocabulary Recognition. The Journal of

- digital policy & management v.10 no.7, pp.167-172, 2012.
- [10] Sang-Yeob Oh. Speech Recognition Optimization Learning Model using HMM Feature Extraction In the Bhattacharyya Algorithm. The Journal of digital policy & management v.11 no.6, pp.199-204, 2013.
- [11] Beaufays, F., Vanhoucke, V., & Strope, B. Unsupervised discovery and training of maximally dissimilar cluster models. Proc. Interspeech, pp. 66-69, 2010.
- [12] Young, S. HTK: Hidden Markov Model Toolkit V3.4.1. Cambridge University, Engineering Department, Speech Group. 1993.
- [13] 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. v.15, no.1, pp.177-183, 2010.
- [14] 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.
- [15] Sang-Yeob Oh. 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, The Journal of digital policy and management. Vol. 13, No. 8 pp. 295-300, 2015

오 상 엽(Oh, Sang Yeob)



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