다중 모드 음악 감정 인식: 새로운 데이터 세트, 방법론 및 비교 분석

R. 팬더[1], R. 말헤이로[1], B. 로차[1], A. 올리베이라[1]그리고 RP 파이바[1],

[1]CISUC – 포르투갈 코임브라 대학교 정보학 및 시스템 센터
{panda, rsmal, bmrocha, apsimoes, ruipedro}@dei.uc.pt

추상적인.우리는 오디오, MIDI 및 가사와 같은 서로 다른 소스의 정보를 결합하여 음악 감정 인식(MER) 문제에 대한 다중 모드 접근 방식을 제안합니다. MIREX 기분 분류 작업에 사용된 감정 태그를 기반으로 AllMusic 데이터베이스를 활용하여 다중 모드 음악 감정 데이터 세트를 자동 생성하는 방법론을 소개합니다. 그런 다음, 획득한 오디오 샘플의 하위 집합에 해당하는 MIDI 파일과 가사를 수집했습니다. 데이터 세트는 MIREX에 정의된 것과 동일한 5개의 감정 클러스터로 구성되었습니다. 오디오 데이터에서 177개의 표준 특징과 98개의 멜로디 특징이 추출되었습니다. MIDI의 경우 320개의 기능이 수집되었습니다. 최종적으로 26개의 서정적 특징이 추출되었다. 우리는 제안된 다중 모드 접근 방식을 평가하기 위해 여러 지도 학습 및 기능 선택 전략을 실험했습니다. 표준 오디오 기능만 사용했을 때 달성된 최고 성능은 44.3%(F-측정)였습니다. 다중 모드 접근 방식을 사용하면 19개의 다중 모드 기능만 사용하여 결과가 61.1%로 향상되었습니다. 멜로디 오디오 기능은 이러한 개선에 특히 중요했습니다.

키워드:음악 감정 인식, 기계 학습, 다중 모드 분석.

1. 소개

현재 음악 저장소에는 개별 사용자의 요구에 맞춰 맞춤화된 고급스럽고 유연한 검색 메커니즘이 부족합니다. 이전 연구에서는 “음악의 가장 중요한 기능은 사회적, 심리적 기능”이라는 사실을 확인했으며, 따라서 “가장 유용한 검색 지표는 이러한 사회적, 심리적 기능에 맞춰 검색을 촉진하는 지표입니다. 일반적으로 이러한 인덱스는 스타일, 분위기 및 유사성 정보에 중점을 둡니다.”[8] 이는 감정을 음악 검색 및 구성의 중요한 기준으로 식별한 음악 정보 행동에 대한 연구에 의해 뒷받침됩니다[4].

음악감정인식(MER) 연구는 최근 몇 년간 주목을 받고 있습니다. 그럼에도 불구하고 이 분야는 특히 오디오 음악 신호의 감정 감지와 관련하여 여전히 많은 한계와 공개적인 문제에 직면해 있습니다. 실제로 현재 오디오 MER 시스템의 현재 정확도는 개선의 여지가 많다는 것을 보여줍니다. 예를 들어 MIR(Music Information Retrieval) 평가 eXchange(MIREX)의 기분 분류 작업에서 가장 높은 분류 정확도를 얻은 것은 67.8%였습니다.

MER의 주요 어려움 중 일부는 노래가 불러일으키는 감정에 대한 인식이 본질적으로 주관적이라는 사실과 관련이 있습니다. 즉, 같은 노래를 들을 때 사람들마다 서로 다른 감정을 인식하는 경우가 많습니다. 게다가 청취자들이 인지된 감정에 동의하더라도 그 설명(예: 사용된 형용사)에 대해서는 여전히 많은 모호함이 있습니다. 또한 음악 요소가 청취자에게 특정한 감정적 반응을 생성하는 방법과 이유가 아직 잘 이해되지 않았습니다[30].

또 다른 문제는 표준적이고 우수한 품질의 오디오 감정 데이터 세트가 부족하다는 것입니다. 이러한 이유로 대부분의 연구에서는 각 저자가 생성한 서로 다른 데이터 세트를 사용하므로 결과를 비교할 수 없습니다. 이 문제를 해결하기 위해 MIREX 기분 분류 데이터 세트라는 몇 가지 노력이 개발되었습니다. 그러나 이 데이터세트는 공개적으로 이용 가능하지 않으며 MIREX 평가에만 사용됩니다.

이 연구의 주요 목표는 MER에 대한 다중 모드 접근 방식이 소위 유리 천장 효과를 깨는 데 어느 정도 효과적일 수 있는지 평가하는 것입니다. 실제로 표준 오디오 기능에만 기반을 둔 대부분의 현재 접근 방식(과거에 우리 팀이 따랐던 방식[22])은 유리 천장에 도달한 것으로 보이며 이는 장르 분류에서도 발생했습니다. 우리의 작업 가설은 MIDI 및 가사와 같은 다양한 소스의 기능과 오디오에서 직접 추출된 멜로디 기능을 사용하면 현재 결과를 개선하는 데 도움이 될 수 있다는 것입니다.

우리의 가설은 타이밍, 역동성, 아티큘레이션, 음색, 음조, 간격, 멜로디, 조화, 조성, 리듬, 모드 등 감정적으로 관련된 몇 가지 특징이 설명된 최근 개요(예: [4], [19])에 의해 동기가 부여되었습니다. 또는 음악적 형태. 이러한 기능 중 다수는 본질적으로 점수 지향적이며 MIDI 영역에서 연구되었습니다. 그러나 이는 활발한 연구 주제(예: 피치 감지 [25])임에도 불구하고 오디오 신호에서 정확하게 추출하는 것이 어려운 경우가 많습니다. 따라서 우리는 일반적으로 사용되는 표준 오디오 기능을 멜로디 오디오 및 MIDI 기능과 결합하면 유리 천장을 깨는 데 도움이 될 수 있다고 믿습니다. 더욱이, 노래 가사는 중요한 감정 정보도 담고 있으므로 [29] 이용되기도 합니다.

이를 위해 동일한 음악에 대한 오디오 신호, MIDI 및 가사 정보를 지원하는 새로운 다중 모드 데이터 세트를 제안하고 MER에서 각각의 중요성과 결합 효과를 연구합니다. 생성된 데이터 세트는 MIREX 기분 분류 작업에서 사용된 것과 동일한 구성, 즉 5개의 감정 클러스터를 따릅니다.

우리는 몇 가지 감독 학습 및 기능 선택 전략을 사용하여 접근 방식을 평가합니다. 이 중 SVM 분류기를 사용하여 최상의 결과를 얻었습니다. 903개 오디오 클립 세트(표준 및 멜로디 오디오 기능만 사용)에서 Fmeasure 64%, 다중 모드 하위 세트(193개 오디오 클립, 가사 및 미디 파일)에서 61.1%를 기록했습니다. .

우리는 이 논문이 MIR/MER 연구 커뮤니티에 관련된 독창적인 기여를 많이 제공한다고 믿습니다.

- MIREX와 유사한 오디오 데이터 세트(903개 샘플)
- MER을 위한 새로운 다중 모드 데이터 세트(193개의 오디오, 가사 및 미디 샘플);
- AllMusic 플랫폼을 기반으로 한 자동 감정 데이터 수집 방법론;
- 오디오, MIDI 및 가사를 결합하여 표준 오디오 기능만으로 얻은 결과를 크게 향상시킬 수 있는 MER의 다중 모드 방법론;
- 범주형 MER에서 멜로디 오디오 기능을 사용한 첫 번째 작업
문제.

본 논문은 다음과 같이 구성된다. 2장에서는 관련 업무를 기술한다. 3장에서는 다음과 같은 방법론을 소개한다. 4장에서는 실험 결과를 제시하고 논의한다. 마지막으로 5장에서는 본 연구의 결론과 향후 연구 결과를 도출한다.

2 관련 업무

감정은 오랫동안 심리학의 주요 연구 주제였으며 수년에 걸쳐 여러 이론적 모델이 제안되었습니다. 이러한 모델은 일반적으로 범주형 모델과 차원 모델이라는 두 가지 주요 그룹으로 나뉩니다. 범주형 모델은 분노, 두려움, 행복 또는 기쁨과 같은 감정의 여러 범주 또는 상태로 구성됩니다.

범주형 패러다임의 예로는 Ekman이 식별한 분노, 두려움, 행복, 슬픔이라는 네 가지 기본 감정에서 파생될 수 있는 감정 모델이 있습니다. 이 네 가지 감정은 다른 모든 감정이 형성되는 기초로 간주됩니다. 생물학적 관점에서 볼 때, 이 아이디어는 기본 감정에 해당하는 신경생리학적, 해부학적 기질이 있을 수 있다는 믿음에서 나타납니다. 심리학적 관점에서 기본 감정은 종종 기본이 아닌 다른 감정의 원시적 구성 요소로 간주됩니다.

널리 알려진 또 다른 범주형 모델은 Hevner의 형용사 순환 모델입니다[6]. 음악 심리학 연구로 가장 잘 알려진 케이트 헤브너(Kate Hevner)는 음악과 감정이 밀접하게 연결되어 있으며 음악은 항상 감정적인 의미를 담고 있다고 결론지었습니다. 그 결과, 저자는 단일 단어를 사용하는 대신 형용사(감정)를 그룹화한 목록을 제안했습니다. Hevner의 목록은 67개의 서로 다른 형용사로 구성되어 있으며 순환 방식으로 8개의 서로 다른 그룹으로 구성되어 있습니다. 이러한 그룹 또는 클러스터에는 동일한 감정 상태를 설명하는 데 사용되는 유사한 의미의 형용사가 포함되어 있습니다.

이 외에도 ISMIR 회의와 함께 개최되는 최신 MER 접근 방식을 매년 비교하는 MIREX 기분 분류 작업에도 범주형 패러다임이 사용됩니다. 이 모델은 감정을 5개의 개별 그룹 또는 클러스터로 분류하며, 각 그룹은 5~7개의 관련 감정(형용사)으로 구성됩니다. 그러나 앞으로 논의할 MIREX 분류법은 심리학적 모델에 의해 뒷받침되지 않습니다.

반면, 차원 모델은 여러 축을 사용하여 감정을 계획에 매핑합니다. 가장 빈번한 접근 방식은 두 개의 축(예: 각성-가치(AV) 또는 에너지-스트레스)을 사용하며 일부 경우에는 3차원(지배)이 사용됩니다[30]. 본 논문에서는 MIREX에서 정의한 5가지 감정 클러스터에 따른 범주형 패러다임을 따릅니다.

연구자들은 적어도 19세기부터 음악과 감정의 관계를 연구해 왔습니다[5]. 이 문제는 20세기에 여러 연구자들이 감정과 모드, 화성, 템포, 리듬, 강약과 같은 특정 음악적 속성 사이의 관계를 조사하면서 더욱 적극적으로 다루어졌습니다[4].

우리가 아는 한, 최초의 MER 논문은 Katayose et al.에 의해 1988년에 출판되었습니다. [9] 그곳에서는 피아노 음악의 다성 녹음의 오디오 특징을 기반으로 한 감정 분석 시스템이 제안되었습니다. 멜로디, 코드, 조, 리듬 특징과 같은 음악 원시 요소를 사용하여 휴리스틱 규칙을 통해 감정을 추정했습니다.

오디오 신호를 사용한 MER에 대한 최초의 작업 중 하나는 2003년 Feng에 의해 수행되었습니다. [삼]. 4가지 감정 범주와 템포와 아티큘레이션이라는 두 가지 음악적 속성만을 사용하여 Feng은 평균 67%의 정확도를 달성했습니다. 이 작업의 주요 제한 사항 중 일부는 23곡만 포함된 매우 작은 테스트 코퍼스, 제한된 수의 오디오 기능(2) 및 카테고리(4)였습니다.

오디오 음악에서 감정 인식을 다루는 다양한 연구 작업(예: [12], [14], [27] 및 [28])에서 감정에 대한 범주적 관점을 사용하는 최초이자 가장 포괄적인 작업 중 하나가 Lu et al.에 의해 제안되었습니다. [14]. 이 연구에서는 범주적 감정을 표현하기 위해 Thayer 모델의 4개 사분면을 사용했으며 강도, 음색 및 리듬 특징을 추출했습니다. 그런 다음 Gaussian Mixture Models를 사용하여 감정을 감지하고 Karhunen-Loeve 변환을 통해 상관 관계 해제 기능을 사용하여 계층적 솔루션과 비계층적 솔루션을 테스트했습니다. 알고리즘이 평균 정밀도 86.3%에 도달했지만 시스템은 클래식 음악 모음에서만 평가되었으므로 이 값은 주의해서 고려해야 합니다.

최근에는 Wang et al. [27]은 음악 태그와 분류기 앙상블을 기반으로 한 특징 벡터의 의미 변환을 이용한 오디오 분류 시스템을 제안했으며 MIREX 2010 기분 분류 작업에서 흥미로운 결과를 얻었습니다.
최근 일부 연구에서는 감정 감지를 위한 다양한 전략을 결합한 다중 모델 접근 방식도 제안했습니다. McVicar 등[17]은 오디오와 노래 가사 연구를 결합하여 둘 사이의 공통 특성을 식별하는 이중 모드 접근 방식을 제안했습니다. 이 전략은 "노래의 의도된 분위기가 작곡가에게 특정 음색, 하모니 및 리듬 기능을 사용하도록 영감을 주어 가사 선택에도 영향을 미친다"는 저자의 가정에 기초합니다. 이 방법을 사용하여 각 오디오 특징과 가사 AV 값 간의 Pearson 상관 계수를 계산한 결과 많은 상관 관계가 통계적으로 매우 유의하지만 절대값이 0.2 미만인 것으로 나타났습니다.

오디오와 가사를 모두 사용하는 다른 이중 모드 작업도 Yang et al에 의해 발표되었습니다. [29]. 저자는 오디오 기능만을 사용하여 발생할 수 있는 감정 분류 한계를 극복하기 위해 의미 정보가 풍부한 가사의 사용법을 탐구합니다. 이러한 한계는 “객체 특징 수준과 감정 인식의 인간 인지 수준 사이의 의미론적 격차”에 기인합니다[29]. 4개의 클래스만을 사용하여 시스템의 정확도는 46.6%에서 57.1%로 향상되었습니다. 저자는 또한 원자가의 분류 정확도를 높이기 위해 가사의 중요성을 강조합니다.

Hu et al [7]의 추가 연구에서는 일부 감정 범주의 경우 가사가 오디오 기능보다 성능이 우수하다는 것을 보여주었습니다. 이러한 경우, 서정적 용어와 범주 사이에 강력하고 명백한 의미적 연관성이 발견되었습니다.
다중 모델 전략은 거의 제안되지 않았지만 우리가 알고 있는 접근 방식 중 MIDI도 사용하는 방법은 없습니다.

3가지 방법

3.1 데이터세트 획득

다중 모드 데이터 세트를 생성하기 위해 우리는 AllMusic 지식 기반을 기반으로 구축하여 MIREX 기분 분류 작업 테스트 베드와 유사한 방식으로 구성했습니다. 여기에는 각각 여러 가지 감정 범주가 포함된 5개의 클러스터가 포함되어 있습니다. 클러스터 1: 열정적, 열광적, 자신감 있음, 떠들썩함, 난폭함; 클러스터 2: 흥겨움, 쾌활함, 재미 있음, 달콤함, 상냥함/좋은 성격; 클러스터 3: 읽고 쓰는 것, 신랄한 것, 아쉬운 것, 씁쓸한 것, 가을의 것, 우울한 것; 클러스터 4: 유머러스함, 바보스러움, 야심적임, 기발함, 기발함, 재치 있음, 씁쓸함; 클러스터 5: 공격적, 불 같은, 긴장/불안, 강렬함, 변덕스러움, 본능적.

MIREX 분류법은 비록 심리적 모델에 의해 뒷받침되지는 않지만 음악 감정 인식 커뮤니티에서 일반적으로 인정되는 유일한 비교 기반이기 때문에 사용됩니다. 또한 Last.FM과 같은 다른 인기 데이터베이스와 달리 주석은 대규모 음악 청취자 커뮤니티(Last.FM에서 발생) 대신 전문가에 의해 수행되기 때문에 AllMusic 데이터베이스를 선택했습니다. 따라서 이러한 주석은 더 신뢰할 수 있습니다. 그러나 주석 과정은 공개되지 않으므로 비판적으로 분석할 수 없습니다.

첫 번째 단계는 자동으로 AllMusic API에 액세스하여 MIREX 무드 태그와 노래 식별자, 아티스트, 제목과 같은 기타 메타 정보가 포함된 노래 목록을 얻는 것이었습니다. 이를 위해 동일한 사이트에서 대부분 30초 mp3 파일인 기존 오디오 샘플을 가져오는 스크립트가 생성되었습니다.

다음 단계는 감정 주석을 만드는 것이었습니다. 이를 위해 MIREX 클러스터에 있는 동일한 무드 태그가 포함된 노래를 선택했습니다. 각 노래에는 두 개 이상의 태그가 있을 수 있으므로 각 노래의 태그는 클러스터별로 그룹화되었으며 결과 노래 주석은 가장 중요한 클러스터, 즉 더 많은 태그가 있는 클러스터(예: 클러스터에서 하나의 태그가 있는 노래)를 기반으로 했습니다. 클러스터 5의 1개 및 3개 태그는 클러스터 5로 표시됩니다. 클러스터 전체에서 거의 균형이 잡힌 총 903개의 MIREX 유사 오디오 클립이 획득되었습니다. 18.8% 클러스터 1, 18.2% 클러스터 2, 23.8% 클러스터 3, 21.2% 클러스터 4 및 18.1% 클러스터 5.

다음으로 Google API를 사용하여 동일한 노래의 가사와 MIDI 파일을 자동으로 검색하는 도구를 개발했습니다. 이 과정에서 가사 정보는 3개 사이트(lyrics.com, ChartLyrics, MaxiLyrics)를 사용했고, MIDI 버전은 4개 사이트(freemidi.org, free-midi.org, midiworld.com, Cool-midi.com)에서 얻었습니다.). 일부 결함이 있는 파일을 제거한 후 가사와 MIDI가 포함된 903개의 원본 오디오 클립을 가로채서 총 764개의 가사와 193개의 MIDI가 생성되었습니다. 실제로 MIDI 파일을 자동으로 가져오는 것이 더 어려운 것으로 나타났습니다.

결과적으로 우리는 903개의 클립이 포함된 오디오 전용(AO) 데이터세트, 764개의 오디오 클립과 가사(여기서는 평가되지 않음)가 포함된 오디오 가사(AL) 데이터세트, 결합된 다중 모드(MM) 데이터세트 등 3개의 데이터세트를 구성했습니다. 193개의 오디오 클립과 그에 상응하는 클립 포함

가사와 미디. 모든 데이터 세트는 클러스터 전체에서 거의 균형을 이루었습니다(최대 및 최소 대표성은 각각 25% 및 13%).

최종 MM 데이터 세트가 의도한 것보다 작더라도(향후 해결할 문제) 이 접근 방식은 AllMusic 주석의 전문적인 인력을 활용하여 음악 감정 데이터 세트를 자동으로 획득하는 이점이 있습니다. 또한 제안된 방법은 이 기사에서 사용된 것과 다른 감정 형용사를 사용하여 다양한 감정 데이터 세트를 생성하는 데 사용할 수 있을 만큼 충분히 일반적입니다.

생성된 데이터 세트는 http://mir.dei.uc.pt/resources/MIREX-like_mood.zip에서 다운로드할 수 있습니다.

3.2 특징 추출

몇몇 저자는 감정 분석과 가장 관련성이 높은 음악적 속성을 연구했습니다. 즉, 주요 모드는 행복이나 엄숙함과 같은 감정 상태와 관련이 있는 반면, 사소한 모드는 슬픔이나 분노와 관련이 있는 것으로 나타났습니다[19]. 단순하고 조화로운 조화는 일반적으로 행복하고 즐겁거나 편안합니다. 반대로 복잡하고 불협화음이 있는 화성은 음악 작품에 불안정성을 조성하기 때문에 흥분, 긴장, 슬픔과 같은 감정과 관련이 있습니다[19]. 최근 개요에서 Friberg [4]는 타이밍, 다이나믹스, 아티큘레이션, 음색, 피치, 간격, 멜로디, 하모니, 조성 및 리듬과 같은 기능을 설명합니다. 해당 목록에 포함되지 않은 다른 일반적인 기능으로는 모드, 음량 또는 음악 형식 등이 있습니다[19].

이전에 언급했듯이 이러한 기능 중 다수는 MIDI 도메인에서 개발되었으며 오디오 신호에서 정확하게 추출하는 것이 어려운 경우가 많습니다. 따라서 표준 오디오 기능과 멜로디 오디오 및 MIDI 기능의 조합을 제안합니다. 이는 결과를 향상시킬 수 있는 잠재력이 있기 때문입니다. 더욱이 노래 가사는 중요한 감정 정보도 담고 있어 악용되고 있다.

표준 오디오(SA) 기능.

의미 있는 음악적 속성을 추출하는 것은 복잡하기 때문에 일반적인 오디오 프레임워크에서 사용할 수 있는 표준 기능을 추출하는 것이 일반적입니다. LLD(낮은 수준 특징 설명자)라고 불리는 이러한 특징 중 일부는 일반적으로 오디오 파형의 단시간 스펙트럼(예: 중심, 확산, 왜도, 첨도, 기울기, 감소, 롤오프와 같은 스펙트럼 모양 특징)에서 계산됩니다. , 플럭스, 대비 또는 MFCC. 템포, 조성 또는 키와 같은 기타 상위 수준 속성도 추출됩니다.

이러한 오디오 기능을 추출하기 위해 여러 오디오 프레임워크를 사용할 수 있습니다. 이 작업에서는 Marsyas, MIR Toolbox 및 PsySound의 오디오 기능이 사용되었습니다.

PsySound 3은 물리적 및 심리음향 알고리즘을 사용하여 녹음을 분석하기 위한 MATLAB 툴박스입니다. 표준 음향 측정뿐만 아니라 심리 음향 및 음악 구현을 사용하여 정확한 분석을 수행합니다.

음량, 선명도, 거칠기, 변동 강도, 피치, 리듬 및 IACC 실행과 같은 모델.

MIR 툴박스는 MATLAB으로 작성된 통합 기능 세트로, 피치, 음색, 조성 등과 같은 음악적 특징을 추출하는 데 특화되어 있습니다[11]. 다양한 저레벨 및 고레벨 오디오 기능을 사용할 수 있습니다.

Marsyas(오디오 신호에 대한 음악 분석, 검색 및 합성)는 MIR 애플리케이션에 특히 중점을 두고 오디오 처리를 위해 개발된 소프트웨어 프레임워크입니다. 템포, MFCC 및 스펙트럼 특징과 같은 특징을 추출할 수 있습니다. 고도로 최적화된 C++ 코드로 작성되었지만 덜 밝은 측면에서는 MER과 관련된 것으로 간주되는 일부 기능이 부족합니다.

Marsyas에서는 프레임 수준 기능에 대한 분석 창이 512개 샘플로 설정되었습니다. MIR 도구 상자는 기본 창 크기 0.05초로 사용되었습니다. 이러한 프레임 수준 기능은 평균 및 분산, 첨도 및 왜도를 계산하여 노래 수준 기능에 통합됩니다. 이 모델은 단시간 특징의 연속 샘플이 독립적이고 가우스 분포이며 더 나아가 각 특징 차원이 독립적이라고 암묵적으로 가정합니다[18]. 그러나 각 특징이 독립적이라는 가정은 올바르지 않다는 것은 잘 알려진 사실입니다. 그럼에도 불구하고 이는 차원의 저주를 해결하는 핵심 문제인 컴팩트함의 장점을 지닌 일반적으로 사용되는 특징 통합 방법입니다[18].

세 가지 프레임워크를 사용하여 총 253개의 특징이 추출되었습니다.

MA(멜로딕 오디오) 기능.오디오에서 멜로디 특징을 추출하려면 이전 멜로디 전사 단계를 거쳐야 합니다. 다성 음악 발췌에서 멜로디 표현을 얻기 위해 Salamon et al이 제안한 자동 멜로디 추출 시스템을 사용합니다. [25]. 그런 다음 추정된 각 주요 멜로디 피치 윤곽에 대해 98개의 특징 세트가 [25]에서와 같이 계산됩니다. 이러한 특징은 피치 범위 및 높이, 비브라토 속도 및 범위 또는 멜로디 윤곽 모양과 같은 멜로디 특성을 나타냅니다.

이러한 기능을 감정 인식에 적용하는 데에는 몇 가지 과제가 있습니다. 첫째, 멜로디 추출은 완벽하지 않습니다. 특히 이 데이터세트의 경우처럼 모든 노래의 멜로디가 명확하지 않은 경우에는 더욱 그렇습니다. 둘째, 이러한 기능은 장르를 분류한다는 매우 다른 목적을 염두에 두고 설계되었습니다. 감정은 매우 주관적이며 노래 내에서 변형되기 쉽습니다. 그럼에도 불구하고 우리는 멜로디의 특징이 우리가 감정을 인식하는 방식에 영향을 미칠 수 있다고 믿습니다. 어쨌든 해당 MIDI 파일에서 추출된 멜로디 특징은 아래와 같이 추출되었습니다.

미디 기능.우리는 문헌([5] 및 [13])과 실험[21]에서 얻은 경험적 결과에 따라 관련성이 있는 것으로 알려진 기능을 얻는 도구 상자를 사용했습니다. 우리는 글로벌 기능에만 집중했습니다(로컬 기능은 고려하지 않았습니다).

MIDI 기능을 추출하기 위해 jSymbolic [16], MIDI Toolbox [2] 및 jMusic [26]의 세 가지 프레임워크가 사용되었습니다. jSymbolic 프레임워크는 278개의 특징(예: 평균 음표 길이 및 음표 밀도)을 추출하고, MIDI 도구 상자는 26개의 특징(예:

멜로디 복잡성 및 키 모드), jMusic은 16가지 특징(예: 클라이막스 위치 및 클라이막스 강도)을 추출합니다.

Friberg의 목록(악기, 강약, 리듬, 멜로디, 질감 및 하모니)과 일치하는 6개 음악 범주에 속하는 총 320개의 MIDI 기능이 추출되었습니다.

서정적 특징.일반적인 가사 분석 프레임워크를 사용하여 서정적 특징을 추출했습니다. 사용된 프레임워크 중 하나인 JLyrics는 Java로 구현되었으며 jMIR 제품군의 오픈 소스 프로젝트에 속합니다[15]. 이 프레임워크는 주로 구조적인(예: 단어 수, 세그먼트 평균 라인 수) 19개의 기능을 추출하지만 몇 가지 의미론적 기능(예: Word Profile Match Modern Blues, Word Profile Match Rap)도 포함합니다.

또한 텍스트 감정 인식을 위한 Java API인 Synesketch 프레임워크[10]를 사용했습니다. Paul Ekman의 모델[1]에 따라 감정을 추출하기 위해 WordNet[20] 기반의 자연어 처리 기술을 사용합니다. 추출된 특징은 행복, 슬픔, 분노, 두려움, 혐오, 놀라움의 무게입니다.

두 가지 프레임워크를 사용하여 총 27개의 서정적 특징을 추출했습니다.

3.3. 분류 및 기능 선택

본 연구에서는 SVM(Support Vector Machines), K-Nearest Neighbors, C4.5 및 Naïve Bayes와 같은 지도 학습 알고리즘을 사용하여 다양한 테스트를 실행했습니다. 이를 위해 Weka(데이터 마이닝 및 기계 학습 플랫폼)와 libSVM이 포함된 Matlab이 모두 사용되었습니다.

분류 외에도 특징 수를 줄이고 결과를 향상시키기 위해 특징 선택 및 순위 지정도 수행했습니다. 이를 위해 Weka 워크벤치를 활용한 Relief 알고리즘[24]이 사용되었습니다. 알고리즘은 순위가 결정되는 기준에 따라 각 기능에 대한 가중치를 출력합니다. Feature Ranking 이후, 획득된 Ranking에 따라 한 번에 하나의 Feature를 추가한 후 결과를 평가하여 실험적으로 최적의 Feature 개수를 결정.

특징 선택 및 분류 모두에 대해 반복된 계층화 10겹 교차 검증(20회 반복)을 통해 결과를 검증하여 평균 획득 정확도를 보고했습니다. 또한 SVM의 경우 그리드 매개변수 검색과 같은 매개변수 최적화가 수행되었습니다.

4 실험 결과

다양한 특징의 출처의 중요성과 감정 분류에서 이들의 조합이 미치는 영향을 평가하기 위해 여러 가지 실험이 실행되었습니다.

오디오 전용(AO) 데이터 세트(표 1 참조)에서 표준 오디오(SA) 기능과 멜로디 오디오(MA) 기능을 사용하는 실험부터 시작합니다. 마지막 열에는 이들의 조합으로 얻은 결과(F-측정값)가 표시됩니다. 모든 기능 세트와 기능 선택(*)이 제시된 후 얻은 F 측정값입니다(사용된 최상의 기능에 대한 자세한 내용은 표 4 참조).

1 번 테이블.오디오 전용(AO) 데이터 세트의 표준 및 멜로디 오디오 기능(F-측정)에 대한 결과입니다.

분류기	SA	엄마	SA+MA
나이브베이즈	37.0%	31.4%	38.3%
나이브베이즈*	38.0%	34.4%	44.8%
C4.5	30.1%	53.5%	55.9%
C4.5*	30.0%	56.1%	57.3%
KNN	38.9%	38.6%	41.5%
KNN*	40.8%	54.6%	46.7%
SVM	44.9%	52.3%	52.8%
SVM*	46.3%	59.1%	64.0%

볼 수 있듯이 SVM 분류기와 특징 선택을 통해 최상의 결과를 얻었습니다. 일반적으로 사용되는 표준 오디오 기능은 멜로디 기능보다 확실히 뒤떨어져 있습니다(F 측정값 59.1%에 대해 46.3%). 그러나 멜로디적 특징만으로는 충분하지 않습니다. 실제로 SA와 MA 기능을 결합하면 결과가 64%까지 더욱 향상됩니다. 또한 중요한 점은 이 성능이 원래 351 SA + MA 기능 세트에서 11개 기능(9 MA 및 2 SA)만을 사용하여 달성되었다는 것입니다. 이러한 결과는 표준 오디오 기능과 멜로디 오디오 기능의 조합이 음악 감정 인식 문제에 중요하다는 초기 가설을 강력하게 뒷받침합니다.

표 2.별도 및 결합된 다중 모드 기능 세트(F-측정값)에 대한 결과입니다.

분류기	SA	엄마	미디	가사	SA+MA	결합된
SVM	35.6%	35.0%	34.3%	30.3%	39.1	40.2%
SVM*	44.3%	55.0%	42.3%	33.7%	58.3	61.2%

표 2는 SVM만 사용하여 각 기능 세트(SA, MA, MIDI 및 가사)를 개별적으로 사용하는 MM 데이터 세트의 결과를 요약합니다. 다시 마지막 열에는 이들의 조합으로 얻은 결과가 표시됩니다.

MM 데이터세트에서 SA와 MA 기능의 조합은 이전과 마찬가지로 결과를 분명히 향상시켰습니다(SA만 사용하는 경우 44.3%에서 58.3%로). 다시 말하지만, 멜로디적 특징은 개선된 결과에 큰 영향을 미칩니다.

SA와 MIDI 기능을 비교하면 성능이 비슷하다는 것을 알 수 있습니다(44.2 대 42.7%). 실제로 차원적 감정 패러다임[23]을 따르는 우리 팀의 이전 연구에 따르면 SA 기능은 각성 예측에 가장 적합하지만 원자가 추정 기능이 부족합니다. 반면에 MIDI 기능은 향상되는 것 같습니다.

원자가 예측은 가능하지만 각성 추정에서는 SA만큼 좋지 않습니다. 따라서 이들의 결합에 따른 보상효과가 발생하는 것으로 보인다.

결합된 다중 모드 기능 세트의 경우 초기에 가정한 대로 결과가 향상되었습니다. SA 및 MA만 사용하는 경우 58.3%에서 61.2%로 향상되었습니다. 이는 추출된 698개 중 19개의 다중 모드 특성만으로 달성되었습니다.

표 3에서는 MM 데이터 세트에서 최상의 결과를 얻을 수 있는 혼동 행렬을 제시합니다. 여기서 클러스터 4는 평균보다 훨씬 낮은 성능(51.5%)을 나타냈으며 다른 모든 클러스터도 비슷한 성능을 얻었습니다. 이는 클러스터 4가 데이터 세트에서 더 모호할 수 있음을 나타냅니다.

표 3.다중 모드 데이터 세트에 대한 혼동 행렬.

	C1	C2	C3	C4	C5
C1	63.6%	15.9%	4.5%	4.5%	11.4%
C2	20.9%	60.5%	11.6%	7.0%	0.0%
C3	4.2%	18.8%	64.6%	8.3%	4.2%
C4	12.1%	18.2%	9.1%	51.5%	9.1%
C5	12.0%	3.0%	12.0%	8.0%	64.0%

앞서 언급했듯이 19가지 기능(SA 5개, MA 10개, MIDI 4개 및 서정적 기능 없음 - 표 4 참조)에서 최상의 결과를 얻었습니다. 선택된 특징에서 관찰된 다양성은 제안된 다중 모드 접근 방식이 우리가 가설을 세웠던 것처럼 멜로디 오디오 특징과 특히 관련성이 있는 음악 감정 분류에 도움이 된다는 것을 시사합니다. 유일한 예외는 서정적 특징이 선택되지 않았다는 것입니다. 이는 사용된 기능(33.7%)으로 달성된 낮은 성능으로 확인되며, 사용된 서정적 프레임워크에 관련 의미 기능이 부족하기 때문에 확실히 설명됩니다. 이 문제는 앞으로 해결될 것입니다.

표 4에는 각 소스의 가장 중요한 5가지 기능이 나열되어 있습니다(MA의 경우 10가지). SA의 경우 선택된 특징은 대부분 조화와 색조와 관련이 있습니다. 하나의 스펙트럼 특징만 선택되었습니다. MA와 관련하여 10개의 상위 특징은 모두 상위 1/3의 더 긴 윤곽선만을 사용하여 계산되었습니다. 대부분은 비브라토와 관련이 있으며 이전 연구에서 장르를 예측하는 데 중요하다고 간주된 것과 유사합니다[25]. MIDI의 경우 중음역과 저음역의 중요성에 대한 특징이 가장 관련성이 높았으며 그 다음에는 전자 악기, 특히 기타가 등장했습니다. 처음 예상했던 것과 달리 조음 기능(예: 스타카토 발생률)은 선택되지 않았습니다. 그 이유는 우리 데이터 세트에서 이러한 공연 스타일의 존재감이 낮았기 때문입니다. 마지막으로 가장 중요한 두 가지 서정적 특징은 두려움과 분노의 무게에 관한 것으로 Synesketch에서 추출되었으나 그 중 어느 것도 선택되지 않았다.

마지막으로 MIREX 2012 기분 분류 작업에서는 유사한 분류 접근 방식을 사용하여 67.8%(최고 결과)를 달성했지만 표준 오디오 기능만 사용했습니다. MIREX 데이터 세트로 얻은 결과와 SAF 기능만 사용하여 이 기사에서 제안한 데이터 세트(46.3%)의 차이는 직접 비교하기는 어렵지만 데이터 세트가 더 어려울 수 있음을 시사합니다.

표 4.각 기능 세트의 상위 5~10개 기능입니다. Avg, std, skw 및 kurt는 각각 평균, 표준 편차, 왜도 및 첨도를 나타냅니다.

기능 세트	기능 이름
SA	1) 고조파 변화 감지 기능(avg), 2) 색조 중심 4(std), 3) 키, 4) 스펙트럼 엔트로피(avg), 5) 색조 중심 3(std)
엄마	1) 비브라토 적용 범위(VC)(skw), 2) VC(kurt), 3) VC(avg), 4) 비브라토 범위(VE)(avg),5) VE(kurt), 6) VC(kurt), 7) 비브라토 레이트(VR)(std), 8) VE(std), 9) VR(avg), 10) VE(skw)
미디	1) 중음역의 중요성, 2) 베이스음역의 중요성, 3) 전자악기 분수, 4) 일렉트릭 기타 분수, 5) 음표 악기의 보급률
가사	1) 공포의 무게, 2) 분노의 무게, 3) 단어 프로필 매치 모던 블루스, 4) Valence, 5) 단어 프로필 매치 랩

5 결론 및 향후 연구

우리는 표준 오디오, 멜로디 오디오, MIDI 및 서정적 기능을 기반으로 MER에 대한 다중 모드 접근 방식을 제안했습니다.

새로운 데이터세트(오디오 전용, 오디오 및 가사, 오디오, 미디 및 가사의 3개 하위 세트로 구성)와 AllMusic 프레임워크를 사용한 자동 획득 전략이 제안되었습니다.

지금까지 얻은 결과는 제안된 다중 모드 접근 방식이 표준 오디오 기능만 사용할 때 감정 분류에서 현재의 유리 천장을 능가하는 데 도움이 된다는 것을 시사합니다.

표준 오디오, 멜로디 및 미디 기능으로 생성된 모델과 비교할 때 사용된 서정적 기능으로 얻은 성능은 훨씬 더 나쁩니다. 이는 아마도 노래 가사에 존재하는 감정을 정확하게 포착하지 못하는 주로 구조적인 특징을 사용한 결과일 것입니다. 앞으로는 감정적 상관관계가 더 강한 의미론적 특징을 사용할 계획입니다.

마지막으로, 가까운 시일 내에 다중 모드 데이터 세트의 크기를 늘릴 계획입니다. 앞서 언급했듯이 MIDI 파일은 자동으로 획득하기가 더 어렵습니다. 따라서 우리는 AllMusic에서 더 큰 오디오 세트를 얻을 것이며, 이를 통해 더 많은 수의 해당 MIDI 아카이브를 얻을 수 있기를 바랍니다.

감사의 말

이 작업은 Fundação para Ciência ea Tecnologia(FCT) 및 Programa Operacional Temático Factores de Competitividade(COMPETE) - 포르투갈의 자금 지원을 받는 MOODetector 프로젝트(PTDC/EIA-EIA/102185/2008)의 지원을 받았습니다.

참고자료

1. Ekman, P.: 인간 얼굴의 감정, Cambridge University Press (1982).
2. Eerola, T., Toiviainen P.: “Matlab의 MIR: Midi 도구 상자” ISMIR(2004).
3. Feng, Y., Zhuang, Y., Pan, Y.: "기분 감지를 통한 대중 음악 검색", Proc. 26일. 국제 ACM SIGIR 회의. 정보 검색 연구 및 개발, vol.
2, 아니. 2, pp. 375–376(2003).
4. Friberg, A.: "디지털 오디오 감정 – 컴퓨터 분석 개요 및 음악의 감정 표현 합성", DAFx, pp.1-6 (2008).
5. Gabrielsson, A., Lindström, E.: “감정 표현에 대한 음악 구조의 영향”, 음악 및 감정: 이론 및 연구, pp.223–248(2001).
6. Hevner, K.: “음악 표현 요소에 대한 실험적 연구”. 미국 심리학 저널, 48(2), pp. 246–268(1936).

7. Hu, X., Downie, J.: “음악 분위기 분류에서 가사가 오디오보다 성능이 뛰어난 경우: 기능 분석,” ISMIR, pp. 619-624 (2010).
8. Huron, D.: "음악 정보 검색의 지각 및 인지적 응용", 음악 정보 검색에 관한 국제 심포지엄(2000).

9. Katayose, H., Imai, M., Inokuchi, S.: “음악의 감정 추출”, Proceedings 9th International Conference on Pattern Recognition pp. 1083–1087 (1988).
10. Krcadinac, U.: 텍스트 감정 인식 및 창의적 시각화, 베오그라드 대학교 졸업 논문(2008).

11. Lartillot O., Toiviainen, P.: "오디오에서 음악적 특징 추출을 위한 Matlab 도구 상자", DAFx-07, p. 237–244 (2007).
12. Liu, D., Lu, L.: “어쿠스틱 음악 데이터로부터 자동 기분 감지”, Int. J. 스트레스 생물학, vol. 8, 아니. 6, pp. 359-377(2003).
13. Livingstone, S., Muhlberger, R., Brown, A., Loch, A.: "음악적 감정 제어: 음악적 감정에 영향을 미치는 감정적 계산 아키텍처", Digital Creativity 18(2007).

14. Lu, L., Liu, D., Zhang, H.-J.: "음악 오디오 신호의 자동 기분 감지 및 추적", IEEE Trans. 오디오, 음성 및 언어 처리, vol. 14, 아니. 1, pp. 5-18(2006).

15. McKay, C.: jMIR을 사용한 자동 음악 분류, Ph.D. 캐나다 맥길대학교 논문(2010).

16. McKay, C., Fujinaga, I.: "jSymbolic: Midi 파일용 기능 추출기", International Computer Music Conference(2006).
17. McVicar, M., Freeman, T.: "서정적 기능과 오디오 기능과 기분의 출현 사이의 상관 관계 마이닝", ISMIR, pp.783-788 (2011).
18. Meng, A., Ahrendt, P., Larsen, J., Hansen, LK: “음악 장르 분류를 위한 시간적 특징 통합”. IEEE 트랜스. 오디오, 음성 및 언어 처리, 15(5), pp. 275–9, (2007).

19. Meyers, OC: 무드 기반 음악 분류 및 탐색 시스템, 석사 논문, MIT(Massachusetts Institute of Technology)(2007).
20. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: “WordNet: 온라인 어휘 데이터베이스,” Int. J Lexicograph, pp. 235-244(1990).
21. Oliveira A., Cardoso, A.: “감정 표현을 위한 음악 시스템”, Knowledge-Based Systems 23, 901–913 (2010).
22. Panda, R., Paiva, RP: "음악 감정 분류: 데이터 세트 수집 및 비교 분석", DAFx-12(2012).

23. Panda, R., Paiva, RP: "thayer 평면에서 기분 재생 목록의 자동 생성: 방법론 및 비교 연구", 제8회 사운드 및 음악 컴퓨팅 컨퍼런스(2011).

24.Robnik-Šikonja, M., Kononenko, I.: “ReliefF 및 RReliefF의 이론적 및 경험적 분석,” 기계 학습, vol. 53, 아니. 1–2, pp. 23–69(2003).
25. Salamon, J., Rocha, B., Gómez, E.: "다성 음악 신호에서 추출한 멜로디 특징을 사용한 음악 장르 분류", ICASSP(2012).
26. Sorensen A., Brown, A.: “JMusic 소개,” 호주 컴퓨터 음악 컨퍼런스, pp. 68–76(2000).

27. Wang, J., Lo, H., Jeng, S.: "Mirex 2010: 의미론적 변환 및 분류자 앙상블을 사용한 오디오 분류", WOCMAT, pp.2-5 (2010).
28. 양덕, 이우: “소프트웨어 에이전트를 이용한 음악 감정의 구분”, ISMIR, pp. 52-58 (2004).

29. Yang, Y., Lin, Y., Cheng, H., Liao, I., Ho, Y., Chen, H.: “다중 모드 음악 감정 분류를 향하여” PCM08, pp. 70-79 (2008).
30. Yang, Y., Lin, Y., Su, Y., Chen, H.: "음악 감정 인식에 대한 회귀 접근 방식", IEEE Trans. 오디오, 음성 및 언어 처리, vol. 16, No. 2, pp. 448-457(2008).