국부 퍼지 클러스터링 PCA를 갖는 GMM을 이용한 화자 식별*

Speaker Identification Using GMM Based on Local Fuzzy PCA

이 기 용**
Ki-Yong Lee

ABSTRACT

To reduce the high dimensionality required for training of feature vectors in speaker identification, we propose an efficient GMM based on local PCA with Fuzzy clustering. The proposed method firstly partitions the data space into several disjoint clusters by fuzzy clustering, and then performs PCA using the fuzzy covariance matrix in each cluster. Finally, the GMM for speaker is obtained from the transformed feature vectors with reduced dimension in each cluster. Compared to the conventional GMM with diagonal covariance matrix, the proposed method needs less storage and shows faster result, under the same performance.

Keywords : Speaker Identification, GMM, Fuzzy clustering, PCA

1. 서 론

대각 공분산 행렬의 GMM은 화자 식별과 화자 확인을 위하여 많이 사용되고 있다 (Reynolds and Rose. 1995). 음성 신호로부터 추출된 특징 벡터들의 요소들이 상관관계가 있기 때문에, 효과적인 근사치를 얻기 위해서 더 많은 양이 혼합성분을 필요로 한다. 또한, 더 높은 차원의 특징 벡터 집합은 화자 인식 시스템의 성능을 향상시킬 수 있다 (Liu and He. 1999). 그러나 특징 벡터 차수와 혼합성분 수의 증가는 몇 가지 문제를 야기시킨다. 먼저, 더 높은 차원의 특징 벡터 집합과 더 많은 수의 혼합 성분의 집합은 classifier를 특징지우는데 더 많은 파라메터와 더 많은 저장공간을 필요로 한다. 계산량과 비용의 증가는 실시간 구현을 어렵게 할 뿐만 아니라, 등록 과정에서 더 많은 양의 음성 데이터를 필요로 한다.

* 본 논문은 2003학년도 숭실대학교 교내학술연구비 지원에 의하여 수행되었습니다.
** 숭실대학교 정보통신전자공학부

특징 벡터의 차원을 감소시키기 위해서, PCA를 이용한 화자 인식 방법이 연구되어 왔다(Ariki, et al. 1996; Seo, et al. 2001; 이윤정 외. 2003). PCA는 특징 벡터 추출방법의 하나로, 특징 벡터의 차원을 감소시키고, 원 특징 벡터의 공간을 더 적은 부분 공간으로 변환하여 특징벡터들 사이의 상관 관계를 감소시킨다.

본 논문에서는, 화자 식별을 위해서 특징 벡터의 차원을 효과적으로 감소시키기 위해서 퍼지 클러스터링을 통한 국부 PCA에 기초를 둔 GMM을 제안한다. 첫번째로, 제안된 방법은 퍼지 클러스터링을 통해서 특징 벡터 공간을 몇 개의 분리된(disjoint) 공간으로 나눈다(Tran and Wagner. 2000; Gustafson and Kessel. 1979; Gath and Geva. 1981). 두 번째로 각각의 클러스터에 PCA를 적용해서 감소된 차원의 새로운 특징 벡터를 얻는다(Seo, et al. 2001). PCA는 p-차원의 특징벡터를 각각의 클러스터 안에서 전체 공분산의 고유벡터를 통해 묶여지는 L차원의 선형 부분 공간으로 변환한다. 제안된 방법의 효율성은 제안된 방법과 기존의 GMM 사이에 비교적인 실험을 통해서 볼 수 있다

2. 퍼지 클러스터링에 기초를 둔 퍼지 PCA

p-차원 공간의 특징 벡터의 집합을 $X=\{x_1,\cdots x_T\}$라 가정하자. 그리고 $U=[u_{jt}]$는 행렬의 원소가 j번째 클러스터 R^j 에서 x_t 의 멤버인 행렬이다. X를 위한 퍼지-분할 (*Fuzzy K-partition*) 공간은 행렬U 의 집합이며 아래와 같다.

$$0\le u_{jt}\le 1, \qquad j=1,2,\cdots,K, \quad t=1,2,\cdots,T$$
$$\sum_{j=1}^{K} u_{jt}=1,\ \forall t, \qquad 0<\sum_{t=1}^{T} u_{jt}<T,\ \forall j \tag{1}$$

위 식에서, $0\le u_{jt}\le 1, \forall j,t$ 이며, 각각의 x_t 는 K 퍼지 클러스터들 사이에 임의의 소속 분포 가지는 것이 가능하다는 것을 의미한다(Bezdek. 1981).

X에서 퍼지 클러스터링을 위한 가장 잘 알려진 목적함수는 *least-squares function*이다. 퍼지 k-평균 함수를 위한 무한 그룹은 함수 J_1으로부터 아래와 같이 생성되어진다

$$J_m(U,C;X)=\sum_{t=1}^{T}\sum_{j=1}^{K}(u_{jt})^m d^2(x_t,c_j) \qquad K\le T \tag{2}$$

$U=[u_{jt}]$는 X의 K-분할이고, $m(>1)$은 각각의 퍼지 멤버 u_{jt} 에서 가중치 이며 퍼지의 정도(*degree of fuzziness*)라고 부른다. c_j 는 j번째 클러스터 R^j 의 중심이다. 식 (2)에서 $d^2(x_t,c_j)$은 x_t 와 c_j 의 거리이며, 아래와 같이 정의된다.

$$d^2(x_t,c_j) = \| x_t - c_j \|_F^2 = (x_t - c_j)^T F_j^{-1}(x_t - c_j) \tag{3}$$

F_j는 j번째 클러스터의 퍼지전체공분산이다.

퍼지 K-평균 알고리즘은 $J_m(U,C;X)$을 위한 최적 쌍의 부분으로써의 행렬 U가 데이터 X의 적당한 분할을 보장한다는 가정하에, U와 C에 대해서 $J_m(U,C;X)$을 최소화 시키는 것에 기초를 둔다. (2)식의 퍼지 목적 함수 $J_m(U,C;X)$를 최소화 시키는 방법은 아래와 같이 주어진다.

$$u_{jt} = \frac{\left[\frac{1}{d^2(x_t,c_j)}\right]^{1/(m-1)}}{\sum_{i=1}^{K}\left[\frac{1}{d^2(x_t,c_i)}\right]^{1/(m-1)}} \tag{4}$$

$$c_j = \frac{\sum_{t=1}^{T}(u_{jt})^m x_t}{\sum_{t=1}^{T}(u_{jt})^m} \tag{5}$$

$$F_j = \frac{\sum_{t=1}^{T} u_{jt}(x_t - c_j)(x_t - c_j)^T}{\sum_{t=1}^{T} u_{jt}} \tag{6}$$

퍼지 PCA는 퍼지 전체 공분산 행렬과 고유치를 계산하여 얻을 수 있다. 변환된 좌표축의 중요성은 고유치 크기로 측정하기 때문에, 가장 큰 고유치들과 관계가 있는 오직 L-개의 주 고유벡터로 얻을 수 있으며, 특징 벡터를 최적으로 변환하는데 사용된다.

등록과 테스트 동안에, GMM을 위한 각각의 입력 벡터는 아래 식으로 변환된다

$$y_{t_j} = \Phi_j x_t, \qquad \text{if } x_t \in R^j \tag{7}$$

$\Phi_j = (\phi_1 \phi_2 \cdots \phi_L)_j$는 j- 번째 클러스터 R^j의 L 개의 고유벡터를 행으로 하는 $L \times p$ 가중치 행렬이다. 벡터 ϕ_i는 F_j의 i번째로 큰 고유치에 대응하는 고유벡터이다. 식(7)의 공분산 행렬은 대각행렬이다.

3. 국부 PCA에 기초를 둔 GMM

j번째 클러스터 R^j에서 T_j의 등록 벡터열에 대해서, 가우시안 혼합성분 밀도는 M_j 성분 밀도의 가중치 합으로 아래와 같이 정의된다.

$$p(y_{t_j}|\lambda)=\sum_{i=1}^{M_j} p_{j,i} b_i(y_{t_j}) \tag{8}$$

위 식에서,

$$b_i(y_{t_j})=\frac{1}{(2\pi)^{\frac{L}{2}}|\Sigma_{j,i}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(y_{t_j}-\mu_{j,i})'\Sigma_{j,i}^{-1}(y_{t_j}-\mu_{j,i})\right\} \text{ 이다.}$$

$\mu_{j,i}$ 는 평균이고 $\Sigma_{j,i}$ 는 분산이다. 혼합성분 가중치 $\sum_{j=1}^{K}\sum_{i=1}^{M_j} p_{j,i}=1$ 을 만족한다. $Y=\{Y_1,\cdots,Y_K\}$가 주어져 있을 때, 화자 모델을 위한 완전한 GMM은 모든 성분 밀도로 얻어진 평균벡터, 공분산 행렬, 혼합 성분 가중치로 파라메터화 된다. 이러한 파라메터들은 아래와 같이 표현된다.

$$\lambda=\{p_{j,i},\mu_{j,i},\Sigma_{j,i}\} \quad i=1,2,\cdots,M_j \text{ and } j=1,2,\cdots,K \tag{9}$$

그리고, GMM의 유사도는 아래와 같이 쓸 수 있다.

$$p(Y|\lambda)=\prod_{t_1=1}^{T_1} p(y_{t_1}|\lambda)\cdot\cdots\cdot\prod_{t_K=1}^{T_K} p(y_{t_K}|\lambda) \tag{10}$$

파라메터 추정은 EM 알고리즘을 사용해서 반복적으로 얻을 수 있다(Seo, et al. 2001). 각각의 EM 반복에서, 모델의 유사도 값의 단조 증가를 만족하는 재추정 식들은 아래와 같이 구할 수 있다.

- 혼합성분 가중치(Mixture Weights)

$$p_{j,i}=\frac{1}{T}\sum_{t_j=1}^{T_j} p(j,i|y_{t_j},\lambda) \tag{11.a}$$

- 평균(Means)

$$\mu_{j,i}=\frac{\sum_{t_j=1}^{T_j} p(j,i|y_{t_j},\lambda)y_{t_j}}{\sum_{t_j=1}^{T_j} p(j,i|y_{t_j},\lambda)} \tag{11.b}$$

- 분산(Variance)

$$\Sigma_{j,i} = \frac{\sum_{t_j=1}^{T_j} p(j,i|y_{t_j},\lambda)(y_{t_j}-\mu_{j,i})(y_{t_j}-\mu_{j,i})^T}{\sum_{t_j=1}^{T_j} p(j,i|y_{t_j},\lambda)} \tag{11.c}$$

j 번째 클러스터의 혼합성분 i의 사후 확률은 아래와 같이 주어진다.

$$p(j,i|y_t,\lambda) = \frac{p_{j,i}b_i(y_{t_j})}{\sum_{i=1}^{M_j} p_{j,i}b_i(y_{t_j})} \tag{12}$$

$L=p$ 이고 $K=1$ 일 때, 제안된 방법은 Liu와 He(1999)의 OGMM 방법과 같다. Liu와 He(1999)에 의해서 제안된 방법은 본 논문에서 제안한 방법의 특별한 경우로써 간주된다.

3. 화자 식별

화자 식별에서, 각각의 화자 S는 GMM $\lambda_1,\cdots,\lambda_S$ 으로 표현되어진다. 화자 식별의 목적은 주어진 특징열에 대해서 사후 확률이 최대가 되는 화자 모델을 찾는 것이다.

$$\begin{aligned}\hat{s} &= \max_{1\le j\le S}\sum_{t=1}^{T}\log p(y_t|\lambda_j) \\ &= \max_{1\le j\le S}\sum_{j=1}^{K}\sum_{t_j=1}^{T_j}\log\left(\sum_{i=1}^{M_j} p_{j,i}b_i(y_{t_j})\right)\end{aligned} \tag{13}$$

4. 실험 및 결과

본 논문에서는 제안된 방법의 효율성을 보여주기 위해서 화자 식별 실험을 했다. 실험에서 사용된 음성 데이터는 4 달 동안에 3 세션에 걸쳐 발성되었다. 100 명의 화자(여자: 50, 남자: 50)은 각각의 세션에서 각자의 문장을 5 번씩 발성했다. 음성 데이터는 11 kHz로 녹음되었고, 12 차 LPC 켑스트럼과 13 차 델타 켑스트럼계수(p=25)로 파라메터화되었다. 윈도우 사이즈는 10 ms의 중첩을 가진 20 ms를 사용하였다. 첫 번째 세션에서 녹음된 발성을 등록과정에 사용하였고, 나머지 데이터는 테스트에 사용하였다.

제안된 방법과 기존 GMM에서 필요로 하는 파라메터의 수는 표 1에 나타나있다. 제안된 방법은 변환 행렬을 저장하기 위한 $K \times L \times p$ 의 저장 공간과 퍼지 클러스터링을 위한 $K \times p$ 의 저장 공간을 더 필요로 한다. 그러나 기존의 GMM 보다 더 적은 파라메터를 필요로 한다. 예를 들어, $M_p = 16$, $M_c = 64$, $K = 2$, $L = 17$, $p = 25$ 인 경우에, 제안된 방법은 1,460 개의 파라메터를 필요로 하는 반면에, 기존의 방법은 3,264 개의 파라메터를 필요로 한다.

표 2는 혼합 성분과 클러스터의 개수에 따른 화자 식별의 성능을 보여주고 있다. M 개 혼합성분과 K 개 클러스터를 가진 퍼지 PCA GMM의 총 혼합성분의 개수는 $M \times K$ 개를 가진 기존의 GMM과 비슷하다. 그러나, M 개 혼합성분과 K 개 클러스터를 가진 퍼지 PCA GMM의 성능이 더 좋다.

그림 1은 화자 모델을 등록할 때 혼합성분의 개수와 화자 식별성능과의 관계를 보여주고 있다. 본 실험에서는 기존의 PCA는 L=17(Liu and He. 1999), 제안된 방법은 L=17, K=2을 사용하였다. 그림 1에서부터, 제안된 방법의 화자 식별 능은 다른 것들 보다 더 좋은 것을 볼 수 있다. 게다가, 기존의 GMM 방법은 기존의 PCA 방법(Liu and He. 1999)이나 제안된 방법보다도 더 낮은 화자 식별 성능을 나타냈다.

그림 2는 변환된 특징 벡터 L차원에 따른 화자 식별 성능을 보여준다. 그림으로부터, 감소된 차원 ($11 \le L \le 25$)을 가지는 제안된 방법은 기존의 GMM인 p=25일 때와 비교해서 같거나 더 좋은 성능을 가진다.

5. 결 론

본 논문에서는, 화자 식별을 위해서 특징 벡터의 차원을 효과적으로 감소시키기 위해서 퍼지 클러스터링을 가진 국부 PCA에 기초를 둔 효과적이 GMM을 제안했다. 제안된 방법은 데이터 공간을 퍼지 클러스터링으로 몇몇의 분리된 클러스터들로 나누고, 각각의 클러스터에서 퍼지 공분산 행렬을 사용해서 PCA를 수행한다.

마지막으로, 화자를 위한 GMM은 각각의 클러스터에서 감소된 차원을 가지고 변환된 특징 벡터들로부터 얻어진다.대각 공분산 행렬을 가진 기존의 GMM과 비교해서, 제안된 방법은 똑같은 성능을 유지하면서, 더 적은 기억 공간을 필요로 하고, 더 빠른 결과를 보여준다. 제안된 방법과 기존의 방법들 사이의 비교 결과는 제안된 방법이 효과적이라는 것을 보여준다.

참 고 문 헌

[1] Seo, C.W., Lee, K.Y. and Lee, J. 2001. "GMM based on Local PCA for Speaker Identification." *Electronics Letters* 37, 24, 1486-1488.

[2] Reynolds, D. and Rose, R. 1995. "Robust text-independent speaker identification using Gaussian mixture speaker models." *IEEE Trans. on SAP*, 3(1), 72-82.

[3] Tran, D. and Wagner, M. 2000. "Fuzzy Entropy Clustering." *Proceedings of the FUZZ-IEEE'2000 Conference*, 1, 152-157.

[4] Gustafson, E.E. and Kessel, W.C. 1979. "Fuzzy clustering with a fuzzy covariance matrix." *Proc. IEEE CCD*, Sandiego, CA, 761-766.

[5] Gath, I. and Geva, A.B. 1981. "Unsupervised optimal fuzzy clustering." *IEEE Trans. on PAMI*, 11(7), 773-781.

[6] Bezdek, J.C. 1981. "Pattern Recognition with Fuzzy Objective Function Algorithms." Plenum Press, New York and London.

[7] Liu, L. and He, J. 1999. "On the use of orthogonal GMM in speaker recognition." *ICASSP99*, 845-849.

[8] Ariki, Y., Tagashira, S. and Nishijima, M. 1996. "Speaker recognition and speaker normalization by projection to speaker subspace." *ICASSP96*, 31

[9] 이윤정, 서창우, 강상기, 이기용. 2003. "화자식별을 위한 강인한 주성분 분석 가우시안 혼합 모델." 한국음향학회지, 22(7), 519~527.

제출일자: 2003. 11. 10.
게재결정: 2003. 12. 15.

▲ 이기용
서울시 동작구 상도 5동 1-1 (우: 156-743)
숭실대학교 정보통신 전자공학부
Tel: +82-2-820-0908 Fax: +82-2-817-4591
E-mail: kylee@ssu.ac.kr

표 1. 제안된 방법과 기존 GMM에서 필요로 하는 파라메터의 수

Proposed GMM	Conventional GMM
$M_o(2L+1)+Kp(L+1)$	$M_c(2p+1)$

표 2. 혼합 성분과 클러스터의 개수에 따른 화자 식별의 성능($L=p$)

M \ K	1	2	3	4
4	91.74	93.72	94.05	94.62
8	91.81	95.73	95.29	95.37
12	93.65	95.85	95.88	95.99
16	93.72	96.36	96.01	96.76
32	95.07	96.53		
64	95.98			

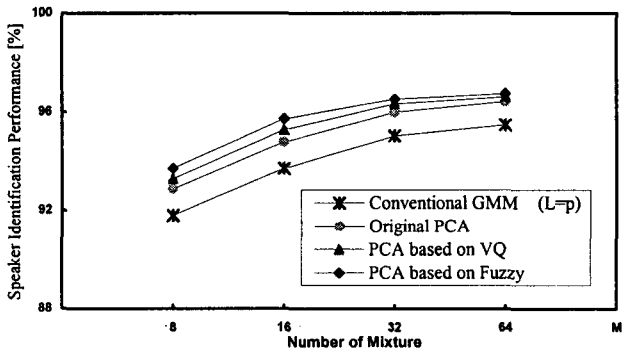

그림 1. 혼합 성분의 개수와 화자 식별률의 관계 (L=17, K=2)

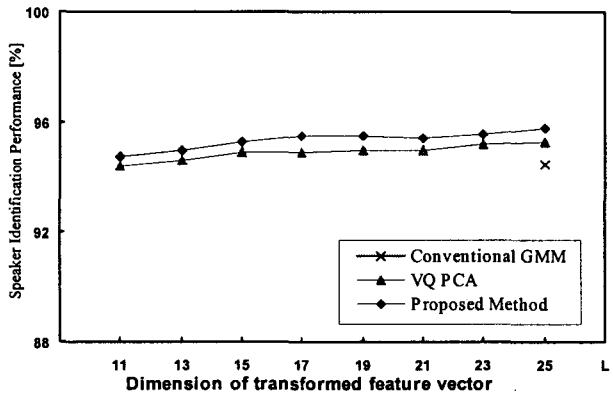

그림 2. 변환된 특징 벡터 차원과 화자 식별률과의 관계 (M=16, K=2)