

벡터모델 기반 바타차랴 거리 측정 기법과 우도 원리 베이시안을 융합한 어휘 인식 모델

오상엽

가천대학교 컴퓨터미디어융합학과

Vocabulary Recognition Model using a convergence of Likelihood Principia Bayesian methode and Bhattacharyya Distance Measurement based on Vector Model

Sang-Yeob Oh

Dept. of Computer Engineering, Gachon University

요 약 어휘 인식 시스템은 구성되어진 모델에서 벗어난 어휘의 입력과 유사한 어휘의 입력은 인식하지 못하거나 유사한 어휘로 인식되어 인식을 저하가 나타나며, 기존의 시스템은 벡터 값을 모델로 만들어 데이터베이스로 구성되어 어휘 인식에 사용하였다. 어휘 인식을 위한 탐색 중에 형성되는 모델은 데이터베이스로 구성되어 있지 않아 인식할 수 없는 단점이 존재한다. 따라서 본 논문에서는 특징 벡터 모델을 기반으로 바타차랴 거리 측정법을 이용한 베이시안 인식 모델을 구성하여 탐색 중에 형성되는 벡터 모델을 인식할 수 있도록 유도하였으며, 위너 필터를 적용하여 인식을 향상시켰다. 2 방법을 융합하여 실험한 결과 향상된 신뢰도로 인해 높은 인식 성능을 확인하였으며, 본 논문에서 제안한 측정법을 이용하여 기존의 방법들에 비하여 평균 98.2%의 성능을 나타내었다.

주제어 : 바타차랴 알고리즘, 거리측정 기법, 베이시안 알고리즘, 인식 모델, 인식을 향상

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. The vector values of the existing system to the model created by configuring the database was used in the recognition vocabulary. The model to be formed during the search for the recognition vocabulary is recognizable because there is a disadvantage not configured with a database. In this paper, it induced to recognize the vector model is formed by the search and configuration using a Bayesian model recognizes the Bhattacharyya distance measurement based on the vector model, by applying the Wiener filter improves the recognition rate. The result of Convergence of two method's are improved reliability experiments for distance measurement. Using a proposed measurement are compared to the conventional method exhibited a performance of 98.2%.

Key Words : Bhattacharyya Algorithm, Distance Measurement Method, Bayesian Algorithm, Recognition Model, Recognition Improve

Received 24 September 2015, Revised 24 October 2015

Accepted 20 November 2015

Corresponding Author: Sang Yeon Oh

(The Society of Digital Policy)

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

어휘 인식, 영상 인식과 같은 패턴 인식에서 거리의 개념은 패턴들이 특정 공간상에서 서로 얼마나 떨어져 있는지를 통하여 패턴들 사이의 비슷한 정도를 측정하기 위한 기준으로 사용한다. 특정 공간상에서 매우 근접한 거리에 있는 두 패턴은 거의 동일한 특징을 가지므로 큰 유사도를 갖는다. 이러한 거리 측정 방법에는 유클리디안 알고리즘, DTW 알고리즘이 사용되고 있다[1]. 유클리디안 알고리즘은 실제 거리를 구하기 위해 사용하며 각각의 개체 특징 값들이 여러 개 일 경우 이들 특징 값들에 의한 두 개체 사이의 유사도를 구하여 확률이 높은 값을 인식한다. DTW 알고리즘은 비선형 시간 정규화를 갖는 패턴 정합 알고리즘으로 공통적이고 균일하게 샘플간격을 갖게 패턴을 시간적으로 샘플링하여 사용한다[2].

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 구성되었던 모델에서 벗어난 어휘의 입력과 유사한 어휘의 입력은 인식하지 못하거나 근사한 어휘로 인식되어지는 경우가 발생하게 되고 이는 인식을 저하로 나타낸다[3]. 따라서 인식을 위한 처리 과정에서 사전 확률 분포는 알 수 없는 형태로 존재하므로 사전 확률을 모델링하여 베이스 정리에 의해 사후 확률 분포를 계산하고 원 신호를 추정하여 인식률을 계산한다[4].

본 논문은 벡터 모델 구성 시 확장성을 반영할 수 있는 바타차라 거리 측정 기법을 이용한 베이지안 인식 모델 구성 방법을 제안하였다. 특정 벡터 모델을 기반으로 바타차라 거리 측정법을 이용한 베이지안 인식 모델을 구성하여 탐색 중에 형성되는 벡터 모델을 이용하여 정확한 학습 모델로 인식하도록 유도하여 인식률을 향상시켰다. 특정 벡터 모델을 기반으로 바타차라 거리 측정법을 이용한 베이지안 인식 모델을 구성하여 제안한 베이지안 인식 모델을 실험한 결과 어휘 인식률에서 98.2%의 인식률을 나타내어 기존의 방법들에 비하여 인식 성능의 향상을 확인하였다.

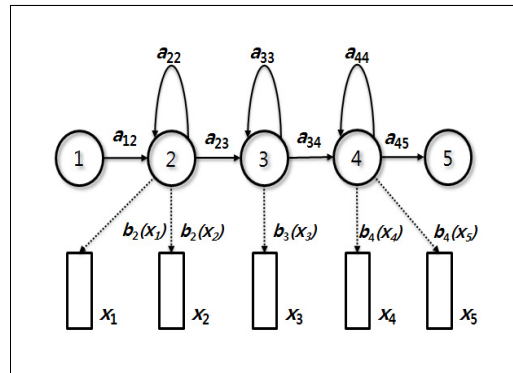
본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 언급하고 3장에서는 바타차라 거리 측정 기법을 이용한 베이지안 인식 모델 구성 방법에 대해 설명하며 4장에서는 시스템 평가를 수행하고 마지막으로 5장에서

결론을 맺는다.

2. 관련연구

2.1 HMM 모델

HMM은 음성 신호와 같이 시간에 따라 변하는 비정상적인 특징의 확률 밀도 함수를 갖는 신호를 모델로 구성하여 사용한다. 모델의 현재 상태는 복수 개의 상태로 이루어지며 각 상태는 확률 밀도 함수로 출력 확률을 의미한다. 모델의 현재 상태는 이전 시간의 모델의 상태에 따라 결정되는 마르코프 모델을 사용하며 상태 천이 확률을 이용하여 상태 사이의 이동 확률을 정의하여 사용한다[5]. 시간에 따라 변하는 확률 모델을 가지는 신호는 관찰된 신호의 값이 어떠한 상태의 확률 밀도 함수에서 발생할 것인지 알 수 없기 때문에 HMM을 이용한 모델링 기법이 효과적으로 사용된다. 관측할 수 없는 음성의 통계적인 특징을 관측 가능한 벡터열을 통해 추정함으로써 음성의 통계적인 변이성을 반영하여 사용한다[6].



[Fig. 1] HMM(Hidden Markov Model)

[Fig. 1]에는 HMM의 상태전이를 나타내었으며 음성 구간의 변이에 의한 상태천이 확률 A_i 와 각 상태에서의 관측 심볼의 출력 확률 B_j , 초기 확률 π 로 표현하며 다음과 같이 나타낸다.

$$\lambda = (A, B, \pi) \tag{1}$$

상태천이 확률 A_i 를 다음과 같이 나타낸다.

$$A = \{a_{ij}\} \quad (2)$$

상태전이 확률 A 는 a_{ij} 로 표시하며 i 상태에서 j 상태로 천이할 확률을 나타낸다. 출력 확률 B_j 를 다음과 같이 나타낸다.

$$B = \{b_j(k)\} \quad (3)$$

출력 확률 B 는 $b_j(k)$ 로 표시하며 j 상태에서 k 가 관측될 확률을 나타낸다. 초기 확률 π 를 다음과 같이 나타낸다.

$$\pi = \{\pi_i, \pi_i = p[q_1 = i]\} \quad (4)$$

초기 확률 π 는 π_i 로 표시하며 초기 상태를 표현하며 상태들은 서로 다른 상태로 연결되어 있으며 각 상태는 $\{1, 2, \dots, N\}$ 로 나타내며 시간 t 에서의 상태는 q_t 와 같이 나타낸다[7].

2.2 DTW 알고리즘

DTW(Dynamic time warping)는 시간의 차이로 길이가 동일하지 않아 비교하기 어려운 시계열 데이터 시퀀스들 간의 유사도를 측정하는 알고리즘으로 사용된다[8]. DTW는 두 시계열 데이터 시퀀스 $P = \{p_1, p_2, \dots, p_m\}$, $Q = \{q_1, q_2, \dots, q_n\}$ 에 대해 비선형 대응으로 처리한다[9]. 임의의 p_i 와 q_j 의 대응 (i, j) 를 w_k 라 할 때 다음과 같이 나타낸다.

$$w_k(i, j) = (p_i, q_j) \quad (5)$$

두 시퀀스의 대응은 정합 경로(warping path) W 는 다음과 같이 나타낸다.

$$W = w_1, w_2, \dots, w_z \quad (6)$$

또한, 정합 경로 W 의 특징을 다음과 같이 나타낸다.

$$w_{k+1} = (i, j) \text{ or } (i+1, j) \text{ or } (i, j+1) \quad (7)$$

정합 경로의 수는 데이터 시퀀스의 길이에 대해 기하급수적으로 증가한다. P 와 Q 의 임의의 데이터 쌍

(p_a, q_b) 의 거리 계산 함수를 $d(p_a, q_b)$ 라 하고, P 와 Q 에 대한 임의의 대응 $w_k = (i, j)$ 의 거리를 $c(w_k) = d(p_i, q_j)$ 로 나타낸다. 정합 경로 W 의 거리를 $C(W)$ 와 $DTW(P, Q)$ 는 다음과 같이 나타낸다.

$$C(W) = \sum_{k=1}^z c(w_k) \quad (8)$$

$$DTW(P, Q) = \min\left(\frac{C(W)}{z} \mid W \in P, Q\right) \quad (9)$$

$DTW(P, Q)$ 는 모든 정합 경로의 거리들 중 최소 거리를 계산하여 거리는 상당히 크게 나타나므로 동적계획법(dynamic programming)에 기초한 계산 방법을 사용하며 다음과 같이 나타낸다.

$$D(i, j) = d(p_i, q_j) + \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\} \quad (10)$$

두 데이터 시퀀스의 유사도는 $DTW(P, Q) = D(m, n)$ 으로 나타낸다.

3. 제안하는 융합인식 모델

3.1 우도 원리에 의한 베이시안 기법

베이시안 기법은 확률을 기반으로 임의의 계수를 사용하여 확률 분포로 추정한다. 알려진 데이터를 이용하여 모르는 데이터에 대한 학습을 반영하는 과정으로 설명된다[9]. 베이시안 확률 기법은 현재 상태만 보고 예측하는 것이 아니라 과거에 있었던 사실이나 예측자의 주관적인 분석까지 함께 넣어서 다음에 일어날 일을 예측하는 이론이다. 이런 특성 때문에 일반적인 확률 분석보다 복잡하지만 정확한 결론을 얻을 수 있다는 장점이 있으며 조건부 확률식은 다음과 같이 나타낸다[10].

$$p(\theta_j|x) = \frac{p(\theta_j)p(x|\theta_j)}{p(x)} \quad (11)$$

θ 는 추정할 모수를 나타내고, $p(\theta_j|x)$ 는 사후 확률, $p(\theta_j)$ 는 사전 확률을 나타내며 $p(x|\theta_j)$ 는 주어진 데이터의 함수인 우도 함수(likelihood function)를 나타낸다.

베이시안 확률 기법은 모델 특징으로부터 얻은 데이

터에 관한 특징과 데이터에 관한 과거의 경험이나 사전 지식 같은 주관적 견해를 수량화한 데이터의 특성을 결합시켜보다 정확한 확률 값을 나타낸다. 사전 확률에 새로운 증거를 대입하여 사후 확률을 얻는다는 의미에서 우도 원리(likelihood principle)를 사용하며 결과가 나왔을 때 그 결과가 나올 수 있는 여러 가능한 가설들의 평가 측도를 이용한다. 해석이 쉽고 모호한 결과를 제시하지 않으며 자연스러운 방법으로 정보 수정이 가능하고 표본 이론을 요구하지 않는다. 확률 기법을 위해서 정확한 계산을 하고자 Markov chain을 사용하는 경우가 있지만 베이지안 확률 기법은 정당한 확률 기법을 위해 근사 이론을 사용하지 않는 장점이 있다. 속성 값이 없는 상태에서 모델을 구축하여 학습을 통해 강인한 모델 구성이 가능하다. 관측 값 $x = (x_1, \dots, x_n)$ 일 때 사후 분포는 다음과 같이 나타낸다.

$$p(\theta_j|x, x_{n+1}) = \frac{p(\theta_j)p(x, x_{n+1}|\theta_j)}{p(x, x_{n+1})} \cdot p(\theta_j|x) \quad (12)$$

3.2 벡터 기반 바타차라 거리 측정 알고리즘

바타차라 거리 측정 알고리즘은 연속적이거나 불연속적인 두 확률 분포의 유사성을 측정하는 통계적 확률 계산으로 오류율을 측정하여 거리를 계산한다[11]. 확률적 계산으로 거리를 측정하는 방법이며 실시간으로 거리 계산이 가능하고 모델의 상태 경로 변이와 최적의 모델에 대한 확률적 모수 추정이 가능하다. 실시간 거리 계산을 위한 비터비 확률 방법이 사용되며 전향과 후향 확률 값에 의해 연산되어진다. 모델의 상태 경로를 확률적 모수 추정 방식으로 사용하며 거리 측정에서 반영되어야 할 오류율 측정을 통해 거리 측정의 인식율을 향상시킨다 [12].

거리 측정에서 반영되어야 할 오류율 측정이 포함되어 있는 바타차라 거리 측정법을 사용하여 거리 측정을 하였으며 다음과 같이 나타낸다.

$$d^p(i, j) = \left[\sum_{k=1}^n \frac{(\mu_{ik} - \mu_{jk})^2}{\sigma_{ik} + \sigma_{jk}} \right] \quad (13)$$

n 은 차수를 나타내고 μ_{ik}, μ_{jk} 는 평균을 나타내며 σ_{ik}, σ_{jk} 는 분산을 나타낸다. 두 모델 사이의 거리 계산을 위

해 모델의 특성을 통해 모델이 겹치는 부분을 측정하며 거리 사이의 범위는 0에서 ∞ 까지의 값을 갖는다. 각각의 모델 특성이 완전히 일치하거나 무한대로 떨어져 있는 거리 측정에 대해 다음과 같이 나타낸다.

$$B_s(p, q) = \sum_{x=1}^n \sqrt{p(x)q(x)} \quad (14)$$

$p(x)$ 는 p 번째 분포의 분산을 나타내며 $q(x)$ 는 q 번째 분포의 분산을 나타내고 p, q 는 서로 다른 두 분포를 나타낸다. 평균과 분산의 혼합 거리 측정은 비슷한 가중치를 갖는 가우시안들이 조합되어 변형되기가 쉬우므로 측정된 거리 데이터를 계산하여 구한다. 가중치 크기 (weighting scalar) B_s 는 가우시안의 가중치 w_1 과 w_2 의 대한 함수로 다음과 같이 나타낸다.

$$B_s = \sum_{x=1}^n \sqrt{\frac{w_1^2 + w_2^2}{w_1^2 w_2^2}} \quad (15)$$

4. 실험결과

본 연구에서는 어휘 인식 실험을 위하여 Aurora 2 데이터베이스[13]를 인식 데이터베이스로 사용하였고 특징 벡터를 이용한 모델을 구축하였으며 훈련과 인식을 위해서는 HTK(Hidden Markov Toolkit)를 사용하였다[14].

본 논문에서 제안한 모델을 이용한 인식을 향상 방법의 성능 검증을 위하여 어휘 인식 실험을 수행하였다. 제안한 2가지 방법의 장점을 살리기 위해 위너 필터(wiener filter)를 사용하여 훈련 과정과 실험 환경의 불일치 문제를 해결하였고 서울 시내 지역명 30개, 지하철역명 30개로 어휘 인식 목록을 구성하여 실험을 수행하였다[15]. 위너 필터는 2부분으로 구성되므로 우도 원리에 의한 베이지안 기법과 바타차라 거리 측정을 적용하였다. 제안한 시스템의 성능 평가를 위하여 기존 방식과 비교 실험을 하였으며 유클리디안(Euclidean) 알고리즘, DTW 알고리즘, 제안 방법에 대해 인식률을 측정된 결과 제안 방법에서 평균 98.2%의 성능을 나타내었다.

<Table 1>은 기존 방식인 유클리디안 알고리즘, DTW 알고리즘, 제안 방법을 실제 환경에서의 실험을 나타내었다.

<Table 1> Non-Noise Environment Recognition Rate

Speech	Recognition Rate (%)		
	Euclidean	DTW	Proposed Method
Speech Dependent	97.3	97.6	98.6
	96.1	96.7	98.1
	96.6	96.9	97.9
Average	96.7	97.1	98.2

<Table 1>에서 보는 것과 같이 잡음이 거의 없는 실내 환경에서 실험한 결과 유클리디안 알고리즘을 이용한 음성 인식을 평균 96.7%로 나타냈으며 DTW 알고리즘을 이용한 음성 인식을 평균 97.1%의 인식을 나타내었고 제안방법의 인식을 평균 98.2%를 나타내었다.

<Table 2> Noise Environment Recognition Rate

Speech	Recognition Rate (%)		
	Euclidean	DTW	Proposed Method
Speech Dependent	84.3	84.9	86.3
	83.9	84.1	86.1
	84.6	85.3	85.7
Average	84.3	84.8	86.0

<Table 2>는 기존 방식인 유클리디안 알고리즘, DTW 알고리즘, 제한 방법을 실외 환경에서의 실험을 나타낸다. 잡음 환경에서 실험한 결과 유클리디안 알고리즘을 이용한 음성 인식을 평균 84.3%로 나타냈으며 DTW 알고리즘을 이용한 음성 인식을 평균 84.8%의 인식을 나타내었고 제안방법의 인식을 평균 86.0%를 나타내었다.

5. 결과

본 논문은 벡터 모델 구성 시 확장성을 반영할 수 있는 바타차랴 거리 측정 기법을 이용한 베이지안 인식 모델 모델 구성 방법을 제안하였다. 특징 벡터 모델을 기반으로 바타차랴 거리 측정법을 이용한 베이지안 인식 모델을 구성하여 탐색 중에 형성되는 벡터 모델을 이용하여 정확한 학습 모델로 인식하도록 유도하여 인식을 향상시켰다.

이를 확인하기 위해 유클리디안 알고리즘, DTW 알고

리즘과 비교하였다. 실내 환경에서 실험한 결과 유클리디안 알고리즘을 이용한 음성 인식을 평균 96.7%로 나타냈으며 DTW 알고리즘을 이용한 음성 인식을 평균 97.1%의 인식을 나타내었고 제안방법의 인식을 평균 98.2%를 나타내었다. 또한 실외 환경(잡음 환경)에서 실험한 결과 유클리디안 알고리즘을 이용한 음성 인식을 평균 84.3%로 나타냈으며 DTW 알고리즘을 이용한 음성 인식을 평균 84.8%의 인식을 나타내었고 제안방법의 인식을 평균 86.0%를 나타내었다.

REFERENCES

- [1] 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.
- [2] Sang-Yeob Oh. Improving Phoneme Recognition based on Gaussian Model using Bhattacharyya Distance Measurement Method. Journal of Korea Multimedia Society. v.14 no.1, pp.85-93, 2011.
- [3] 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.
- [4] 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.
- [5] A. Srinivasan, Speech Recognition Using Hidden Markov Model, Applied Mathematical Sciences, vol. 5, no. 79, pp. 3943-3948, 2011.
- [6] 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.
- [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. 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.
- [9] Le Hoang Linh, Nguyen Thanh Hai, Ngo Van Thuyen, Tran Thanh Mai, Vo Van Toi. MFCC-DTW Algorithm for Speech Recognition in an Intelligent Wheelchair. IFMBE proceedings v.46, pp.417-421, 2015.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] Beaufays, F., Vanhoucke, V., & Strophe, B. Unsupervised discovery and training of maximally dissimilar cluster models. Proc. Interspeech, pp. 66-69, 2010.
- [14] Young, S. HTK: Hidden Markov Model Toolkit V3.4.1. Cambridge University, Engineering Department, Speech Group. 1993.
- [15] 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.

오 상 엽(Oh, Sang Yeob)



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