DOI: 10.3745/KIPSTA.2010.17B.4.263

상황인지 음악추천을 위한 음악 분위기 검출

이 종 인[†] · 여 동 규[††] · 김 병 만[†††]

요 약

상황인지 음악추천 서비스를 제공하기 위해서는 무엇보다 상황 또는 문맥에 따라 사용자가 선호하는 음악의 분위기를 파악할 필요가 있다. 음악 분위기 검출에 대한 기존 연구의 대부분은 수작업으로 대표구간을 선정하고, 그 구간의 특징을 이용하여 분위기를 판별한다. 이러한 접근 방법은 분류 성능이 좋은 반면 전문가의 간섭을 요구하기 때문에 새로운 음악에 대해서는 적용하기 어렵다. 더욱이, 곡의 진행에 따라 음악 분위기가 달라지기 때문에 음악의 대표 분위기를 검출하는 것이 더욱 어려워진다.

본 논문에서는 이러한 문제점들을 보완하기 위해 음악 분위기를 자동으로 판별하는 새로운 방법을 제안하였다. 먼저 곡 전체를 구조적 분석 방법을 통하여 비슷한 특성을 갖는 세그먼트들로 분리한 후 각각에 대해 분위기를 판별한다. 그리고 세그먼트별 분위기 파악 시 Thayer 의 2차원 분위기 모델에 기초한 회귀분석 방법으로 개인별 주관적 분위기 성향을 모델링하였다. 실험결과, 제안된 방법이 80% 이상의 정확도를 보였다.

키워드 : 상황인지 음악 추천; 음악 장르 분류; 음악 구조 분석; 대표구간 탐지; 내용기반 음악 특징 추출

Detection of Music Mood for Context-aware Music Recommendation

Jongln Lee[†] · Dong-Gyu Yeo[††] · Byeong Man Kim[†††]

ABSTRACT

To provide context-aware music recommendation service, first of all, we need to catch music mood that a user prefers depending on his situation or context. Among various music characteristics, music mood has a close relation with people's emotion. Based on this relationship, some researchers have studied on music mood detection, where they manually select a representative segment of music and classify its mood. Although such approaches show good performance on music mood classification, it's difficult to apply them to new music due to the manual intervention. Moreover, it is more difficult to detect music mood because the mood usually varies with time.

To cope with these problems, this paper presents an automatic method to classify the music mood. First, a whole music is segmented into several groups that have similar characteristics by structural information. Then, the mood of each segments is detected, where each individual's preference on mood is modelled by regression based on Thayer's two-dimensional mood model. Experimental results show that the proposed method achieves 80% or higher accuracy.

Keywords : Context-aware Music Recommendation; Musical Genre Classification; Musical Structure Analysis; Salient Segment Detection; Content-based Musical Feature Extraction

1. 서 론

사용자의 현재 상태 (Situation) 나 행위 (Behavior) 를 파악하는 것은 상황 인지 (Context-aware) 컴퓨팅의 핵심 요소이다. 사용자의 현 행위나 상황을 파악하게 되면 사용자에게 보다 지능적인 서비스를 제공할 수 있게 된다. 현재는 상황에 무관하게 사용자 프로파일에 기록된 관심사 정보와 일치하는 내용, 혹은 과거의 평가정보가 유사한 성향의 다른 사람들이 추천한 내용을 추천하는 형태를 띠고 있다. 보다 사용자 중심의 고급 서비스를 제공하기 위해서는 현재의 상태 (상황 또는 문맥) 에 맞추어 추천해 줄 수 있는 기능이 필요하다.

추천 대상은 책에서부터 영화에 이르기까지 다양하나 그 중 음악은 우리의 일상생활과 매우 밀접한 관계를 맺고 있고 여러 상황에서 다양한 용도로 쓰이고 있기 때문에 상황인지 추천을 적용하기에 가장 적당한 콘텐츠이다. 상황별 음악 추천이 가능하기 위해서는 무엇보다도 다양한 상황에 따라 각 개인이 좋아하는 음악의 특성을 파악하는 일이 중

※ 이 논문은 2008년도 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2008-313-D00906).
† 정 회 원 : 슈어소프트테크(주)
†† 정 회 원 : 금오공과대학교 컴퓨터공학과(공학박사)
††† 정 회 원 : 금오공과대학교 교수
논문접수 : 2009년 10월 12일
수 정 일 : 1차 2010년 4월 26일, 2차 2010년 6월 7일
심사완료 : 2010년 6월 7일

요하다. 여러 가지 음악 특성들 중에서 분위기는 사람의 감정과 밀접한 관계를 맺고 있어 심리 치료 분야 등에서도 사용되어지고 있다. 음악의 내용을 기반으로 분위기를 탐지할 수 있다면 상황이나 행위에 맞는 지능적인 서비스를 제공할 수 있게 된다. 예를 들어, 사용자가 흥분되어 있는 상태라면 차분하고 편안한 분위기의 음악을 추천하거나 들려주어 심리적 안정을 유도해줄 수 있다. 또한, 현재 재생중인 음악의 분위기에 맞게 조명을 자동으로 조정하는 감성 조명 시스템으로도 확장할 수 있을 것이다.

분위기 추출에 관한 초창기의 연구들 [1-3] 은 일반적인 기계 학습/판별 방법을 사용하였으나, 이러한 방법은 음악을 하나의 분위기로 판단하기 때문에 정확성이 떨어지는 문제가 있고, 또한 개인의 주관적인 느낌과 이질감을 반영하지 못하는 문제가 있었다. 단일 분위기로 판정하기 때문에 발생하는 불확실성을 해결하기 위해 [4] 의 연구에서는 퍼지 기반의 학습/분류 방법을 사용하였으나, 이 또한 개인에게 느껴지는 음악의 분위기에 대한 주관적 성향을 해결하기에는 한계가 있었다.

또한, 하나의 음악은 전체 내용이 동일한 분위기 특성을 유지하기 보다는 중간 중간에 다른 분위기로 변화하며, 음률의 변화 또한 다양하다. 따라서 음악의 분위기를 탐지하기 위해서는 전체 음악을 의미 있는 몇 개의 부분으로 나누고 각 부분들에 대하여 독립적인 분위기를 탐지하는 기법이 필요하다. 하지만, 기존 연구들에서는 이러한 특성을 고려하지 않고 각 연구의 필요성에 따라 음악의 일부분을 전문가의 수작업을 통해 잘라내어 사용하거나 임의로 설정된 구간 (예를 들어 음악의 시작 후 30초 구간부터 30초 길이의 구간) 을 사용하였다. 이러한 방법들은 새로이 출판되는 음악에 적용시키기에 무리가 있으며 변화가 많은 음악의 특성상 정확도가 떨어지는 단점이 있다.

본 논문에서는 이러한 문제점을 해결하기 위하여, 수동이 아닌 자동으로 음악 자체의 내용을 바탕으로 한 구조 분석 기법을 통하여 음악을 의미 있는 구간들로 나누고, 각 구간들의 독립적인 분위기를 탐지하는 방법을 제안하였다. 또한, 전문가가 아닌 일반 사용자의 개인적 성향을 학습하여야 하는데, 비전문가인 경우 분위기에 대한 평가를 직접 지정하기가 어렵기 때문에 Thayer 의 2 차원 분위기 모델 [5] 을 응용한 분위기 형용사를 제공하고 사용자가 느끼는 분위기들을 여러 개 선택하도록 하였다. 이렇게 입력된 분위기 형용사들을 바탕으로 회귀학습 (Regression Training) 을 통하여 개인의 음악 분위기에 대한 성향을 모델링하였다.

본 논문은 다음과 같이 구성된다. 2 장에서는 기존의 음악 분위기 탐지를 위한 음악 분위기 분류방법, 음악 분위기 탐지 방법, 내용기반 음악 분류 분야 등에서 사용된 음향 특징, 그리고 음악 구조 분석 연구들에 대해 살펴본다. 3 장에서는 본 논문에서 제안한 음악 구조 분석 및 세그먼트 방법, 음향 특징 추출 방법 그리고 음악 분위기 분류를 위한 학습 방법에 대하여 살펴본다. 4 장에서는 성능평가 척도와 이를 바탕으로 한 성능 평가에 대하여 살펴보고, 5 장에서는 결론 및 향후 연구 방향에 대하여 살펴본다.

2. 관련연구

음악의 분위기 탐지 연구들 중, [6] 에서는 팝음악에 대하여 감정 (Sentiment) 판별시스템을 제안하였는데, 이를 위해 단선율의 음향데이터가 먼저 음악 코드 데이터로 변환되고 이로부터 멜로디, 리듬, 하모니, 형식 등을 추출하여 감정 추출에 이용하였다. [7] 의 연구에서는 음악의 분위기를 분류하기 위해 퍼지 분류기를 사용하였으며 템포, 세기, 피치변화, 음조 밀도 (Note density), 음색 (Timbre) 등의 특징을 사용하였다.

위에 열거한 두 연구는 나름대로 의미를 갖고 있으나 음향 데이터로부터 유용한 특징을 추출하기가 어려운 관계로 MIDI 또는 기호적 (Symbolic) 표현을 사용하고 있다. 하지만, 많은 실세계의 음악이 기호적 표현으로 되어있지 않기 때문에 음향 데이터를 기호적 표현으로 번역하기 위한 연구들이 진행되어지고 있지만, 대다수의 연구들이 단선율의 음향에 대해서만 좋은 성능을 보이고 복합음이 혼재한 음악의 경우 일반화된 성능을 나타내지 못하고 있기에 번역상의 오류가 분위기 탐지 성능에 개입될 여지가 있다.

다른 방법으로 음향 데이터로부터 직접적인 하위 레벨의 특징을 추출하여 음악의 분위기를 탐지하는 방법이 있다. 이 방법은 일반적인 기계학습 및 분류방법을 적용하여 사용할 수 있지만, 다음 세 가지 문제가 고려되어야 한다.

(가) 일반화된 음악 분위기 분류법 - 일반적인 학습 분류 방법을 위해서는 분류의 척도가 되는 클래스의 정의가 필요하다. 하지만 음악 분위기라는 특성상 획일화된 분류법이 존재하지 않고 주관적 차이가 존재하기 때문에, 널리 일반화된 분류법을 사용하여야한다.

(나) 내용기반 분류를 위한 특징 추출 - 음향 데이터를 그대로 사용하는 것보다 탐지 모델에 알맞은 특징 추출이 이루어져야한다.

(다) 음악의 규칙적인 변화를 고려 - 음악은 하나의 일관된 특성을 지니지 않고 다양하게 변화되는 특성을 지닌 연속된 부분들의 집합으로 이루어진다. 따라서 음악을 변화하는 특성에 맞게 나눈 다음 각기 분위기를 탐지하는 방법이 필요하다.

(가) 의 문제에 있어 내용기반 음악 분위기 탐색 연구들의 분위기 분류법과 함께 해당 분류법에 적합한 탐지 모델에 대한 연구들을 2.1 절과 2.2 절에서 살펴보고, (나) 의 문제에 대해서 기존 내용기반 음악 분류 연구들에서 사용한 특징들을 2.3 절에서 살펴보겠다. 마지막으로 (다) 의 문제를 해결하기 위한 음악 구조 분석에 관한 연구들을 2.4 절에서 살펴보겠다.

2.1 음악 분위기 분류법

Hevner[8]는 피실험자가 음악으로부터 느껴지는 감정을 67 개의 형용사로 축약한 뒤 비슷한 감정을 표현하는 8 개의 집합군으로 나누어 표현하였다. Farnsworth [9] 는 Hevner 의 감정 형용사법을 기반으로 의미가 중첩되거나 표현이 모호한 형용사에 대한 개정과 함께 10 개의 감정 형용사 그룹으로 재 할당한 분류법을 제안하였다.

위 두 분류법의 경우, [9] 에서 일부 형용사에 대해 개정을 거쳤고 이후 분위기 연구들에서 개정하여 사용하였지만 같은 그룹에 속한 형용사적 의미가 너무 비슷하여 이들 사이를 구분하기가 어렵고 분류에 대한 이론적 근거가 미비하였다. 1990년대 후반에 Thayer 는 2 차원 분위기 분류법을 제안하였다[5]. 이 방법은, Hevner 의 형용사법과 달리 분위기는 두 요소, 즉 스트레스와 에너지에 의해 결정된다는 이론을 채택하였으며 이에 따라 분위기를 Contentment, Depression, Exuberance, Anxious/Frantic 4 개로 분류하였다.

지금까지의 음악 분위기 탐지에 관한 연구들은 Hevner 의 형용사 체크리스트 법과 Thayer 의 2 차원 분류법을 기초로 하여 이루어졌다. 각 연구들은 다양한 학습 모델을 사용하여 분위기를 탐지하였는데 다음 절에서 각 연구들의 학습 모델들을 살펴보겠다.

2.2 분위기 탐지를 위한 학습 모델

[1, 10] 의 연구에서는 전통적인 클래스 기반 기계 학습 모델을 사용하여 음악의 분위기를 탐지하였다. 위 연구들에서는 음악의 분위기 분류 클래스를 위해 [9] 의 분류법에 3 개의 분위기 그룹을 추가한 총 13 개의 분위기 분류그룹으로 두고 각 그룹을 하나의 분위기 클래스로 가정하여 이를 학습 한 후 분위기를 탐지하는 모델을 사용하였다.

[2] 의 연구에서는 음악 분위기 탐지를 위해 Thayer 의 2 차원 분위기 모델을 사용한 분류법을 사용하고, 분류법의 이론적 배경에 맞추어 에너지와 스트레스 두 가지를 표현하는 음악 특징을 각각 추출하여 계층형 분류 모델에 적용하여, 비계층형 분류 모델에 비해 높은 성능을 나타냄을 보여주었다.

[4] 의 연구에서는 음악에 대한 분위기를 단일의 클래스로 분류하는 것은 각 분위기 클래스에 따른 문화적, 개인적, 그리고 클래스를 나타내는 형용사적 문제로 인해 모호함이 발생하게 된다는 것을 지적하였고, 이와 같은 클래스간의 애매성을 회피하기 위해 단일 클래스로 분류하는 대신에 각 클래스에 속한 정도로 표현하는 퍼지 기반의 분위기 탐지 방법을 제안하여 실험하였다.

[11, 12] 에서는 개인화 서비스를 제공하는 시스템인 경우, 퍼지 방법을 사용하면 개인의 주관적 성향을 제대로 처리하지 못할 수 있음을 지적하고, 이를 해결하기 위해 분위기 클래스를 사용하는 것이 아닌 Thayer 의 2 차원 분위기 모델의 각 축의 값을 직접 -1~1 사이의 실수로 두어 사용하였다. AV (Arousal/Valence) 계수라 불리는 2차원 벡터로 이루어진 이 값은 각 값이 실수로 이루어지기 때문에 두 개의 회귀 분석기를 통해 학습이 가능하게 된다.

2.3 내용기반 음악 분류를 위한 음향 특징들

음악의 음향 특징을 추출하여 음악을 분류하고자하는 연구들은 90년대 후반부터 지금까지 활발히 이루어지고 있는 분야이다. 기존의 음성 인식이나 자연어 인식 등의 연구 분야에서 좋은 성능을 나타내었던 특징들과 함께, 음악 분류의 주체인 장르나 스타일 등에 맞춘 새로운 특징 추출 방법에 관한 연구들이 진행되어 왔는데, 다음과 같은 특징들이 사용되어졌다.

(가) 음색 특징 (Timbral Feature)

ANSI (American National Standard Institute) 에서는 음향의 3 요소로 음고 (Pitch), 소리의 강도 (Loudness), 음색 (Timbre) 을 정의하고 있다. 음고와 강도의 경우는 청취자의 민감도와 관련하여 음의 높이 및 세기의 척도가 되고 있다. 음색의 경우는 같은 음고와 강도를 가지는 음에 대해서 음 자체의 특성을 구분하는 척도가 되고 있는데, 최근의 음악 장르 분류를 위한 음향 특징 추출 연구의 모태가 된 [13] 에서 음색 특징을 사용하여 좋은 성능을 보여 주었다. 음색 특징으로는 주파수 스펙트럼 (Spectrum) 의 여러 분산학적 특징을 나타내는 Spectral Shape 특징들, 사람의 청각 모델에 기반한 MFCC, 그리고 시간 레벨의 신호에서 추출하는 ZCR 이 있다.

(나) MPEG-7 Basic Spectral Descriptors 특징

MPEG-7 에서는 음향 신호의 특징 추출을 위해 하모니, 피치, 선명도, 음색 등 음향의 특성을 기술하는 저수준의 오디오 특징인 LLD (Low Level Descriptors) 들을 정의하였다. 이 LLD들은 6개의 그룹 (Basic, Basic Spectral, Spectral Basis, Signal Parameters, Timbral Temporal, Timbral Spectral) 으로 분류되어 지는데, [14, 15] 의 연구에서는 Basic Spectral Descriptors 특징을 기존 음색 특징과 비교하였으며 이를 음악 장르 판별을 위한 특징으로 사용하였다.

(다) Spectral Contrast

[16]에서는 옥타브 밴드 (Octave Band) 기반의 특징인 Spectral Contrast 라는 새로운 특징을 제안하여 바로크 (Baroque), 로맨틱 (Romantic), 팝 (Pop), 재즈 (Jazz) 그리고 락 (Rock)의 5 개 장르에서 MFCC 와 비교하여 더 좋은 성능을 나타내는 것을 보여주었다.

(라) 하모니(Harmony) 특징

음악에 있어 음 자체의 멜로디와 하모니는 중요한 요소이다. [13] 의 연구에서 음악의 전역적인 멜로디의 특징을 추출하기 위해서 피치 히스토그램 (Pitch Histogram) 방법을 제안하였다. 피치 히스토그램을 구하기 위해서는 음악 신호를 작은 프레임으로 나눈 뒤, 프레임 별로 [17] 에서 제안한 다중 피치 탐색 방법을 사용하여 전체의 피치 히스토그램을

계산한다.

(마) 리듬 (Rhythm) 특징

리듬은 통상적으로 음향보다 음악에 있어 중요한 요소로 여겨진다. 일반적으로 음악의 기본 리듬에 대한 척도로 BPM (beats-per-minute) 을 사용하며, 각 음표 (Note) 의 시간적 길이는 BPM 에 맞추어 진다. [13] 에서는 음악 전체의 BPM 정보와 BPM 에 맞춘 음표길이의 분산정도, 그리고 전체적인 BPM 에 알맞은 박자를 가진 연관 정도를 표현하기에 알맞은 비트 히스토그램을 제안하였다.

(바) DWCHs (Daubechies Wavelet Coefficient Histogram)

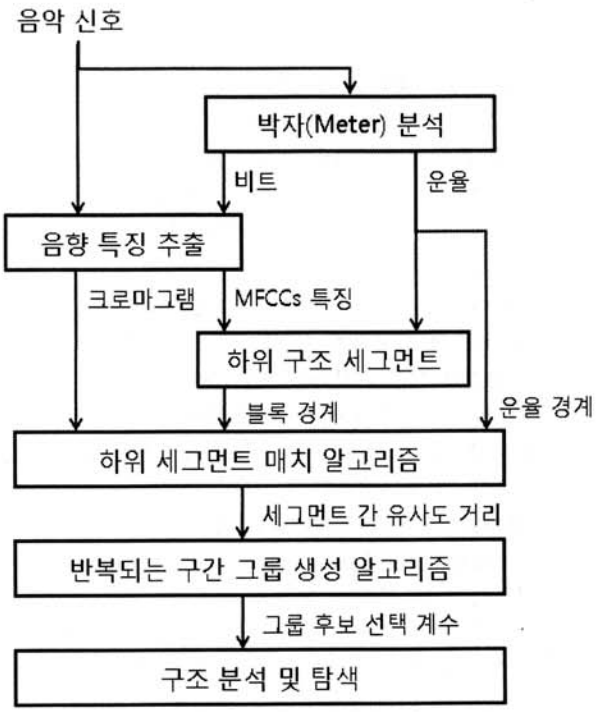
[18] 에서는 DWCHs 라 불리는 웨이블릿 히스토그램에 기초한 새로운 음악 특징 추출 방법을 제안하였다. DWCHs 는 기존 특징 추출에 많이 사용된 퓨리에 변환과 달리 주파수 대역에 따른 다해상도 (Multi-resolution) 특징을 나타내는 다우비치 웨이블릿 계수를 히스토그램 함으로써 국부적인 정보와 전역적인 정보를 모두 표현하게 된다. [18] 의 실험 결과에 따르면 음색적 특징과 DWCHs 를 함께 사용 하였을 시, 장르 판정에 있어 높은 성능을 얻을 수 있음을 알 수 있다.

2.4 음악 구조 분석 연구들

음악 구조 분석 연구에서는 가사구간 탐색, 악기의 변화 구간 탐색과 같은 음색의 변화 구간 탐색 방법 [19, 20], 음악 온셋 (Onset) 추출을 한 뒤 음표 리듬 패턴을 통한 구조 분석 방법 [21], 특징 벡터간의 유사도 매트릭스의 클러스터링을 통한 방법 [22, 23] 그리고 유사 멜로디의 반복적인 구조 탐색을 통한 방법 [24], 그리고 HMM (Hidden Markov Model) 을 통하여 유추한 가상의 코드 (Chord) 순서열의 규칙성과 유사도를 이용하는 방법 [25, 26, 27] 등을 사용하고 있다. 이러한 음악 구조 분석 연구를 바탕으로 한 응용분야로 이미지의 썸네일처럼 음악에서도 노래를 대표할 수 있는 부분을 추출하는 연구들 [28, 29] 이 있다.

반복 구간 탐색 기반의 음악 구조 분석에 관한 연구들은 기본적으로 음악을 일정한 구간으로 나눈 후 각 구간들 간의 유사도를 계산하여 높은 유사도를 지니는 구간들을 반복 구간으로 판별하는 방법을 사용하고 있다. 이러한 방법들은 우선적으로 유사도를 비교할 일정한 구간을 탐색하는 방법에 초점이 맞추어진 경향이 있는데, 이를 위해 온셋 정보나 비트 정보 추출을 통한 경계점 파악 알고리즘이나, 음향 특징의 급격한 변화 구간 탐색을 통한 경계점 탐색 알고리즘이 쓰인다. [23] 에서 사용한 반복 구간 파악을 통한 음악 구조 분석 방법이 이러한 방법을 종합적으로 사용한 대표적 예이며 (그림 1)에 도시하였다.

반복구간 탐색 기반의 구조 분석방법은 음악이 전체적으로 비슷한 음률로 이루어질 경우 탐색의 성능이 많이 떨어지는 것으로 알려져 있다. 이를 방지하기 위해 구간의 경계점 탐색과 같은 부가적인 방법을 추가하는 방식으로 성능 저하를 회피하는 방법들이 쓰인다. 또한 가장 기초가 되는 기본 음향 특징 추출 프레임의 크기에 따라 값의 부정확성이 발생하여 실제로 유사구간의 유사도 계산에 있어서도 오류가 발생할 여지가 있으므로 일반적으로 박자 분석과 같은 방법으로 최하위 프레임의 경계를 결정하나 이러한 방법 또한 박자 분석 방법이 좋은 성능을 나타낸다는 가정을 바탕으로 하고 있다.



(그림 1) 반복구간 탐색 기반 음악 구조 분석 시스템

[30]에서 처음으로 제안한 가상의 코드 상태열 (State Sequence) 기반의 음악 구조 분석방법에서는 여러 내외부적 요인들에 의해 생성되는 노이즈에 견고한 특징을 추출하기 위해 1 차적으로 추출된 음향 특징을 가상의 음악 코드 상태로 라벨링하여 연속적인 순차열을 생성한 후, 이를 이용하여 반복적이고 유사한 상태열을 파악하여 음악 구조 분석을 하였다. 노이즈에 견고한 가상의 상태열을 추출하기 위해 HMM 알고리즘을 사용하였으며, HMM 의 은닉 상태를 음악의 가상 코드 라벨로 사용하였다. [25, 26, 27] 의 연구들에서는 이러한 HMM 알고리즘을 통해 추출된 가상 코드를 Low-Level State Type 또는 Timbre-Type 이라 칭하고 있다.

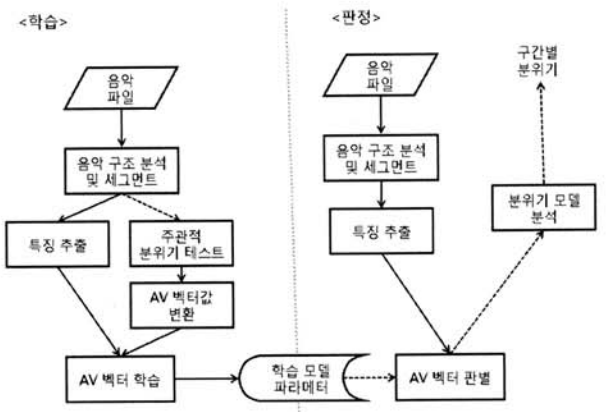
3. 음악 구조 분석을 이용한 음악 분위기 탐지 시스템

기존 분위기 판별에 관한 연구들에서는 곡 전체의 분위기 탐지에 관해 고려된 연구가 드물다. [2] 와 [4] 의 경우 분위기 탐색에 있어 전체의 곡을 여러 분위기의 시퀀스

(Sequence) 로 보고 곡의 일부분의 분위기를 개별적으로 탐지하는 방법을 사용하였다. [4] 는 음악을 일정한 간격의 세그먼트로 나눈 뒤 각 세그먼트에 대한 퍼지 벡터를 추출하는 간략한 방법을 사용하였고, [2] 의 경우 Thayer 의 2 차원 모델의 이론적 특징을 기초로 한 분위기 탐색 모델에 바탕을 두었다. [2] 에서는 음악 전체에 대하여 우선 분위기 학습의 1 차 판별 조건인 음의 자극도 (에너지) 에 기반하여 곡 전체 에너지에 대한 평균을 기초로 에너지가 평균보다 낮은 지역과 높은 지역으로 나누었다. 그리고 2 차 판별 조건인 리듬과 음색 특징이 변화하는 구간을 기초로 하여 한 곡의 음악을 여러 파트로 나눈 뒤 각 파트의 분위기를 탐색해 나가는 방법을 사용하였다.

위 두 연구 중, [4] 는 이미 분위기 판정 모델 학습이 완료된 시점에서 단순히 음악을 일정 구간으로 잘라가며 각 구간의 분위기를 탐지하였고 [2] 는 분위기 판정 모델의 이론적 특성에 맞추어 구간의 경계를 세그먼트 하는 방법을 사용하였다.

그러나 음악은 일정한 음악학적 구조에 맞추어 비슷한 구간이 반복되고 일정한 규칙을 지니고 있다. 본 논문에서는 이러한 음악 구조적 특징에 기반을 두어 우선 음악 구조를 분석한 뒤 이를 바탕으로 음악 분위기를 탐지하는 방법에 대한 연구를 수행하였다. 전체적인 시스템의 구조는 (그림 2) 와 같다.

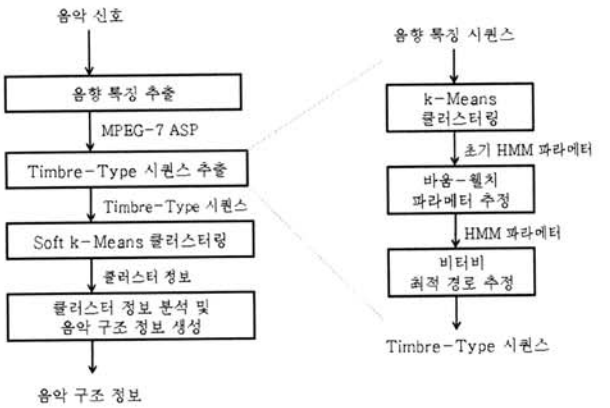


(그림 2) 시스템 전체 구조도

3.1 음악 구조 분석 및 세그먼트

음악의 구조 분석을 통한 세그먼트를 위하여 우선 음악 구조 정보를 추출하는 방법이 필요하다. 본 논문에서는 상태열 기반의 유사 구간 클러스터링 방법을 사용하였는데, (그림 3) 에서 보는 바와 같이 음악 특징 벡터 추출, Timbre-Type 시퀀스 추출, Timbre-Type Soft k-Means 클러스터링 방법을 통하여 유사 구간을 클러스터링 한다.

가상 코드 상태열 기반 음악 구조 분석방법에서는 음향의 특징들을 각각 음악의 가상 코드 값으로 매핑하여야 하기 때문에 음악의 코드 즉 멜로디를 표현하기 알맞은 특징의 선택이 필요하다. 본 논문에서는 멜로디를 표현하는데 있어



(그림 3) 유사 구간 클러스터링 방법을 통한 구조 분석

다른 특징에 비해 성능이 좋다고 알려진 MPEG-7 의 ASP (Audio Spectrum Projection) 특징을 1차 음향 특징으로 사용하였다.

1차 음향 특징은 각 프레임 별로 MPEG-7의 1/8 옥타브의 해상도를 가지는 ASE (Audio Spectrum Envelope) 를 추출 한 후, PCA (Principal Component Analysis) 차원축소 알고리즘을 통하여 상위 20 프로젝션 (ASP) 을 계산하여 사용한다. 하지만 PCA 알고리즘에 의해 정규화 된 ASP 값으로는 각 프레임의 에너지 차이에 대한 정보가 사라지기 때문에 각 프레임별 파워 스펙트럼 값의 L2-Norm 을 구하여 멜로디와 에너지 모두를 표현하는 총 21 차의 음향 특징 벡터 $\vec{x}=(x_1,...,x_{20},x_{21})$ 를 추출한다. 프레임별 음향 특징 벡터 $(\vec{x})$ 들이 추출되면 이를 N 개의 상태를 가지는 Timbre-Type $(q \in \{q_1,q_2,...,q_N\})$ 시퀀스로 매핑하기 위하여 N 개의 상태로 구성된 HMM 을 사용하게 된다. 프레임의 시간 수순에 의한 음향 특징 벡터열, 즉 관측열 $\vec{x}(1),\vec{x}(2),...\vec{x}(t),...,\vec{x}(T)$ 에 해당하는 은닉 상태열을 Timbre-Type 시퀀스 $q(1),q(2),...,q(t),...,q(T)$ 로 사용하게 된다.

음향 특징 벡터열을 HMM 에 적용시키기 위한 파라메터의 설정은 k-Means 클러스터링을 통해 N 개의 클러스터로 나눈 뒤, N 개의 은닉 상태 $(q_j, 1 \le j \le N)$ 의 관측 데이터 $(\vec{x})$ 에 대한 확률 $b_j(\vec{x})=p(\vec{x}|q_j)$ 는 가우시안 분산 $N(\vec{x},\vec{\mu}_j,\Sigma)$[1]으로, 초기 상태 확률 $\pi \in \{\pi_1,\pi_2,...,\pi_N\}$ 와 은닉 상태 치환 확률 $\{a_{ij}|1 \le i,j \le N\}$ 는 $1/N$ 로 초기화 한 뒤 관측열에 맞춘 파라메터 설정법인 바움-웰치 파라메터 추정법[2] 을 사용하여 HMM 의 모델 파라메터를 관측열에 최적화시켜 구하게 된다. HMM 의 모델이 구해진 후에는 최종적으로 비터비 최적 경로 추정 알고리즘을 통하여 관측열 $\vec{x}(1),\vec{x}(2),...\vec{x}(t),...,\vec{x}(T)$ 에 대하여 최적의 확률을 가지

1) 여기서 μ_j 는 각 은닉 상태 q_j 에 해당하는 음향 특징벡터 $\vec{x}$ 의 평균으로 구하고 공분산 Σ는 음향 특징 벡터열 전체의 공분산을 계산하여 모든 상태 확률 함수에서 공유한다.

2) 파라메터를 갱신할 시 은닉 상태 확률 $b_j(\vec{x})$ 은 $N(x,\mu_j,\Sigma)$ 을 사용, 평균만을 갱신한다.

는 은닉 상태열 $q(1), q(2), ..., q(t), ..., q(T)$ 를 추출하여 이를 Timbre-Type 시퀀스로 사용한다.

Timbre-Type 시퀀스가 추출된 후 Timbre-Type 시퀀스 정보를 이용하여 음악의 특징적 구간 (Segment) 으로 나누기 위해서, 히스토그램 기반 Soft k-Means 클러스터링 방법을 통해 M 개의 세그먼트 종류로 클러스터링하여 나누어지는 구간 생성법을 사용하였다. 히스토그램 기반 Soft k-Means 클러스터링을 위해서는 우선 Timbre-Type $q(1), q(2), ..., q(t), ..., q(T)$ 상태열에 대하여 W의 크기를 가지는 윈도우를 한 스텝씩 이동시키며 윈도우 내의 Timbre-Type에 대한 누적 히스토그램 $y(1), y(2), ..., y(t), ..., y(T)$ 를 생성한다. 각 y 의 값은 N 차원 벡터로 각 차원은 윈도우 내의 해당 Timbre-Type 의 발생 횟수를 나타내게 된다. 데이터 히스토그램 $y(1), y(2), ..., y(t), ..., y(T)$ 가 생성된 후에는 각 히스토그램에 세그먼트 라벨 $m\ (1 \le m \le M)$ 을 할당하여 세그먼트 시퀀스 $s(1), s(2), ..., s(T)$ 를 생성하게 된다. 본 논문에서는 반복적인 실험을 통하여 얻어낸 최적의 값인 7 과 12 를 각각 W 와 M 의 값으로 사용하였다.

세그먼트 라벨을 할당하기 위해서는 우선 각 세그먼트의 참조 히스토그램 h_m 를 랜덤하게 생성하고 이를 이용하여 각 데이터 히스토그램과 각 세그먼트의 참조 히스토그램 간의 유사도, 즉 세그먼트 소속 신뢰도를 식 (1) 과 같이 계산한다. 일단, 이 신뢰도를 기초로 하여 각 데이터 히스토그램에 대해서 소속 신뢰도가 가장 높은 세그먼트의 라벨을 할당한다. 동시에 국지적인 특성을 고려하여 데이터 히스토그램의 세그먼트 소속 신뢰도를 식 (2) 와 같이 갱신한다. 갱신된 소속 신뢰도를 바탕으로 각 데이터 히스토그램의 세그먼트 라벨을 재할당한다. 갱신된 소속 신뢰도와 데이터 히스토그램을 이용하여 각 세그먼트의 참조 히스토그램을 갱신한다. 이러한 과정을 지정된 횟수만큼 또는 할당 단계의 변화가 없을 때까지 반복한다. 본 논문의 실험에서는 국지화 이웃거리 상수로 $\lambda = 0.02$ 로, 루프 수렴 타협 관련 상수로 $\beta_0 = 100$, $\beta_{final} = 0.1$ 로 설정하여 충분히 반복되도록 하였다. d_{KL}은 쿨백-라이블러 발산(Kullback-Leibler divergence)을 나타낸다.

$$r_m(t) = \frac{\exp(-\beta d_{KL}(h_m, y(t)))}{\Sigma_{m'} \exp(-\beta d_{KL}(h_{m'}, y(t)))} \quad (1)$$

$$r'_m(t) = r_m(t) \exp(-\lambda n_m(t)) \quad (2)$$

3.2 음향 특징 추출

음악 분위기 학습 및 판정을 위한 음향의 특징으로는 2.3 절에서 나열한 기존 장르 및 스타일 분류분야에서 좋은 성능을 나타낸 음향 특징들인 Spectral Shape, MFCC, ASF, Spectral Contrast, Pitch Histogram, Beat Histogram, DWCHs 를 사용하였다. 음악의 특징 추출은 다음과 같이 단계 1~5 의 과정을 거쳐 이루어진다.

단계 1: 음악 파일로 부터 특징추출을 위한 신호특징을 담고 있는 PCM (Pulse-code modulation) 의 샘플을 읽어들인다.

단계 2: 음악 특징 추출이 일정한 크기의 프레임단위로 이루어지므로 이를 위해 일정한 크기의 프레임 크기만큼 잘라서 다음 모듈로 전달한다.

단계 3: 주파수 도메인의 특징을 추출하기 위하여 윈도우 함수 적용과 퓨리에 변환을 한다.

단계 4: Spectral Shape, MFCC, ASF, Spectral Contrast, Pitch Histogram, Beat Histogram, DWCHs 특징을 각각 추출한다.

단계 5: 추출한 모든 특징을 하나의 데이터로 통합하여 최종적인 특징 포맷으로 만든다.

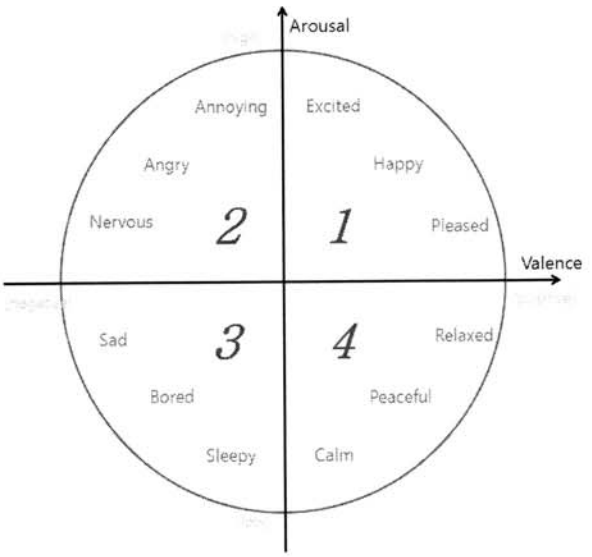
3.3 분위기 학습 및 판별 방법

3.3.1 개인별 음악 분위기 데이터 수집

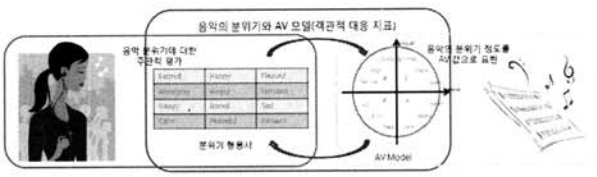
곡에 대한 분위기 평가는 특성상 사람의 인지와 개인적 취향, 환경/문화적 차이에 의해 개인별로 달라질 수밖에 없다. 이러한 특성을 무시하고 공통된 분위기 판별 모델을 구축하여 사용하게 되면 성능이 떨어질 수밖에 없다. 따라서 본 논문에서는 개인별로 분위기에 대한 평가정보를 받고 이를 기반으로 개인별 분위기 판별 모델을 구축하여 사용하였다. 분위기 판별 모델은 회귀분석 방법을 사용하였다.

[11, 12]에서처럼 음악 분위기 형용사가 아닌 분위기를 나타내는 직접적인 AV 계수를 피실험자가 입력하는 방식을 사용할 수도 있지만 일반적으로 음악에 대한 전문성이 없는 일반인이 직접 음악의 분위기를 나타내는 AV 계수를 입력하는데 무리가 있다고 판단되어 본 실험에서는 피실험자가 청취한 음악의 일부에 대해 느낀 분위기를 주어진 분위기 형용사 항목 중에서 선택하고, 이를 토대로 내부적으로 AV 계수로 변환하여 사용하는 방법을 채택하였다. 본 논문에서는 분위기를 나타낼 수 있는 대표적인 형용사 12 개를 선정하여, 사용자가 선택한 분위기와 AV 매핑을 위하여 세분화된 Thayer 의 2 차원 분위기 모델의 형태인 (그림 4)와 같이 설정하였다. 분위기 형용사와 AV 계수 사이의 대응 지표를 설정하는 간략한 개요는 (그림 5)와 같다.

사용자가 비전문가의 입장에서 분위기 평가정보를 입력하기 편리하도록 각 피실험자에게 (그림 6)의 사용자 평가 프로그램을 제공하여 각 피실험자의 음악에 대한 분위기 평가 항목을 수집하였다. (그림 6)의 평가 폼을 보게 되면 각 분위기 형용사와 함께 AV 맵 또한 제공하고 있다. 이는 피실험자가 형용사만으로 분위기를 선택할 시 본 논문에 사용하는 분위기 분류법인 AV 맵과 차이가 나는 평가가 이루어질 수 있으므로 이를 방지하기 위하여 사전에 각 피실험자들에게 간단한 AV 분위기 분류법에 대하여 교육을 한 뒤, 음악의 AV 강도를 고려하여 분위기 형용사를 선택하게끔 하였다. 각 음악 구간에 대한 평가는 총합이 5 인 평가값을 피실험자가 여러 형용사에 걸쳐 분배하는 방식으로 평가하도록 하였는데, 이때 평가값은 동시에 5 개 이상의 형용사에



(그림 4) 세분화된 분위기로 개정된 Thayer 의 2 차원 모델



(그림 5) 음악 분위기와 AV 모델 대응 개념

분배하지 못하고 또한 평가값이 상반된 형용사에 함께 분배되는 것을 회피하도록 권고하였다.

3.3.2 주관적 분위기 학습 및 판별 방법

음악의 구조 정보와 각 구조의 사용자의 주관적 테스트 평가 정보, 그리고 구간별 추출된 음향 특징을 이용하여 분위기를 학습하기 위해서 본 논문에서는 회귀 모델을 기초로 하여 학습하도록 하였다. 회귀 모델을 위한 학습 알고리즘으로는 이진 클래스 분류에 있어 좋은 성능을 보여준 SVM (Support Vector Machine) 을 회귀 학습에 적용한 SVR (Support Vector Regression) 을 사용하였다.

(그림 6) 사용자 평가 프로그램 - 재생 및 평가 폼

비전문가인 사용자가 선택한 평가 정보를 내부적으로 AV 계수로 변화하기 위하여 본 논문에서는 Thayer 의 2 차원 분위기 모델의 각 차원 축인 A (Arousal) 와 V (Valence) 의 값을 -1 과 1 사이의 값으로 두고 각각에 대하여 SVR 을 통한 회귀 분석 학습과 평가를 수행하였다. 하지만 3.3.1 절에 기술한 바와 같이 수집된 개인별 음악 분위기 평가 데이터가 [11, 12] 와는 다르게 AV 를 직접 입력하는 대신에 AV 계수의 강도를 고려하여 분위기 형용사에 평가하는 방식을 사용하였기 때문에 삼각함수 공식을 적용한 식 (3)과 식 (4)의 방법을 사용하여 AV 값 (A_{val}, V_{val}) 을 계산하여 사용하였다. 사용자가 선택한 각각의 형용사에 대하여 고유한 값을 부여하기 위하여, (그림 6) 의 평가폼에 위치한 각 형용사의 각도위치정보를 삼각함수를 이용하여 형용사에 부여한 가중치 점수를 반영한 AV 값으로 계산하였다.

최종적으로는 구축된 A와 V의 SVR 모델을 이용하여 새로운 음악 구간에 대한 $A_{predict}$ 와 $V_{predict}$ 값을 예측하게

$$A_{val} = \frac{\begin{array}{l}(\sin(75^\circ)\times Excited_{val} + \sin(45^\circ)\times Happy_{val} + \sin(15^\circ)\times Pleased_{val} + \\ \sin(345^\circ)\times Relaxed_{val} + \sin(315^\circ)\times Peaceful_{val} + \sin(285^\circ)\times Calm_{val} + \\ \sin(255^\circ)\times Sleepy_{val} + \sin(225^\circ)\times Bored_{val} + \sin(195^\circ)\times Sad_{val} + \\ \sin(165^\circ)\times Nervous_{val} + \sin(135^\circ)\times Angry_{val} + \sin(105^\circ)\times Annoying_{val})\end{array}}{5} \quad (3)$$

$$V_{val} = \frac{\begin{array}{l}(\cos(75^\circ)\times Excited_{val} + \cos(45^\circ)\times Happy_{val} + \cos(15^\circ)\times Pleased_{val} + \\ \cos(345^\circ)\times Relaxed_{val} + \cos(315^\circ)\times Peaceful_{val} + \cos(285^\circ)\times Calm_{val} + \\ \cos(255^\circ)\times Sleepy_{val} + \cos(225^\circ)\times Bored_{val} + \cos(195^\circ)\times Sad_{val} + \\ \cos(165^\circ)\times Nervous_{val} + \cos(135^\circ)\times Angry_{val} + \cos(105^\circ)\times Annoying_{val})\end{array}}{5} \quad (4)$$

되는데 응용에 따라 ($A_{predict}$, $V_{predict}$) 쌍을 새로운 음악 구간에 대한 분위기로 사용하거나 또는 분위기 형용사에 매핑하여 사용하여도 된다.

4. 실험 및 성능 평가

본 논문에서 사용하는 음향 특징들은 아래 <표 1>과 같다. 각 특징들은 입력 도메인의 종류 (주파수 도메인, 신호 도메인) 에 맞게 별도의 최종 후처리 과정을 거친다. Pitch Histogram 과 Beat Histogram 그리고 DWCHs 의 경우는 별도로 지정된 후처리 과정을 수행하였고 ASF 에 대해서는 MPEG-7 의 권고사항에 따른 후처리 방법을 사용하였다. 이 방법에서는 프레임으로부터 추출된 특징들의 평균과 표준편차를 사용한다. 그리고 나머지 프레임기반 특징들에 대해서는 텍스쳐 윈도우 (Texture Window) 방법을 사용하여 후처리를 하였다. 텍스쳐 윈도우 방법은 분석 윈도우 (Analysis Window) 단위로 퓨리에 변환을 적용한 후 추출한 국부적인 특징들에 대하여 텍스쳐 윈도우 단위로 표준 분산적 특징 (평균, 표준편차) 을 추출한 후 다시 이를 이용하여 전체의 표준 분산적 특징을 추출하게 된다.

음원 데이터베이스로 락, 메탈, 리듬 앤 블루스, 발라드, 재즈 등의 다양한 장르로 구성된 팝 음악과 국내 음악 그리고 뉴에이지로 이루어진 100 개의 음악을 사용하였다. 음악 데이터의 포맷은 범용적으로 사용되어 지는 음악 포맷인 44,100 Hz 샘플링레이트 스테레오 채널의 MP3 파일 형식이며 이로부터 구조분석 방법을 사용하여 구간들을 추출하였다.

분위기에 대한 주관적인 평가를 위해 음원 데이터베이스 내의 각 음악에 3.1 절의 구조 분석방법을 적용하여 추출한 319 개의 구간을 피실험자에게 제공하였다. 이때 동일한 곡에 속하는 구간들을 연속해서 들려주었을 시 피실험자가 이전에 들은 구간에 대한 느낌때문에 잘못된 평가를 내릴 수 있다. 이러한 상황을 방지하기 위해 구간들을 무작위 순서로 제공하였다.

실험에 참여한 피실험자는 두 그룹으로 나눠지며 모두 사전에 Thayer 의 2 차원 분위기 모델에 대한 이론적 내용에 대하여 교육을 실시한 뒤 이중 Group-1 은 3.3.1 절에 언급한 권고사항을 고려하여 평가를 수행하도록 하였고 Group-2 는 제약사항을 꼭 지킬 필요 없이 본인의 느낌을 바탕으로 자유롭게 평가하도록 하였다.

〈표 1〉 실험에 사용하는 음향 특징

특징	프레임 분리 설정		추출 도메인		후처리
	프레임 크기	홉 사이즈	윈도우 함수	도메인	
Spectral Shape	2,048 (46ms)	1,764 (40ms)	Hamming	주파수	Texture Window (15/10)
MFCC	2,048 (46ms)	1,764 (40ms)	Hamming	주파수	Texture Window (15/10)
ASF	2,048 (46ms)	1,764 (40ms)	Hamming	주파수	기본 평균 / 표준편차
Spectral Contrast	2,048 (46ms)	1,764 (40ms)	Hamming	주파수	Texture Window (15/10)
Pitch Histogram	1,024 (23ms)	882 (20ms)	Hamming	혼합	피치 히스토그램
Beat Histogram	131,072 (약 3초)	44,100 (1초)	X	웨이블릿	비트 히스토그램
DWCHs	524,288 (약 12초)	X	X	웨이블릿	히스토그램 분산 특징(모멘트)

4.1 SVR 학습 평가

4.1.1 평가 척도

일반 클래스 학습이 아닌 연속적인 값을 학습하기 위하여 본 논문에서 사용한 회귀 학습의 평가를 위해서는 MSE (Mean Square Error) ϵ을 추정하는 방법을 사용할 수 있다. 학습을 위한 N 개의 입력 데이터로 (반응변수, 설명변수) 인 (y_i, x_i), $i=1,2,...,N$ 가 주어진다면, 이를 분석하여 새로운 설명변수 $\hat{x}$에 대한 반응 변수 $\hat{y}$를 추론하는 회귀 모형을 $R(\bullet)$이라 하고 이에 회귀 모형 $R(\bullet)$에 대한 MSE 를 구하는 식은 식 (5) 와 같다.

$$\epsilon = \frac{1}{N}\sum_{i=1}^{N}(y_i - R(x_i))^2 \tag{5}$$

하지만 회귀 분석의 경우 MSE 를 학습 평가에 사용하기보다는 학습된 회귀 모형과 주어진 입력 데이터간의 분산학적인 적합도를 나타내는 R^2 결정계수를 사용하는 것이 일반적이다. R^2 결정계수를 구하는 식은 식 (6) 과 같다.

$$R^2 = 1 - \frac{N \times \epsilon}{\sum_{i=1}^{N}(y_i - \bar{y})} \tag{6}$$

식 (6)에서 $\bar{y}$는 학습데이터로 주어진 반응변수 y_i 들의 평균값이다. 식 (6)에서 $\bar{y}$와 MSE ϵ을 구하는 과정을 생략하고 반응변수 y_i와 회귀 모델 $R(\bullet)$ 만을 사용하여 R^2 결정계수를 구하는 식으로 유도된 공식은 식 (7)이다. R^2 값이 1 에 가까울수록 회귀 모델이 설명변수와 반응변수 사이의 관계를 잘 모델링하고 있다는 의미이다.

$$R^2 = \frac{(N\Sigma R(y_i)y_i - \Sigma R(y_i)\Sigma y_i)^2}{N\Sigma(R(y_i))^2 - (\Sigma R(y_i))^2 \times N\Sigma y_i^2 - (\Sigma y_i)^2} \tag{7}$$

본 논문에서는 개인별 분위기 학습에 대한 성능의 평가로 R^2 결정계수를 평가의 척도로 사용하였다. 즉, 각 피실험자

별 학습데이터를 사용하여 Arousal 에 대한 SVR 인 $SVR_{arousal}$ 과 Valence 에 대한 SVR 인 $SVR_{valence}$ 를 구축하고 이를 테스트 데이터에 적용시켜 R^2 결정계수를 계산하였다.

4.1.2 교차 검증

클래스 기반의 학습이나 회귀 분석 등의 성능 평가를 위해서는 학습데이터와 테스트데이터를 나누어 성능평가가 이루어진다. 하지만 주어진 데이터들이 적어 학습데이터와 테스트데이터를 나누기 곤란하거나 주어진 데이터만을 최대한 활용하여 가장 적절한 성능평가를 하기 위해서는 일반적으로 교차 검증 (Cross-validation) 을 사용하여 평가를 하게 된다. 본 논문에서는 총 319 개의 데이터 집합에 관하여 SVR 의 성능 평가를 위하여 일반적으로 사용되는 5-fold Cross-validation 을 사용하여 SVR 의 학습 성능을 살펴보았다.

4.1.3 그리드 파라메터 선택

SVM 과 SVR 의 경우 초기 설정 파라메터에 따른 성능의 차이가 큰 특징이 있다. 특히 비선형 커널을 사용할 경우 커널 파라메터에 따른 성능의 편차는 더욱 크게 된다. 이러한 특징으로 인해 주어진 데이터 집합에 대한 최적의 성능을 유도하기 위해서는 최적의 파라메터를 찾는 것이 중요하다. 그리드 파라메터 선택 (Grid Parameter Selection) 방법은 SVM 과 SVR 의 최적의 파라메터를 찾기 위하여 사용하게 되는데 그 방법은 현재 SVM/SVR 타입에 따른 파라메터와 커널 함수의 파라메터의 최소값과 최대값을 지정한 후 일정 단위로 증가시키며 각각의 파라메터를 적용하여 성능 평가를 한 뒤 성능이 가장 좋은 파라메터를 선택하는 방법이다.

본 논문에서도 가장 좋은 성능을 탐색하기 위해 SVR 의 타입은 $\epsilon-SVR$ 과 $\nu-SVR$ 을, 그리고 SVR 의 커널 타입으로는 선형 커널과 RBF 커널을 사용하여 각 타입별 그리드 파라메터 선택 방법을 적용하여 최적의 파라메터와 그 성능을 살펴보았다.

4.1.4 실험 결과

각 사용자별 SVR 과 커널의 타입별 최적의 파라메터와 학습 성능에 대한 실험결과를 <표 2>와 <표 3>에 나타내었다. <표 2>와 <표 3>에서 Group-1 은 3.3.1 절의 권고사항을 지키며 평가한 피실험자 그룹이며 Group-2 는 권고사항을 신경 쓰지 않고 평가한 피실험자 그룹을 나타낸다.

<표 2>에서 보는 바와 같이 $SVR_{arousal}$ 의 경우는 전체적으로 0.6~0.7 사이의 R^2 값을 보여주는 반면에 $SVR_{valence}$ 의 경우는 <표 3>에서 보는 바와 같이 0.1~0.5 사이의 비교적 낮은 값과 함께 개인별 편차가 크게 나오고 있다. 이는 2차원 분위기 분류법에 있어 음의 활력 즉 에너지에 관한 특징인 Arousal 의 경우는 피실험자들의 개별적인 판단과 음악으로부터 추출한 특징간의 관계가 회귀 분석을 통하여 일정 이상의 모델링이 가능하나, 음의 안정도를 나타내는 Valence 의 경우는 개인의 음악적 이해도에 따라 평가 항목의 일관성이 부족하여 회귀 분석을 통한 모델링이 제대로 되지 않음을 의미한다.

앞에서 Group-1 과 Group-2 로 나누어 실험한 이유는 분위기 정보 입력 시 3.3.1 절의 제약사항이 주는 효과를 파악하기 위해서이다. <표 2>와 <표 3>에서 보는 바와 같이 Arousal 에 대해서는 제약사항이 부정적으로 작용하는 것으로 보이나 Valence 에 대해서는 반대로 긍정적으로 작용하는 것으로 보인다. 이는 학습 성능과 연관이 있는 것으로 보이며, 앞에서 언급하였듯이 Valence 의 경우는 개인의 음악적 이해도에 따라 평가 항목의 일관성이 부족하여 제약사항을 가하지 않았을 경우 그 정도가 더 심해지는 것으로 보

〈표 2〉 피실험자별 $SVR_{arousal}$의 최적의 파라메터 성능

피실험자		SVR[1]			SVR[2]			SVR[3]				SVR[4]			
		파라메터		R^2	파라메터		R^2	파라메터			R^2	파라메터			R^2
		lnC	lnϵ		lnC	lnν		lnC	lnϵ	lnγ		lnC	lnν	lnγ	
Group 1	A	-4	-1	0.6346	-4	-1	0.6319	0	-1	-4	0.6396	-1	-1	-3	0.6352
	B	-4	-3	0.6304	-4	-1	0.6286	0	-2	-3	0.6621	0	-1	-3	0.6527
	C	-3	-3	0.6296	-3	-1	0.6248	0	-5	-3	0.6394	1	-1	-5	0.6249
Group 2	D	-4	-1	0.6846	-4	-1	0.6822	0	-2	-3	0.6981	0	-1	-3	0.6961
	E	-4	-2	0.6970	-4	-1	0.6863	0	-6	-3	0.7106	0	-1	-3	0.6995

SVR[1] : $\epsilon-SVR$, 선형커널
SVR[2] : $\nu-SVR$, 선형커널
SVR[3] : $\epsilon-SVR$, RBF커널
SVR[4] : $\nu-SVR$, RBF커널

〈표 3〉 피실험자별 $SVR_{valence}$의 최적의 파라메터 성능

피실험자		SVR[1]			SVR[2]			SVR[3]				SVR[4]			
		파라메터		R^2	파라메터		R^2	파라메터			R^2	파라메터			R^2
		lnC	lnϵ		lnC	lnν		lnC	lnϵ	lnγ		lnC	lnν	lnγ	
Group 1	A	-3	-2	0.2491	-2	-1	0.2393	0	-3	-1	0.3079	1	-1	-1	0.3061
	B	-3	-2	0.2534	-3	-1	0.2471	0	-4	-3	0.2558	1	-1	-5	0.2502
	C	-2	-2	0.4840	-2	-1	0.4449	0	-2	-2	0.5268	1	-1	-2	0.5300
Group 2	D	-3	-1	0.1098	-3	-1	0.1081	0	-2	-2	0.1827	0	-1	-2	0.1733
	E	-3	-1	0.2449	-3	-1	0.2384	0	-4	-1	0.3214	0	-1	-1	0.3195

SVR[1] : $\epsilon-SVR$, 선형커널
SVR[2] : $\nu-SVR$, 선형커널
SVR[3] : $\epsilon-SVR$, RBF커널
SVR[4] : $\nu-SVR$, RBF커널

인다.

4.2 AV 공간 정보를 이용한 분위기 유사도 평가

4.2.1 평가 척도

실생활에서 음악의 내용을 바탕으로 한 분위기 탐지 문제에 있어 음악 분위기와 관계한 AV 계수 값을 정확히 예측하는 문제보다는 실제 사용자가 느끼는 음악의 분위기를 파악하는 방법이 더 중요할 것이다. 즉, 피실험자로부터 음악 분위기 정보를 입력 받을 시 AV 값을 직접 입력 받지 않고 대신에 분위기 형용사를 이용한 것처럼, 분위기 판별 성능을 평가할 시에도 이러한 점을 고려할 필요가 있다. 이를 위해서 본 논문에서는 SVR 회귀 분석기를 통해 예측된 AV 벡터 값과 피실험자가 입력한 분위기 정보가 얼마나 비슷한 분위기 형용사 그룹에 있는지를 측정하고자 하였다.

같은 분위기 형용사 그룹에 속하는 AV 벡터 값일수록 두 벡터가 이루는 각은 0도에 가까워지기 때문에 본 논문에서는 분위기 유사도 측정 방법으로 코사인 유사도 공식을 응용한 식 (8) 을 사용하였다. 식 (8) 은 -1~1 의 범위의 값을 갖는 코사인 유사도를 0~1 사이의 값으로 매핑하기 위한 공식으로 1 에 가까울수록 유사함을, 0 에 가까울수록 유사 관계가 적음을 나타낸다.

$$AV_{predictrates}=\frac{\cos\theta+1}{2},\ \cos\theta=\frac{AV_{origin}\cdot AV_{predict}}{\|AV_{origin}\|\,\|AV_{predict}\|} \quad (8)$$

여기서, AV_{origin} 은 사용자가 평가한 분위기 값으로부터 식 (3) 과 식 (4) 에 의해 계산된 AV 의 값이며 $AV_{predict}$ 는 SVR 모델에 의해 예측된 AV 의 값이다.

4.2.2 실험 결과

앞 절의 분위기 유사도 척도를 사용하여 <표 2>와 <표 3>의 결과 중 개인별 가장 좋은 성능을 나타낸 SVR 타입과 파라메터를 사용할 경우의 분위기 유사도와 가장 낮은 성능을 나타낸 SVR 타입과 파라메터를 사용할 경우의 분위기 유사도를 살펴보았다. 그 결과는 <표 4> 와 같다.

<표 4>에서 보듯이 AV 계수의 2 차원 분위기 분류공간에서의 유사도 측정법인 코사인 유사도 공식을 이용하였을 시 전체적으로 82% 이상의 유사도를 보여주고 있다. 이는 독립적으로 회귀 분석 모델을 사용하여 AV 계수를 독립적으로 예측한 뒤 다시 예측된 A 와 V 를 사용한 분위기 분류 공간의 공간적 정보를 사용하여 분위기를 판별하게 되면 좋은 성능을 보여줄 수 있음을 나타낸다. 또한, 유사도 평가 척도를 사용했을 경우는 3.3.1 절의 제약사항의 효과가 크게 반영되지 않음을 알 수 있다.

〈표 4〉 개인별 SVR 모델로부터 추출된 AV 계수의 분위기 유사도

피실험자		분위기 탐지율	
		최소 R^2 파라메터 선택	최대 R^2 파라메터 선택
Group 1	A	0.8273	0.8427
	B	0.8474	0.8522
	C	0.8732	0.8859
Group 2	D	0.8308	0.8422
	E	0.8289	0.8486

5. 결론 및 향후과제

본 논문에서는 음악의 분위기 탐지를 위하여 기존 내용기반 음악 추천 및 분류 연구들에서 사용하는 수작업을 통한 일정길이 선택 방식이 아닌 실제 우리가 접하고 있는 음악

을 구조분석 방법을 통하여 의미 있는 구간으로 나누고 이를 바탕으로 개인별 음악 분위기를 탐지하는 방법에 대해 연구하였다.

음악 분위기를 판별함에 있어 음악 분위기의 이산적인 클래스를 고려한 전통적인 방법을 사용하지 않고 Thayer 의 2 차원 모델에 기반하여 먼저 각 AV 모델의 차원별 회귀 모델을 학습한 후 이를 이용하여 AV 값을 예측하고 이를 분위기 형용사와의 공간적 유사도를 고려하는 방법을 사용하였다. 그리고 개인별 분위기 판별 성능을 수정된 코사인 유사도 공식을 통하여 분석하였다. 실험 결과, 개인별 평가에 맞춘 음악 분위기 판별율은 개인별로 편차가 있기는 했으나 전체적으로 평균 유사도가 82%에서 89%정도의 높은 유사도를 보여주었다.

본 논문의 주요한 기여사항은 다음과 같다.

- 음악 구조 분석 기법을 적용하여 세그먼트 추출을 자동화하였다.
- 변화하는 음악의 특징을 고려하여 단일의 분위기 클래스가 아닌 각 부분들의 독립된 분위기를 탐지하는 기법을 제안하였다.
- 기존의 연구들은 개인의 주관적 성향을 고려하지 못하였으나, 본 논문에서는 개인별 상황을 인지하여 음악을 추천하기 위한 맞춤식 학습이 가능하도록 하였다.
- 비전문가의 특성을 고려하여, 음악의 특징을 2 차원 맵을 통하여 쉽게 평가할 수 있는 기법을 제안하였다.
- 기존의 연구들이 SVR 에 대한 학습 평가에서 그쳤으나, 본 논문에서는 더욱 정확한 평가를 위하여 AV 계수 값이 아닌 실제 사용자가 느끼는 음악의 분위기를 파악하여 평가하기 위한 AV 공간 정보를 이용한 평가 방법을 제안하였다.

본 논문에서는 상황인지 음악 추천을 위한 전단계로서 음악의 음향 데이터에서 분위기 정보 (AV 값) 를 추출하는 연구를 수행하였다. 향후, 상황별로 사용자가 좋아하는 음악들의 AV 특성을 파악하고 이를 바탕으로 상황정보를 고려한 내용 기반 상황 인지 음악 추천 방법에 대하여 연구할 계획이다.

참 고 문 헌

[1] T. Li and M. Ogihara, "Detecting Emotion in Music," Proc. of the International Symposium on Music Information Retrieval(ISMIR), pp.239-240, Washington D.C., USA, 2003.

[2] L. Lu, D. Liu and H. Zhang, "Automatic Mood Detection and Tracking of Music Audio Signals," IEEE Trans. on Audio, Speech, and Language Processing, Vol..14, pp.5-18, 2006.

[3] Y. Feng, Y. Zhang and Y. Pan, "Popular Music Retrieval by Detecting Mood," Proc. of ACM SIGIR 2003, pp.375-376, 2003.

[4] Y.H. Yang, C.C. Liu and H.H. Chen, "Music Emotion Classification: a Fuzzy Approach," Proc. of ACM Multimedia 2006 (ACM MM'06), pp.81-84, Santa Barbara, CA, USA, 2006.

[5] R.E. Thayer, "The Biopsychology of Mood and Arousal", Oxford University Press, 1989.

[6] H. Katayose, M. Imai and S. Inokuchi, "Sentiment Extraction in Music," Proc. of International Conference Pattern Recognition, Vol.2, pp.1083-1087, 1998.

[7] D. Liu, N. Zhang and H. Zhu, "Form and Mood Recognition of Johann Strauss's Waltz Centos," Chinese Journal of Electronics, Vol.12, Part.4, pp.587-593, 2003.

[8] D. Hevner, "Experimental Studies of the Elements of Expression in Music," American Journal of Psychology, Vol.48, pp.246-268, 1936.

[9] P.R. Farnsworth, "The Social Psychology of Music", The Dryden Press, 1958.

[10] T. Li and M. Ogihara, "Content-based Music Similarity Search and Emotion Detection," Proc. of ICASSP '04, Vol.5, pp.705-708, 2004.

[11] Y.H. Yang, Y.F. Su, Y.C. Lin and H.H. Chen, "Music Emotion Recognition: the Role of Individuality," Proc. of ACM SIGMM International Workshop on Human-centered Multimedia 2007, pp.13-21, Augsburg, Germany, 2007.

[12] Y.H. Yang, C.C. Liu and H.H. Chen, "A Regression Approach to Music Emotion Recognition," IEEE Trans. on Audio, Speech, and Language Processing, Vol.16, pp.448-457, 2008.

[13] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals," IEEE Trans. on Speech and Audio Processing, Vol.10, No.5, pp.293-302, 2002.

[14] J. J. Burred and A. Lerch, "A Hierarchical Approach to Automatic Musical Genre Classification," Proc. of the 6th International Conference on Digital Audio Effects (DAFx-03), 2003.

[15] J. J. Burred and A. Lerch, "Hierarchical Automatic Audio Signal Classification," Journal of the Audio Engineering Society, Vol.52, No.7/8, pp.357-365, 2004.

[16] D. Jiang, L. Lu, H. Zhang, J. Tao and L. Cai, "Music Type Classification by Spectral Contrast Feature," Proc. of ICME '02, Vol.1, pp.113-116, 2002.

[17] T. Tolonen and M. Karjalainen, "A Computationally Efficient Multipitch Analysis Model," IEEE Trans. on Speech Audio Processing, Vol.8, pp.708-716, Nov. 2000.

[18] T. Li, M. Ogihara and Q. Li, "A Comparative Study on Content-based Music Genre Classification," Proc. of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.282-289, 2003.

[19] Y. K. Kim and Y. Brian, "Singer Identification in Popular Music Recordings Using Voice Coding Features," Proc. of International Conference on Music Information Retrieval, 2002.

[20] T. Zhang, "Automatic Singer Identification," Proc. of IEEE International Conference on Multimedia and Expo, IEEE CS

Press, 2003.
[21] X. Shao, N.C. Maddage, C. Xu and M.S. Kankanhalli, "Automatic Music Summarization Based on Music Structure Analysis," Proc. of ICASSP'05, Vol.2, pp.1169-1172, 2005.
[22] Y. Shiu, H. Jeong and C.-C. J. Kuo, "Musical Structure Analysis Using Similarity Matrix and Dynamic Programming," Proc. of SPIE, Multimedia systems and applications, Vol.3, pp.398-409, 2005.
[23] J. Paulus and A. Klapuri, "Music Structure Analysis by Finding Repeated Parts," Proc. of ACM AMCMM'06, pp.59-67, 2006.
[24] M. Goto, "SmartMusicKIOSK: Music Listening station with Chorus-Search Function," Proc. of the 16th annual ACM symposium on User Interface Software and Technology, pp.31-40, 2003.
[25] S. Abdallah, K. Nolad, M. Sandler, M. Casey and C. Rhodes, "Theory and Evaluation of a Bayesian Music Structure Extractor," Proc. of 6th International Conference on Music Information Retrieval London, UK, Sept. 2005.
[26] M. Levy, M. Sandier and M. Casey, "Extraction of High-Level Musical Structure From Audio Data and Its Application to Thumbnail Generation," Proc. of ICASSP'06, Vol.5, pp.13-16, Toulouse, France, May 2006.
[27] M. Levy, M. Sandier and M. Casey, "Structural Segmentation of Musical Audio by Constrained Clustering," IEEE Trans. on Audio, Speech, and Language Processing, Vol.16, pp.318-326, 2008.
[28] L. Lu and H. Zhang, "Automated Extraction of Music Snippets," Proc. of the 11'th ACM International Conference on Multimedia, pp.140-147, 2003.
[29] T. Zhang and R. Samadani, "Automatic Generation of Music Thumbnails," Proc. of IEEE International Conference on Multimedia and Expo, pp.228-231, 2007.
[30] G. Peeters, "Deriving Musical Structure from Signal Analysis for Music Audio Summary Generation: "Sequence" and "State" Approach," In Lecture Notes in Computer Science, Vol.2771, pp.143-166. Springer-Verlag, 2004.

이 종 인

e-mail : inisphier.sof@gmail.com
2007년 금오공과대학교 컴퓨터공학부(학사)
2009년 금오공과대학교 소프트웨어공학과 (공학석사)
2009년~현 재 슈어소프트테크(주)
관심분야 : 인공지능, 패턴인식, 소프트웨어 공학

여 동 규

e-mail : sylot@kumoh.ac.kr
1999년 금오공과대학교 컴퓨터공학과(학사)
2001년 금오공과대학교 컴퓨터공학과(공학석사)
2010년 금오공과대학교 컴퓨터공학과(공학박사)
관심분야 : 정보보호, 디지털 워터마킹, 디지털 포렌식, 인공지능 등

김 병 만

e-mail : bmkim@kumoh.ac.kr
1987년 서울대학교 컴퓨터공학과(학사)
1989년 한국과학기술원 전산학과(공학석사)
1992년 한국과학기술원 전산학과(공학박사)
1992년~현 재 금오공과대학교 교수
1998년~1999년 미국 UC, Irvine 대학 방문교수
2005~2006 미국 콜로라도 주립대학 방문교수
관심분야 : 인공지능, 정보검색, 정보보안