화자 식별을 위한 GMM의 혼합 성분의 개수 추정

Estimation of Mixture Numbers of GMM for Speaker Identification

이 윤 정* · 이 기 용*
Youn-Jeong Lee · Ki-Yong Lee

ABSTRACT

In general, Gaussian mixture model(GMM) is used to estimate the speaker model for speaker identification. The parameter estimates of the GMM are obtained by using the expectation-maximization(EM) algorithm for the maximum likelihood(ML) estimation. However, if the number of mixtures isn't defined well in the GMM, those parameters are obtained inappropriately. The problem to find the number of components is significant to estimate the optimal para meter in mixture model. In this paper, to estimate the optimal number of mixtures, we propose the method that starts from the sufficient mixtures, after, the number is reduced by investigating the mutual information between mixtures for GMM. In result, we can estimate the optimal number of mixtures. The effectiveness of the proposed method is shown by the experiment using artificial data. Also, we performed the speaker identification applying the proposed method comparing with other approaches.

Keywords: Speaker Identification, GMM, Mutual Information, number of mixtures

1. 서 론

음성 데이터 집합에서 추출된 특징벡터를 이용하여, 화자인식을 하기 위해서는 각 화자의 특성을 나타낼 수 있는 가우시안 혼합 모델을 주로 사용한다(Reynolds and Rose. 1995). 이때, 가우시안 혼합 모델은 여러 개의 확률 밀도 함수로 구성된다. 주어진 특징벡터에서 미지의 데이터를 추정하기 위해서는 최대 유사도(Maximum Likelihood: ML)를 갖기 위하여 Expectation-Maximization(EM) 알고리즘이 사용된다. 일반적으로, EM 알고리즘은 혼합 모델 개수를 미리 알고 있다고 가정하여 왔다. 그러나, 고정된 혼합 성분들을 사용하여 파라미터를 추정할 경우에

* 숭실대학교 정보통신 전자공학부

로그 유사도 함수는 국부 최대값을 가진다고 알려져 있다 (Kehtarnavaz and Nakamura. 1998). 데이터가 적은 경우에는 로그 유사도 함수의 국부 최대값을 가지는 파라미터는 효용성이 있지만, 데이터가 많은 경우에는 문제가 발생하게 된다. 또한, 주어진 특징벡터에서 고정된 혼합 성분의 개수는 미지의 파라미터를 추정하는데 너무 많거나, 오히려 파라미터를 추정하기에 부족할 수가 있다. 따라서, 혼합 모델의 개수의 적절한 선택은 효율적으로 정확한 기본 확률밀도함수의 추정을 하기 위해 중요하다. 최적의 혼합 성분 개수를 구하기 위해 Schwarz 척도, 베이시안 정보 척도(BIC) 등이 사용되어 왔다. 이 방법들은 확률 함수의 성분의 수의 영향에 초점이 맞추어져 있다. 그러나, 혼합 성분의 개수가 증가하면, 확률 함수의 유사도 함수도 같이 증가하는 문제가 있다 (Kehtarnavaz and Nakamura. 1998, Fraley and Raftery. 1998, Mclachlan and Peel. 1998 and Paclik and Novovicova. 1998).

본 논문에서는 주어진 음성의 특징벡터의 분석을 위하여 초기에 충분히 많은 혼합 모델로부터, 두 개의 혼합 모델의 상호 정보량을 이용하여 상호 정보량이 많은 경우 혼합 모델 개수를 점차 감소 시키는 방법을 제안하였다. 주어진 특징벡터에서 양수의 상호 정보량을 갖는 혼합 모델을 제거하여 최적의 혼합 모델을 선택하여, 화자 식별의 성능을 높일 수 있었다. 혼합 모델 개수를 감소시키는 경우에는 두 개의 혼합 성분의 평균을 취하는 병합 방법과, 혼합 성분들 간의 상호 정보량이 가장 큰 성분을 제거하는 방법을 적용하였다(Yong and Zwolinski. 2001).

2. 가우시안 혼합 모델(GMM)

상태열 N 개의 특징벡터를 $X = \{x_1, ..., x_N\}$, $x_t \subseteq R^d$ 라 하자. GMM 유사도는

$$p(X|\lambda) = \prod_{t=1}^{N} p(x_t|\lambda) \tag{1}$$

로 구할 수 있다. $p(x_t|\lambda)$ 는 가우시안 혼합 성분 밀도이고, 혼합 성분(mixture)의 확률 밀도값으로 가중된 합이다(Reynolds and Rose. 1995).

$$p(x_t|\lambda) = \sum_{i=1}^{M} p_i b_i(x_t) \tag{2}$$

여기에서,

$$b_i(x_t)=\frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_i|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x_t-\mu_i)^T\Sigma_i^{-1}(x_t-\mu_i)\right\}$$

이고, μ_i는 평균벡터이고, Σ_i는 분산행렬이다. 혼합성분의 가중치는 $\sum_{i=1}^{M}p_i=1$를 만족한다. 화자모델을 위한 파라미터 λ는 각 가우시안 혼합 성분 밀도의 가중치, 평균벡터, 공분산 행렬로 구성된다.

$$\lambda=\{p_i,\mu_i,\Sigma_i\},\quad i=1,\cdots,M \tag{3}$$

GMM의 확률 값을 최대로 하기 위해 최대유사도 알고리즘을 사용하여 화자모델 파라미터 λ를 찾는다. ML을 위한 파라미터는 EM을 반복적으로 계산하여 구할 수 있다. 파라미터 추정은 반복적인 EM 알고리즘을 사용하여 얻을 수 있다. 다음의 재추정식은 GMM을 위한 화자 모델의 유사도를 단조 증가시킨다.

- 성분의 가중치(mixture weight)

$$\overline{p}_i=\frac{1}{N}\sum_{t=1}^{N}p(i\mid x_t,\lambda) \tag{4.a}$$

- 평균 벡터(mean vector)

$$\overline{\vec{\mu}}_i=\frac{\sum_{t=1}^{N}p(i\mid x_t,\lambda)\,x_t}{\sum_{t=1}^{N}p(i\mid x_t,\lambda)} \tag{4.b}$$

- 분산 행렬(variance matrix):

$$\overline{\Sigma}_i=\frac{\sum_{t=1}^{N}p(i\mid x_t,\lambda)\,x_t^2}{\sum_{t=1}^{N}p(i\mid x_t,\lambda)}-\vec{\mu}^2{}_i \tag{4.c}$$

i번째 혼합 성분의 사후확률(A posterior probability)은

$$p(i|x_t,\lambda)=\frac{p_ib_i(x_t)}{\sum_{i=1}^{M}p_ib_i(x_t)} \tag{5}$$

로 구할 수 있다.

3. 혼합 성분들간의 상호 정보량 측정

상호 정보 이론은 시스템에서 엔트로피를 최소화하는데 목적이 있고, 두 개의 혼합 성분들 사이에 공유된 정보량을 측정하기 위해 사용되고, 상호 정보는 초기 불확실성(uncertainty)과 조건 불확실성 사이의 차이로부터 얻어진다. 혼합 성분 공간을 θ 라 하고, $\mu_i \in \theta$, $\mu_j \in \theta$ 를 만족하는 두 개의 혼합 성분이 존재한다고 가정하자. 여기에서, p_i 는 μ_i 의 사전 확률 값으로 정의된다. μ_i 와 μ_j 의 상호 관계는

$$\varphi(\mu_i, \mu_j) = p_{i,j} \log \frac{p_{i,j}}{p_i p_j} \tag{8}$$

로 측정된다. 위의 식을 이용하여, 혼합 모델들의 공유된 상호 관계를 측정 할 수 있다.

여기에서, $p(i)$는 i-번째 혼합성분의 확률 값이고, $p(i,k)$는 혼합성분 i, j의 결합 확률(joint probability) 값이다. 만약 혼합성분 i와 혼합성분 j가 통계적으로 독립(independent)이면, $\varphi(\mu_i, \mu_j) = 0$ 이고, 통계적으로 의존적(dependent)이면 $\varphi(\mu_i, \mu_j) > 0$ 인 값을 갖는다. 만약 $\varphi(\mu_i, \mu_j) < 0$ 이면, 혼합성분 i와 j는 매우 적게 의존적임을 나타낸다(Yang, and Zwaliuski. 2001). 이로부터, 혼합 모델의 공유된 정보의 상호 정보량이 크다면, 추정된 확률 밀도함수에 중대한 영향을 주지 않고도 그 성분을 제거할 수 있다. 반면에, 두 개의 혼합 성분들 사이에 공유된 정보가 적으면, 두 개의 성분은 서로에게 독립적인 존재이므로, 즉, 혼합 모델의 상호 정보량이 적은 경우에 혼합 모델은 다른 혼합 모델에 대하여 통계적으로 독립이므로, 이 성분은 시스템 확률 밀도함수에 중요한 역할을 하므로 제거 될 수 없다. 따라서, 모든 혼합 성분들의 다른 혼합 성분에 대한 혼합 성분 상호 정보량 $\varphi(\mu_i, \mu_j) < 0$ 이 될 때까지 혼합 성분의 개수를 점차 감소시키면서 최적의 혼합 성분의 개수를 얻을 수 있다.

혼합성분의 개수를 감소시키는 경우에, 본 논문에서는 아래의 두 가지 방법을 각각 사용하였다.

(a) 제거 방법

두 혼합 성분의 상호 관계가 양의 값을 가질 경우, 전체 혼합 성분 중에서 가장 큰 상호 관계를 갖는 것이 $\varphi(\mu_\alpha, \mu_\beta)$라면, α 와 β 중에서 하나를 제거하고, 남은 혼합 성분들로 EM 알고리즘을 사용하여 평균 및 분산을 재추정한다.

(b) 병합방법

두 혼합 성분의 상호 관계 $\varphi(\mu_\alpha, \mu_\beta)$가 양의 값을 가질 경우, 두 혼합 성분의 가중치, 평균

과 분산을 다음과 같이 계산할 수 있다.

$$\vec{\mu}_{i*} = \frac{p_i \vec{\mu}_i + p_j \vec{\mu}_j}{p_{i*}} \tag{11.a}$$

$$\Sigma_{i*} = \frac{p_i \Sigma_i + p_j \Sigma_j}{p_{i*}} \tag{11.b}$$

여기에서, $p_{i*} = p_i + p_j$ 이다.

두 혼합 성분의 상호 관계가 양의 값을 가질 경우, 혼합 성분의 개수를 감소시킬 때 위의 두 가지 방법을 적용하여 감소된 혼합 성분의 평균과 분산을 얻을 수 있다.

4. 알고리즘

상호 정보량을 이용하여 최적의 혼합 성분 개수를 찾는 방법은 다음 순서로 구할 수 있다.

① 충분히 큰 혼합 성분의 개수를 설정한 다음, 주어진 데이터의 확률 밀도 함수를 계산하기 위하여, 가우시안 혼합 모델의 초기 가중치, 평균벡터, 공분산 행렬을 계산한다.

② 가우시안 혼합 모델이 최대 유사도비를 갖도록 파라미터들이 수렴될 때까지 EM 알고리즘을 이용하여 반복 수행한다. 이 과정에서 상호정보량을 계산하기 위한 $p(\mu_i)$, $p(\mu_j)$ 값이 얻어진다.

③ 시스템의 상호 정보량을 측정하고 정보는 가장 큰 양수 값을 갖는 혼합 성분을 제거하여 혼합 성분의 개수를 하나 감소시킨다.

④ $\varphi(\mu_i, \mu_j) < 0$ 이면, 혼합 성분의 개수를 감소시키는 것을 정지하고, 최적의 혼합 성분 개수가 결정한다.

⑤ 만약 적어도 하나의 혼합 성분의 상호관계가 $\varphi(\mu_i, \mu_j) > 0$ 이면, 가장 큰 양수의 상호 정보량을 갖는 혼합 성분을 제거한다. 유사도값이 수렴할 때까지 ②~⑤의 과정을 반복한다.

5. 실험 및 결과

본 논문에서 제안한 방법을 검증하기 위하여 두 가지 데이터를 사용하였다. 하나는 혼합성분이 4 개인 2 차원 가우시안 분포를 가진 2000 샘플을 생성시킨 데이터이다.

$$0.25N\left[x\left|\begin{pmatrix}2.2\\1.9\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right.\right]+0.25N\left[x\left|\begin{pmatrix}1.4\\0.9\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right.\right]$$
$$+0.25N\left[x\left|\begin{pmatrix}2.4\\1.2\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right.\right]+0.25N\left[x\left|\begin{pmatrix}1.3\\2\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right.\right]$$

그림 1은 생성한 데이터에서 초기 혼합 성분 개수 10 개로 시작하여 상호정보량을 측정하여 혼합 성분 개수를 감소시켜 최적의 개수를 구한 결과이다. 여기에서, 혼합성분을 제거한 방법과 병합 시킨 방법의 평균은 거의 원래의 평균과 거의 동일한 값임을 확인할 수 있었다. 혼합성분들의 상호 정보량은 표 1에 나타내었다.

다음으로, 200 명(남자 100 명, 여자 100 명)의 화자가 발성한 한국어 문장 종속 연속음 "열려라 참깨" 음성 데이터이다. 수집된 음성은 한 화자당 1 회에 5 번씩 발성한 뒤, 1 주 간격의 시간차를 가지고 3 주에 걸쳐서 수집하였다. 개인별 전체 발성된 데이터 수는 15 개이고, 각 화자의 10 개 파일을 학습과정에, 나머지 1,000 개의 참여 데이터를 성능을 위하여 사용하였다. 16 kHz로 샘플링하였고, 음성 분석을 위하여 해밍창이 사용되었고, 한 프레임은 50% 중첩된 256 샘플을 사용하였다. 특징벡터로는 12 차 MFCC 켑스트럼과 12 차 델타 켑스트럼과 델타 에너지를 포함하여 전체 25 차를 사용하였다. 그림 2는 GMM에서 혼합성분의 개수에 따른 화자 식별률과 제안된 방법에서의 화자 식별률을 나타낸 것이다. 실험에서 GMM 방법은 일정한 수준의 성능에 접근하면 혼합 성분의 개수가 증가해도 더 이상 성능이 향상되지 않는데, 이는 학습 데이터가 충분하지 못했기 때문에 나타난 현상이라 할 수 있다. GMM 방법에서, 최적의 혼합 성분개수는 15 개일 때 가장 높은 화자 식별률이 얻어졌다. 제안된 방법에서는 그림 2의 경우 혼합 성분 개수는 제거한 경우 평균 17 개, 병합 시킨 경우 평균 19 개로 최적화되었다. 실제 데이터에서는 제거 방법과 병합 방법에서의 개수 차이가 발생하였는데, 여기에서 개수를 감소시킬 때, 혼합성분들을 병합 한 방법이 더 좋은 결과를 보였다. 이 경우, GMM에서 얻어진 최적의 화자 식별률과 비슷한 성능을 얻을 수 있었다.

6. 결 론

본 논문에서는 화자식별을 위한 가우시안 혼합 모델의 혼합 성분의 개수를 결정하기 위한 방법을 제안하였다. 이는 충분히 큰 혼합 성분개수로부터 각 혼합 성분들간의 상호 관계가 0이 나올 때까지, 즉 혼합성분들간의 관계가 모두 독립적으로 측정될 때까지, 가장 큰 양수의 값을 갖는 혼합 성분을 제거하는 과정을 반복 수행하는 방법이다. 최적의 혼합 성분을 구한 결과, GMM에서 얻어진 최적의 화자 식별률과 비슷한 성능을 보임을 알 수 있다.

감사의 글

본 논문은 2004학년도 숭실대학교 교내학술연구비 지원에 의하여 수행되었습니다.

참 고 문 헌

[1] Fraley, C. & Raftery, A. E. 1998. "How many clusters? Which clustering method? Answers via model-based cluster analysis." *The Computer Jounal,* vol.41, no.8, pp. 578~588.

[2] Kehtarnavaz, N. & Nakamura, E. 1998. "Generalization of the EM algorithm for mixture density estimation." *Pattern Recognition Letters(19),* pp. 133~140.

[3] Mclachlan, G. J. & Basford, K. E. 1998. "Mixture models inference and applications to clustering." *Marcel Dekker.*

[4] Reynolds, D. & Rose, R. 1995. "Robust text-independent speaker identification using Gaussian mixture speaker models." *IEEE Trans. on SAP,* vol. 3, no. 1, pp. 72~82.

[5] Yong, Z. R. & Zwolinski, M. 2001. "Mutual information theory for adaptive mixture models." *IEEE Trans. on PAMI,* vol.23, no.4.

[6] Paclik, P. & Novovicova, J. 1998. "Number of components and initialization in Gaussian Mixture Model for pattern recognition." *In Proceedings of the 14th ICPR,* pp 886~890, Australia.

[7] Mclachlan, G., Peel, D. 2000. "*Finite Mixture Models.*" New York: John Wiley & Sons.

접수일자: 2004. 4. 30
게재결정: 2004. 6. 15

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

▲ 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

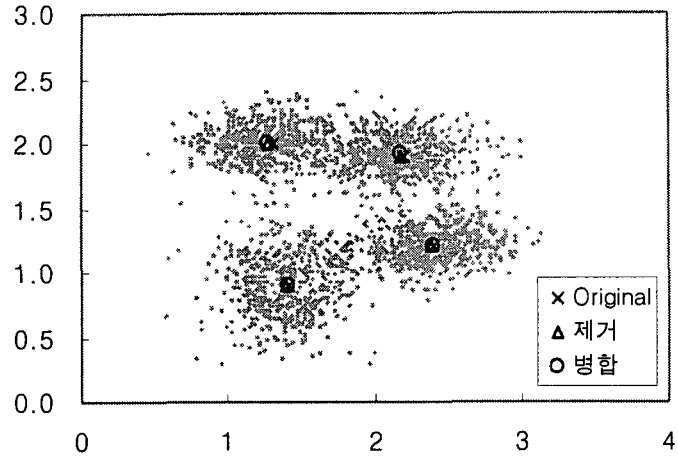

그림 1. 생성 데이터에서의 추정된 평균

표 1. 혼합 성분의 상호 정보량 관계

i	$\mu(i)$	$\varphi(i,j$	j
1	(1.363,0.909)	0.338	9
2	(1.562,1.177)	0.207	4
3	(2.151,1.923)	-0.010	
4	(1.668,1.075)	0.148	2
5	(1.259,2.007)	-0.042	
6	(1.628,0.954)	0.148	2
7	(2.147,1.586)	0.181	10
8	(2.426,1.190)	0.024	10
9	(1.416,0.870)	0.342	1
10	(2.259,1.485)	0.174	7

(a) 제거방법 이용한 결과

i	$\mu(i)$	$\varphi(i,j)$
1	(1.396,0.900)	-0.0370
2	(2.392,1.201)	-0.0376
3	(1.279,2.003)	-0.0397
4	(2.179,1.905)	-0.0496

(b) 병합방법 이용한 결과

i	$\mu(i)$	$\varphi(i,j)$
1	(1.400,0.901)	−0.037
2	(2.174,1.909)	−0.050
3	(1.276,2.004)	−0.039
4	(2.395,1.206)	−0.038

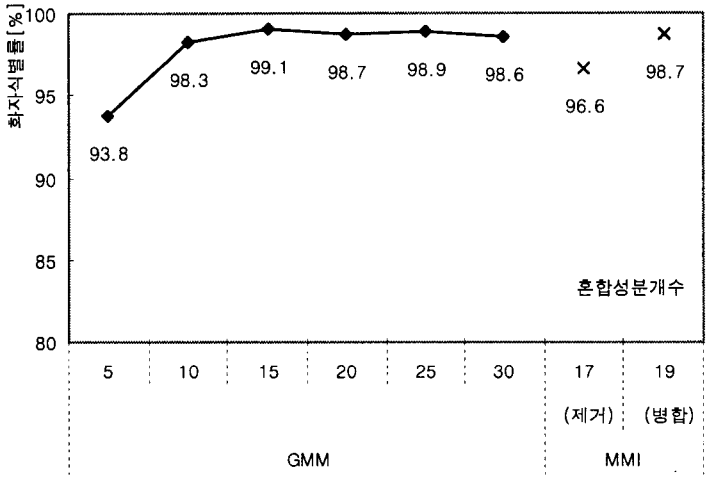

그림 2. GMM과 제안된 방법(MMI)에서 혼합성분 개수에 따른 화자 식별률[%]