

화자적응 시스템의 성능향상을 위한 FCM 알고리즘에 대한 연구

방기덕*, 전선도, 강철호
광운대학교 전자통신공학과

A Study on FCM Algorithm for the Performance Improvement of Speaker Adaptation System

Bhang, Ki-Duck* , Jun, Sun-Do , Kang, Chul-Ho
Kwangwoosn University
E-Mail : carrot@dspalpha.kwangwoon.ac.kr

요약

기존의 반연속 HMM의 파라미터들 중에서 평균 벡터와 분산 행렬은 Maximum Likelihood Estimation 방법을 사용하여 학습한다. 본 논문에서는 평균 벡터를 위하여 Fuzzy c-means(FCM) 알고리즘을 사용하였고 분산 행렬을 위하여 FCM 알고리즘의 평균 벡터를 적용, 변형한 새로운 함수를 사용하여 화자적응에 적용하였다[1]. 이러한 평균 벡터와 분산 행렬의 추정 방법은 새로운 화자에 대한 적응 능력을 갖는다. 제안한 방법을 적용한 한국어 격리 단어에 대한 컴퓨터 모의 실험 결과 새로운 화자에 대해 적응함을 확인하였다.

I. 서론

화자중속 음성인식 시스템에서는 인식의 성능에 크게 영향을 미치는 시스템의 인자들(acoustic prototypes, HMM 인자, 신경회로망의 가중치 등)이 인식의 대상이 되는 특정 화자에 맞게 구성되어 질 수 있기 때문에 좋은 인식 결과를 얻을 수 있다. 그러나, 화자독립 음성인식 시스템에서는 서로 다른 화자들 사이의 성도 길이, 구강 크기 등의 해부학적인 차이와 액센트, 발성 속도와 크기 등 화자의 개별적인 발성 습관의 차이가 존재하기 때문에 인식될 음성이 학습에 사용된 음성 샘플에 의해 잘 특성화되지 못할 때 시스템의 성능을 상당히 저하시킨다.

음성인식 시스템에서 인식 대상 화자가 다른 화자들로부터 학습된 음성인식 시스템에 적용하기 위해서는

상당히 많은 데이터를 전체 시스템에 적용 시켜야 하는 문제점이 발생한다. 화자독립 및 화자중속 음성인식 시스템에서는 새로운 화자에 대해서 이러한 불필요한 재적용 과정을 피하기 위해 인식을 처음 수행하기 전에 소수의 단어나 단문의 문장 등의 적용화 문장을 발성하여 이 음성 데이터를 자동으로 적용할 수 있는 화자적응 학습기능을 인식시스템에 추가한다.

이러한 화자 적응화를 통하여 서로 새로운 화자특성을 고려해 줄으로써 인식시스템의 인식률을 높일 수 있다. 화자 적응 방법으로는 VQ(Vector Quantization) 코드북의 Bayesian learning 방법, VQ mapping 방법, Unified modeling 이론 등이 있다[1].

본 논문에서는 화자 적응을 FCM 알고리즘을 이용하여 설계하였다. 이 방법은 새로운 화자에 대해서 HMM 파라미터와 VQ 코드북을 적응 데이터에 의해서 재추정하는 방법이다.

특히 본 논문에서는 FCM 알고리즘에 의해서 평균벡터를 구하고 Maximum likelihood estimation의 분석형태를 도입하여 분산 벡터를 구하였다.

이러한 방법의 적용화 공식을 사용하여 실험한 결과 적용화 전과 적용화 후의 인식률을 비교하여 인식률의 향상됨을 알 수 있었다.

II. 반연속 HMM의 기본 이론

이산 HMM보다 작은 코드북을 사용하고 연속 HMM보다 적은 계산량이 필요하도록 두 경우를 결합한 것을 반연속 HMM (Semi-Continuous HMM: SCHMM)이라

한다. 이 경우는 벡터 양자화 코드워드를 가우시안 분포의 평균치들로 생각하며, 각 분포의 공분산 행렬의 대각선 값들을 코드북에 포함시키게 된다. 즉, 크기 L 인 코드북에서 각 코드워드에 해당하는 D차 평균값 μ 와 공분산 행렬의 주대각선 성분 \sum 가 D개 주어지게 된다. 각 코드워드마다 공분산 행렬값이 주어지므로 일반적인 벡터 양자화의 경우와는 달리 유클리디안 (Euclidean) 거리 대신 마할라노비스 (Mahalanobis) 거리를 사용하게 된다.

이와같은 경우, 확률 밀도 행렬 B는 상태 j에서 l번째 코드워드에 해당하는 가우시안 성분을 발견할 상대적인 크기가 되므로 확률 밀도 행렬 요소 b_{jl} 은 연속 밀도 HMM의 C_{jl} (상태 j에서 l번째 가우시안 성분의 상대적인 크기) 와 같은 역할을 한다. 그러면 상태 j에서 관찰값 O_t 를 발견할 확률은

$$p_j(o_t) = \sum_{i=1}^L b_{ji} p(o_t | \mu_i, \Sigma_i) \quad (2-1)$$

로 주어진다.

반연속 HMM을 이용하여 음성의 학습 데이터를 잘 표현하기 위해서는, 반연속HMM의 모수 재추정 (parameter reestimation) 과정이 필요하다. 이것은 모수가 주어졌을 때, 관찰열을 발견할 확률을 반복적으로 최대화시키는 것으로서 EM(Expectation Maximization) 알고리즘이라 한다. 또한 주어진 반연속 HMM 모수들로부터 하나의 관찰열에 대응되는 가장 적합한 상태열을 찾는 방법으로 Viterbi 알고리즘이 있다.

반연속 HMM에서 재추정해야 할 변수들은 행렬, 상태 천이 A행렬, 출력 확률 행렬 B 및 코드북의 μ, Σ 값들이다. 이들은 다음의 재추정식으로부터 구한다.[2]

$$\pi_i = \gamma_i(i) \quad (2-2)$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \gamma_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (2-3)$$

$$b_j(k) = \frac{\sum_{t=1}^{T-1} \xi_t(j, k)}{\sum_{t=1}^{T-1} \gamma_t(j)} \quad (2-4)$$

$$\mu_j = \frac{\sum_{t=1}^{T-1} \xi_t(j) o_t}{\sum_{t=1}^{T-1} \xi_t(j)} \quad (2-5)$$

$$\sum_j = \frac{\sum_{t=1}^{T-1} \xi_t(j) (o_t - \mu_j) (o_t - \mu_j)^t}{\sum_{t=1}^{T-1} \xi_t(j)} \quad (2-6)$$

여기에서 중간 변수는 식(2-7), (2-8), (2-9), (2-10)과 같다.

$$\gamma_t(i, j) = P(s_t = i, s_{t+1} = j | O, \lambda) \quad (2-7)$$

$$\gamma_t(i) = P(s_t = i | O, \lambda) \quad (2-8)$$

$$\xi_t(i, k) = P(s_t = i, o_t = v_k | O, \lambda) \quad (2-9)$$

$$\xi_t(k) = P(o_t = v_k | O, \lambda) \quad (2-10)$$

III. FCM 클러스터링 알고리즘

퍼지 클러스터링 알고리즘에서 가장 일반적인 것이며 $x = \{x_1, x_2, \dots, x_n\}$ 는 실공간 R^K 안에 포함되어 있는 유한 데이터 집합이라고 가정하자. 정수 C, ($2 \leq C \leq n$)는 *priori* 로 알려진 클러스터의 숫자이다.

FCM 클러스터링 알고리즘은 X를 C 퍼지 클러스터로 분할하는 것을 시도하고 x안의 가까운 성분은 유사한 classification(membership value)을 가질 것이다. 그리고 다른 성분은 다른 classification을 가질 것이다.

$d_{ik} = \|x_k - v_i\|$ 에서 $\|x_k - v_i\|$ 는 x_k 와 v_i 사이의 거리를 나타낸다. 이것은 클러스터의 형태에 의해서 결정되어지는데 그 형태가 원형일 경우는 유클리디안 디스턴스를 선택한다[3][5].

membership value μ_{ik} 는 다음 식 (2-11), (2-12)의 조건에 의해 제약받는다.

$$\mu_{ik} \in [0, 1] \quad (2-11)$$

$$\sum_{i=1}^c \mu_{ik} = 1, \forall k \quad (2-12)$$

Lagrangian multiplier method를 사용하여 이러한 제약을 해결하고, 다음 식 (2-13), (2-14)를 얻을 수 있다.

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c (d_{ik}/d_{jk})^{2/m-1}}, \forall i, k \quad (2-13)$$

$$V_i = \frac{\sum_{k=1}^n \mu_{ik}^m x_k}{\sum_{k=1}^n \mu_{ik}^m}, \forall i \quad (2-14)$$

$m \in [1, \infty]$ 는 weight exponen이며 클러스터들 사이

의 멤버십 분할 범위에 영향을 미치며 실험적으로 선택하여진다. $V = \{V_1, V_2, \dots, V_C\}$ 는 cluster center의 집합이고 코드북 벡터로 사용된다[4]. 위의 주된 결과들은 다음의 FCM clustering algorithm 안에 포함된다[6].

IV. 제안한 FCM 알고리즘을 이용한 반연속 HMM 적용 방법

기존의 maximum likelihood estimation에서 사용한 평균 벡터와 분산 확률의 재추정 공식은 식 (2-5), (2-6)을 사용하여 적용되어 사용하였다[1]. 본 논문에서 제안한 재추정 공식은 위에서 3장에서 설명한 FCM 알고리즘에서의 클러스터 센터 식(2-14)을 평균 벡터의 재추정 공식으로 사용하며 분산 확률은 다음의 해를 재추정 된다.

$$\bar{V}_i = \frac{\sum_{k=1}^n \mu_{ik}^m x_k}{\sum_{k=1}^n \mu_{ik}^m}, \quad \forall i \quad (2-15)$$

을 평균벡터의 재추정 식으로[3]

$$\sigma_i = \frac{\sum_{k=1}^n \mu_{ik}^m x_k (v_i - x_k)(v_i - x_k)}{\sum_{k=1}^n \mu_{ik}^m}, \quad \forall i \quad (2-16)$$

을 분산 확률의 재추정 공식으로 사용하였다. 이 공식을 Maximum likelihood estimation 유도 방법을 적용하여 구한 함수이다.

V. 실험 방법 및 결과

실험은 한국어 10단어(서울, 인천, 수원, 안양, 의정부, 포천, 용인, 일산, 평택, 성남)를 사용하여 10명의 화자가 2번 발음하였다.

VQ 코드북을 위하여 FCM 클러스터링 알고리즘을 사용한 반연속 HMM을 인식기로 사용하여 실험하였는데 FCM 알고리즘의 초기 값으로는 K-means에 의한 평균 벡터와 분산행렬을 사용하였고 FCM 알고리즘에 의해서 학습되어졌다. 100번의 학습에 의해서 실험한 결과가 표-1에 나타내었다

본 논문에서 제안한 적응화 공식을 사용하여 4명의 화자에 의해서 인식실험을 한 결과가 표-2이다.

적용화 실험을 위해서 4명의 화자에 대해서 각각 10번, 추가 10번, 추가 20번, 추가 30번 실험한 결과가 표-3이다.

또한 적응화 실험을 통해서 기존의 화자에 대한 인식률 향상을 실험한 결과가 표-4에 나타내었다.

<표-1>

화자	인식률(%)	화자	인식률(%)
jsd	70	kdw	60
jsh	80	akm	50
myu	80	kjhi	50
bkd	80	ksjj	50
kbt	60	ksy	60

<표-2>

화자	인식률(%)
kkw	50
lcn	60
pkj	50
sdv	50

<표-3>

화자	반복			
	10반복	+10반복	+ 20반복 (2번째발음)	+ 30반복 (2번째발음)
kkw	70	70	70(80)	70(90)
lcn	80	80	80(80)	90(80)
pkj	70	80	80(70)	80(90)
sdv	80	90	90(70)	90(70)

<표-4>

화자	인식률(%) (인식률 향상)	화자	인식률(%) (인식률 향상)
jsd	70(0)	kdw	70(10)
jsh	80(0)	akm	60(10)
myu	80(0)	kjhi	70(20)
bkd	80(0)	ksjj	70(20)
kbt	70(10)	ksy	80(80)

표-1에서는 평균 64%의 인식률을 나타내며 표-2에서는 평균 52.5%의 인식률을 나타내었고, 표-3에서는 반복회수가 증가함에 따라 각각 75%, 80%, 80%(77.5%), 82.5%(82.5%)로 나타났다.

또한, 기존의 화자에 대해 다시 인식실험을 한 결과 표-4에서와 같이 73%의 인식률을 나타내었으며 이는

적용전에 비해 9% 증가하였다.

IV. 결론 및 고찰

본 논문에서 제안한 FCM을 이용한 평균 벡터와 분산 벡터의 추정에 의해 적응화 인식 실험을 하였다. 학습에 참여하지 않은 4명의 화자를 적응화 실험 하였다. 실험 결과는 화자를 적응함으로써 인식률이 증가함을 확인하였다. 또한 애초에 학습에 참여한 화자에 대한 인식 실험에서도 적응하기전의 인식 결과보다 높은 인식률을 보임으로써 화자의 특성을 잃지 않을 뿐만 아니라 보다 높은 인식률을 보임을 알 수 있다.

현재 실험은 10 단어로 제한 하였으나 앞으로 단어의 확장이 요구되며 또한 더 많은 적응을 수행하여야 할 것이다.

참 고 문 헌

- [1] X.D. Huang, Y.Ariki, M.A. Jack. "Hidden Markov Models for Speech Recognition"
- [2] Lawrence Rabiner, Biing-Hwang Juang, "Fundamentals of Speech Recognition"
- [3] J.-S. R. JANG, C.-T. SUN, E. MIZUTANI. "Neuro-Fuzzy and Soft Computing"
- [4] Nicolas B. Karayiannis, James C. Bezdek. "An Integrated Approach to Fuzzy Learning Vector Quantization and Fuzzy c-Means Clustering"
- [5] BART KOSKO, "Fuzzy Engineering" ,PRENTICE HALL
- [6] Chung-Woei and Chung-Chung Chju and Po-Chiang Lu, "Codebook design for vector quantization of images based on fuzzy C-means clustering algorithm", Optical engineering, vol. 36 No 2, February 1997