

음악 장르 분류를 위한 새로운 자동 Taxonomy 구축 알고리즘

New Automatic Taxonomy Generation Algorithm for the Audio Genre Classification

최택성*, 문선국*, 박영철**, 윤대희*, 이석필***

Tacksung Choi*, Sunkook Moon*, Youngcheol Park**, Daehee Youn*, Seokpil Lee***

*연세대학교 전기전자공학과, **연세대학교 컴퓨터정보통신공학부, ***전자부품연구원(KETI) 디지털미디어 연구센터
(접수일자: 2008년 1월 24일; 채택일자: 2008년 3월 14일)

본 논문에서는 음악 장르 분류를 위한 새로운 자동 Taxonomy 구축 알고리즘을 제안한다. 제안된 알고리즘은 모든 가능한 노드들의 분류 확률을 예측하여 예측된 분류 성능값이 가장 좋은 조합을 Taxonomy로 구축하는 것이다. 제안된 알고리즘에서의 분류 확률 예측은 훈련 데이터를 k-fold cross validation을 이용하여 분류기에 적용함으로써 이루어진다. 제안된 알고리즘을 기반으로한 분류 성능 측정은 2 클래스로 이루어진 각각의 노드에 2개 범주 분류에 효과적인 support vector machine을 적용함으로써 이루어진다. 제안된 알고리즘의 성능 검증을 위해 음색, 리듬, 피치 등 오디오 신호의 특징을 나타내는 다양한 파라미터를 오디오 신호로부터 추출하여 제안된 알고리즘과 기존의 다중 범주 분류기들을 이용하여 분류 성능을 평가하였다. 다양한 실험결과 제안된 알고리즘은 기존의 알고리즘에 비하여 5%에서 25% 정도의 분류 성능이 향상된 것을 확인할 수 있었고 특히 낮은 차원의 특징벡터를 이용한 분류 실험에서는 10%에서 25% 향상된 좋은 성능을 보였다.

핵심용어: 특징벡터 선택 알고리즘, 장르 분류, 계층구조, 분류학, 래퍼 알고리즘

투고분야: 음향 신호처리 분야 (1,2)

In this paper, we propose a new automatic taxonomy generation algorithm for the audio genre classification. The proposed algorithm automatically generates hierarchical taxonomy based on the estimated classification accuracy at all possible nodes. The estimation of classification accuracy in the proposed algorithm is conducted by applying the training data to classifier using k-fold cross validation. Subsequent classification accuracy is then to be tested at every node which consists of two clusters by applying one-versus-one support vector machine. In order to assess the performance of the proposed algorithm, we extracted various features which represent characteristics such as timbre, rhythm, pitch and so on. Then, we investigated classification performance using the proposed algorithm and previous flat classifiers. The classification accuracy reaches to 89 percent with proposed scheme, which is 5 to 25 percent higher than the previous flat classification methods. Using low-dimensional feature vectors, in particular, it is 10 to 25 percent higher than previous algorithms for classification experiments.

Keywords: Feature selection algorithm, Genre classification, Hierarchy, Taxonomy, Wrapper algorithm

ASK subject classification: Acoustical Signal Processing (1,2)

I. 서론

자동 음악 정보검색 연구분야에서 최근의 중요한 이슈는 자동 음악 장르 분류 시스템이다. 일반적으로 장르 분류 시스템은 크게 특징벡터 추출과 장르 분류의 두 가

지 과정으로 이루어진다 [1]-[3]. 첫 번째 과정에서 음악 신호로부터 오디오의 피치, 음색, 박자 등의 특징을 추출하며 두 번째 과정에서 추출된 특징벡터들을 이용하여 입력된 음악의 장르를 분류한다.

최근에 음악신호의 특징벡터 추출과 장르 분류에 관한 많은 연구가 진행되어왔다 [2][10]. 그러나 대부분의 연구가 상대적으로 적은 수의 장르 데이터베이스에 대해서 진행되었고, 대부분의 알고리즘이 다중 범주 분류에 관

한 것이었다. 다중 범주 분류는 음악장르를 독립적으로 다루어서 장르들 간의 관계를 정의할 수 없다는 단점이 있으며, 또한 음악산업이 커지고 음악 장르의 수가 늘어나면서 장르들 사이의 경계가 모호해진다는 점 때문에 한계를 보이고 있다 [15].

최근에 연구에서는 계층 구조를 이용하여 분류 성능을 높이는 방법이 제시되고 있다 [4]. 계층 분류는 범주를 유사한 범주끼리 묶어 노드를 만들고, 노드들을 묶어 계층 구조를 만들어 각각의 노드들마다 분류를 하는 구조로 되어있다. 계층 분류를 이용하는 방법에서는 계층 구조를 (hierarchical structure) 구축하는 과정이 매우 중요한데 [4][5], 구조를 수동으로 구축하는 경우, 정확성은 높지만 비용이 많이 들고, 또한 장르는 내재적으로 잘 정의되어 있지 않아서 분류학자들조차 일관성이 없는 분류 결과를 보이기도 한다 [13].

이런 수동 분류의 문제를 해결하기 위해 혼동 매트릭스 (confusion matrix)를 이용하여 계층을 클러스터링 하는 자동 분류법을 사용한다 [4]. 그러나 이 방법은 분류 되어야 할 장르 범주가 증가할수록 장르들 간의 혼동 정도가 심화되어 범주를 명확하게 구축할 수 없고, 분류성능 또한 좋지 않다는 단점이 있다 [4]. 따라서 수동구축에 들어가는 비용을 절감할 수 있고 범주가 증가하여 혼동 정도가 심화 되어도 분류 성능 열화가 없는 자동 분류 알고리즘 개발이 필요 하다.

본 논문에서는 이러한 단점들을 극복하기 위하여 새로운 taxonomy 구축방법을 제안하였다. 제안된 알고리즘은 오디오 장르의 계층구조를 만들 때 모든 가능한 장르 쌍들의 분류 확률 예측을 기반으로 자동으로 구축하여 자동분류 시스템의 분류 성능을 높인다. 분류 확률 예측을 위해 모든 가능한 클러스터 쌍들을 나누어 놓고 wrapper algorithm [16]을 이용하여 각각의 클러스터 쌍들을 가장 잘 분리할 수 있는 특징벡터를 추출한다. 추출된 특징벡터를 이용하여 훈련 데이터에 대해 k-fold cross validation을 통해 모든 가능한 클러스터에서의 분류 성능을 예측함으로써 가장 높은 예측 성능치에 해당하는 쌍들을 묶어 노드를 형성하는 방법을 사용한다. 이렇게 구축된 모든 노드들에 2-class support vector machine (SVM) 분류기를 적용하여 훈련시키며, 테스트 데이터에 대해 동일한 2-class SVM으로 분류를 실행한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서 사용될 특징벡터들에 대해서 살펴본다. 다음으로 제안된 알고리즘을 3장에서 설명한다. 제안된 알고리즘의 성능 평가를 위하여 다양한 실험을 4장에서 진행하고, 마지막으로 결

론을 5장에 언급한다.

II. 특징벡터의 추출

특징벡터의 추출은 오디오를 구간별로 분석할 때 오디오 신호의 특징을 나타내도록 간결하게 표현된 파라미터를 계산하는 과정이다. 음악 범주 분류 알고리즘 등 패턴 인식 시스템에서 특징을 잘 나타내어 주는 파라미터를 디자인하는 작업은 가장 주요한 문제라 할 수 있다.

본 논문에서 사용한 특징벡터들은 크게 음색 표현 특징벡터, 리듬의 특성을 나타내는 특징벡터, pitch의 특징을 나타내는 특징벡터로 구분될 수 있다 [2][6][12].

특징벡터 추출 과정은 다음과 같다. 먼저 음색 표현 특징벡터들은 프레임 단위로 추출된다. 이 특징벡터는 20ms의 Hamming window를 인접한 프레임과 50%씩 중복되게 이동하면서 추출하여 프레임 가장자리에서 표현되는 특징의 손실을 최대한 억제한다. 리듬의 특성을 나타내는 특징벡터와 피치의 특징을 나타내는 특징벡터들은 곡 단위로 추출된다. 이 특징벡터들은 히스토그램 기반의 특징벡터들로서 한 곡의 리듬과 피치 정보를 프레임별로 추출하여 이들의 히스토그램으로부터 곡의 특징을 추출한다는 특징이 있다.

2.1. 음색 표현 특징벡터

음색으로 표현되는 특징벡터는 음악-음성 구분 알고리즘과 음성 인식 등의 분야에서 기본적으로 사용되는 특징벡터이다 [6]. 사용된 특징 파라미터와 차수를 표 1에 정리하였다.

2.2. 리듬 표현 특징벡터

음악의 범주 중 특히 음악 장르는 리듬에 따라 분류되는 경향이 강한 것을 관찰할 수 있다 [2]. 본 연구에서는 참고 문헌 [2]의 방법과 같이 리듬 히스토그램을 형성한 후 히스토그램의 특징을 나타내는 6차의 파라미터 (RHYTHM)를 추출하여 사용하였다.

2.3. 피치 표현 특징벡터

본 논문에서는 다중 피치 검출 알고리즘을 사용하여 피치를 추출하였다 [12]. 이렇게 추출된 피치는 악보에서 사용되는 값으로 바꾸고 곡 전체에 대하여 히스토그램을 구성한 후 6차의 피치 파라미터 (PITCH)를 추출하였다 [2].

표 1. 음색 표현 특징 파라미터들과 차수
Table 1. Timbral feature vectors and order.

차수	특징 파라미터
12*2=24	Mel frequency Cepstral coefficients (MFCC) [7]
12*2=24	Autocorrelation coefficient (ACC) [8]
1*2=2	Root mean square (RMS)
1*2=2	MAX
1*2=2	Time-domain zero-crossing (TDZC)
4*2=8	Temporal moment (TMM) [8]
2*2=4	AR coefficient (ARC) [8]
5*2=10	Spectral crest factor (SCF)
1*2=2	Spectral rolloff (SR)
1*2=2	Spectral flux (SF) [2]
1*2=2	Spectral slope (SS)
1*2=2	Spectral decrease (SD) [8][9]
1*2=2	Spectral sharpness (SHARP)
1*2=2	Spectral spread (SPREAD) [9]
6*2=12	Spectral contrast (SC) [10]
5*2=10	Spectral flatness measure (SFM) [8][9]
6*2=12	VALLEY [10]
6*2=12	Octave band signal intensity (OBSI)
5*2=10	Octave-band signal intensity ratio (OBSIR) [11]
37*2=74	Relative specific loudness (Nrel) [9]
1*2=2	Harmonic ratio (HR)
1*2=2	Modified harmonic ratio (MHR) [5]

2.4. 특징벡터 결합법

본 연구에서 사용된 특징벡터는 크게 프레임 기반의 특징벡터와 히스토그램 기반의 특징벡터로 나눌 수 있다. 프레임 기반의 특징벡터는 2.1절의 음색표현 특징벡터이고, 히스토그램 기반의 특징벡터는 2.2절과 2.3절의 리듬 및 피치 표현 특징벡터이다. 프레임 기반의 특징벡터들은 20ms 윈도우 안에 포함된 오디오 신호에 따라 매우 급격하게 변할 수 있다. 이렇게 변화가 심한 특징벡터가 장르 등을 표현하는 특징벡터로서 의미가 있기 위해서 여러 프레임 동안 특징벡터의 통계적 특성을 추출하여 고차원의 특징벡터를 구성하였다. 이를 위해 여러 프레임이 모인 구간을 texture window로 정의한다 [2]. 본 논문에서는 참고문헌 [2]에서 제시한 texture window를 1초에 해당하는 100개의 프레임으로 두고 평균과 분산을 구하였다. 따라서 각 파라미터 마다 2개의 고차원 파라미터가 추출된다. 예를 들어 2분 분량 음악의 경우 최초 30초와 마지막 30초를 제외한 중간부분 60초 동안 10ms 간격으로 특징벡터가 추출되며 이를 이용하여 다시 100개의 프레임 동안 모아 한 곡당 60개의 특징벡터가 추출되도록 하였다. 이 때 프레임 기반 특징벡터와 히스토그램

기반 특징벡터는 서로 다른 시간단위에서 추출되었기 때문에 이를 맞춰 주어야 한다 [2]. 본 연구에서는 한 곡에 대하여 하나의 특징벡터가 나오도록 설계하였다. 프레임 기반 특징벡터는 1곡 전체에 대하여 하나의 특징벡터로 표현되어야 한다. 이를 위해 한 곡 전체에 대하여 texture window를 통해 구한 고차원 특징벡터의 평균을 구하여 하나의 특징벡터를 추출하였다.

III. 제안된 자동 계층 분류 시스템

최근에 많은 연구들은 미리 정의된 계층 taxonomy들을 기반으로 진행되었다. 그러나 장르는 내재적으로 잘 정의 되어 있지 않고 사람들마다 정의하는 장르는 다르다. Pachet과 Cazaly는 3개의 인터넷 장르 taxonomy들을 비교하였다 [13]. 비교 결과 70개의 단어들이 3개의 taxonomies에서 공통적이었다. 심지어 70개의 공통된 단어들 중에 많이 사용되는 “Rock”이나 “Pop”과 같은 용어들도 공통된 정의로 사용되지 않고 있었다. 게다가 분류 학자들조차 정의하는 장르 분류가 달랐다 [15].

이를 해결하기 위하여 기존의 자동 계층분류 연구들은 단지 혼동 매트릭스를 이용하였다 [4]. 그러나 이 방법은 분류 되어야 할 장르 범주가 증가할수록 장르들간의 혼동 정도가 심화되어 범주를 명확하게 구축할 수 없고 분류성능 또한 좋지 않게 된다.

본 논문에서는 오디오 장르의 계층구조를 만들 때 모든 가능한 장르 쌍들의 분류 확률 예측을 기반으로 자동으로 구축하는 알고리즘을 제안하고 이를 분석한다.

3.1. 자동적인 장르 계층의 구축 알고리즘

본 연구의 목표는 주어진 특징 벡터들을 사용하여 최상의 분류 성능을 가지는 taxonomy를 자동으로 구축하는 것이다. 핵심적인 아이디어는 분류 성능을 예측하여 장르들간의 관계를 추론하는 것이다. 제안하는 자동 장르 계층 구축 알고리즘을 다음과 같다.

먼저 훈련데이터로부터 특징벡터를 추출하여 데이터 베이스를 구축하고 이를 D라 한다. D는 M개의 범주를 가지고 있고 N차수의 특징벡터로 구성된 $M \times N$ 차원의 행렬이다. 그리고 M_l 은 다음과 같이 정의한다.

$$M_l = \begin{cases} M & , \text{if } M = \text{even} \\ M+1 & , \text{o/w} \end{cases} \quad (1)$$

여기서 l 은 레벨을 지칭하는 변수이다.

첫 번째 과정으로 M_l 개의 범주 중에 $M_{l/2}$ 개를 추출하면 총 $M_l C_{M_l/2}$ 개의 후보 조합이 생성된다.

선택된 범주의 조합과 선택되지 않은 범주의 조합을 각각 $S_j = [s_{1,j}, s_{2,j}, \dots, s_{M_l/2,j}]$ 와 $U_j = [u_{1,j}, u_{2,j}, \dots, u_{M_l/2,j}]$ 라 지칭 하고 그 조합에 포함된 범주들로 구성된 D의 행 벡터를 $d_{s_{i,j}}$ 와 $d_{u_{i,j}}$ 라 하고 그 조합으로 구성된 D 하위 행 렬은 수식 (1)과 같다.

$$D_{s_j} = \begin{bmatrix} -d_{s_{1,j}} - \\ -d_{s_{2,j}} - \\ \vdots \\ -d_{s_{M_l/2,j}} - \end{bmatrix}, \quad D_{u_j} = \begin{bmatrix} -d_{u_{1,j}} - \\ -d_{u_{2,j}} - \\ \vdots \\ -d_{u_{M_l/2,j}} - \end{bmatrix} \quad (2)$$

여기서 $i=1,2,\dots,\frac{M_l}{2}$ 이고, $j=1,2,\dots, M_l C_{M_l/2}$ 이다.

다음으로 각각 생성된 S_j 와 U_j 를 묶어서 후보 노드들을 생성한다.

$$N(j) = [S_j, U_j] \quad (3)$$

이때 가능한 노드들은 모두 2개의 클러스터로 나뉘고 전체 $M_l C_{M_l/2}$ 개 만큼 생성된다.

이렇게 하여 만들어진 2개의 클러스터로 이루어진 각각의 후보 노드 들에서 각각의 훈련 데이터베이스를 k등분하여 k-fold cross validation 과정을 진행하여 분류 성능을 예측한다.

$$r(j) = eval(D_{N(j)}), \quad j=1,2,\dots, M_l C_{M_l/2} \quad (4)$$

$$D_{N(j)} = [D_{s_j}, D_{u_j}]$$

위 식에서 $eval(\cdot)$ 은 k-fold cross-validation 을 나타낸다. 이 과정에서는 훈련 데이터베이스 만을 사용하기 때문에 나중에 실제 테스트 데이터의 분류 성능에 영향을 주지 않는다.

여기서 k-fold cross validation과정 [14]은 다음과 같다. 훈련데이터 베이스를 k등분하여 $(k-1)/k$ 등분은 트레이닝 데이터로 훈련시킨다. 나머지 $1/k$ 등분은 테스트 데이터로 사용한다. 테스트 데이터의 위치를 k번 변화해 가면서 분류성능을 측정하고 이의 평균값을 분류 성능 예측 값으로 사용한다. 여기서 훈련과 테스트는 2개의 범주 구분에 강인한 SVM을 사용한다. 위와 같이 예측된 분류성능 예측 값 중에 가장 높은 값을 가지는 노드를

top-level taxonomy로 구축한다.

$$Taxo(l) = \arg \max_{N(j)} (r(j)) \quad (5)$$

결국 이와 같은 과정을 거쳐서 M 개의 범주가 한 쌍의 클러스터로 재배열 된다. 위와 같이 구축된 각각의 클러스터들에는 각각 $M/2$ 개의 범주들이 존재하는데 이러한 $M/2$ 개의 범주들을 다시 가능한 모든 2쌍의 클러스터 세트 로 묶어 새로운 하위 노드를 생성한다. 여기서는 $(k-1)$ 개의 후보 노드들이 생긴다. 각각의 노드들에서는 상위 노드의 경우와 마찬가지로 k-fold cross validation 과정을 진행하여 분류 성능을 예측하며, 예측 값 중에 가장 높은 값을 가지는 클러스터 쌍을 second level taxonomy 로 구축한다. 결국 과정 (2)~(5)를 반복하여 최하단 level 의 클러스터들 모두가 범주 조합이 아닌 1개의 범주일 때 끝나게 된다. 이를 위해서는 레벨을 지칭하는 변수 l 과 각 노드별 범주의 수 M_l 을 $l=l+1, M_l = M_l/2'$ 로 조절하며 루 프를 실행하면 된다. 다시 말하면 M 개의 범주 들을 최상의 분류 성능을 가지는 한 쌍의 클러스터로 이루어진 노드들로 나누어 가면서 L 개의 레벨을 가지는 계층구조를 구축하는 것이다.

각각의 노드들을 반드시 2개의 클러스터들로 나눌 필요는 없다. 그런데 여기서 노드를 한 쌍, 즉 두 개의 범주 조합으로 나누는 이유는 2개의 범주 분류에 매우 강인하다고 알려져 있는 SVM [18]을 각각의 노드마다 이용하여 분리 성능을 최대화 시키기 위해서이다.

3.2. 특징벡터 선택 알고리즘

범주 분류 성능 예측 시에 일반적으로 특징벡터의 차수가 많으면 많을수록 실제 데이터의 분류성능이 높아진다고 알려져 있다. 그러나 훈련 시킬 데이터베이스의 길이가 유한하기 때문에 어느 정도 차수이상에서는 오히려 분류성능이 떨어지는 문제 즉 차원의 저주 문제가 발생하게 되고 또한 많은 차수를 사용하면 효율성도 떨어지게 된다 [15]. 따라서 정해진 차수에서 최적의 특징벡터 결함을 선택할 수 있는 특징벡터 선택 알고리즘이 필요하다.

일반적으로 각기 다른 평가 기준을 가지고 만들어진 특징벡터 선택 알고리즘은 크게 3종류로 나뉘어 진다 [16]. 필터 모델, wrapper 모델, 혼합 모델이다. 이 중에서 wrapper 모델은 가장 연산량이 많지만 사전에 mining 알고리즘을 사용하기 때문에 가장 성능이 좋고 훈련 데이터를 k-fold cross validation과 mining 알고리즘을 사용하여 분류 성능을 예측하므로 실제 데이터의 분류성능값

과 가장 유사한 값을 얻어 낼 수 있다.

Wrapper 모델은 훈련 데이터를 통해서 인식률을 예측할 때 훈련 데이터를 k등분으로 나누어 (k-1)개는 훈련 데이터 k번째 등분은 test 데이터로 나누어 실제 인식률을 측정한다. Wrapper 알고리즘은 특징 파라미터 하나씩 분류기에 적용시켜보고 가장 분류성능이 좋은 파라미터들을 선택해 나가는 것이기 때문에 분류성능 예측과 동시에 특징 파라미터가 선택되는 것이다. 결국 wrapper 알고리즘은 k-fold cross validation과 mining 알고리즘으로 구성되므로 3.1절의 k-fold cross validation을 wrapper 알고리즘으로 바꾸면 제안된 알고리즘 구축시에 인식률 값을 예측함과 동시에 특징벡터를 선택할 수 있다.

본 논문에서는 분류 성능 예측시에 wrapper 알고리즘을 사용함으로써 실제 데이터의 분류 성능을 잘 예측하여 가장 성능이 좋은 특징벡터 들을 선택하였다.

IV. 실험

4.1. 실험 구성

제안된 알고리즘의 실험을 위해 사용된 음악 DB는 ISMIR2004 Audio Description Contest에서 장르 구분 대회를 위해 제공된 음악을 사용하였다 [17]. 이 데이터베이스는 classical (1), electronic (2), jazz_blues (3), metal_punk (4), rock_pop (5), world (6)의 6개의 범주로 구성되어 있으며 training set과 development set이 각각 729 곡으로 구성되어 있다. Training set는 classical이 320곡, electronic이 115곡, jazz_blues가 26곡, metal_punk가 45곡, rock_pop이 101곡, world가 122곡으로 이루어져 있다. 오디오 파일들은 44100 Hz, 16-bit, 모노 형식으로 저장되었다. 특징벡터를 선택할 때와 모델을 만들 때에는 training set을 사용하였고 테스트를 할 때에는 development set을 사용하였다. 특징벡터를 추출할 때에는 training set과 development set의 모든 곡에서 최종적으로 하나의 특징벡터를 추출하였다. 각 음악에서 최초 30초부터 음악이 끝나기 30초 전까지의 오디오 신호를 대상으로 특징벡터를 추출하였다. 프레임 기반의 특징벡터는 texture window마다 추출된 고차원의 특징벡터를 한 곡 전체에 대하여 평균과 분산을 구하여 히스토그램 기반의 특징벡터와 결합하였다.

특징벡터 결합 시에 texture window 기반의 특징벡터 모두를 사용하여 한 곡 전체에 대하여 평균과 분산을 구

하는데 texture window 기반 모든 표본을 직접 사용하면 평균과 멀리 떨어진 표본의 영향을 많이 받게 되는 경우가 생길 수 있다 [14]. 따라서 표준편차의 3배 (정규 분포로 가정할 경우 평균에서 가장 먼 1%)가 넘는 표본은 제외하였다. 성능 평가는 전체 샘플 중 범주를 제대로 찾는 샘플의 비를 측정하여 평가하였다.

4.2. 선택된 특징벡터들과 구축된 taxonomy

제안된 알고리즘을 사용하여 구축된 taxonomy는 그림 1에 나타내었다. 자동적으로 장르 계층을 구축할 때 앞서 언급한대로 wrapper 알고리즘을 사용하여 특징벡터 선택과 분류 성능 예측을 동시에 시행한다. 따라서 wrapper 알고리즘을 이용하면 각각의 노드에서 가장 분류성능이 좋은 특징벡터들을 추출할 수 있다. 특징벡터의 차수에 따른 분류 성능 측정을 위하여 wrapper 알고리즘에서 특징벡터의 차수를 변화시켜가면서 실험하였다.

구축된 taxonomy의 각각의 노드에서 선택된 특징벡터의 부분집합을 표 2에 나타내었다.

표에 정리된 변수 이름은 표 1과 동일하며 변수 뒤의 번호는 해당 변수의 차수 인덱스를 나타낸다. 편의상 30차 까지만 도시하였다.

각각의 노드마다 선택된 특징벡터들이 차이가 있기 때문에 이 결과를 분석하는 것이 직관적으로 크게 의미가 없지만 표에서 확인할 수 있듯이 가장 많이 선택된 특징벡터는 기존의 장르 분류 연구에서는 많이 사용되지 않았던 *Nrel* 계수이다. 이로부터 바크 (Bark) 밴드에서의 에너지 비가 장르 분류에 중요한 특징벡터임을 알 수가 있다. 다음으로는 *TMM*이 많이 선택되었는데 이로부터 시간적인 파형의 모멘트 또한 장르 분류 성능에 많은 영향을 준다는 것을 알 수 있다. 마지막으로 많은 논문에서 다중 범주 분류에 사용했던 *MFCC*는 어느 정도 선택 되기

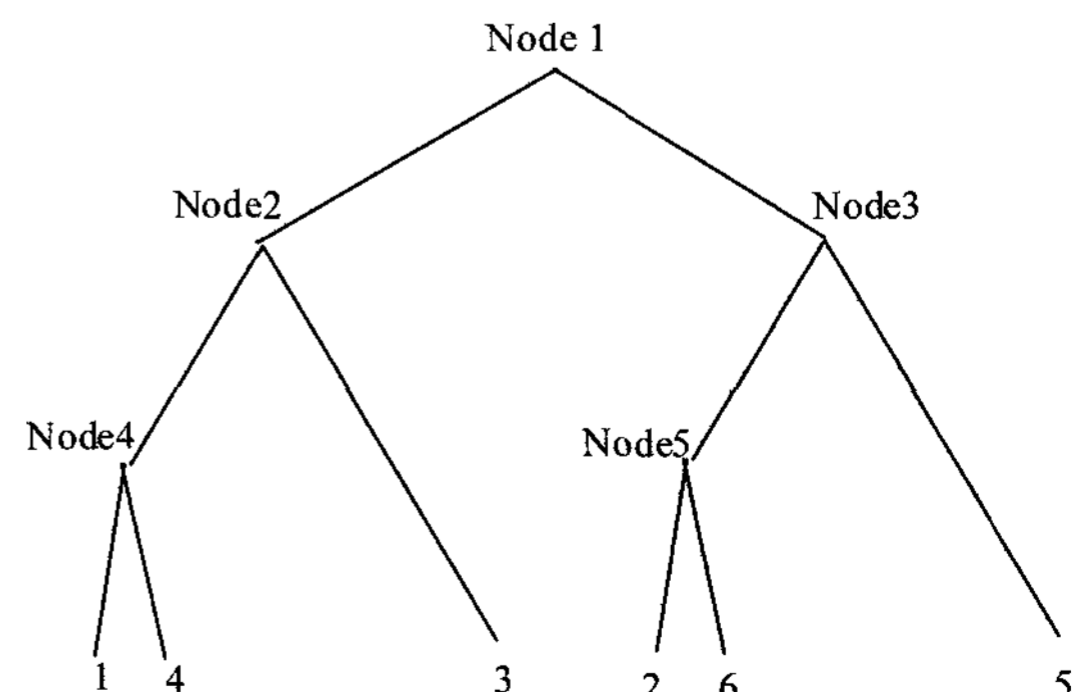


그림 1. 구축된 장르 계층
Figure 1. Generated genre taxonomy.

표 2. 구축된 taxonomy의 노드 각각에서 선택된 특징벡터
Table 2. The selected feature vector at each node of generated taxonomy.

Node	선택된 특징벡터
Node1	TMM1, ..., TMM6, Nrel1, ..., Nrel5 ACC1, ACC2, OBSI1, SD1, SD2, MFCC1, ..., MFCC7, SF1, SF2, OBSIR1, SCF1, HR1, SFM1, Max1
Node2	TMM1, TMM2, TMM3, SFM1, SFM2, SR1, SC1 MFCC1, ..., MFCC5, Nrel1, ..., Nrel9 TDZC1, ARC1, ACC1, Valley1, OBSI1, OBSI2, SFM1, SFM2, SCF1
Node3	TMM1, ..., TMM4, MFCC1, ..., MFCC9 SD1, SD2, ACC1, ACC2, Nrel1, ..., Nrel3 Valley1, SCF1, SCF2, ARC1, ARC2, OBSIR1, HR1, Spread1, SR1, Max1
Node4	ACC1, ..., ACC4, SFM1, ..., SFM4, Valley1, Valley2, TMM1, TMM2, SCF1, ..., SCF3, Sharp, SFM1, SFM2, TDZC1, Spread, SD1, SC1, ..., SC3, Nrel1, ..., Nrel6
Node5	TMM1, ..., TMM3, Nrel1, ..., Nrel10, Spread1, Valley1, ..., Valley4, OBSIR1, ..., OBSIR4, MFCC1, ..., MFCC3, SD1, SD2, SS1, SS2, ACC1

는 하지만 상대적으로 *Nrel* 계수 보다는 덜 선택되었다. 이로부터 한번에 여러 장르를 분류할 때 중요한 특징벡터와 한번에 2개의 범주만을 분류할 때 중요한 특징벡터는 차이가 있음을 알 수 있다.

4.3. 분류기 및 파라미터 결정

제안된 알고리즘의 효율성을 측정하기 위하여 실험에서 사용한 기존의 다중 범주 분류기는 다음과 같다. Multi-class support vector machine [18], K-nearest neighbor (KNN) [14], Gaussian mixture model (GMM) [19]과 Neural network (NN) [20]와 같은 다양한 다중 범주 분류기와 비교하였다. 제안된 알고리즘의 성능을 측정하기 위하여 2 클래스의 분류에 적합한 one-versus-one support vector machine을 사용하였다.

제안된 계층 구조 시스템에서는 매 노드마다 특징벡터 선택을 하여야 하므로 기존의 다중 범주 분류기 보다 훨씬 많은 차수의 특징벡터를 사용하게 된다. 이는 공정한 실험이 되지 못한다. 따라서 다중 범주 분류기와 똑 같은 차수를 사용하기 위해서 제안된 알고리즘에서는 최상위 레벨에서 다중 범주 분류기와 똑 같은 차수의 특징벡터를 선택한 후에 아랫단의 모든 노드에서는 최상위 레벨에서 선택된 특징벡터만을 가지고 그 중에서 더 적은 차수의 특징벡터만을 선택하는 방식으로 실험을 진행하였다. 이렇게 하면 제안된 알고리즘에서 필요한 모든 특징벡터를 이미 최상위 레벨에서 추출하였기 때문에 추가적인 특징벡터가 필요 없게 된다.

표 3. 각각의 노드에서의 혼동 매트릭스
Table 3. The confusion matrix at each node.

Node1/구성된 범주	1,4,3	2,6,5
1,4,3	92%	8%
2,6,5	5%	95%
Node2/구성된 범주	1,4	3
1,4	100%	0%
3	1%	99%
Node3/구성된 범주	2,6	5
2,6	98%	2%
5	24%	76%
Node4/구성된 범주	1	4
1	100%	0%
4	1%	99%
Node5/구성된 범주	2	6
2	88%	12%
6	4%	96%

표 4. 분류기별 분류 성능

Table 4. Classification performance according to classifiers.

차수/분류기	30차	50차	100차
KNN	64%	71%	75%
GMM	58%	60%	65%
NN	71%	74%	79%
SVM_multi	74%	80%	84%
2-class SVM based on proposed algorithm	83%	84%	89%

4.4. 분류실험 결과

제안된 알고리즘으로부터 분류 실험을 진행하여 얻어진 각각의 노드에서의 혼동 매트릭스를 표 3에 나타냈다.

표 3에서 각각의 Node1~Node5는 제안된 알고리즘으로 구축된 그림 1의 Taxonomy에서의 Node들이다. 여기서 각각의 Node들은 그림 1에서와 같이 크게 2 클래스로 구성된다. 표에서 구성된 범주라는 것은 각각의 클래스가 이루어진 장르 범주들이다. 2 클래스로 이루어진 각각의 노드에서의 분류 성능을 측정하여 표 3의 혼동매트릭스를 구성하였다. 표에서 보듯이 Node3에서 클래스 5와 클래스 (2, 6)의 분류성능이 좋지 못한데 특징벡터 분석 결과 world (6)가 rock_po (5)와 상당히 유사한 특성을 보였기 때문이다.

또한 분류기별 전체적인 분류 성능을 표 4에 정리하였다. 표 4에서 2~5열의 결과들은 각각의 분류기를 사용하여 6개의 장르를 한번에 즉 다이렉트하게 분류하였을 때의 결과이다.

표 4에서의 최하단의 2-class SVM based on proposed

algorithm은 제안된 알고리즘을 사용하여 구축된 자동 Taxonomy를 support vector machine을 사용하여 분류 성능을 측정하는 것이다.

다양한 특징벡터의 차수에 대하여 실험을 하였으나 차수가 증가함에 따라서 분류성능이 증가하는 경향에는 변함이 없었기 때문에 대표적으로 30, 50, 100차의 차수에 대해서만 도시하였다.

제안된 알고리즘을 이용한 분류성능은 최하단의 레벨까지의 누적 분류 성능이다. 즉 최상위 레벨에서부터 최하단의 노드에 있는 분류성능을 차례로 곱하여 얻어진 값이다. 성능 비교 결과 모든 차수에 대해서 제안된 알고리즘을 이용한 분류 시스템은 기존의 분류 시스템들보다 성능이 우수함을 보였다.

또한 30차를 사용한 제안된 알고리즘의 성능은 83%를 보였으며 이는 100차를 사용한 기존의 다중 범주 분류기들보다 우수한 성능을 보였다. 따라서 제안된 알고리즘은 적은 차수 즉 적은 연산량을 사용하여 기존의 분류기들보다 매우 좋은 성능을 보임을 확인하였다. 이는 저전력의 응용분야에 제안된 알고리즘이 매우 효과적임을 보여주었다.

그러나 표 3에서 보듯이 Node3에서의 성능이 다른 노드에 비해서 떨어져서 전체 Taxonomy의 분류성능에 영향을 주어 추가적인 분류 성능 향상을 방해하였다. 이는 기본적으로 2절에서 추출된 특징벡터들은 기존에 미리 정해지거나 많이 알려진 장르 분류에 적합한 것이 많았기 때문이다. 그런데 제안된 알고리즘을 이용하여 구축된 Taxonomy의 각각의 노드는 기존에 미리 알려진 서로 다른 장르를 묶어서 또 하나의 클래스를 이루었기 때문에 기존의 특징벡터만을 사용하면 분류 성능 향상에 한계가 있음을 알 수 있다.

V. 결론

본 논문에서는 계층적인 오디오 장르에 따른 오디오 신호의 자동 분류 시스템을 제안하였다. 제안된 알고리즘은 모든 가능한 노드들의 분류 확률을 예측하여 예측된 분류 성능값이 가장 좋은 조합을 Taxonomy로 구축하는 것이다. 제안된 알고리즘에서의 분류 확률 예측은 모든 가능한 장르 쌍들을 나누어 놓고 각각의 장르 쌍들을 가장 잘 분리할 수 있는 특징벡터를 wrapper algorithm을 이용하여 추출하여 추출된 특징벡터를 이용하여 훈련 데이터를 k-fold cross validation을 이용하여 분류기에 적

용시키어 모든 가능한 장르 쌍들의 분류 성능 값을 예측하였다. 최종적으로 후보가 되는 노드중에서 예측값이 가장 좋은 조합을 Taxonomy로 구축하였다.

제안된 알고리즘을 이용한 분류의 성능 검증 결과 기존의 다중 범주 분류 방법들 보다 5%에서 25% 정도의 성능 향상을 이루어 좋은 성능을 보임을 확인하였다.

그러나 표 3에서 보듯이 Node3에서의 성능이 다른 노드에 비해서 떨어져서 전체 Taxonomy의 분류성능에 영향을 주어 추가적인 분류 성능 향상을 방해하였다. 이를 개선하고 추가적인 분류 성능 향상을 얻어내기 위해서는 구축된 Taxonomy에서 각각의 노드에 적합한 새로운 장르 의존적인 특징벡터 개발이 필요하다. 향후 각각의 노드에서의 장르 의존적인 특징벡터를 개발하여 추가적인 분류 성능 향상을 위한 연구가 필요하다.

참고 문헌

1. L. Lu and H. Zhang, "Content analysis for audio classification and segmentation," *IEEE Trans. on Speech and Audio Process.*, 10(5), 504-516, Sep. 2002.
2. G. Tzanetakis and P. Cook, "Musical Genre Classification of audio signals", *IEEE Trans. on Speech and Audio Process.*, 10(4), 293-302, July 2002.
3. C. Yang, *Database retrieval based on spectral similarity*, (Stanford Univ. Database Group, Stanford, CA, Tech. Rep. 2001-14, 2001).
4. Tao Li and Mitsunori Ogihara, "Music genre classification with taxonomy," *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 197-200, 2005.
5. Juan Jose Burred and Alexander Lerch, "A hierarchical approach to automatic musical genre classification," *Proc. of the 6th Int. Conference on Digital Audio Effects (DAFX-03)*, London, UK, Sept. 8-11, 2003.
6. E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1331-1334, 1997.
7. Beth Logan, "Mel Frequency Cepstral Coefficients for music modeling," in *Proc. of the First International Symposium on Music Information Retrieval (ISMIR)*, 2000.
8. S.Essid, G.Richard, and B.David, "Instrument Recognition in Polyphonic Music Based on Automatic taxonomies," *IEEE Trans. Audio, Speech, and Lang. Process.*, 14(1), 68-80, Jan. 2006.
9. G. Peeters, "A large set of audio fetures for sound description (similarity and classification) in the CUIDADO project," *CUIDADO I.S.T. Project Report*, 2004.
10. D.-N. Jiang, L. Lu, H.-J. Zhang, J.-H. Tao, and L.-H. Cai, "Music type classification by spectral contrast feature," *Proc. of IEEE Int. Conf. on Multimedia and Expo (ICME02)*, Lausanne Switzerland, Aug, 2002.
11. S. Essid, G. Richard and B. David, "Musical instrument recognition based on class pairwise feature selection," *Proc.*

5th Int. Conf. Music Information Retrieval (ISMIR), Barcelona, Spain, Oct. 2004.

12. T. Tolonen and M. Karjalainen, "A computationally efficient multipitch analysis model," *IEEE Trans. Speech, Audio Process.*, 8(6), 708-716, Nov. 2000.
13. F. Pachet and D. Cazaly, "A taxonomy of musical genres," *Proc. Content-based Multimedia Information Access (RIAO)*, Paris, France, 2000.
14. P. A. Devijver and J. Kittler, *Pattern Recognition: A statistical approach*. (New York, Prentice-Hall, 1982).
15. J.-J. Aucouturier and F. Pachet, "Representing music genre: A state of the Art," *J. of New Music Research*, 32(1), 83-93, 2003.
16. Huan Liu and Lei Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. on Knowledge and Data Eng.*, 17(4), April 2005.
17. http://ismir2004.ismir.net/genre_contest/index.htm
18. V. Vapnik, "The nature of statistical learning theory," New York: Springer-Verlag, 1995.
19. D. A. Reynolds and R. C. Rose, "Robust test-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech, Audio Process.*, 3(1), 47-60, Nov. 1996.
20. S-Y. Kung and J-N. Hwang, "Neural networks for intelligent multimedia processing," *Proceedings of the IEEE*, 86(6), 1244-1272, June 1998.

저자 약력

• 최택성 (Tacksung Choi)

2003년 2월: 중앙대학교 공과대학 전기전자공학과 (공학사)
 2005년 2월: 연세대학교 공과대학 전기전자공학과 (공학석사)
 2005년 9월 ~ 현재: 연세대학교 공과대학 전기전자공학과 박사과정
 *주요연구분야: 디지털 신호처리, 오디오 신호처리, 3-D 오디오, 인공 진화계, 오디오 장르 분류

• 문선국 (Sunkook Moon)

2006년 2월: 연세대학교 공과대학 전기전자공학과 (공학사)
 2006년 3월 ~ 현재: 연세대학교 공과대학 전기전자공학과 석사과정
 *주요연구분야: 디지털 신호처리, 오디오 신호처리, 오디오 장르 분류

• 박영철 (Youngcheol Park)

1986년 2월: 연세대학교 전자공학과 (공학사)
 1988년 2월: 연세대학교 전자공학과 (공학 석사)
 1993년 2월: 연세대학교 전자공학과 (공학 박사)
 현재: 연세대학교 컴퓨터정보통신공학부 교수
 *주요연구분야: 디지털 신호처리, 오디오 신호처리, 음성 신호처리, 적응 신호처리

• 윤대희 (Daehee Youn)

1977년 2월: 연세대학교 전자공학과 (공학사)
 1979년 8월: Kansas State Univ. (공학 석사)
 1982년 2월: Kansas State Univ. (공학 박사)
 현재: 연세대학교 전기전자공학부 교수
 *주요연구분야: 디지털 신호처리, 적응 신호처리, 음성 신호처리, 오디오 신호처리

• 이석필 (Seokpil Lee)

1992년 8월: 연세대학교 전기공학과 (공학 석사)
 1997년 8월: 연세대학교 전기공학과 (공학 박사)
 현재: 전자부품연구원 방송통신융합연구센터 센터장
 *주요연구분야: 차세대디지털방송, 디지털 신호처리, 메타데이터 처리