

논문 2010-47CI-3-5

진보된 다단계 특징벡터 기반의 분류기 모델

(Advanced Multistage Feature-based Classification Model)

김재영*, 박동철**

(Jae-Young Kim and Dong-Chul Park)

요약

본 논문에서는 다단계 특징벡터 기반의 분류기 모델(Multistage Feature-based Classification Model: MFCM)의 성능을 향상시킨 진보된 형태의 MFCM (Advanced MFCM: AMFCM)을 제안하는데, AMFCM은 MFCM과 같이 주어진 데이터에서 추출된 전체의 특징벡터를 연결하여 이용하지 않고, 같은 성질의 특징벡터들끼리 모아서, 각각의 국지적 학습기를 통하여 분류에 이용한다. 그러나, AMFCM은 MFCM에서 사용되는 각각의 국지적 분류기를 위한 각 특징벡터의 분류기여도를 더욱 섬세하게 조정하여 최종적인 분류의 정확도를 높이는 방안을 제안한다. 제안된 AMFCM의 성능을 검증하기 위하여, 음악장르 분류의 문제에 대한 실험을 진행하였다. 또한, 국지적 분류기로 Self-Organizing Map과 중심 신경망을 사용하여 실험을 수행하였는데, 제안된 AMFCM은 사용된 국지적 분류기의 종류와 사용된 군집의 개수에 따라 기존의 MFCM에 비해 평균 8% - 15% 이상의 성능향상을 보여 준다.

Abstract

An advanced form of Multistage Feature-based Classification Model(AMFCM), called AMFCM, is proposed in this paper. AMFCM like MFCM does not use the concatenated form of available feature vectors extracted from original data to classify each data, but uses only groups related to each feature vector to classify separately. The proposed AMFCM improves the contribution rate used in MFCM and proposes a confusion table for each local classifier using a specific feature vector group. The confusion table for each local classifier contains accuracy information of each local classifier on each class of data. The proposed AMFCM is applied to the problem of music genre classification on a set of music data. The results demonstrate that the proposed AMFCM outperforms MFCM by 8% - 15% on average in terms of classification accuracy depending on the grouping algorithms used for local classifiers and the number of clusters.

Keywords : Multistage feature, Classification model, Pattern classification, Music genre classification

I. 서론

수집된 데이터로부터 각 특징들을 추출하여, 그 데이터를 대표할 수 있는 특징벡터를 이용하는 방법은 방대한 데이터를 효율적으로 관리하며, 연산량을 줄이는 방

법으로 패턴인식 분야에서 사용된 가장 기본적인 데이터처리 방법의 하나이다. 음성 데이터에 대해, 각 데이터의 고유한 특성을 반영하기 위한 특징벡터를 추출하는 많은 연구가 진행되어 왔다^[1~8]. 음악 데이터에서의 특성추출을 위하여 Music Fish Group^[3]은 소리의 세기, 높이, 밝기, 대역폭, 화성과 같은 특성들을 음악 데이터에서 추출하도록 제안하였다. 또한, Foote^[4]는 MFCC (Mel-scaled Frequency Cepstral Coefficients)를 추출하는 방법을 제안하였고^[5~6], Li^[7]는 이산 웨이블렛 변환을 사용하여 특성을 추출하였으며, Tzanetakis와 Cook^[8]은 음색의 성질, 리듬의 성질, 음의 높낮이 정보와 같은 특성들을 추출하여 이용하였다.

여러 방법을 이용하여 추출된 특징벡터들은 각 성분

* 정희원, 명지대학교 정보공학과
(Dept. of Information Eng., Myong Ji University)

** 정희원, 명지대학교 전자공학과
(Dept. of Electronics Eng., Myong Ji University)

※ 본 연구는 2010년도 명지대학교 교책중점연구소 지원 및 2010년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2010-0009655).

접수일자: 2010년4월5일, 수정완료일: 2010년4월30일

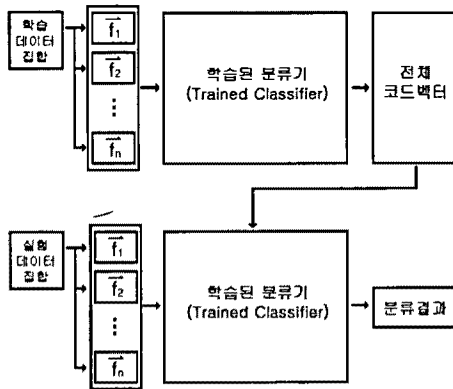


그림 1. 기존의 분류 모델
Fig. 1. Conventional Classification Model.

이 서로 다른 특성과 크기를 가지므로 각각의 특징벡터의 성분을 같은 공간상에서 구별 없이 분류기의 입력벡터로 사용하는 것은 효과적이지 않을 수도 있다. 이를 보완하기 위한 여러 가지 방안들이 강구되고 있는데, 그 중 정규화 방법은 서로 다른 특징벡터의 성분들을 정규화 된 크기로 만들어 이용하는 방법이다. 하지만, 각 특성 성분이 여전히 같은 공간상에 구별 없이 사용되고 있으므로, 제한된 정도의 성능향상만을 기대할 수 있다. 이 문제를 해결하기 위하여 여러 가지 방법들을 통해 추출된 다양한 특성요소들을 이용하면서, 각 특성요소를 효과적으로 이용하여, 분류기에 이용하는 분류모델인 MFCM (Multistage Feature-based Classification Model)^[2]이 제안되었는데, 다양한 특징벡터를 효율적으로 분류기에 이용할 수 있는 방안으로 기존의 방안에 비해 인식성능의 향상을 보일 수 있었다.

즉, MFCM은 전체의 특징요소들을 같은 의미를 가지는 요소들의 몇 개의 그룹으로 나누고, 각각의 그룹에 속하는 특징요소들을 이용해서 여러 개의 특징벡터를 구성한다. 구성된 특징벡터들은 독립적으로 서로 다른 공간상에서 국지적 분류기에 이용되므로, 각 특징벡터의 고유의 특성들을 최대한 보장받게 된다.

본 논문의 II장에서는 제안된 분류모델의 배경이 된 MFCM에 대해 살펴보고, III장에서는 기존 방법의 문제점을 개선하여 제안된 AMFCM에 대해 살펴본다. IV장에서는 실험에서 국지적 분류기로 사용될 Self-Organizing Map^[9]과 중심신경망 알고리즘^[10~11]중 중심신경망에 대해 간단히 알아보고, V장에서는 실험을 통해 제안된 분류모델의 성능을 살펴보고, VI장에서는 결론을 내린다.

II. 다단계 특징벡터 기반의 분류기 모델

MFCM^[2]은 그림 2 에서 주어진듯이 각 특성의 학습데이터(특징벡터)를 국지적 학습기에서 각각 학습시키고, 학습데이터에 대해 각 특징벡터에 의한 국지적 정확도를 산출하여 기여도를 계산한다. 즉, 주어진 데이터 \vec{x} 에 대하여 서로 다른 N 개의 특징추출기에 의해 추출한 특징벡터 \vec{f} 는 같은 특성을 가지는 N개의 소그룹 특징벡터 \vec{f}_i ($i = 1, 2, \dots, N$)로 나누어진다. 즉,

$$\vec{f} = [\vec{f}_1^T : \vec{f}_2^T : \dots : \vec{f}_N^T]^T \quad (1)$$

학습과정에서는 각각의 특징벡터 \vec{f}_i ($i = 1, 2, \dots, N$)를 이용해서 N 개의 국지적 분류기 C_i 를 독립적으로 구성하게 된다. 또한, 각각의 국지적 분류기 C_i 에 학습에 사용하지 않은 검증데이터를 이용하여, 국지적 분류기 C_i 의 분류정확도 α_i 를 구하며, N 개의 국지적 분류기의 전체 분류기에 대한 분류기여도 ω_i 를 다음의 식에 의하여 정의한다.

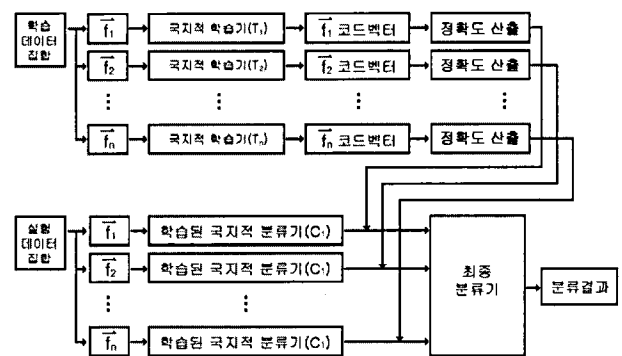


그림 2. 다단계 특징벡터 기반의 분류기 모델^[2]
Fig. 2. Multistage Feature-based Classification Model.

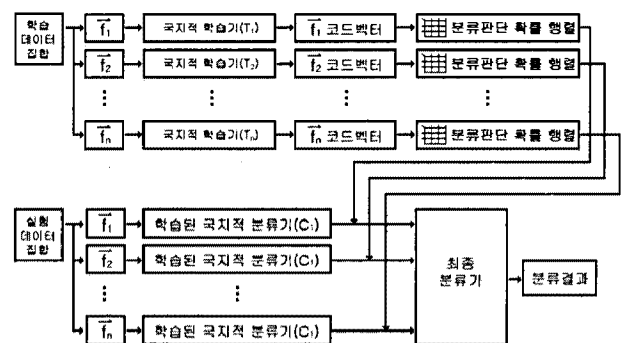


그림 3. 진보된 다단계 특징벡터 기반의 분류기 모델
Fig. 3. Advanced Multistage Feature-based Classification Model.

$$\omega_i = \alpha_i / \sum_{j=1}^N \alpha_j \quad (2)$$

주어진 데이터 \vec{x} 에 대해, 학습된 국지적 분류기, $\{C_1, C_2, \dots, C_N\}$, 각각의 학습된 분류기의 기여도, $\{\omega_1, \omega_2, \dots, \omega_N\}$,가 주어져 있을 때, 데이터 \vec{x} 를 M 개의 클래스 중 하나의 클래스에 속하는 분류의 문제가 있다고 가정한다. 주어진 데이터 \vec{x} 에서 추출한 \vec{f}_i ($i = 1, 2, \dots, N$)가 대응하는 국지적 분류기 C_i ($i = 1, 2, \dots, N$)를 통과하면, 데이터 \vec{x} 와 M 개의 군집과의 거리는 다음의 식으로 표현된다.

$$\{d_{i1}(\vec{x}), d_{i2}(\vec{x}), \dots, d_{iM}(\vec{x})\}, i = 1, 2, \dots, N \quad (3)$$

한편, 각각의 국지적 분류기의 기여도를 고려하여, \vec{x} 와 각 군집과의 거리를 다음의 식으로 구한다.

$$d_i(\vec{x}) = \sum_{j=1}^N \frac{d_{ji}(\vec{x})}{\omega_j}, i = 1, 2, \dots, M \quad (4)$$

여기서 $d_{ji}(\vec{x})$ 는 국지적 분류기 j 에서 \vec{x} 와 군집 i 사이의 거리를 나타내며, ω_j 는 국지적 분류기 j 의 기여도를 나타낸다. 따라서, \vec{x} 에 대한 분류 결과, $Class(\vec{x})$ 는 다음의 식으로 정해진다.

$$Class(\vec{x}) = \underset{i}{\operatorname{argmin}} d_i(\vec{x}), \forall i = 1, 2, \dots, M \quad (5)$$

III. 진보된 다단계 특징벡터 기반의 분류기 모델

MFCM에서 국지적 분류기, $\{C_1, C_2, \dots, C_N\}$,에 대한 각각의 학습된 분류기의 기여도, $\{\omega_1, \omega_2, \dots, \omega_N\}$ 는 각각의 국지적 분류기 모델의 정확도가 얼마인가를 계산하는 방식이었다. 그런데, 국지적 분류기 $C_i, i = 1, \dots, N$,가 주어진 데이터에 대해서 분류를 $j, j = 1, \dots, N$,로 했을 때, 그 국지적 분류기 $C_i, i = 1, \dots, N$,가 이전에 어떤 분류 판단을 내렸는가를 고려해서 그 판단의 기여도를 측정하는 것이 더욱 정확할 것이다. 즉, 각각의 분류기에 다음과 같은 분류 판단 확률 행렬을 가지고 있다고 하면, 여기서 P_{ij}^k 는 k 번째 국지적 분류기가 Class i 의 데이터를 Class j 로

판단하여 분류했던 확률이다. 다음의 식(6)의 분류판단 확률 행렬은 주어진 학습데이터를 통해 구해진 것이다. 물론 실시간에서 사용할 때, 분류판단 확률 행렬은 실시간으로 갱신된다.

$$\begin{bmatrix} P_{11}^k & P_{12}^k & \dots & P_{1N}^k \\ P_{21}^k & P_{22}^k & \dots & P_{2N}^k \\ \vdots & & \ddots & \\ P_{N1}^k & \dots & P_{NN}^k \end{bmatrix}, \quad (6)$$

$$\sum_{j=1}^N P_{ij}^k = 1$$

이 경우, 기존의 다단계 분류기는 k 번째 국지적 분류기의 분류정확도를 다음의 형태로 구한 것이다. 즉,

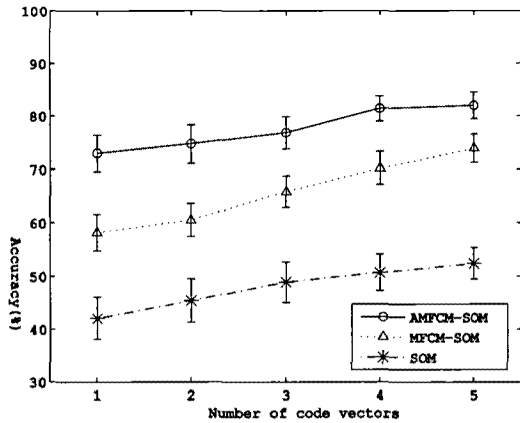
$$\alpha_k = \frac{\sum_{i=1}^N P_{ii}^k}{\sum_{i=1}^N \sum_{j=1}^N P_{ij}^k} \quad (7)$$

본 논문에서는 위의 분류판단 확률 행렬을 최적으로 이용하여, 각 국지적 분류기의 판단 결과에 대한 최적의 분류정확도를 구하여 이용하는 분류기를 제안한다. 이는 각각의 국지적 분류기가 과거에 어떤 판단을 내렸었고, 어느 클래스의 데이터에서는 특별히 좋은 성능을 보이고, 어느 클래스의 데이터에는 매우 좋지 않은 성능을 보였는가에 대한 정보를 이용하는 방법이다. 이는 매우 자연스런 현상으로 각각의 특징 추출기에서 추출된 특징벡터가 각 클래스를 표현하는 방법이 다를 수 있을 때, 각 클래스마다의 분류성능이 다르다는 것을 이용하는 것이다.

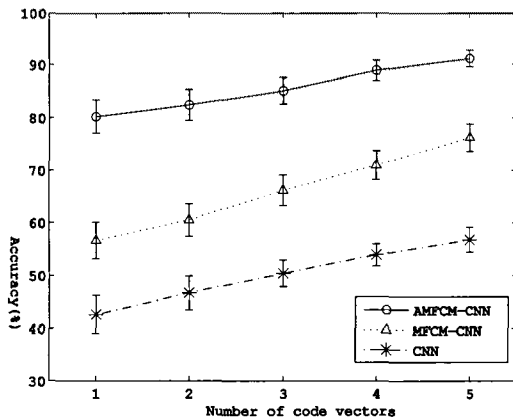
IV. 실험에 사용된 군집화 알고리즘

현재까지 많은 군집화 알고리즘이 개발 되어 왔으며 많은 분야에 쓰이고 있다^[7~11]. 그중에 Self-Organization Map (SOM)^[9]과 Centroid Neural Network (CNN)^[10]을 본 실험에서 사용하여, 제안된 AMFCM의 성능을 검증하고자 하는데, SOM은 수많은 논문에서 소개 되었으며, 여기서는 CNN 알고리즘^[10]에 대해서 간단히 살펴본다.

중심 신경망(CNN)은 k-means와 SOM 같은 기존 알고리즘이 매 학습마다 연결강도(weight)를 수정하는 것



(a) SOM 기반 분류기



(b) CNN 기반 분류기

그림 4. 코드벡터 수에 따른 각 알고리즘의 분류정확도
Fig. 4. Classification accuracy of different algorithms with various numbers of code vectors.

과 달리, 승자와 패자의 연결강도를 각각 설정하여 현재와 이전 학습에서 뉴런의 상태에 대한 정보를 이용하여, 가중치를 변화시킨다. 즉, 이전 학습에서 승자로 선택되지 않았지만, 현재의 학습에서 입력데이터에 가장 가까운 뉴런을 승자라 하며, 반대로 이전 학습에서 승자였지만 현재의 학습에서 입력데이터에 가장 가깝지 않다면 현재 학습에서는 패자라고 한다. 이때 각 뉴런이 승자 또는 패자의 상태가 변할 경우에만 연결강도를 수정한다.

중심 신경망은 k-means 알고리즘과 자기 조직화 지도 같은 기존 경쟁 알고리즘 등과 비교하여 몇 가지 장점을 가지고 있다. 중심 신경망은 사전에 학습 계수나 전체 학습 반복수(epoch) 등 동작을 위한 변수를 설정할 필요가 없으며, 더불어 기존의 군집화 알고리즘에서

표 1. CNN 기반의 여러 분류모델의 분류 정확도 (%)
Table 1. Classification accuracy of different algorithms with CNN(%).

	CNN	MFCM	AMFCM
클래식	98	99	99
컨츄리	99	91	89
힙합	84	73	91
재즈	11	73	95
블루스	92	90	99
메탈	40	60	91
네추럴	66	100	98
포크	21	79	99
오페라	12	60	86
트로트	41	40	65
평균	56.4	76.5	91.2

문제점으로 지적되었던 초기 연결강도 설정의 문제까지 효과적으로 해결함으로써 자동적으로 준최적의 군집화 결과를 안정적으로 도출하는 장점이 있다.

V. 실험 및 결과

우리는 제안된 분류 모델을 이용하여 클래식, 컨트리, 힙합, 재즈, 블루스, 메탈, 네추럴, 포크, 오페라, 트로트의 10가지 음악장르를 분류하기 위한 실험을 수행하였다. 음악 데이터베이스는 오디오 CD나 인터넷 사이트에서 다양하게 수집되었으며, 각각의 데이터들은 다른 포맷과 다른 주파수로 구성 되어 있으므로 같은 조건에서 특징벡터를 얻기 위하여 22,050 Hz, 16 bit mono 파일들로 동일하게 변환하였다. 실험을 위해서 10개 장르의 음악을 각 30초의 길이로 각 장르 당 100개씩, 총 1,000개 데이터를 수집하였으며, 수집된 데이터의 총 시간은 500분이다.

데이터에서 특징벡터들을 추출하기 위하여 Tzanetakis와 Cook^[6]이 개발한 공개 소스 코드인 Marsyas 프로그램을 사용 하였다. 하나의 음악 데이터에서 30차원의 Timbral texture, 17차원의 Pitch contents, 9차원의 FFT 계수, 10차원의 MFCC 계수, 24차원의 DWT 계수의 총 90차원 특징벡터를 추출했다. 또한 제안된 분류모델을 비교하기 위하여 여러가지 특성 전부를 이용하여 일렬로 연결한 특징벡터를 학습하는 기존 방법(90차원)과 이 논문에서 제안한 각 특성들을 각각 학습시키는 분류모델을 서로 비교 분석했다.

이 실험에 사용된 음악 데이터베이스의 경우 각 장르

표 2. AMFCM-CNN의 분류판단 확률 행렬의 예 (%)
Table 2. Examples of confusion matrix of AMFCM with CNN (%).

(a) MFCC

	cl	co	hh	jz	bl	mt	nt	fk	op	tr
cl	33	18	0	39	0	0	0	0	0	0
co	0	69	2	18	1	0	0	0	0	0
hh	0	8	39	0	24	0	0	12	2	5
jz	18	20	0	48	0	0	0	0	4	0
bl	0	0	30	0	60	0	0	0	0	0
mt	0	0	0	0	1	80	6	2	0	1
nt	0	0	0	0	0	1	89	0	0	0
fk	0	0	0	0	0	0	0	51	0	39
op	1	0	17	2	0	0	0	2	66	2
tr	0	0	5	0	34	46	0	3	0	2

(b) DWTC

	cl	co	hh	jz	bl	mt	nt	fk	op	tr
cl	84	0	0	6	0	0	0	0	0	0
co	3	32	0	31	0	4	15	1	4	0
hh	0	7	51	2	0	0	0	0	30	0
jz	11	25	0	43	0	0	0	0	11	0
bl	6	0	0	0	27	12	0	26	0	19
mt	1	0	0	0	33	33	0	3	0	20
nt	0	0	0	0	0	0	77	10	0	3
fk	6	0	0	0	22	9	4	20	0	29
op	1	0	0	0	2	0	30	28	0	29
tr	12	0	0	0	0	4	0	25	0	49

주: cl(클래식), co(컨트리), hh(힙합), jz(재즈),
bl(블루스), mt(메탈), nt(네추럴), fk(포크),
op(오페라), tr(트로트)

의 100개 중 무작위로 추출된 90개의 데이터는 학습데이터로 사용하였으며 나머지 10개의 데이터를 실험데이터로 사용하였다. 그림 4는 음악 데이터베이스에서 코드벡터 수의 증가에 따른 각 알고리즘의 정확도를 나타낸다.

각 수치는 10번의 반복 실험의 평균과 표준편차를 나타내고 있다. 그림 4에서와 같이 음악 데이터에서 추출한 다양한 특징벡터를 일렬로 연결한 (concatenated) 후 학습한 기존 분류모델은 각 코드벡터에서 60%의 결과 값을 넘지 못하며, 다양한 특징벡터의 국지적 분류기 정확도를 이용한 분류모델은 77% 이하의 분류 정확도를 보여준다. 반면, 제안된 분류모델은 약 85%의 정확도를 산출해 내며, 특히 제안된 방법에 CNN을 국지적 분류기로 이용한 실험에서 14개의 코드벡터의 경우 91.2%의 뛰어난 결과를 보여준다.

표 1은 코드벡터의 수가 14개 일 때 CNN 방법에서의 최종 정확도를 나타낸다. 제안된 분류모델에서 특히 분류가 잘 되지 않던 메탈, 오페라, 트로트의 장르가 기존 방법보다 월등히 좋은 성능을 보인다. 표 2는 제안된 분류모델에 CNN을 국지적 분류기로 이용했을 때 최종 분류판단 확률 행렬의 예를 보여준다. 표 2에서 보듯이 MFCC 특징벡터에 의한 국지적 분류기는 네추럴, 메탈, 오페라 등에서 상대적으로 신뢰도 있는 분류 결과를 내지만, 클래식, 힙합트로트 등의 음악에서는 분류 성능이 좋지 않음을 알 수 있다. 반면에, DWTC 특징벡터에 의한 국지적 분류기는 클래식, 네추럴, 트로트에서 상대적으로 신뢰도 있는 분류 결과를 내고, 오페라, 포크, 메탈, 블루스의 분야에서는 매우 낮은 분류 성능을 보이고 있다. 따라서 제안된 AMFCM은 이렇게 다른 분야에 각기 다른 성능의 분류 성능을 가지는 각각의 특징벡터에 의한 국지적 분류기의 분류 결과를 최적으로 이용하여 전체적 분류 성능을 극대화한다.

VI. 결 론

본 논문에서는 진보된 형태의 MFCM을 이용한 분류 모델을 제안하였으며, SOM 과 CNN 알고리즘과 기존의 분류모델인 MFCM과 새로이 제안된 분류모델인 AMFCM에 적용하여 음악장르 분류의 문제에 대해 비교 실험하였다. 기존의 분류모델에서는 각 데이터 군에서 추출된 여러 특성들을 하나의 벡터로 하기 때문에 고유의 특성들에 대해 간섭을 일으켜 좋은 결과 값을 가질 수 없었으나, 제안된 분류모델을 사용 하였을 경우에는 각 특성을 최대한 보장하므로 상대적으로 높은 분류 정확도를 얻을 수 있었다. 또한 하나의 데이터에서 추출된 특성들은 그 데이터에 대해 똑같은 영향을 미치는 것이 아니라 각 특징벡터마다 그 데이터를 표현할 수 있는 정도가 다르므로, 제안된 분류모델에서는 학습 데이터의 분류판단 확률 행렬의 산출을 통해 각 특성의 기여도를 측정하여 차등적인 가중치를 주었다. 실험에서는 10가지 음악장르를 분류하기 위해 다양한 특징벡터들과 알고리즘들을 사용하였다. 제안된 분류모델을 이용하여 음악 데이터를 분류한 경우 14개의 코드벡터의 경우, 단순 나열된 특징벡터에 비해 평균적으로 32.3%, MFCM 보다는 10.9%의 개선된 인식 성능의 향상을 얻을 수 있었다. 앞으로 실험에서 사용된 특성들 이외의 다양한 장르의 데이터와 특성들을 수집하여 더

좋은 분류결과 값을 도출 할 수 있을 것이며, 또한 실험에 사용되지 않은 신경망 알고리즘을 국지적 분류기로 적용하여 향상된 결과 값을 얻을 수 있을 것으로 예상된다.

참 고 문 헌

[1] K. Kim, S. Kim, J. Jeon, K. Park, "Quick Audio Retrieval Using Multiple Feature Vectors," IEEE Trans. Consumer Electronics. V. 52, No.1, 2006.

[2] 송영수, 박동철, "다단계 특징벡터 기반의 분류기 모델," 전자공학회논문지, 46권, CI편 1호, 121-127쪽, 2009.

[3] E. Wold, T. Blum, D., Keislar, and J. Wheaton, "Content-based classification, search, and retrieval of audio", IEEE Trans. Multimedia, V.3, No.3, pp 27-36. 1996.

[4] J. Foote, "Content-based retrieval of music and audio", Proc. SPIE, Multimedia Storage and Archiving Systems, pp 138-147. 1997.

[5] 정성윤, 김민성, 손종목, 배건성, "PCA-optimized 필터뱅크 기반의 MFCC 특징과라미터 추출 및 한국어 4연숫자 전화음성에 대한 인식실험," 전자공학회논문지, 제41권 SP편, 6호, 279-2839쪽, 2004년 6월

[6] 박찬웅, "은닉 마코프 모델과 캡스트럴 계수들에 따른 한국어 속삭임의 인식 비교," 전자공학회논문지, 제43권 IE편, 제2호, 22-29쪽, 2006.

[7] G. Li, and A. Khokar, "Content-based indexing and retrieval of audio data using wavelets", Proc. Int. Conf. Multimedia Expo. pp 885-888. 2000.

[8] G. Tzanetakis, and P. Cook, "Musical genre classification of audio signals", IEEE Trans. Speech Audio Process. V. 10, No.5, pp. 293-302, 2002.

[9] T. Kohonen, "The Self-Organizing Map", Proc. IEEE, Vol. 78. pp 1464-1480. 1990.

[10] D. Park, "Centroid Neural Network for Unsupervised Competitive Learning", IEEE Trans. Neural Networks, V.11, pp.520-528, 2000.

[11] D.-C. Park et al., "Centroid Neural Network with a Divergence Measure for GPDF Data Clustering ", IEEE Trans. Neural Networks, V.19, No. 6, pp 948-957. 2008.

저 자 소 개



김 재 영(정회원)
 2008년 명지대학교 정보공학과
 학사 졸업.
 2010년 명지대학교 정보공학과
 석사 졸업 예정.
 <주관심분야 : 신경망, 인공지능,
 신호처리>



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.
 2010년 현재 명지대학교 전자공학과 교수.
 <주관심분야 : 지능컴퓨팅, 신호처리>