

논문 2009-46CI-3-1

분산커널 기반의 퍼지 c-평균을 이용한 음악 데이터의 장르 분류

(Classification of Music Data using Fuzzy c-Means with Divergence Kernel)

박 동 철*

(Dong-Chul Park)

요 약

본 논문은 효율적인 음악 데이터의 분류를 위한 방법으로 분산커널 기반의 퍼지 c-평균을 이용한 분류기 모델을 제안한다. 분산 커널 기반의 퍼지 c-평균은 주어진 오디오 데이터에서 추출된 특징벡터의 평균과 공분산 정보를 동시에 이용하여 기존의 평균값만을 사용하는 방식에 비해 성능을 월등히 향상시킬 수 있는 장점이 있다. 사용된 방식은 확률적 분포로 주어지는 데이터 사이의 거리를 분산거리척도로 측정하고, 복잡한 분류 경계를 단순화 시키는데 효율적인 커널 개념을 사용함으로써 분류의 정확도를 극대화시킬 수 있는 장점이 있다. 제안하는 분류기의 성능을 평가하기 위하여 고전음악, 컨트리음악, 힙합, 재즈의 4개의 장르 음악데이터를 총 1200개 수집하여 실험을 진행하였다. 실험의 결과 제안된 분산커널 기반의 퍼지 c-평균을 이용하는 분류기는 기존의 방식과 비교하여 분류정확도에서 평균적으로 17.73% - 21.84%의 성능향상을 보여준다.

Abstract

An approach for the classification of music genres using a Fuzzy c-Means(FcM) with divergence-based kernel is proposed and presented in this paper. The proposed model utilizes the mean and covariance information of feature vectors extracted from music data and modelled by Gaussian Probability Density Function (GPDF). Furthermore, since the classifier utilizes a kernel method that can convert a complicated nonlinear classification boundary to a simpler linear one, the classifier can improve its classification accuracy over conventional algorithms. Experiments and results on collected music data sets demonstrate that the proposed classification scheme outperforms conventional algorithms including FcM and SOM 17.73% - 21.84% on average in terms of classification accuracy.

Keywords: Classification model, Clustering algorithm, Music genre, Fuzzy c-Means, Self-Organizing Map

I. 서 론

멀티미디어 데이터베이스(Database: DB)는 음악, 담화, 또는 연설 등의 여러 형태의 많은 오디오 녹음을 데이터 파일의 형태로 가지고 있다. 최근 들어 폭발적으로 증가하는 멀티미디어 데이터의 양은 오디오 데이터

를 데이터 형태에서 자동적으로 직접 분류하거나 검색할 수 있는 방법의 출현을 기다리고 있었다. 그러나 오디오 데이터 파일은 보통 파일의 이름, 파일의 형태(format), 샘플링 비율 등의 기본적인 필드만을 가지고 있는 일반적 데이터 파일로 취급되어왔기 때문에, 사용자로서는 필요한 내용의 파일을 찾기 위해 직접 들어보는 수 밖에 없었다. 기본적으로, 소리를 들어보지 않고 오디오 데이터를 분류하거나 검색하는 문제는 특징의 추출과 추출된 특징을 이용해 분류하는 두 가지 요소의 패턴인식 문제로 다루어질 수 있다. 이에 대한 접근 방법으로 지난 십 수년 동안 연구의 결과가 발표되었는

* 정회원, 명지대학교 정보공학과
(Dept. of Information Eng., Myong Ji University)
※ 본 연구는 한국과학재단 특정기초연구
(R01-2007-000-20330-0) 지원에 의한 것임.
접수일자: 2009년3월27일, 수정완료일: 2009년5월4일

데, 1996년도에 Music Fish Group의 Word 등에 의해 제안된 방법에서는 음원에서 음의 세기, 음률의 고저, 밝기, 대역폭, 화성과 같은 음향학적인 특징을 추출하였다. 이들 다차원의 특징들로부터 계산된 평균과 공분산을 이용하여 특징공간(feature space)을 구성하였다^[1]. 거의 동시에 Sanders는 zero-crossing rate 와 신호의 에너지를 간단하게 이용한 특징벡터를 사용하여 음악 데이터와 대화 데이터를 분류해내는 방법을 제안하였고^[2], Foot는 Mel-scaled Frequency Cepstral Coefficients (MFCCs)를 사용하는 방법^[3-4]을 제안하였으며, Li와 Khokar는 Discrete Wavelet Transform (DWT)를 사용하여 특징 추출을 시도하였다^[5]. 그러나 이러한 노력에서 음향학적인 특징 자체가 직접적으로 음악 데이터를 모델링하는 데는 이용되지 않았다. Tzanetakis와 Kook는 리듬과 pitch와 같이 음악 데이터에만 있는 특수한 특징들을 추출하여 음악 데이터를 모델링하는 방법을 제안하였는데^[6], 이러한 특징을 이용한 후속 연구가 진행되어 왔다^[7]. 그러나 이들 Tzanetakis 와 Kook의 특징벡터를 이용하는 방법에서 공통적으로 나타나는 문제점은 오디오 신호의 통계적인 정보를 충분히 이용하지 않았다는 데 있다. 즉, 오디오 신호가 확률적 데이터로 모델링되는 것이 더욱 적절하며, 데이터의 평균과 분산이 각각 직접 특징벡터의 일부분으로 아무런 구별 없이 사용되었기 때문에, 오디오/음악 데이터에서만 나타나는 통계적인 특성이 무시된 채 사용되었다.

본 논문에서 사용하고자하는 커널 방식은 일정 차원의 입력데이터를 더 높은 차원을 가지는 특징공간으로 맵핑시켜, 원래 입력데이터의 공간에서는 복잡한 비선형 분류의 문제를 선형 분류의 문제로 단순화 시키려는 노력으로 여러 가지의 군집화 문제에 적용되어왔다^[9]. 서포트 벡터 머신(Support Vector Machine: SVM)에 커널 방식을 이용한 예 [9]는 매우 특징적이고 성공적인 응용인데, 필요한 계수를 정하는 문제와 복잡한 학습이 그 유용성을 제한하고 있다. 이를 극복하기 위한 대안으로 커널 방식은 입력데이터를 고차원의 특징공간으로 맵핑하는 목적으로만 사용하고, 군집화 기능은 최적의 군집화 알고리즘에 맡기는 방법으로, 커널 방식의 장점과 군집화 알고리즘의 장점만을 이용하는 것으로 여러 가지의 전통적 군집화 알고리즘에 성공적으로 적용되었다^[10-13]. 본 논문에서는 분산커널과 퍼지 c-평균(Fuzzy c-Means: FcM) 알고리즘^[14-16]을 이용하여, 오디오 신호의 통계적 특성을 좀 더 추출하여 이용하는

분류기를 제안한다. 제안되는 분류기는 군집화에서 보장된 성능을 보이는 FcM과 오디오 신호의 통계적 특성을 모델링하기에 적합한 분산 척도, 커널 방식의 장점을 모두 결합하는 형태로 각 요소 알고리즘의 장점을 이용하는 분류기이다. 특히, 가우시안 확률 밀도함수(Gaussian Probability Density Function: GPDF)의 형태로 주어지는 오디오 신호를 사용하기 위하여, 두 개의 확률분포 사이의 거리척도로 분산거리를 사용하는 분산 기반의 커널 방식은 오디오 신호의 특징을 효율적으로 이용하는 방식이므로 오디오 신호의 분류 정확도에서 향상된 결과를 기대할 수 있다.

본 논문은 다음과 같이 구성되어 있다. 논문의 II장에서는 제안된 방법의 기반이 되는 FcM과 커널 기반의 FcM에 대해 간단히 살펴보고, III장에서는 분산 커널 기반의 FcM을 요약하여 설명한다. IV장에서는 제안된 방법의 성능을 평가하기 위하여, 몇 가지의 실제 수집된 음악 데이터에 대한 장르 판별 실험과 결과를 몇 가지 알고리즘과 비교 분석하며, V장에서는 결론을 내린다.

II. 커널 기반의 퍼지 c-평균 알고리즘

2.1 퍼지 c-평균 알고리즘

일반적으로 군집화 알고리즘의 목적은 유사한 객체를 한 집단으로 모으고, 상이한 객체를 분리해 내는 것을 목적으로 한다. Bezdek이 다음과 같은 목적함수 $J_m, 1 < m < \infty$, 을 정의하여 처음으로 퍼지 ISO-DATA를 일반화시키고, 목적함수 J_m 에 대한 수렴을 증명하였다^[14].

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ki})^m (d_i(x_k))^2 \quad (1)$$

여기서 $d_i(x_k)$ 는 입력 데이터 x_k 와 군집 i 의 중심인 v_i 사이의 거리, μ_{ki} 는 데이터 x_k 가 군집 i 에 속할 멤버십 값, m 은 가중치 지수, 그리고 n 과 c 는 각각 입력 데이터와 군집의 수를 나타내고 있다. 참고로 FcM에서는 거리척도로 Euclidian 거리를 사용한다.

Bezdek은 식 (1)의 목적함수를 최소화시키는 조건으로 다음의 두 식을 유도하였다^[14-15]:

$$\mu_{ki} = \frac{1}{\sum_{j=1}^c \left(\frac{d_i(x_k)}{d_j(x_k)} \right)^{\frac{2}{m-1}}} \quad (2)$$

$$v_i = \frac{\sum_{k=1}^n (\mu_{ki})^m x_k}{\sum_{k=1}^n (\mu_{ki})^m} \quad (3)$$

FcM은 식 (2)와 (3)을 한번 씩 교대로 반복적으로 사용하여 최적의 군집중심을 찾는다.

$$\mu_{ik} = \frac{1}{\sum_{j=1}^C \left(\frac{1 - K(x_k, v_j)}{1 - K(x_k, v_i)} \right)^{\frac{1}{m-1}}} \quad (8)$$

III. 분산커널 기반의 퍼지 c-평균 알고리즘

2.2 커널 기반의 퍼지 c-평균 알고리즘

FcM이 여러 가지 응용에서 성공적으로 사용되고 있지만, 입력 데이터에 존재하는 군집들 사이의 경계가 비선형적으로 복잡할 경우에는 좋은 성능을 보여주지 못하는 단점이 있다. 이를 극복하기 위한 대안으로 Mercer 정리를 이용하여 일정 차원의 입력데이터를 더 높은 차원을 가지는 특징 공간으로 맵핑시켜, 원래 입력데이터의 공간에서는 복잡한 비선형 분류의 문제를 선형 분류의 문제로 단순화 시키는 커널 방식이 있어 왔다^[8]. 커널 방식을 이용하기 위해, 먼저 맵핑 함수 Φ 를 이용하는 목적함수를 다음과 같이 표현한다:

$$F_m = \sum_{i=1}^C \sum_{k=1}^N \mu_{ik}^m \|\Phi(x_k) - \Phi(v_i)\| \quad (4)$$

커널을 적용시키면 식 (4)는 다음과 같이 다시 정리될 수 있다:

$$F_m = 2 \sum_{i=1}^C \sum_{k=1}^N \mu_{ik}^m (1 - K(x_i, v_k)) \quad (5)$$

여기서 $K(x, y)$ 는 특징공간에서 두 벡터 x 와 y 의 dot product 를 계산하는 커널 함수이다.

한편, 두 벡터의 커널을 계산하기 위하여, 보통 다음의 가우시안 커널이 널리 이용된다:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right) \quad (6)$$

이제 위의 식(5)를 최소화 하는 v 를 구하기 위하여 Lagrange multiplier를 사용하면, 다음의 식을 얻는다:

$$v_i = \frac{\sum_{k=1}^N \mu_{ik}^m K(x_k, v_i) x_k}{\sum_{k=1}^N \mu_{ik}^m K(x_k, v_i)} \quad (7)$$

그리고 멤버쉽 값은 다음과 같이 얻어진다:

대부분의 전통적인 커널기반의 군집화 알고리즘들은 확정적인 데이터에 적용하기 위해 고안되었지 때문에 확률분포를 가지는 데이터를 이용하는 문제에 적용하는 데는 적절하지 않을 수 있다. 본 논문에서는 오디오 데이터로부터 추출되는 데이터가 확률분포를 갖기 때문에 전통적인 커널기반의 군집화 알고리즘들을 사용할 수 없으므로, 이를 극복하기 위한 방안으로 오디오 데이터로부터 추출된 두 확률분포 데이터 사이의 거리를 측정하기 위하여 분산거리를 이용하는 방법에 기반한 커널을 이용하는 알고리즘을 제안한다. 즉, GPDF 데이터의 군집화에 대한 문제에서는 각 군집의 대표(protoytype)는 확정적 데이터로 표현되기 보다는 평균 벡터와 공분산 매트릭스로 표현되는 것이 더욱 적절하며, 두 개의 GPDF 데이터 사이의 커널을 계산하기 위하여서는 분산 기반의 커널이 사용되는 것이 유리하다. 분산 기반의 커널은 가우시안 커널의 확장형인데, 일반적으로 가우시안 커널은 식 (6)과 같이 두 확정적 벡터사이의 가중된 Euclidean 거리를 사용하지만, 분산 기반의 커널은 다음의 식 (9)에서와 같이 분산거리를 사용한다:

$$DK(g_x, g_y) = \exp(-\alpha D(g_x, g_y) + \beta) \quad (9)$$

여기서 $DK(g_x, g_y)$ 는 두 가우시안 분포 데이터인 g_x 와 g_y 사이의 분산거리이고, α 와 β 는 임의의 상수이다. $DK(g_x, g_y)$ 와 유사한 거리척도가 이미 SVM 문제에 성공적으로 사용되었다.

몇 가지의 분산거리척도를 고려하여 평가한 결과, Bhattacharayya 거리척도를 선정하였는데, Bhattacharayya 거리척도를 사용하여 두 GPDF 데이터의 유사도 척도를 다음과 같이 정의 한다:

$$D(G_i, G_j) = \frac{1}{8} (\mu_i - \mu_j)^T \left[\frac{\Sigma_i + \Sigma_j}{2} \right]^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \frac{\left| \frac{\Sigma_i + \Sigma_j}{2} \right|}{\sqrt{|\Sigma_i| |\Sigma_j|}} \quad (10)$$

여기서 μ_i 와 Σ_i 는 가우시안 분포 데이터 G_i 의 평균 벡터와 공분산 매트릭스를 각각 나타낸다.

일반적인 커널기반 FcM에서 식 (7)의 군집 대표의 갱신식을 유도하기 위하여 Lagrange multiplier 기법을 사용하였는데, 분산커널 기반의 FcM에서도 Lagrange multiplier 기법을 사용하여 다음과 같이 유도된다:

$$m_{v_i} = \frac{\sum_{k=1}^N \mu_{ik}^m DK(x_k, v_i) m_{x_k}}{\sum_{k=1}^N \mu_{ik}^m DK(x_k, v_i)} \quad (11)$$

$$\Sigma_{v_i} = \frac{\sum_{k=1}^N \mu_{ik}^m DK(x_k, v_i) \Sigma_{x_k}}{\sum_{k=1}^N \mu_{ik}^m DK(x_k, v_i)} \quad (12)$$

여기서 m_{v_i} 와 m_{x_k} 는 각각 군집 대표인 v_i 와 입력벡터 x_k 의 평균을 나타내며, Σ_{v_i} 와 Σ_{x_k} 는 각각 v_i 와 x_k 의 공분산을 나타내고, $DK(x_k, v_i)$ 는 두 가우시안 분포 x_k 와 v_i 사이의 분산 기반 커널 함수를 나타낸다.

또한, 군집 대표 값의 갱신식은 식 (8)을 유도하는 같은 방법으로 다음과 같이 얻어진다^[12]:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{1 - DK(x_k, v_i)}{1 - DK(x_k, v_j)} \right)^{\frac{1}{m-1}}} \quad (13)$$

FcM과 분산 기반의 커널방식을 결합하여 개발된 분산

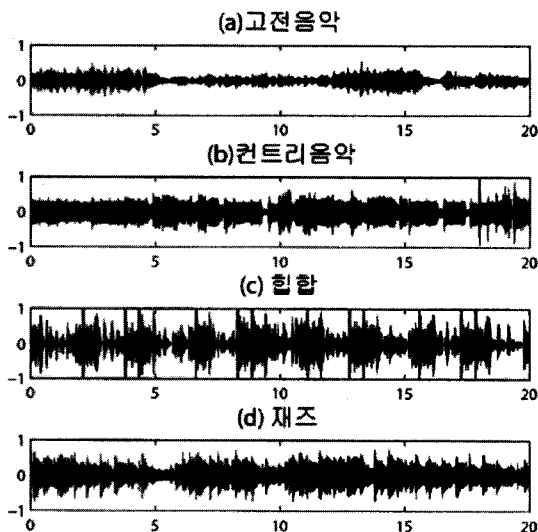


그림 1. 각 장르의 음악데이터의 예
Fig. 1. Example of music data for different genres.

커널기반의 FcM은 GPDF 데이터의 군집화를 수행함에 있어, 퍼지 군집화 기법과 커널 방식의 장점들을 유용하게 결합한 알고리즘이다.

IV. 실험 및 결과

본 논문에서 제안하는 분산커널 기반의 FcM을 이용한 음악 데이터의 장르 분류기에 대한 유용성을 평가하기 위하여, 오디오 CD와 웹 사이트등에서 1200개의 음악 신호를 수집하였는데, 그림 1은 이들 데이터의 20초 분량의 예이다. 데이터는 4개의 음악 장르 데이터인데, 고전음악, 컨트리음악, 힙합, 재즈로 각각 300개의 데이터로 이루어져있다. 각 음악 신호 데이터는 30초 길이의 발췌곡으로, 전체 음악 신호 데이터는 10시간 분량이다. 수집된 음악 데이터는 22,050Hz, 16비트, 모노 오디오 파일로 변환되어 특징 검출에 사용되었다.

각 장르의 데이터 300개는 다시 250개의 무작위로 추출된 학습데이터 부분과 나머지 50개의 검증데이터 부분으로 나뉘어 사용된다. 실험은 50개의 서로 다른 학습데이터/검증데이터 쌍으로 진행되었다. 즉, 무작위로 추출된 250개의 학습데이터와 나머지 50개의 검증데이터 쌍을 만들어 실험을 진행하고, 다시 무작위로 250개의 학습데이터를 추출하고, 나머지 50개의 검증데이터 쌍에 대해 실험을 진행하는 방법으로 50번의 실험을 진행하였다. 본 논문에서 보고되는 각 알고리즘의 정확도는 이러한 50번의 실험의 결과에서 보인 각각의 정확도에 대한 평균과 표준편차로 표현된다.

각 데이터의 특징추출의 과정에서 프레임 크기는 23ms (512 샘플)과 43 프레임의 윈도우를 사용하였으며, 각 발췌곡은 한 개의 특징벡터를 추출하는데 사용하였다. 특징 추출을 위한 도구로는 Tzanetakis와 Cook에 의해 개발된 Marsyas가 사용되었는데^[6], 추출된 특징은 9개의 FFT(Fast Fourier Transform) 계수와 10개의 MFCC (Mel-Frequency Cepstrum Coefficient)로 이루어져 있어서, 총 19차원의 평균벡터와 19x19 공분산

표 1. 여러 모델의 평균정확도(표준편차) (%):
(코드벡터의 수= 7,8,9)

Table 1. Average classification accuracy of different algorithms (%): (7,8,9 code vectors)

	고전	컨트리	힙합	재즈	평균 정확도
SOM	93.1	35.2	81.6	62.9	68.21(7.82)
FcM	91.9	44.3	72.7	76.4	71.32(8.02)
분산커널FcM	94.8	61.5	98.2	97.3	87.94(8.65)

매트릭스가 실험에서 사용되었다. 각 장르를 위해 각 장르에 속하는 특징벡터들을 몇 개의 코드벡터를 이용해 학습의 결과로 군집화되는데, 각 코드벡터는 각기 평균과 공분산 매트릭스를 갖는 한 개의 군집을 표현한다. 학습된 군집화에 결과에 대한 검증의 과정에서 각 음악 데이터의 장르는 Bayesian 분류기를 이용하여 판별되었는데, 이때, Bayesian 분류기는 검증용 음악 데이터가 4가지 장르에 속할 확률을 계산하여, 확률적으로 가장 큰 장르를 결정한다. 즉,

$$Genre(x) = \arg \max_i P(x|v_i) \tag{14}$$

$$P(x|v_i) = \sum_{i=1}^m c_i \mathfrak{N}(x, \mu_i, \Sigma_i) \tag{15}$$

$$\mathfrak{N}(x, \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} e^{-0.5(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)} \tag{16}$$

여기서 M 은 코드벡터의 수, c_i 는 코드벡터 i 의 가중치, d 는 특징벡터의 차원, μ_i 와 Σ_i 는 i 번째 그룹의 평균 벡터와 공분산 매트릭스를 각각 나타낸다.

실험에서 사용된 계수들은 다음과 같다:

- 식 (9)에서 $\alpha = 2.0, \beta = 1.0$
- 식(11)-(13)에서 $m = 2.0$
- 식 (16)에서 $d = 19$

실험에서는 식 (9)의 분산거리식에서 $0.5 \leq \alpha \leq 5.0, -2 \leq \beta \leq 2.0$ 의 범위의 값들에 대해 조사해 보았으나, 어떤 유의한 결론을 내릴 수 없는 미세한 성능의 차이만

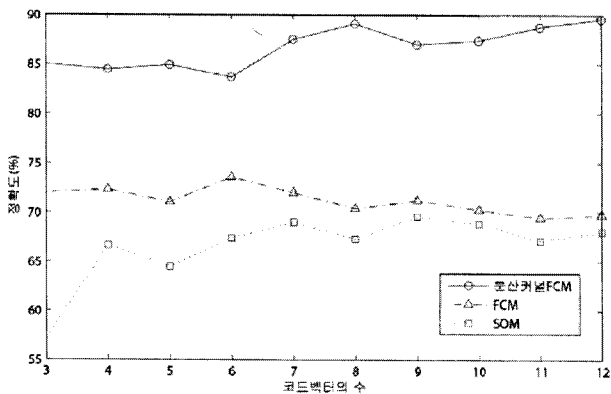


그림 2. 코드벡터 수에 따른 각 알고리즘의 분류 정확도

Fig. 2. Classification accuracy of different algorithms with various numbers of code vectors.

을 보였음을 밝혀둔다.

제안된 방법의 성능을 평가하기 위하여, 코드벡터의 수를 3개-12개로 변화시켜 가면서 SOM (Self-Organizing Map), FcM의 전통적 방법과 비교하였다. 코드벡터의 수가 7개 이상일 경우 안정된 결과 값을 보이는데, 이들 중 코드벡터의 의 수를 7,8,9 개로 한 경우에 대한 평균 정확도는 SOM, FcM, 분산커널 FcM에 대해 각각 68.21%, 71.32%, 87.94%가 관측되었다. 그림 2는 코드벡터의 수에 따른 각 알고리즘의 분류 정확도를 나타내며, 표 1은 코드벡터의 수가 7,8,9개일 경우에 각 알고리즘의 각 장르에 대한 평균 분류 정확도를 나타낸다. 실험에서 SOM과 FcM은 알고리즘의 특성상 19차원 데이터에서 공분산 정보를 사용하지 못하고, 평균에 관한 정보만을 이용할 수 밖에 없었다. 표 1의 실험 결과에서, 제안된 분산커널 FcM가 SOM과 FcM에 비해 각각 19.73%, 16.62%의 성능 향상을 보이고 있는데, 이는 Euclidean 거리척도를 사용함으로 인해 데이터의 공분산 정보를 사용할 수 없는 SOM 이나 FcM에 비해, 분산커널 FcM은 공분산 정보를 이용하는 것에 기인하는 것으로 추정된다. 표 1에서 보듯이, 분산커널 FcM은 재즈의 경우에는 거의 완벽한 분류 정확도를 보이고 있지만, 컨트리 음악의 경우에는 가장 나쁜 결과를 보이는 분산커널 FcM이 61.5%, SOM이나 FcM은 약 35%-45%의 매우 부정확한 분류 결과를 보인다. 이 결과에 대한 이유는 분류의 혼동표(Confusion table)인 표 2로 유추할 수 있다. 즉, 표 2는 분산커널 FcM이 7-9개의 코드벡터에 대한 분류 결과에 대한 분석을 기초로 작성된 것으로, 분류의 혼동표를 나타내는데, “컨트리 음악”을 “고전음악”으로 혼동하는 경우가 매우 많음을 보이고 있다. 이러한 경향은 SOM이나 FcM에서도 역시 나타난다. 반면에, 분산커널 FcM은 “고전음악”, “재

표 2. 제안된 분류모델에 대한 혼동 테이블 (코드벡터의 수 7,8,9 개 결과 평균(표준편차))

Table 2. Confusion table for the proposed model with 7,8,9 code vectors: average(standard deviation)

		분류결과			
		고전	컨트리	힙합	재즈
입력	고전	94.81 (8.29)	0.27 (0.19)	0.31 (0.97)	4.61 (3.98)
	컨트리	37.4 (6.91)	61.52 (4.51)	0.11 (1.43)	0.97 (1.89)
	힙합	0.00 (0.00)	0.28 (1.21)	98.2 (8.35)	1.52 (2.57)
	재즈	1.48 (2.92)	1.21 (0.88)	0.01 (0.32)	97.3 (1.24)

즈”, “힙합”의 경우에는 매우 높은 분류 정확도를 보이는데, 이는 “고전음악”, “재즈”, “힙합”은 서로 매우 다른 음악적/데이터적 특성을 보이지만, “컨트리 음악”과 “고전음악”이 어느 면에서는 매우 유사한 음악적/데이터적 특성을 나타내고 있기 때문인 것으로 유추된다. 유사한 결과와 좀 더 자세한 분석이 보고되고 있다^[17]. 차후의 연구에서는 “컨트리 음악”과 “고전음악”을 분리해 내는데 유용한 특징의 추출에 대한 노력을 통해, 이들 두 장르 사이의 분류 정확도를 높이는 결과를 도출해야 할 것이다.

V. 결 론

오디오 신호 데이터의 특징벡터에서 평균과 공분산 정보를 동시에 이용하여 효율적인 음악장르 분류를 수행하는 방안이 본 논문에서 제안 되었다. 분산 거리 척도와 커널 방식을 FcM에 적용하며, 오디오 데이터에서 특징을 추출하는 방법으로는 Tzanetakis와 Cook에 의해 개발된 Marsyas가 사용되었는데, 9차원의 FFT 계수와 10차원 MFCC를 사용한다. 제안된 방법의 성능을 평가하기 위하여, 고전음악, 컨트리음악, 힙합, 재즈의 4개의 장르 음악데이터를 각각 300개, 총 1200개를 수집하여 실험을 진행하였다. 코드벡터의 수를 3개 -10개로 변화시켜 가면서 SOM, FcM의 전통적 방법과 비교하였는데, 제안된 분산커널 FcM가 SOM과 FcM에 비해 각각 19.73%, 16.62%성능 향상을 보였다. 이는 Euclidean 거리척도를 사용함으로 인해 데이터의 공분산 정보를 사용할 수 없는 SOM 이나 FcM에 비해, 분산커널 FcM은 공분산 정보를 이용하는 것에 기인하는 것으로 추정된다. 분산커널 FcM은 “고전음악”, “재즈”, “힙합”의 경우에는 매우 높은 분류 정확도를 보이며, “컨트리 음악”과 “고전음악”의 분류에서는 정확도가 상대적으로 낮는데, 이는 “고전음악”, “재즈”, “힙합”은 서로 매우 다른 음악적/데이터적 특성을 보이지만, “컨트리 음악”과 “고전음악”이 어느 면에서는 매우 유사한 음악적/데이터적 특성을 나타내고 있기 때문인 것으로 유추된다. 이러한 성향은 SOM과 FcM에서도 나타나고 있음으로 보아, 알고리즘의 문제라기보다는 FFT 계수와 MFCC로 구성되는 특징벡터의 문제라고 추정된다. 따라서, 앞으로의 연구에서는 FFT계수와 MFCC이외에 더욱 유용한 특징의 추출에 초점이 모아져야 할 것이다.

참 고 문 헌

- [1] E. Wold, et al., “Content-based classification, search, and retrieval of audio”, *IEEE Tr. Multimedia*, V.3, No.3, pp 27-36. 1996.
- [2] J. Saunders, “Real time discrimination of broadcast speech/music”, *Proc. of ICASSP*, pp. 993-996, 1996
- [3] J. Foote, “Content-based retrieval of music and audio”, *Proc. SPIE, MSAS*, pp 138-147. 1997.
- [4] 정성운, 김민성, 손종목, 배건성, “PCA-optimized 필터뱅크 기반의 MFCC 특징파라미터 추출 및 한국어 4연숫자 전화음성에 대한 인식실험”, *전자공학회논문지*, 제41권 SP편, 6호, 279-2839쪽, 2004
- [5] G. Li, and A. Khokar, “Content-based indexing and retrieval of audio data using wavelets”, *Proc ICME*, pp 885-888, 2000.
- [6] G. Tzanetakis, and P. Cook, “Musical genre classification of audio signals”, *IEEE Trans. Speech Audio Process.*, V. 10, No.5, pp. 293-302, 2002.
- [7] D. Turnbull, C. Elkan, “Fast Recognition of Musical Genres Using RBF Networks”, *IEEE Trans. Knowl. Data Eng.*, V. 17, No. 4, pp. 580-584, 2005.
- [8] K. Muller, et. al, “An Introduction to Kernel-Based Learning Algorithms”, *IEEE Tr. Neural Networks*, V. 12, No. 2, pp.181-201, 2001.
- [9] N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machine*, Cambridge Univ. Press, Cambridge, 2000.
- [10] J. Chen, C. Chen, “Fuzzy Kernel Perceptron”, *IEEE Tr. Neural Networks*, V. 13, No. 6, pp. 1364-1373, 2002.
- [11] S. Chen, D. Zhang, “Robust Image Segmentation using FCM with Spatial Constraints Based on New Kernel-Induced Distance Measure”, *IEEE Tr. SMC*, V. 34, No.4, pp.1907-1916, 2004.
- [12] Y.-S. Song, et al, “Fuzzy C-Means Algorithm with Divergence-Based Kernel”, *LNAI*, V. 4223, pp. 99-108, 2006.
- [13] D.-C. Park, et al., “Classification of Audio Signals Using a Bhattacharyya Kernel-Based Centroid Neural Network”, (Accepted for PAKDD 2009).
- [14] J. Bezdek, “A convergence theorem for the fuzzy ISODATA clustering algorithms”, *IEEE Tr. PAMI*, V. 2. pp 1-8. 1980.
- [15] R. Hathaway, J. Bezdek, Y. Hu, “Generalized fuzzy c-means clustering strategies using Lp norm distances”, *IEEE Tr. Fuzzy Systems*, V.8,

No. 5, pp. 576-582, 2000.

- [16] 박현욱, 오준택, 손영호, 김욱현, “엔트로피 기반의 가중치와 분포크기를 이용한 향상된 FCM 알고리즘,” *전자공학회논문지*, 제43권 SP편, 제4호, 1-8 쪽, 2007.
- [17] D.-C. Park, “Classification of audio signals using Fuzzy C-means with divergence-based Kernel”, *Pattern Recognition Letters*, (accepted for publication)

— 저 자 소 개 —



박 동 철(정회원)
 1980년 서강대학교 전자공학과
 학사 졸업.
 1982년 한국과학기술원 전기 및
 전자공학과 석사 졸업.
 1990년 Univ. of Washington,
 Seattle, Dept. of Electrical
 Eng. 박사 졸업.

2009년 현재 명지대학교 정보공학과 교수.

<주관심분야 : 지능컴퓨팅, 신호처리>