한국음향학회지 제24권 제3호 pp. 141-149 (2005)

GMM을 위한 점진적 k-means 알고리즘에 의해 초기값을 갖는 EM알고리즘과 화자식별에의 적용

EM Algorithm with Initialization Based on Incremental k-means for GMM and Its Application to Speaker Identification

이 윤 정*, 서 창 우**, 한 헌 수*, 이 기 용*
(Younjeong Lee*, Changwoo Seo**, Hernsoo Hahn*, Kiyong Lee*)

*숭실대학교 정보통신공학과, **(주)인스모바일 기술연구소

(접수일자: 2004년 7월 14일; 수정일자: 2005년 1월 24일; 채택일자: 2005년 3월 7일)

개개인의 음성을 이용한 화자식별에서, 화자 모델을 추정하는데 가우시안혼합모델이 주로 사용된다. 최대 우도 추정을 갖는 가우시안 혼합모델의 파라미터 추정은 Expectation-Maximization (EM)을 사용하여 얻을 수 있다. 그러나, EM 알고리즘은 초기값에 상당히 민감하고, 혼합성분의 개수를 미리 알고 있어야 하는 단점이 있다. 본 논문에서는, EM 알고리즘의 문제점을 해결하기 위하여 가우시안 혼합모델을 위한 점진적 k-means 알고리즘에 의한 초기값을 갖는 EM 알고리즘을 제안한다. 제안된 방법은 혼합성분의 개수를 점진적 k-means방법을 이용하여 한번에 하나씩 혼합성분을 추정하여 최적의 혼합성분이 얻어 질 때까지 이를 반복 수행한다. 하나의 혼합성분이 추가될 때 마다, 새로 얻어진 혼합성분과 이전에 구한 혼합성분들간의상호 관계를 각각 측정한다. 이로부터, 통계적으로 독립인 최적의 혼합성분 개수를 추정할 수 있다. 제안된 방법의 성능을 확인하기 위하여 임의의 생성 데이터와 실제 음성을 사용하였다. 실험 결과에서, 제안된 방법이 기존의 방법보다 화자 식별 성능이 우수 하였으며, 또한 성능을 유지하면서도 계산량 감소의 효과까지 볼 수 있었다.

핵심용어: 가우시안 혼합모델, EM 알고리즘, k-means, 상호관계, 화자 식별

투고분야: 음성처리 분야 (2.5)

In general, Gaussian mixture model (GMM) is used to estimate the speaker model from the speech 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, the EM algorithm has such drawbacks that it depends heavily on the initialization and it needs the number of mixtures to be known. In this paper, to solve the above problems of the EM algorithm, we propose an EM algorithm with the initialization based on incremental k-means for GMM. The proposed method dynamically increases the number of mixtures one by one until finding the optimum number of mixtures. Whenever adding one mixture, we calculate the mutual relationship between it and one of other mixtures respectively. Finally, based on these mutual relationships, we can estimate the optimal number of mixtures which are statistically independent. The effectiveness of the proposed method is shown by the experiment for artificial data. Also, we performed the speaker identification by applying the proposed method comparing with other approaches.

Keywords: Gaussian Mixture Model, EM Algorithm, k-means, Mutual Relationship, Speaker Identification
ASK subject classification: Speech Signal Processing (2.5)

I. 서 론

화자인식은 미리 등록한 사용자의 음성과 서비스에 접속을 요구하는 음성을 비교하여, 두 신호성분의 패턴의 우도 (likelihood)를 측정함으로써 올바른 사용자인지를 판별하는 것이다. 이와 같은 패턴 인식, 음성과 영상 신호 분석, 통계적 분석과 같은 다양한 분야에서 주어진 관측 데이터부터 미지의 파라미터를 추정하기 위하여 혼합모델 (mixture model)이 많이 사용되고 있으며, 음성

책임저자: 이 윤 정 (youn@ssu.ac.kr)
156-743 서울시 동작구 상도동
숭실대학교 정보통신학과 전시관 402호
(전화: 02-817-4591; 팩스: 02-817-4591)

파형에 내재된 화자의 고유 정보를 사용하여 화자를 자동으로 인식하는 화자식별을 위하여 GMM[1,2]방법을 사용하고 있다. 각 화자의 신원을 모델링 하기 위하여 화자식별에서는 주로 GMM을 사용하고[3], 화자 모델을 위한 GMM 파라미터가 최대 우도 값으로 수렴되도록 하기 위하여 EM 알고리즘을 반복 사용한다.

그러나, EM 알고리즘은 파라미터의 초기값에 매우 민감하므로, 적절하지 못한 초기값을 가지게 되면 국부해(local minima)로 빠지는 문제점이 있다[2,4]. 이러한 초기화 문제를 해결하기 위하여, 하나의 랜덤 값을 선택하거나 여러 개의 랜덤 값의 조합을 초기값으로 설정하는 방법, 가장 높은 우도를 갖는 마지막 추정을 선택하는 방법[5], 그리고, 클러스터링 알고리즘에 의한 초기화[6]등과 같은 여러 가지 방법들이 제안되어 왔다.

또한, EM 알고리즘에서는 혼합성분의 수를 미리 알고 있다고 가정하였으므로, 실제 데이터에 이를 적용하기에는 문제가 있다. 화자 인식의 성능을 높이기 위해서는 많은 혼합성분이 필요하다. 그러나 많은 혼합성분은 많은 음성데이터를 필요로 하고, 실시간 구현이 어려울 뿐 아니라, 데이터를 과잉조정 (overfit)할 수 있다. 반면에 너무 적은 혼합성분을 사용한다면, 모델 추정을 위한 데이터가 부족하게 된다. 따라서, 최적의 혼합모델은 전역해 (global value)를 찾기 위한 효율적이고 정확한 추정을 위하여 상당히 중요한 변수이다. 이러한 문제를 해결하기 위하여, 여러 가지 방법들이 제안되어 왔다. 예를 들어 정보이론에 의한 접근 방법인 AIC (Akaike' s information criterion)[7], 베이시안 척도를 이용한 Schwarz의 BIC (Bayesian inference criterion)[8], 완전한 (complete) 우도를 이용한 방법인 ICL (integrated classification likelihood)[9], 그리고, 상호 정보량을 이용하여 혼합성분을 점차 감소시켜가는 MI (Mutual Information)[10,11] 방법들이 있다. 이들 방법에서 각각의 기준 (criterion)에 따라 최적 혼합성분의 수가 결정 된다. 전형적인 방법에서는 최적의 혼합성분의 수는 적절한 M_{min}과 M_{max}사이에 존재하는 모든 후보 값들에 대하여 테스트 함으로써, 기준에 해당하는 M 값을 선택하게 된다. 그 결과 계산량이 많이 소요되고, 또한 각각의 방법들의 로그-우도 값이 파라미터의 수에 의하여 영향을 받으므로, 정확한 혼합성분의 수를 추정하기 어렵다는 문제점이 있다.

이러한 문제를 해결하기 위하여, 본 논문에서는 점진적 k-means 방법을 기반으로 EM 알고리즘의 초기값을 설정하고, 상호 관계를 계산하여 서로 독립인 최적의 혼합성분 개수를 결정하는 방법을 제안한다. 먼저, 한 단계에 하나의 혼합성분을 추가시키는 전역 (global) 검색 절차를 사용하여[12] 평균을 추정함으로써 EM알고리즘을 초기화 시킨다. 다음으로, 로그-우도 값이 전역해 (global value)로 수렴할 때까지 EM 알고리즘을 반복 수행한다. 우도가 수렴되면, 새롭게 추가된 혼합성분과 이전의 다른 혼합성분들과의 상호 관계를 측정하여 이들의 관계가 통계적으로 독립인지 종속인지를 판단한다. 하나 이상의 양의 상호관계를 갖는 혼합성분이 나타나면, 이들은 서로에게 종속적이라는 의미이므로, 바로 이전의 서로 독립적인 관계였던 혼합성분들을 GMM 모델을 위한 파라미터로 결정한다. 제안된 방법은 모든 후보 값 M에 대해 반복 실험하지 않고, 최적의 개수가 결정되면 바로 알고리즘의 수행을 멈추므로, 기존의 방법들과 비교할 때, 계산량이 크게 감소한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 최대 우도 학습을 위한 EM 알고리즘의 초기화를 갖는 혼합성분 모델에 대하여 설명한다. 다음 3장에서는 혼합성분들간의 상호 관계를 측정하여 최적 혼합성분 개수를 추정하는 방법에 대하여 제시하고, 4장에서 제안된 방법의 알고리즘을 자세히 설명한다. 실험 결과는 5장에 나타내었고, 마지막 6장에서 결론을 내린다.

II. GMM을 위한 EM 알고리즘의 초기화와 점진적인 k-means 알고리즘

T 개의관측열 $X=\{x_1,\Lambda,x_T\}, x_t\in R^d$로부터, GMM은 M 개의 혼합성분 밀도의 가중화된 합으로 정의된다.

$$p(x_t|\theta)=\sum_{i=1}^{M} w_i b_i(x_t) \tag{1}$$

위의 식에서 $b_i(x_t)$는 혼합성분의 밀도 함수로, d-차원을 갖는 가우시안 함수로 표현된다.

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

여기에서, μ_i, Σ_i는 각각 i-번째 혼합성분의 평균 벡터, 공분산 행렬이다. 그리고, 가중치 w_i는 $\sum_{i}^{M} w_i = 1$라는 강제 조건을 만족한다. 결과적으로, GMM은 $\theta = \{w_i, \mu_i, \Sigma_i\}_{i=1}^{M}$로 나타낼 수 있다[3]. 벡터 X의 T개 관측 열로부터 파라미터를 추정하기 위해서, GMM우도는 다음 식으로 정의 할 수 있다.

$$P(X|\theta) = \prod_{i=1}^{T} p(x_i|\theta) \tag{3}$$

GMM의 우도를 최대화 시키는 파라미터 θ를 추정하기 위하여 주로 ML 추정이사용된다. 그러나, GMM의 우도를 구하는 식은 비선형 함수이므로 우도를 직접적으로 최대화 시키는 것은 불가능하다. 따라서, EM 알고리즘을 반복적으로 사용하여 ML 추정을 하게 된다[3, 4]. 그러나, EM 알고리즘은 초기값에 상당히 민감하므로, 우리는 점진적 k-means 알고리즘에 의하여 초기값을 설정하여 EM 알고리즘을 사용하고자 한다. 제안된 EM 알고리즘은 우도가 수렴될 때까지 E, M 단계를 번갈아 가면서 새로운 모델을 추정하게 된다. 먼저, 임의의 Q-함수를 다음 식으로 가정한다.

$$Q(\theta, \hat{\theta}) = \sum_{i=1}^{T}\sum_{i=1}^{M} p(i|x_t, \hat{\theta})\{\log w_i + \log b_i(x_t)\} \tag{4}$$

점진적 k-means알고리즘에 의한 혼합성분의 초기화 (Initialization of mixtures): 주어진 데이터로부터, 점진적인 k-means 알고리즘을 사용하여, 차례대로 $\mu_k\,(k = 1, \Lambda\,, M)$을 추정하고, 이를 EM 알고리즘의 초기값으로 설정한다. μ_k로부터 EM 알고리즘이 초기화 된다면, 랜덤하게 구한 초기값으로 수렴 시킬 때 보다 더 정확하고 빠르게 국부해(local minima)로 빠지지 않고, 전역해로 구할 수 있다. 점진적 k-means 알고리즘이란 각각의 모든 데이터들을 하나의 혼합 성분의 평균으로 간주하고, 다른 데이터 값들과의 유클리안 거리의 합을 계산하여 혼합 성분의 개수를 한 단계에 하나씩 증가시켜가는 방법이다. 혼합성분의 평균 μ_k를 구하기 위하여, 각각의 모든 관측 벡터 X_t가 이를 위한 후보로 사용 되고, 모든 데이터 후보들 중에서 다른 각각의 데이터들과 유클리안 거리의 제곱(squared Euclidean distances)의 합의 함수를 기준으로 하여 선택하게 된다.

$$E_n^1 = \sum_{t=1}^{T} \exp\left\{-\|x_t - x_n\|^2\right\} \quad , \quad 1 \le n \le T \tag{5}$$

위의 식은 첫 번째 혼합성분의 평균값을 추정하기 위한 함수로, n번째의 관측 벡터 X_n을 기준으로 주위에 얼마나 많은 데이터들이 분포 되어 있는지를 측정하는 밀도 (density) 함수이다. 기존의 점진적인 k-means방법은 유클리안 거리의 제곱함수를 사용하기 때문에, 첫 번째 혼합 성분의 평균 값을 측정할 경우, 가장 작은 거리 값을 갖는 데이터를 평균으로 추정하게 되어 잘못된 초기값을 얻게 된다. 즉, 데이터가 많이 분산되어 있는 경우에 최소 거리 값을 갖는 데이터를 찾게 되어 첫 번째 혼합 성분의 평균으로 적당하지 않다. 그러나 밀도 함수를 사용하게 되면, 데이터들 중에서 가장 높은 밀도 값을 가지는 데이터의 위치를 혼합 성분의 평균으로 간주하게 되므로, 초기 혼합 성분의 위치를 정확하게 추정할 수 있다. 따라서, 가장 높은 밀도를 갖는 n* 번째 관측 데이터 값을 선택하여 첫번째 혼합성분의 평균 μ_1으로 결정한다.

$$\mu_1 = x_{n^*}, \text{ where } n^* = \underset{1 \le n \le T}{\arg\max}\, E_n^1 \tag{6}$$

(k-1)-번째 혼합성분을 구한 다음, k-번째 기준을 결정하기 위해서, 식 (5)의 수식을 이용하여 각 관측 데이터의 기준을 다음과 같이 수정한다.

$$E_n^k = \sum_{t=1}^{T}\left\{\sum_{m=1}^{k-1} I(x_t \in C_m)\|x_n - \mu_m\|^2 + I(x_t \in C_n)\|x_t - x_n\|^2\right\},$$
$$2 \le k \le M, 1 \le n \le T \tag{7}$$

위의 수식에서 C_m은 평균 μ_m을 갖는 m번째 혼합성분을 의미하고, C_n은 평균이 X_n인 혼합성분으로 정의한다. 이때, X가 참이면, $I(X) = 1$이고, 그렇지 않으면, $I(X) = 0$이다[12]. 예를 들어, 식(7)에서 $I(x_t \in C_n) = 1$이면, nearest-neighborhood (NN)에 의하여 (k-1)개 혼합성분의 평균 μ_m들과 X_n중에서 X_t와 가장 가까운 거리에 있는 혼합성분의 평균은 X_n임을 의미한다. 식 (7)을 사용하여 거리 척도 E_n^k를 계산하고, 최소 기준을 만족하는 k-번째 혼합성분을 다음 식으로부터 구한다.

$$\mu_k = x_{n^*}, \text{ where } n^* = \underset{1 \le n \le T}{argmin} E_n^k \tag{8}$$

기존의 전형적인 k-means 알고리즘에 의해 결정된 초기값은 이상치에 의해 평균이 좌우되어, 데이터가 존재 하지 않는 곳에서 평균의 위치가 결정될 수 있다. 그러나 이렇게 얻어진 혼합성분의 평균값들은 모두 실제 관측 데이터로부터 얻어지는 값들이다. 따라서, 각 데이터들을 혼합성분의 후보로 사용하는 점진적인 k-means 알고리즘의 경우에는 실제 데이터가 존재 하는 곳에서 혼합성분의 평균들이 결정되므로, 이상치가 데이터에 존재 하더라도 좀 더 강인하게 된다.

E-단계(E-Step): 로그-우도의 조건부 기대치 (conditional expectation)를 계산하기 위해서 관측열 X에 대하여 현재 $\hat{\theta}$를 추정한다. EM 알고리즘의 E-단계는 혼합성분이 어떻게 할당되었는지를 평가하기 위해 사후 확률 (a posteriori probability)을 사용한다. 베이시안 이론 (Bayes' theorem)에 따라, i-번째 혼합성분의 x_t에 대한 사후 확률 $\zeta_{it} = p(i \mid x_t, \hat{\theta})$는 다음과 같이 정의 할 수 있다.

$$\zeta_{it} = p(i \mid x_t, \hat{\theta}) = \frac{p_i b_i(x_t)}{\sum_{h=1}^{M} p_h b_h(x_t)} \tag{9}$$

x_{n^*}에 의하여 혼합성분의 평균 μ_k가 EM알고리즘의 초기값으로 결정이 되면, 다음의 M 단계에서 파라미터들이 재 추정된다.

M-단계 (M-Step): E-단계 이후에, M-단계는 임의의 $Q(\theta, \hat{\theta})$ 함수를 최대화 시키는 단계이다. 따라서, 로그-우도를 단조 증가 시키는 파라미터들, 즉, 혼합성분의 가중치 (Mixture Weight), 공분산 (Variances)을 다음 수식으로 재계산 한다.

$$\text{Mixture Weight: } p_i = \frac{1}{T}\sum_{t=1}^{T}\zeta_{it} \tag{10}$$

$$\text{Variances: } \Sigma_i = \frac{\sum_{t=1}^{T}\zeta_{it}x_t^2}{\sum_{t=1}^{T}\zeta_{it}} - \mu_i^2 \tag{11}$$

점진적 k-means 알고리즘에 의하여 초기화된 EM 알고리즘을 반복 사용하여, 관측 데이터의 GMM을위한 파라미터들을 구할 수 있다.

III. 상호 관계에 의한 최적 혼합성분의 개수 추정

3장에서는, 점진적 k-means 알고리즘에 의해 혼합성분이 추가되는 과정에서 두 개의 혼합성분들 사이의 상호 관계를 측정하고 최적의 혼합성분의 수를 결정 하는 방법에 대하여 살펴 본다. 너무 많은 혼합성분의 수를 사용할 경우에는 데이터를 분석하는데 과잉맞춤 (overfit)할 수 있고, 반면에 너무 적은 혼합성분의 수를 사용한 경우에는 데이터 분석에 필요한 개수보다 적어 부적당한 국부해 (local minima)가 얻어지게 된다. 그러므로, 잘 조정된 혼합모델의 개수 추정에 관한 문제는 EM 알고리즘에서는 상당히 중요한 부분이다. 즉, 기존의 EM 알고리즘에서는 미리 지정된 고정의 혼합성분 개수를 사용하였지만, 실제 관측 데이터는 몇 개의 혼합성분 개수가 적당한지 알 수 없다. 따라서, EM 알고리즘을 좀더 효과적으로 정확하게 계산하기 위하여, 혼합성분이 새로 추가될 때 마다, 새로 추가된 혼합성분과 이전에 존재하던 (k-1)개의 혼합성분들을 각각 상호 관계 (mutual relationship)를 측정한다. 이렇게 구해진 상호 관계는 혼합성분들이 서로 통계적으로 독립 (statistically independent)인지 종속적 (statistically dependent)인지를 결정하는데 사용된다[10,11].

한번에 하나의 혼합성분이 추가될 때, 새로 추가된 k-번째혼합성분이 이전에 구한 (k-1)개의 혼합성분들과 각각의 상호 관계가 여전히 독립적이라면, (k+1)번째 혼합성분을 다시 추정한다. 반대로 상호관계가 통계적으로 종속적이면(즉, 적어도 하나 이상의 상호 관계 값이 양수라면), 이전에 추정하였던, (k-1)개의 혼합모델일 때 최적의 독립적인 혼합모델이 되므로, 혼합성분의 개수를 (k-1)개로 결정한다.

i-번째와 k -번째 혼합성분의 상호 관계는 다음과 같이 계산할 수 있다.

$$\varphi(i,k) = p(i,k)\log\frac{p(i,k)}{p(i)p(k)}, \quad i = 1, \cdots, k-1 \tag{12}$$

여기에서, $p(i)$는 i-번째 혼합성분의 확률 (probability)이고, $p(i,k)$는 i-번째와 k-번째 혼합성분의 결합 확률 (joint probability)이다.

$$p(i)=\frac{1}{T}\sum_{t=1}^{T}\zeta_{it} \tag{13}$$

$$p(i,k)=\frac{1}{T}\sum_{t=1}^{T}\zeta_{it}\zeta_{kt} \tag{14}$$

여기에서 ζ_{it}는 EM 알고리즘의 E단계에 얻은 사후 확률 값이다.

혼합성분의 상호 관계는 음수 (-), 영 (0), 양수 (+)의 세 가지 값으로 구분 된다. 여기에서, $\varphi(i,k)$값이 영 (0)이면, i-번째와 k-번째 혼합성분이 통계적으로 독립임을 의미한다: $p(i,k)=p(i)p(k)$. 만약 $\varphi(i,k)$가 양수 (>0) 이면, i-번째와 k-번째 혼합성분은 통계적으로 서로 종속적임을 나타낸다. 또한, 음수(<0)인 상호관계가 된다면, i-번째와 k-번째 혼합성분은 거의 관련성이 없으므로, 독립적이라고 볼 수 있다. 따라서, $\varphi(i,k)$가 양수 값으로 나타나면, i-번째와 k-번째 혼합성분 중 하나는 추정된 모델에 손실 없이 제거할 수 있다는 의미가 된다.

따라서, 최적의 혼합성분의 개수를 추정하기 위하여, 새로운 혼합성분이 추가될 때 마다 상호 관계를 측정하고, 혼합성분들의 상호 관계가 종속관계로 확인 될 때까지 반복 수행한다. 제안한 방법은 다음 장에 자세히 설명한다.

IV. 알고리즘

주어진 관측데이터로부터 GMM을 추정하기 위하여 최적의 혼합성분 개수를 갖는 점진적 k-means 알고리즘에 의해 초기화된 향상된 EM 알고리즘을 사용한다. 혼합성분이 점진적 k-means 알고리즘에 의해 새로 추가되면, 추가된 혼합성분의 평균을 이용하여 EM 알고리즘의 초기값으로 설정하고, EM 알고리즘을 반복 수행하여 최대 우도 값을 갖는 혼합성분의 가중치와 공분산을 추정하고, 우도가 일정한 값으로 수렴되도록 한다. 우도 수렴 후에, 추정된 혼합성분의 상호 관계를 구하여, 서로가 통계적으로 독립인지 종속인지를 판별한다. 여기에서 적어도 하나 이상의 양의 상호 관계가 측정된다면, 서로 종속적인 혼합성분이 존재함을 의미하므로 새로운 k-번째 혼합성분을 추가하는 과정을 중지하고, 독립적인 관계를 유지했던 이전 단계의 $(M=k-1)$개의 혼합성분들로 GMM을 위한 파라미터 모델이 결정된다.

① $n=1,\cdots,T$, k=1일 때, 각 데이터 n을 중심으로 분포된 밀도 함수 E_n^1를 계산하고, 첫 번째 혼합성분의 평균 μ_1를 결정한다.

② 또 다른 하나의 혼합성분을 추가하기 위하여 점진적인 k-means 알고리즘에 기반한 최소 거리를 갖는 E_n^k를 계산한다. 따라서, 혼합성분의 개수가 하나 추가되어 (k=k+1)이 된다.

③ k-개의 혼합성분에 대한 초기값이 주어지면, 주어진 혼합성분들의 모델 파라미터를 재추정하기 위하여 식 (9), (10), (11)을 사용하여 우도가 일정한 값으로 수렴할 때까지 EM 알고리즘을 번갈아 가면서 수행한다. 이 과정에서, 혼합성분들의 상호 관계를 측정을 위한 $p(i), p(k), p(i,k)$ 값들도 구한다.

④ $i=1,\cdots,k$ 인 경우에, 식 (12)를 사용하여 k-번째 추가된 혼합성분과 이전의 혼합성분들간의 상호 관계 $\varphi(i,k)$를 측정한다. 만약 $\varphi(i,k)\le 0$이면 ②단계로 되돌아간다. 그러나, 만약 $\varphi(i,k)>0$이면, k-번째에 새로

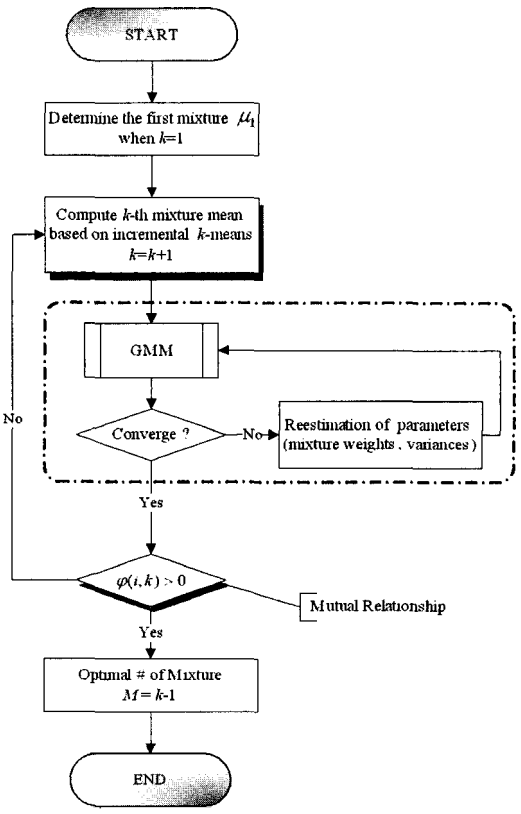


그림 1. 제안된 알고리즘
Fig. 1. The proposed Algorithm.

추가된 혼합성분이 이전의 i-번째 혼합성분과 종속적인 관계임을 나타내므로, 혼합성분의 개수 추가 과정을 멈추고 최적의 혼합성분의 개수를 $M(=k-1)$으로 설정한다.

즉, 위의 과정을 수행하면서, 각 혼합성분을 추가시키는 단계마다 상호 관계를 측정하여 양의 값이 나온다면, 바로 이전 단계에서의 결과가 혼합성분간의 관계가 최적의 독립적인 관계임을 나타내므로, M(=k-1)으로 최적의 모델링을 하게 된다. 제안된 방법은 다음의 그림 1과 같이 나타낼 수 있다.

V. 실험 결과

본 논문에서는 임의의 생성된 데이터(artificial data)와 실제 음성 데이터를 사용하여 제안된 방법의 성능을 확인하였다.

먼저, Cheung이 사용한 2-차원을 갖는 3개의 혼합성분의 가우시안 분포를 갖는 2000개의 데이터 샘플을 생성하였다[13].

[Case 1] :

$$0.3N\left[x,\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}0.15 & 0.05\\0.05 & 0.25\end{pmatrix}\right]+0.4N\left[x,\begin{pmatrix}1\\2.5\end{pmatrix},\begin{pmatrix}0.15 & 0\\0 & 0.15\end{pmatrix}\right]+0.3N\left[x,\begin{pmatrix}2.5\\2.5\end{pmatrix},\begin{pmatrix}0.15 & -0.1\\-0.1 & 0.15\end{pmatrix}\right] \quad (15)$$

여기에서, $aN[x|\mu,\Sigma]$는 a는 혼합성분의 가중치이고, μ는 평균, Σ는 분산을 나타낸다. 생성된 전체 데이터에 대한 검색을 통하여 첫 번째 혼합성분을 추정하고, 최적의 혼합성분이 얻어 질 때까지 새로운 혼합성분을 증가 시켰다. 이 데이터에서는 $k=4$일 때 상호 관계가 영보다 크게 나타났다. 따라서, $k=4$일 때 혼합성분들이 종속적 관계로 판정되었다. 표1은 [Case 1]의 경우에 혼합성분이 차례로 추가되는 과정을 나타낸 것이다. 세 번째 단계까지는 서로의 상호 관계가 모두 음수로 독립적인 관계를 유지하고 있지만, 네 번째 혼합성분이 얻어졌을 때, 이전 혼합성분들과의 상호 관계 값에서, 첫 번째 혼합성분과의 상호 관계가 0.1255로 나타났다. 즉, 네 번째 얻어진 혼합성분과 첫 번째 얻어진 혼합성분들이 종속적 관계로 판단되므로, 이 데이터의 분포를 표현

표 1. 혼합 성분들의 상호 관계 [Case 1]
Table 1. The mutual relationship between mixtures [Case 1].

# of mixture(k)	Mean of the i-th Mixture	i	$\varphi(i,k)$
1	(1.0788,2.4606)	1	~
2	(2.5416,2.4477)	1	-0.0611
3	(0.9467,0.9378)	1	-0.0603
		2	-0.0017
4	(0.9013,2.8942)	1	0.0266
		2	-0.0170
		3	-0.0060

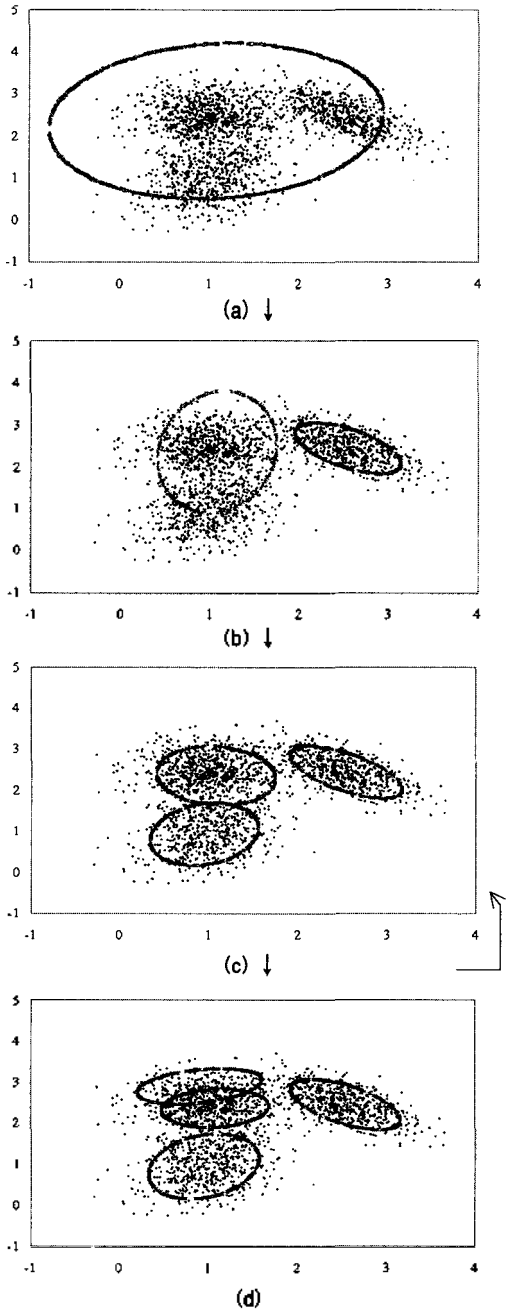
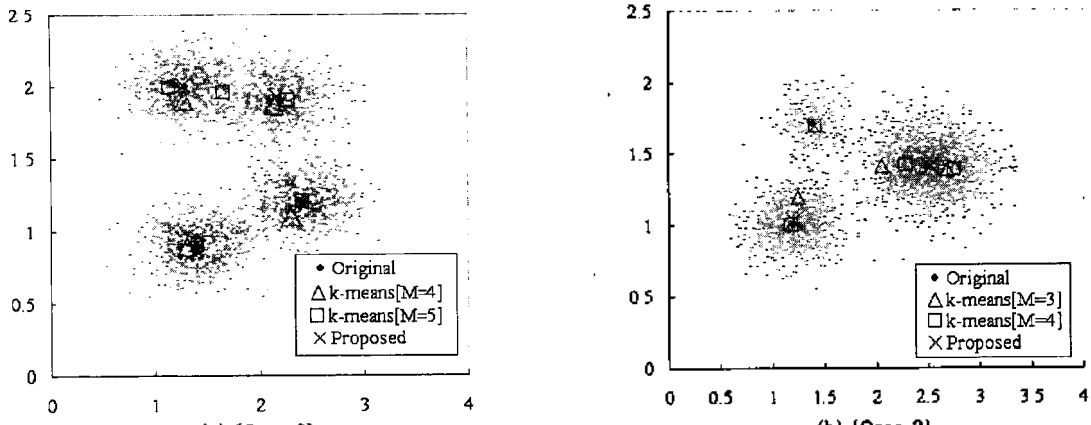


그림 2. 가우시안 혼합모델 추정 과정 [Case 1]
Fig. 2. The process of the estimated Gaussian mixtures [Case 1].



(a) [Case 2] (b) [Case 3]

그림 3. 가중치에 따른 혼합 성분의 평균 분포 [Case 2, 3]
Fig. 3. Distribution of mean of mixtures as weight [Case 2, 3].

하는데 정보의 손실 없이 네 번째 얻어진 혼합성분을 제거할 수 있었다. 따라서, 위의 생성 데이터의 추정된 최적의 혼합성분의 수는 3개로 결정되고, 이는 생성 데이터 (15)식의 분포와 일치함을 확인하였다.

그림 2는 [Case 1]의 경우에 최적의 혼합성분이 어떻게 얻어 지는지에 관한 과정을 나타낸 것이다. 그림 2-(a)는 첫 번째 혼합성분에 대하여 나타낸 것이고, 그림 2-(b), (c), (d)는 혼합성분이 추가됨에 따라, 변화과정을 각각 나타낸 것이다. 그림 2-(d)에서 추가된 혼합성분은 첫 번째 얻어진 혼합성분과 중첩되는 부분이 많으므로 통계적으로 종속적 관계임을 시각적으로 보여준다. 따라서, 마지막 단계에서 얻어진 혼합성분을 제외한 세 번째 단계에서 얻어진 혼합성분들로 최종 모델 파라미터를 추정한다.

다음으로 초기값 설정에 따라 혼합성분의 모델 추정에 어떤 영향을 주는지 확인하기 위하여 기존의 k-means 방법과 제안된 방법의 성능을 살펴 보았다. 여기에 사용된 데이터는 다음의 두 가지 데이터로 하나는 혼합성분의 가중치가 동일한 경우[Case 2]와 다른 가중치를 갖는 경우[Case 3]이다. 각각은 다음의 분포를 갖는다.

[Case 2]:

$$0.25N\left[x;\begin{pmatrix}2.2\\1.9\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right]+0.25N\left[x;\begin{pmatrix}1.4\\0.9\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right]$$
$$+0.25N\left[x;\begin{pmatrix}2.4\\1.2\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right]+0.25N\left[x;\begin{pmatrix}1.3\\2\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right] \quad (16)$$

[Case 3]:

$$0.1N\left[x;\begin{pmatrix}1.4\\1.7\end{pmatrix},\begin{pmatrix}0.04 & 0\\0 & 0.02\end{pmatrix}\right]+0.3N\left[x;\begin{pmatrix}1.2\\1.0\end{pmatrix},\begin{pmatrix}0.06 & 0\\0 & 0.02\end{pmatrix}\right]+0.6N\left[x;\begin{pmatrix}2.5\\1.2\end{pmatrix},\begin{pmatrix}0.09 & 0\\0 & 0.03\end{pmatrix}\right] \quad (17)$$

그림 3은 가중치에 따른 혼합성분의 평균 분포를 나타낸 것이다. 그림 3의 (a)는 혼합성분의 가중치가 동일한 경우[Case 2] 이고, (b)는 혼합성분의 가중치가 하나의 혼합성분에 많이 분포 되어 있는 경우이다[Case 3]. 그림3-(a)의 경우에는 혼합성분의 개수가 $M=4$일 때는 기존의 k-means 알고리즘과 제안한 방법이 모두 비슷한 결과를 얻었지만, $M=5$인 경우에는 k-means 알고리즘만이 추정된 평균의 위치는 원본 데이터의 평균과는 거리가 멀게 나타났다. 그림 3-(b)은 혼합성분이 $M=3$이거나 $M=4$이든 상관없이, k-means 알고리즘의 경우에는 추정된 값이 원본과 거리가 멀게 나타났다. 즉, k-means 알고리즘은 혼합성분의 개수가 정확하지 않거나 데이터가 불균일하게 분포되어 있는 경우에 정확한 평균값을 얻지 못한 반면에, 제안한 방법은 정확한 모델을 추정할 수 있었다.

마지막으로, 실제 음성 데이터를 이용하여 화자 식별을 수행하였다. 실험에 사용한 음성데이터는 한국인 남자 100명, 여자 100명이 각각 3회 방문하여 5번 발성한 총 15개의 발성 문장이다. 각 화자를 모델링 하기 위한 학습과정으로 각 화자마다 10개의 발성 문장을 사용하였고, 화자 식별 테스트를 위하여 각 화자 별 나머지 5개의 문장을 사용하였다. 음성 데이터는 16kHz로 샘플링 하였고, 음성 분석을 위하여 10ms 중첩을 가진 20ms 해밍창을 이용하였으며, 실험에 사용된 파라미터는 캡스트럼+델타켑스트럼+델타에너지로 구성된 25차 MFCC이다. 화자 식별에서 제안된 알고리즘의 성능을 확인하기 위하여 전형적인 GMM과 AIC, BIC, ICL, MI 방법을 적용하였다. 여기에서 전형적인 GMM의 혼합성분의 수는5, 10, 15, 20, 25, 30개로 각각 고정시켜 성능을 얻었다.

표 2. 화자 식별률(%)
Table 2. The performance of speaker identification (%).

Algorithm	Number of mixtures	Performance (%)
The Original GMM with the fixed number	5	93.8
	10	98.3
	15	99.1
	20	98.7
	25	98.9
	30	98.6
AIC	28	95.5
BIC	21	96.9
ICL	20	98.1
MI (removing mixture)	19	98.7
The proposed method	18	98.9

나머지 방법들은 $M_{min}=2$, $M_{max}=30$로 두고 사이에 존재하는 모든 혼합성분에 대하여 각각 테스트 하여 최적의 혼합성분을 추정하였다.

제안된 방법과 기존의 방법들에 대한 결과를 표 2에 나타내었다. 전체 화자에 대하여 고정된 개수의 혼합성분을 사용하여 화자 식별을 수행하였다. 고정된 혼합 성분을 사용한 경우에, 15개의 혼합성분을 사용한 경우에 가장 좋은 99.1%의 화자식별률을 보였다. 즉, 전형적인 GMM 방법의 결과에서 보여 주듯이, 화자인식의 성능은 일정한 수준의 성능에 접근하면 혼합성분의 개수가 증가해도 더 이상 성능이 향상되지 않는데, 이는 너무 많은 혼합성분을 사용하는 경우에는 종속관계에 있는 혼합성분들이 과잉상태로 존재하기 때문에 성능이 일정하게 나타나는 것이다. 나머지 기존의 방법들은 각 화자별로 혼합성분의 개수를 추정하였으며, 전체 화자에 대한 평균 혼합성분 개수로 표에 나타내었다. AIC방법에서는 28개의 혼합성분이 추정되어 95.5%의 화자 식별 성능을 얻었고, BIC 방법은 21개의 혼합성분을 이용하여 96.9%, ICL방법은 20개의 혼합성분과 98.1% 성능을 얻었다. 혼합성분의 개수를 감소시켜 가면서 구하는 MI 방법은 19개의 혼합성분과 98.7% 화자 식별률을 보였으며, 마지막으로 제안된 방법은 18개의 혼합성분이 추정되었고, 98.9%의 성능을 보였다. 따라서, 혼합성분의 개수를 추정하는 방법들 중에서 제안된 방법의 성능이 가장 높게 나타났다.

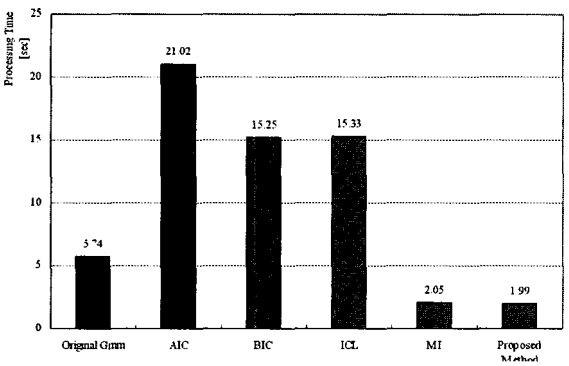


그림 4. 개인별 모델 추정을 위한 처리 시간 비교
Fig. 4. Comparison of the processing time for the estimation of model parameters by individual.

그림 4는 계산량 측면에서 제안된 방법과 기존의 방법을 비교하기 위하여 개인별 모델 추정을 위한 처리 시간을 나타낸 것이다. 실험은 Pentium 4 CPU 3.4Ghz PC 환경에서 수행되었고, 처리시간은 화자 한 명당 모델 추정에 소요되는 시간을 의미한다. 그림 4에서, 전형적인 GMM 방법은 위에서 사용한 각각 5, 10, 15, 20, 25, 30개 일 때 수행한 결과의 총 시간이고, 다른 AIC, BIC, ICL 방법들은 혼합성분 2개부터 시작하여 30개까지 하나씩 증가 시켜가면서 가장 최적의 혼합 성분을 결정하는데 소요된 시간이다. 이에 반해 MI 방법은 30개의 혼합성분으로부터 하나씩 감소시키면서 최종 모델을 결정하는데 소요된 시간이고, 제안한 방법은 2개의 혼합 성분부터 시작하여 혼합성분을 하나씩 증가 시키면서 얻어진 시간이다. MI 방법과 제안된 방법은 상호관계를 측정하므로, 최적의 혼합성분으로 판단되면 알고리즘을 정지한다. 즉, $M_{min}=2$부터 $M_{max}=30$까지 전체에 대한 실험 없이 제안된 방법은 단지 $M_{min}=2$부터 최적 혼합성분이 얻어 질 때 까지만 추정하게 되므로 최적 해를 얻는 시간을 줄일 수 있다. 따라서, 제안된 방법에서 계산에 소요된 처리 시간이 가장 적음을 확인할 수 있었다.

VI. 결 론

본 논문에서는 화자인식을 위하여 GMM의 파라미터를 정확하게 추정하기 위하여 점진적 k-means알고리즘에 의해 초기화된 EM 알고리즘방법을 제안하였다. EM 알고리즘이 초기값에 상당이 민감하지만, 점진적인 k-means알고리즘에 의해 초기화된 EM 알고리즘은 전역 최적점으로 수렴되도록 보장된다. 또한, 점진적 k-means알고리즘을 사용함에 따라 한번에 하나씩 혼합성분을 증가 시키고, 최적의 혼합성분을 위한 척도로 새로 추가된 혼합성분과 이전 혼합성분들의 상호 관계를 측정하여 판단한다. 양의 상호 관계를 가지는 혼합성분

이 나타나면, 통계적으로 이전에 얻어진 모델들과 종속적인 관계이므로, 이전 단계에서 추정되었던 개수가 하나 적은 모델들로 최적의 혼합성분이 결정된다. 제안된 방법은 임의로 생성한 데이터로부터 성능의 정확성을 확인할 수 있었고, 이를 화자 인식에적용한 결과 기존의 방법보다 더 높은 화자 식별률을 얻을 수 있었다. 또한, 최적의 혼합성분을 추정하는데 있어서, 모든 후보 혼합성분들에 대한 실험을 하지 않으므로 계산시간을 많이 줄일 수 있었다.

감사의 글

This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center (BERC) at Yonsei University.

참 고 문 헌

1. Paclik, P., Novovičová, J., "Number of components and initialization in Gaussian Mixture Model for pattern recognition", In Proceedings of the 14th ICPR, Australia, 886-890, 1998.
2. Figueiredo, M.A.T., Jain, A.K., "Unsupervised Learning of finite mixture models, " IEEE Trans. on PAMI., 24 (3), 381-396, 2002.
3. Reynolds, D.A., Rose, R., "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Trans. on SAP, 3 (1), 72-82, 1995.
4. Dempster,A., Laird, N., Rubin, D., "Maximum likelihood from incomplete data via the EM algorithm," J.Roy.Statist. Soc. Ser., B39, 1-38, 1977.
5. Richardson, S., Green, P., "Bayesian Approaches to Gaussian Mixture Modelling,"IEEE Trans. on PAMI, 2, 243-252, 1997.
6. Mclachlan, G., Peel, D., *Finite Mixture Models*, (John Wiley & Sons, New York, 2000).
7. Akaike, H., "Information theory and an extension of the maximum likelihood principle", In second International Symposium on Information Theory, eds. V.N. Petrov and F. Csaki, Budapest: Akailseoniai-Kiudo, 267 -281, 1973.
8. Schwarz, G., "Estimating the Dimension of a Model," Annals of Statistics, 6, 461-464, 1978.
9. C. Biernacki, G.Celeux and G. Govarert., "Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood," Technical Report 3,521, Inria ,1998.
10. Yang, Z.R., Zwolinski, M., "Mutual information theory for adaptive mixture models," IEEE Trans. on PAMI , 23 (4), 396-403, 2001.
11. 이윤정, 이기용, "화자 식별을 위한 GMM의 혼합성분의 개수 추정", 한국음성과학회, 11 (2), 237~246, 2004.
12. Likas,A., Vlassis, N., Verbeek, J., "The Global k-means clustering algorithm," Pattern Recognition 36, 451-461, 2003.
13. Cheung,Y., "k*-means: A new generalized k-means clustering algorithm," Pattern Recognition Letters 24, 2883-2893, 2003.

저자 약력

• 이 윤 정 (Younjeong Lee)

2001년 2월: 숭실대학교 정보통신과(공학사)
2003년 2월: 숭실대학교 정보통신과(석사)
2003년 3월~현재: 숭실대학교 정보통신과(박사과정)
*주관심분야: 음성신호처리, 화자인식, 멀티미디어, 신경망

• 서 창 우 (Changwoo Seo)

1996년 2월: 창원대학교 전자공학과(공학사)
1998년 2월: 창원대학교 전기전자제어공학부(석사)
2003년 2월: 숭실대학교 정보통신전자공학부(박사)
2000년 3월~2003년 5월: ㈜웹프로텍 음성개발팀 팀장
2003년 5월~현재: ㈜인스모바일 기술연구소 S/W 개발팀 책임연구원
※주관심분야: 음성신호처리, 멀티미디어, 모바일

• 한 헌 수 (Hernsoo Hahn)

1991년: university of southern California(공학박사)
1992~현재: 숭실대학교 정보통신전자공학부 교수
1994년: 일본기계기술 연구소 객원 연구원
1998년: 숭실대학교 어학원장
1999년: 숭실대학교 정보통신 전자공학부 학부장
※주관심분야: 비젼을 이용한 로봇제어, 알골영상처리기술, 센서 융합

• 이 기 용 (Kiyong Lee)

1991년 2월: 서울대학교 전자공학과(박사)
1991년 9 월~1997년 8월: 국립창원대학교 조교수
1994년 8월~1995년 6월: 일본 와세다대학 초빙연구원
1996년 1월~3월: 영국 에딘버러대학 박사후과정
1997년 6월~ 8월: 독일뮨헨공대 초빙연구원
1997년 9월~현재: 숭실대학교 정보통신전자공학부 부교수
※주관심분야: 음성신호 향상, 화자인식, 음성인식, 신경망