한국음향학회지 제26권 제5호 pp. 199~206 (2007)

경계 차감 클러스터링에 기반한 클러스터 개수 추정 화자식별

Speaker Identification with Estimating the Number of Cluster Based on Boundary Subtractive Clustering

이 윤 정*, 최 민 정**, 서 창 우***, 한 헌 수****
(Younjeong Lee*, Minjung Choi**, Changwoo Seo***, Hernsoo Hahn****)
국방과학연구소*, (주)인스모바일 기술연구소**, (주)에스씨디정보통신연구소***,
숭실대학교 정보통신전자공학부****

(접수일자: 2007년 5월 14일, 수정일자: 2007년 6월 25일, 채택일자: 2007년 7월 11일)

본 논문에서는 화자식별을 위한 특징벡터의 새로운 클러스터링 방법을 제안한다. 제안된 방법은 클러스터 센터에 대한 초기값 설정과 클러스터 개수에 대한 사전 정보 없이 클러스터링이 가능하다. 각 클러스터 센터는경계 차감 클러스터링 알고리즘으로 한 번에 한 개의 클러스터 센터가 추가됨으로써 순차적으로 구해지며, 클러스터 개수는 클러스터간의 상호관계를 조사하여 결정된다. 인공 생성 데이터 및 TIMIT 음성을 이용하여 실험한 결과로부터 제안된 방법이 기존의 방법보다 우수함을 확인하였다.

핵심용어: 화자식별, 클러스터링, 차감 클러스터링, K-평균 알고리즘, 상호관계

투고분야: 음성 처리 분야 (2,5)

In this paper, we propose a new clustering algorithm that performs clustering the feature vectors for the speaker identification. Unlike typical clustering approaches, the proposed method performs the clustering without the initial guesses of locations of the cluster centers and a priori information about the number of clusters. Cluster centers are obtained incrementally by adding one cluster center at a time through the boundary subtractive clustering algorithm. The number of clusters is obtained from investigating the mutual relationship between clusters. The experimental results for artificial datum and TIMIT DB show the effectiveness of the proposed algorithm as compared with the conventional methods.

Key words: Speaker identification, Clustering, Subtractive clustering, K-means algorithm, Mutual relationship

ASK subject classification: Speech Processing (2,5)

I. 서 론

화자식별 (speaker identification)은 발성된 음성 신호가 등록된 화자들 중에서 어떤 화자인지를 골라내는 것이다. 화자식별을 위하여 학습단계 (training stage)에서 화자의 특징벡터 (feature vector)들을 몇 개의 분리된 클러스터로 구분하여 화자모델을 구성하고, 테스트 단계 (testing stage)에서 입력된 화자의 특징벡터와 모든 화자모델간의 유사도를 측정한다.

책임저자: 이 윤 정 (youn@add.re.kr)
138-110 서울 송파구 거여동 산25번지 국방과학연구소
(전화: 02-3400-2684; 팩스: 02-403-3512)

클러스터링의 목적은 주어진 데이터 집합을 구분되는 클러스터로 나누는 것으로 k-means및 fuzzy c-means (FCM) 알고리즘 등이 잘 알려져 있다. 이러한 기존의 클러스터링 알고리즘의 잘 알려진 문제점은 실제 데이터로 작업할 때 클러스터 센터의 부적절한 초기값이 성능에 영향을 주는 것과, 클러스터의 개수는 항상 사전에 정의해야 하는 것이다 [1, 2]. 초기값 설정 문제점을 해결하기 위하여, mountain clustering [3, 4], global k-means [5] 방법 등 많은 방법이 제안되어 있으나 근본적으로 클러스터 개수를 안다는 전제가 필요하며, 데이터 수가 많을 경우 계산량이 증가하는 문제가 발생한다.

이러한 문제점을 해결하기 위해서 수정된 mountain clustering [6], 비용함수 (cost function) [7] 및 클러스터의 상호관계 (mutual relationship) [8]에 근거한 새로운 클러스터링 방법을 제안하였다. 본 논문에서 제안하는 방법은 경계 차감 클러스터링 방법으로 한 번에 하나씩 클러스터 센터를 추가하여 클러스터가 점진적으로 구해지며, 추정된 클러스터간의 상호관계를 조사하여 최적의 클러스터 개수가 결정된다. 즉, 한 개의 클러스터로 시작하여, 클러스터 간의 상호관계가 최초로 양의 값을 보일 때까지 반복함으로써 최적 클러스터 개수를 얻을 수 있다. 또한 제한된 전체 유사성 (bounded total similarity)을 측정하여 클러스터의 센터로 사용되는 각각의 데이터에 대하여, 모든 다른 데이터들과의 거리를 계산하는 것이 아니라 센터 후보의 경계 안에 속하는 데이터들에 대한 경우만 거리를 계산함으로써 클러스터 추정 속도를 향상시켰다. 제안된 알고리즘의 최적의 클러스터 개수 추정은 화자모델의 초기값으로 사용되었다.

II. k-means 알고리즘

이 절에서는 하나의 클러스터로부터 점진적으로 하나씩 증가 시켜가는 증가방법에 대하여 설명한다. T개의 d-차원 특징벡터 열 $X=\{x(1),\cdots,x(T)\}, x(t)\in R^d$ 가 주어졌을 때 다음의 알고리즘들을 적용할 수 있다.

2.1. 증가 k-means 알고리즘

k-means 알고리즘은 클러스터의 개수가 정해진 상태에서 사용할 수 있는 방법으로 모델 생성을 위하여 가장 잘 알려져 있다 [1, 5]. 이 알고리즘은 몇 개의 분리된 영역에서 임의의 클러스터 센터 (center)와 모든 입력 특징벡터들과의 거리를 비교하여 가장 작은 값을 갖는 센터에 대하여 각각의 특징벡터들을 그 성분의 요소로 하여 센터를 다시 측정하는 방법이다. 여기에서 가장 가까운 성분의 센터와 각각의 데이터 사이의 거리를 측정하여 더한 값을 식 (1)의 척도 (criterion) F로 표현한다. 이 방법은 센터 μ_k가 변화가 없을 때까지 반복 수행하여 클러스터의 센터를 추정한다.

$$F(\{C_1,\cdots,C_M\})=\sum_{k=1}^{M}\sum_{j=1}^{T_k}\|\mu_{kj}-\mu_k\| \qquad (1)$$

여기에서 M은 클러스터링 개수, T_k는 클러스터 k에 속한 데이터들의 개수이고, μ_{kj}는 각각의 특징벡터들 중 k번째 클러스터의 j번째 데이터로 간주된다. μ_k는 k번째 클러스터의 센터로 식 (2)와 같이 구할 수 있다.

$$\mu_k=\frac{1}{T_k}\sum_{j=1}^{T_k}\mu_{kj}\quad,\quad k=1,\cdots,M \qquad (2)$$

일반적으로 증가 k-means 알고리즘은 클러스터링 에러 (clustering error) 관점에서 최적해 (optimal solution)를 찾는 방법이다. 각 단계마다 클러스터링 에러를 최소화하기 위하여 초기값으로부터 클러스터의 센터가 계속 이동되면서 처리되기 때문에 클러스터링 센터의 초기 위치에 상당히 민감하다는 단점이 있다 [1, 5]. 그러므로 최적해에 근접한 값을 얻기 위해서는 초기값을 변화시키면서 여러 번 수행해야 한다. 그러나 이러한 클러스터링 방법은 데이터의 분포가 한 곳에 몰려있는 경우에는 클러스터의 개수가 정확하더라도 정확한 센터를 구하기 어렵다. 이 경우 클러스터의 센터와 데이터 사이의 거리가 가장 가까운 곳에서 센터의 위치가 결정되므로, 데이터의 분포가 집중되어있지 않는 곳에서 센터가 얻어지는 경우가 발생할 수 있다.

2.2. 고속 글로벌 k-means 알고리즘 (Fast Global k-means algorithm)

글로벌 k-means 알고리즘의 경우 데이터가 증가할수록 계산량이 증가되는 단점을 보완하기 위하여 고속 글로벌 k-means 알고리즘 방법이 제안되었다 [6]. 이 방법은 클러스터를 추정하기 위하여 글로벌 k-means 알고리즘의 질을 저하시키지 않고 계산량을 감소시킬 수 있다. 고속 글로벌 k-means 알고리즘은 전체 T개의 데이터와 k-1개의 클러스터들 사이의 최소 거리를 사용하므로 최종 클러스터를 추정할 때까지 k-means 알고리즘을 수행하지 않는다. 대신 클러스터의 위치로 할당 (allocation)이 가능한 모든 $x(t)$가 최소 클러스터링 에러 값을 갖도록 식 (3), (4)를 이용하여 새로운 클러스터의 센터를 t^*에서 구할 수 있다 [5, 9, 10].

$$F_k(t)=\sum_{n=1}^{T}\max\left(d_{k-1}^n-\|x(t)-x(n)\|^2,0\right),\quad 1\le t\le T \qquad (3)$$

여기에서,

$$\mu_k = x(t_k^*), \quad t_k^* = \arg\max_{1\le t\le T} F_k(t) \tag{4}$$

이다. 식 (3)에서 d_{k-1}^n은 $x(n)$이 $k-1$개의 클러스터 센터 중에서 가장 가까운 거리, 즉 $x(n)$이 속한 클러스터의 센터와의 거리이다. 따라서 고속 글로벌 k-means 알고리즘을 사용할 경우 글로벌 k-means 알고리즘에서 각각의 데이터 $x(n)$에서의 클러스터와의 거리를 매번 계산하지 않고, $k-1$개의 클러스터와 가장 가까운 거리를 나타내는 d_{k-1}^n과 새로운 클러스터와의 거리만 비교하므로 계산량이 감소하게 된다.

III. 경계 차감 클러스터링 알고리즘 (Boundary Subtractive Clustering Algorithm)

앞에서 설명한 증가 k-means 알고리즘, 고속 글로벌 k-means 알고리즘의 경우에는 균등하게 분포되어 있는 데이터에 대해서는 추정된 클러스터의 센터는 원본(original) 데이터의 센터와 상당히 근접한 결과를 보인다. 그러나 데이터의 분포가 균일하지 않을 경우에는 정확한 클러스터의 개수를 알고 있더라도 데이터의 분포를 정확하게 추정할 수 없다. 또한 모델링을 위한 데이터들 사이에 이상치 (outlier)들이 존재할 경우에 이상치에 의하여 클러스터들의 센터가 왜곡될 가능성이 높다. 이를 해결하기 위하여 주어진 데이터 분포에서 가장 높은 밀도를 갖는 데이터의 위치를 클러스터 센터로 결정하는 경계 차감 클러스터링 알고리즘을 제안한다.

3.1. 클러스터 센터 추정

경계 차감 클러스터링 알고리즘은 최적의 클러스터 추정을 위하여 계산량이 감소된 수정된 글로벌 k-means 알고리즘을 사용한다. 앞 절에서 제시된 방법들은 데이터와 클러스터들 센터와의 최소 클러스터링 에러를 갖는 것을 목표로 하지만 차감 클러스터링 방법은 밀도 함수인 지수함수 (exponential function)를 사용하므로 클러스터의 센터 주변에 있는 데이터들의 빈도수에 의하여 결정된다. 지수함수의 특성은 데이터들 사이의 거리가 멀어질수록 유사성이 매우 떨어지므로 클러스터의 경계 바깥쪽 데이터들의 경우 전체 유사성에 미치는 영향이 적다. 따라서 클러스터의 센터로 사용되는 후보에 이웃한 데이터들에 대해서만 유사성을 계산하고, 경계 밖의 데이터들은 제외 할 수 있다. 제한된 영역 안의 데이터만을 고려하여 클러스터를 추정하는 경계 차감 클러스터링 알고리즘은 다음과 같다.

특징벡터 열 $X = \{x(1),\cdots,x(T)\}$, $x(t)\in R^d$ 인 데이터가 주어져 있을 때 각각의 데이터는 $x(t)$이고, 최적의 클러스터의 개수를 M이라고 가정하자. 만약 데이터 X를 M개의 클러스터링으로 모델링하고자 하면 전체 상이 목적 함수 (total dissimilarity objective function)를 최소화하기 위하여 클러스터의 센터인 μ_k를 찾아야 한다. 여기에서 전체 유사도 측정함수 $J(\mu)=\sum_{k=1}^{M} J(\mu_k)$를 최대화하는 μ_k를 찾기 위하여 $x(t)$와 μ_k 사이의 유사성을 측정하여 사용한다 [2, 6]. 데이터들의 밀도가 가장 높은 지점을 찾기 위하여 데이터들의 유사성은 식 (5)와 같이 나타낼 수 있다.

$$J(\mu_k) = \sum_{t=1}^{T} \exp\left(-\frac{\|x(t)-\mu_k\|^2}{\rho}\right), \quad t=1,\cdots,T, \tag{5}$$

여기에서 $\mu=(\mu_1,\cdots,\mu_M)$이고, ρ는 정규화를 위한 데이터들의 분산으로 식 (6)에 의해 구해진다.

$$\rho = \frac{1}{T}\sum_{t=1}^{T}\|x(t)-\bar{x}\|^2, \quad \bar{x} = \frac{1}{T}\sum_{t=1}^{T} x(t) \tag{6}$$

위의 식 (5), (6)을 바탕으로, 첫 번째 클러스터에 대한 데이터 $x(t)$의 전체유사도 $F_1(t)$는 모든 데이터에 대한 유클리안 놈 (norm)으로 식 (7)과 같이 표현할 수 있다.

$$F_1(t) = \sum_{n=1}^{T} I(x(n),x(t))\exp\left(-\frac{\|x(n)-x(t)\|^2}{\alpha}\right), \quad t=1,\cdots,T, \tag{7}$$

식 (7)로부터 가장 큰 값을 갖는 $F_1(t)$를 식 (8)과 같이 선택하면 첫 번째 클러스터의 센터가 결정된다.

$$\mu_1 = x(t_1^*), \quad t_1^* = \arg\max_{1\le t\le T} F_1(t) \tag{8}$$

여기서 α는 스케일 파라미터이며, $I(x,\mu)$의 값은 식 (9)와 같이 정의된다.

$$I(x,\mu) = \begin{cases} 1, & if \quad x\in B_\mu \\ 0, & if \quad x\notin B_\mu \end{cases} \tag{9}$$

식 (9)에서 x는 데이터를 의미하고, μ는 클러스터의 센터이며, B_μ는 μ를 중심으로 주변 경계영역(boundary region)을 나타낸다. 따라서 데이터 x가 B_μ안에 포함되면 $I(x,\mu)=1$이고 그렇지 않으면 $I(x,\mu)=0$이 된다. 경계영역 B_μ는

$$B_\mu = \kappa \cdot \rho \tag{10}$$

으로 얻어진다. 여기에서 κ는 범위의 정도 (degree of boundary)를 나타내고, ρ는 데이터 샘플의 분산이다.

$(k-1)$번째 클러스터가 추정되면 k 번째 클러스터에 대한 전체 유사도 $F_k(t)$는 식 (11)로 표현할 수 있다.

$$F_k(t) = F_{k-1}(t) + I(x(t),\mu_{k-1})\left[F_{k-1}(t_{k-1}^*)\exp\left(-\frac{\|x(t)-\mu_{k-1}\|^2}{\beta}\right)\right], \quad t = 1,\cdots,T \tag{11}$$

여기서 β는 스케일 파라미터이고, k 번째 클러스터의 센터는 아래 식으로 얻을 수 있다.

$$\mu_k = x(t_k^*), \quad t_k^* = \arg\max_{1\le t\le T} F_k(t) \tag{12}$$

식 (7)과 (11)에서 인접하게위치한 클러스터 선택을 방지하기 위해서 β가 α 보다 큰 값을 사용한다.

3.2. 클러스터 개수 추정

경계 차감 클러스터링 알고리즘으로 구한 클러스터 개수가 최적인지는 클러스터간의 상호관계를 측정하여 알 수 있다. 먼저, 새로 추가한 k번째 클러스터의 입장에서 이전 $(k-1)$개의 클러스터들과의 상호관계를 측정한다. 클러스터 i와 k 간의 상호관계는 다음과 같이 정의 된다 [8, 9].

$$\varphi(i,k) = p(i,k)\log\frac{p(i,k)}{p(i)p(k)}, \quad i,k = 1,\cdots,M \tag{13}$$

여기서 $p(i)$는 i번째 클러스터의 확률 (probability)이고, $p(i,k)$는 i번째와 k번째 클러스터의 결합확률 (joint probability), $p(i\,|\,x(t))$는 i번째 클러스터를 위한 사후확률 (a posteriori probability)이다.

$$p(i) = \frac{1}{T}\sum_{t=1}^{T} p(i\,|\,x(t)) \tag{14}$$

$$p(i,k) = \frac{1}{T}\sum_{t=1}^{T} p(i\,|\,x(t))p(k\,|\,x(t)) \tag{15}$$

$$p(i|x(t)) = \frac{\exp\left(-\frac{\|x(t)-\mu_i\|^2}{\alpha}\right)}{\sum_{i=1}^{k}\exp\left(-\frac{\|x(t)-\mu_i\|^2}{\alpha}\right)} \tag{16}$$

클러스터의 상호관계는 음수 (−), 영 (0), 양수 (+)의 세 가지 값으로 구분된다. 여기서 $\varphi(i,k)$값이 영 (0)이면 i번째와 k번째 클러스터는 통계적으로 독립임을 의미한다: $p(i,k)=p(i)p(k)$. 만약 $\varphi(i,k)$가 양수 (>0)이면 i번째와 k번째 클러스터는 통계적으로 서로 종속적임을 나타낸다: $p(i,k)>p(i)p(k)$. 또한 음수 (<0)인 상호관계가 측정되면 i번째와 k번째 클러스터는 약한 종속적 관계로 간주된다: $p(i,k)<p(i)p(k)$. 따라서, $\varphi(i,k)$가 양수 값으로 나타나면, i번째와 k번째 클러스터 중 하나는 추정된 모델에 영향을 주지 않고 제거할 수 있다는 것이다.

3.3. 경계 차감 클러스터링 알고리즘

클러스터의 상관관계를 적용하여 경계 차감 클러스터링 알고리즘은 다음과 같이 정리 할 수 있다.

- ▶ Step 1: $x(t), t=1,\cdots T$가 주어질 경우, 먼저 $k=1$로 초기화한다. 데이터로부터 전체 유사성 측정 함수인 $F_1(t)$는 식 (7)을 이용하여 계산하고, 식 (8)에 의하여 클러스터의 센터가 되는 μ_1를 구한다.
- ▶ Step 2: $k=k+1$로 증가시킨다.
- ▶ Step 3: k 번째 클러스터를 추정하기 위하여 식 (11)의 $F_k(t)$를 구하고 최대값을 갖는 식 (12)의 μ_k를 구한다.
- ▶ Step 4: $i=1,\cdots,k$에 대하여 확률 $p(i)$, $p(k)$와 결합확률 $p(i,k)$를 계산한다
- ▶ Step 5: 추가된 k번째 클러스터와 $i=1,\cdots,k-1$에 대하여 상호관계 $\varphi(i,k)$를 측정한다.
- ▶ Step 6: 만약 $\varphi(i,k)>0$인 경우가 적어도 하나이상 존재하면 클러스터의 추정을 멈추고, 최적의 클러스터 개수는 $M=k-1$로 정한다. 그렇지 않으면 다시 Step 2를 수행한다.

IV. 실험 및 결과

제안된 알고리즘의 성능확인을 위하여 증가 k-means 알고리즘, 고속 글로벌 k-means 알고리즘과 제안된 알고리즘의 성능을 비교하였다. 실험에 사용한 데이터는 정확한 추정을 확인하기 위하여 인공적인 (artificial) 데이터와 실제 음성 신호로부터 추출된 2차원 특징벡터 열을 이용하였다. 또한 화자 식별 성능 확인을 위하여 TIMIT DB를 사용하였고, 화자모델은 GMM (Gaussian Mixture Model) [9, 10]을 적용하였다. 제안된 알고리즘에서는 클러스터링 시 효율적인 차감을 위하여 본 실험에서는 $\alpha = 0.1, \beta = 1, \kappa = 2$ 로 사용하였다.

4.1. 실험 데이터

4.1.1. 인공 생성 데이터

실험을 위하여 Case 1과 Case 2의 분포를 갖는 2차원 가우시안 분포를 가진 2000 개의 데이터를 생성하였다. 제안된 알고리즘의 특성을 비교하기 위하여 인공 생성 데이터는 동일한 가중치를 갖는 경우 (Case 1)와 가중치가 다른 경우 (Case 2)를 사용하였다.

Case 1:

$$0.25N\left[x\middle|\begin{pmatrix}2.2\\1.9\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right]+0.25N\left[x\middle|\begin{pmatrix}1.4\\0.9\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right]$$
$$+0.25N\left[x\middle|\begin{pmatrix}2.4\\1.2\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right]+0.25N\left[x\middle|\begin{pmatrix}1.3\\2.0\end{pmatrix},\begin{pmatrix}0.07 & 0\\0 & 0.02\end{pmatrix}\right] \quad (17)$$

Case 2:

$$0.1N\left[x\middle|\begin{pmatrix}1.4\\1.7\end{pmatrix},\begin{pmatrix}0.04 & 0\\0 & 0.02\end{pmatrix}\right]+0.3N\left[x\middle|\begin{pmatrix}1.2\\1.0\end{pmatrix},\begin{pmatrix}0.06 & 0\\0 & 0.02\end{pmatrix}\right]$$
$$+0.6N\left[x\middle|\begin{pmatrix}2.5\\1.2\end{pmatrix},\begin{pmatrix}0.09 & 0\\0 & 0.03\end{pmatrix}\right] \quad (18)$$

여기에서 $aN[x|\mu,\Sigma]$는 a는 가중치이고, x는 데이터 벡터, μ는 평균 벡터, Σ는 공분산 행렬을 나타낸다.

4.1.2. 실제 음성 데이터 (TIMIT DB)

TIMIT DB는 총 630명으로 남자 438, 여자 192명으로 구성되어 있고, 각각의 화자는 10개의 문장을 발성하였다. TIMIT DB에서 한 화자의 음성 데이터는 동일한 두 개의 문장 SA1, SA2와 다른 화자와 다른 문장인 각 3개의 SI문장과 5개의 SX문장으로 구성된 10개로 이루어져 있다. 여기에서 학습을 위하여 8개의 문장을 사용하였고 나머지 2개의 문장을 테스트에 적용하였다. DB 구성 및 실험 환경은 표 1에 나타내었다. 실험을 위한 음성의 특징벡터 추출은 20차 MFCC를 사용하였다.

표 1. TIMIT DB를 위한 파라미터
Table 1. Parameters for the TIMIT DB.

종류	값
Dimension of MFCC	20
Sampling	16khz
FFT	512
WINSIZE with 50% overlap	20ms
Number of Mixtures	8~64

4.2. 실험 결과

4.2.1. 인공 생성 데이터

Case 1의 경우는 표 2 (a) 에 클러스터의 추정과정과 상호관계를 설명하였다. 표 2 (a)에서 k=2, 3, 4일 때 까지는 모두 음의 상호관계가 측정되었고, k=5가 되었을 때 다섯 번째 클러스터와 세 번째 클러스터의 상호관계가 0.2812로 측정되어 다섯 번째 클러스터는 세 번째 클러스터와 서로 종속관계임을 알 수 있었다. 따라서 Case1의 경우 최종적으로 4개의 클러스터가 얻어진다.

경계 차감 클러스터링 알고리즘을 이용한 Case 2에 대한 클러스터 추정과정은 표 2 (b)에 나타내었다. 표 2 (b)에서 클러스터의 개수가 k=2, 3인 경우에는 음의 상호관계를 갖지만, k=4가 되면 네 번째 클러스터와 세 번째 클러스터의 상호관계가 0.0976이 되어 최적의 클러스터 개수는 3개로 결정된다.

Case 1, 2의 데이터에 대하여 클러스터링 에러값(E)은 그림 1에 나타내었고, 다음과 같이 구할 수 있다.

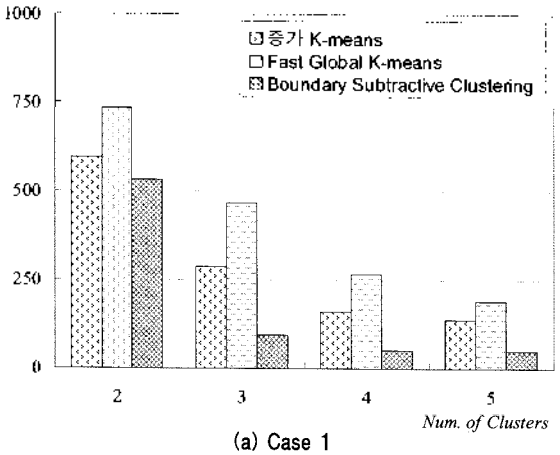
표 2. 경계 차감 클러스터링 알고리즘을 사용한 클러스터들의 상호관계
Table 2. Mutual Relationship of Clusters using Boundary Subtractive Clustering Algorithm.

(a) Case 1

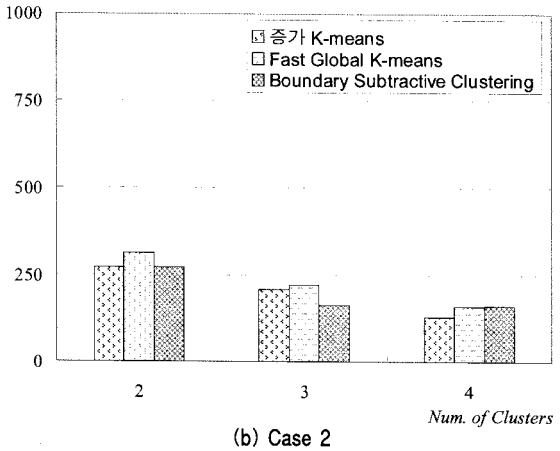
k	μ_k	$\varphi(i,k)$	i
1	(2.1870,1.9043)	~	1
2	(1.3892,0.8963)	- 0.0793	1
3	(1.2572,1.9998)	- 0.0455	1
		- 0.0004	2
4	(2.3877,1.1988)	- 0.0095	1
		- 0.0248	2
		- 0.0007	3
5	(1.2963,2.0014)	- 0.0188	1
		- 0.0003	2
		0.2812	3
		- 0.0006	4

(b) Case 2

k	μ_k	$\varphi(i,k)$	i
1	(2.5034,1.2009)	~	1
2	(1.2050,1.0011)	-0.0328	1
3	(1.3911,1.7572)	-0.0084	1
		-0.0087	2
4	(1.4030,1.6805)	-0.0076	1
		-0.0079	2
		0.0976	3



(a) Case 1



(b) Case 2

그림 1. 클러스터링 에러(E)
Figure 1. Error of Clustering(E).

$$E = \sum_{t=1}^{T} \left\| x(t) - \mu_{k_{\min}} \right\|^2 \tag{19}$$

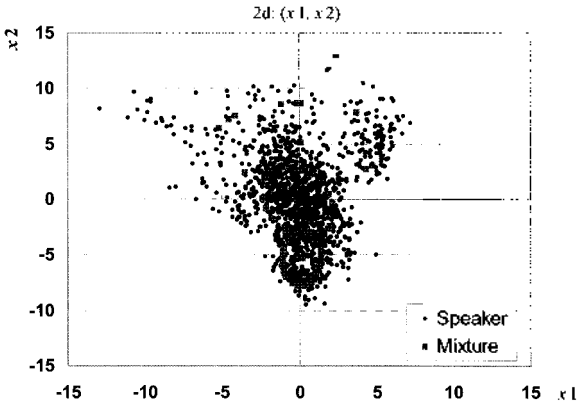
$$k_{\min} = \arg\min_{1 \le k \le M} \left\| x(t) - \mu_k \right\|^2, t = 1, \cdots, T \tag{20}$$

여기에서 $\mu_{k_{\min}}$ 는 k의 클러스터 중에서 $x(t)$와 가장 가까운 $k_{\min}$ 번째 클러스터의 평균이다. 그림 1에서, 기존 방법에서 보다 제안된 방법의 경우가 가장 작은 클러스터링 에러를 가짐을 알 수 있다.

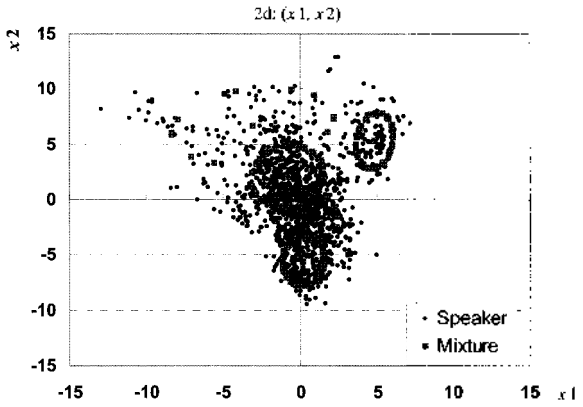
4.2.2. TIMIT DB

이 실험에서는 제시된 알고리즘들이 실제 음성신호에서 어떻게 클러스터가 추정되는지 확인한다. TIMIT DB의 MFCC 20차원 특징벡터 중에서 1, 2차 특징벡터를 사용하여 한 화자에 대하여 각 특징벡터의 분포와 추정된 클러스터의 윤곽선 (contour)을 그림 2에 나타내었다. 그림 2에서 (a)는 증가 k-means 알고리즘, (b)는 고속 글로벌 k-means 알고리즘, 그리고 (c)는 경계 차감 클러스터링 알고리즘을 사용한 경우에 대하여 추정된 클러스터의 윤곽선을 나타낸 것이다. (a), (b) 방법의 결과는 추정된 공분산 범위가 크게 측정되어 정확한 분포를 얻을 수 없지만, (c)의 경계 차감 클러스터링 알고리즘을 사용한 경우 적절하게 클러스터가 추정됨을 확인 할 수 있다.

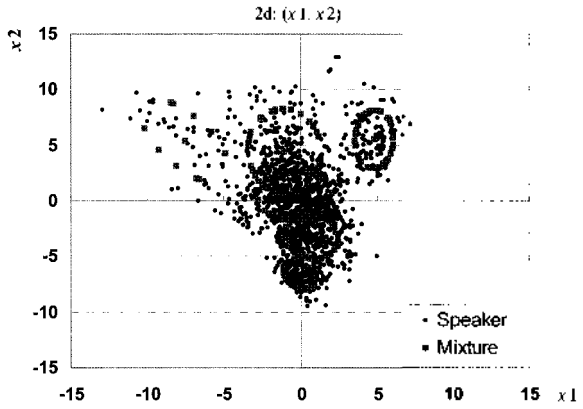
앞에서 임의의 생성데이터와 음성 신호에서 추출된 2차원 특징벡터들을 이용하여 기존의 알고리즘 및 제안된 방법의 성능을 확인하였다. 이를 바탕으로 TIMIT DB에



(a) 증가 k-means 알고리즘



(b) 고속 글로벌 k-means 알고리즘



(c) 경계 차감 클러스터링 알고리즘

그림 2. 2차원 특징벡터의 클러스터 추정
Figure 2. Estimating of clusters in 2-dimensional feature vectors.

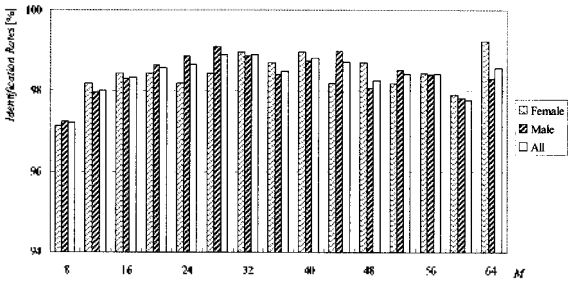


그림 3. 화자 식별 성능(%)
Figure 3. Speaker Identification Rates.

대하여 클러스터의 개수를 추정하여 화자모델을 생성한 후 화자식별 실험을 하였다. 클러스터의 개수를8에서 64까지 증가시키면서 화자식별 실험 결과를 그림 3에 나타내었다. TIMIT DB에서 여자의 경우 32개의 혼합성분에 대하여 98.96%, 남자의 경우 28개 일 때 99.09%, 전체적으로는 28개 일 때 98.90%로 가장 높은 성능을 보였다. 이와 같이 화자식별 성능은 각 화자에 따라 최적의 클러스터 개수가 다르기 때문에 실제 음성신호를 직접 입력 받아 사용하는 시스템에 대해서는 사용자의 음성신호에 대한 사전 지식이 없으므로 클러스터의 개수를 미리 알 수 없다. 따라서, 화자식별의 성능을 향상시키기 위하여 화자모델을 정확하게 추정할 수 있도록 최적의 클러스터 개수 추정 방법을 사용해야 한다.

표 3은 각 알고리즘에 대하여 추정된 최적의 클러스터들의 센터와 화자식별 성능을 나타낸 것이다. 표에서 증가 k-means 방법의 경우 20개일 때 98.57%의 성능을 보였고, 고속 글로벌 k-means 알고리즘의 경우 15개의 클러스터가 추정되었고 98.26% 성능을 보였다. 마지막으로 경계 차감 클러스터링 알고리즘의 경우 16개의 클러스터가 추정되었고, 98.74%의 성능을 보였다. 실험 결과로부터 임의의 데이터뿐만 아니라 실제 음성을 사용하는 화자모델 추정에 사용되는 화자식별 알고리즘에서도 제안된 알고리즘의 성능이 우수함을 확인 할 수 있었다.

표 3. 화자식별 성능 비교
Table 3. Comparision of Speaker identification Rates.

Algorithm	Number of Clusters (M)	Identification Rates(%)
증가 k-means 알고리즘	12	98.02
	16	98.33
	20	98.57
고속 글로벌 k-means 알고리즘	15	98.26
경계 차감 클러스터링 알고리즘	16	98.74

V. 결 론

본 논문에서는 최적의 클러스터 개수를 추정하기 위하여 경계 차감 클러스터링 알고리즘을 제안하였다. 제안된 알고리즘에서 각 클러스터 센터는 한 단계에 하나씩 클러스터 센터가 추가됨으로써 순차적으로 구해지며, 클러스터 개수는 클러스터간의 상호관계를 측정하여 결정된다. 따라서 클러스터 센터에 대한 초기값 설정과 클러스터 개수에 대한 사전 정보 없이 최적화된 클러스터링이 가능하다.

실험 결과로부터 임의의 데이터뿐만 아니라 실제 음성을 사용하는 화자모델 추정에 사용되는 화자식별 알고리즘에서도 제안된 알고리즘의 성능이 우수함을 확인 할 수 있었다.

참 고 문 헌

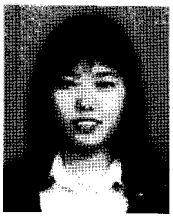
1. Lozano, J.A., Pena, J.M., and Larranaga, P., "An empirical comparison of four initialization methods for the k-means algorithm," Pattern Recognition Letters, 20 1027-1040, 1999.
2. Gath and Geva, A.B., "Unsupervised optimal fuzzy clustering," IEEE Trans. Pattern and Machine Intelligence, 11 (7) 773-778, 1989.
3. Ryager, R., Filev, D.P., "Approximate clustering via the mountain method," IEEE Trans on Systems, Man, and Cybernetics, 24, 1279-1284, 1994.
4. Zadeh, L.A., "Similarity Relations and fuzzy Orderings," Information Science 3, 177-200, 1971.
5. Likas, A., Vlassis, N. and Verbeek, J.J., "The global k-means clustering algorithm," Pattern recognition 36, 451-461, 2003.
6. Chiu, S.L., "Fuzzy model identification based on cluster estimation," J. of Intelligent and Fuzzy sys., 2 (3), 267-278, 1994.
7. Kothari, R., Pittas, D., "On finding the number of clusters," Pattern Recognition Letters 20, 405-416, 1999.
8. Yang, Z. R. and Zwaliuski, M., "Mutual information theory for adaptive mixture model," IEEE Trans. Pattern and Machine Intelligence, 23 (4) 396~403, 2001.
9. Lee, Y., Lee, J. and Lee, K.Y., "The Estimating Optimal Number of Gaussian Mixtures Based on Incremental k-means for Speaker Identification," International Journal of Information Technology, 12 (7) 2006.
10. 이윤정, 서창우, 한한수, 이기용, "GMM을 위한 점진적 K-means 알고리즘에 의해 초기값을 갖는 EM알고리즘과 화자식별에의 적용," 한국음향학회지, 24 (3) 117~126, 2005.

저자 약력

• 이 윤 정 (Younjeong Lee)

2001년 2월 : 숭실대학교 정보통신전자공학부(공학사)
2003년 2월 : 숭실대학교 정보통신공학과(공학석사)
2006년 2월 : 숭실대학교 정보통신공학과(공학박사)
2006년 1월~현재 : 국방과학연구소 선임연구원
※ 주관심분야 : 음성신호처리, 화자인식, 잔술데이터링크

• 최 민 정 (Minjung Choi)



2003년 2월 : 숭실대학교 정보통신전자공학부(공학사)
2005년 2월 : 숭실대학교 정보통신공학과(공학석사)
2005년 1월~현재 : ㈜인스모바일 기술연구소 주임연구원
※ 주관심분야 : 음성신호처리, 화자인식, 모바일

• 서 창 우 (Changwoo Seo)

1996년 2월 : 청원대학교 전자공학과(공학사)
1998년 2월 : 청원대학교 전기전자제어공학부 (공학석사)
2003년 2월 : 숭실대학교 전자공학과(공학박사)
2000년 3월~2003년 4월 : ㈜웹프로텍 음성개발팀 팀장
2003년 5월~2005년 6월 : ㈜인스모바일 기술연구소 책임연구원
2005년 7~현재 : ㈜에스씨디 정보통신연구소 책임연구원
※ 주관심분야 : 음성신호처리, 멀티미디어, 모바일

• 한 헌 수 (Hernsoo Hahn)

1991년 : University of Southern California 컴퓨터 공학과 (공학박사)
1992년~현재 : 숭실대학교 정보통신전자공학부 교수
1994년 : 일본기계기술 연구소 객원연구원
2006년 : 오스트리아 비엔나 공과대학 객원교수
※ 주관심분야 : 로봇 비전, 얼굴검출 및 인식, 물체 추적