

Music Information Retrieval 기술 동향

□ 번가람, 김무영 / 세종대학교

I. 서론

최근 초고속 네트워크의 발달과 함께 디지털 음악 시장이 커짐에 따라 인터넷이나 스마트 디바이스로부터 디지털 음악을 쉽게 다운로드 할 수 있게 되었다. 이러한 방대한 양의 음악을 효과적으로 관리해야 할 필요성 또한 점차 커지게 되었고, 여기에 발맞추어 음악 정보 처리(music information retrieval)라는 분야에 대한 연구가 활발히 진행되고 있다[1-2]. 음악 정보 처리 기술은 대용량의 음악 검색 시스템이나 음악 추천 시스템 등에 적용될 수 있다. 예를 들어, 오디오 신호의 정보를 분석해서 실제 음악의 장르나 무드를 자동 분류해 주거나 음악의 분위기나 장르에 맞는 태그를 붙여주는 등 여러 기술에 응용될 수 있다.

다양한 음악 정보 처리 기술 알고리즘과 최신 동향은 MIREX 평가회 결과를 보면 확인할 수 있다.

MIREX는 Music Information Retrieval Evaluation eXchange의 약자로 음악 정보 처리 시스템에 사용되는 알고리즘들을 평가하기 위해 매년 개최되는 국제적 평가회이다. 미국 일리노이 대학의 Stephen Downie 교수 연구실 주관으로 2005년부터 매년 개최되고 있다. 평가 결과는 국제 학술대회인 International Society for Music Information Retrieval (ISMIR) Conference에서 발표되고 있다. 따라서, ISMIR에서 발표되는 MIREX 평가회 결과를 통해 음악 정보 처리 기술의 최신 동향을 확인 할 수 있다. 이에 본 논문에서는 2013년도 MIREX 기술 동향을 살펴봄으로써 음악 정보 처리 기술의 최근 동향을 소개하고자 한다.

II 장에서는 2013년도 MIREX 동향과 대표적인 평가 task를 요약하고, III 장에서 음악 정보 처리 기술의 미래를 짚어보고 결론을 맺고자 한다.

II. 본 론

MIREX에는 총 16개의 평가 task가 있으며 2012년도와 2013년도 MIREX 평가회에 참여한 알고리즘 수는 각 task 별로 <표 1>과 같다[3-6].

<표 1> 2012년도와 2013년도 MIREX 평가회 Task별 참여한 알고리즘 수

Task	2012	2013
Audio Classification (Train/Test)	20	23
Audio Tag Classification	9	N/A
Audio Music Similarity and Retrieval	10	8
Symbolic Melodic Similarity	6	5
Audio Onset Detection	10	11
Audio Key Detection	6	3
Real-time Audio to Score Alignment	3	2
Query by Singing/Humming	5	9
Audio Melody Extraction	5	4
Multiple Fundamental Frequency Estimation & Tracking	12	6
Audio Chord Estimation	11	4
Query by Tapping	2	2
Audio Beat Tracking	20	20
Structural Segmentation	9	8
Audio Tempo Estimation	4	11
Discovery of Repeated Themes & Sections	-	10

각 평가 task에 대한 상세한 설명은 다양한 참고문헌들을 참고하길 바라며 [1,2,4,6,7], 본 논문에서는 대표적인 평가 task에 대해서만 살펴보고자 한다. 대표적인 평가 task로는 audio classification, audio tag classification, audio music similarity and retrieval, query by singing/humming 등이 있다. 본 장에서는 대표적인 평가 task들에 대한 2013년도 MIREX 평가 결과와 기술을 요약해보고자 한다.

1. Audio Classification

Audio classification task에서는 입력 오디오 신

호로부터 장르, 무드, 작곡가를 분류하는 연구가 진행되고 있으며, <표 1>에서 보듯이 다른 task와 비교될 정도로 많은 수의 알고리즘들이 제출되고 있다.

장르 분류(audio genre classification) 기술은 미국 팝 음악 장르 분류와 라틴 음악 장르 분류로 나뉜다. 미국 팝 음악 장르 분류에는 MIREX 2007 장르 데이터베이스가 사용되었는데, 이 데이터베이스는 blues, jazz, country/western, baroque, classical, romantic, electronica, hip-hop, rock, hardRock/metal의 10개 장르로 총 30초 길이 7000곡으로 구성되어 있다. 라틴 음악 장르 분류에는 MIREX 2007 장르 데이터베이스와는 다른 종류의 데이터베이스가 사용되었는데, axe(악세), bachata(바차타), bolero(볼레로), forro(포로), gaucha(가우차), merengue(메링게), pagoda(파고테), sertaneja(세르타네자), tango(탱고)의 10개 라틴 장르로 총 3227곡으로 구성되어 있다. 미국 팝 음악 장르 분류에서는 Ming-Ju와 Jyh-Shing이 제안한 알고리즘[8]이 76%로, 라틴 음악 장르 분류에서는 Aggelos가 제안한 알고리즘[9]이 78%로 가장 우수한 성능을 보였다. Ming-Ju와 Jyh-Shing은 Gabor filter를 이용하여 visual 특징을 추출, 장르 분류를 하였고, Aggelos는 리듬 특징을 이용한 장르 분류를 하였다. 이 외에도 높은 성능을 가진 알고리즘들을 보면 주로 리듬이나 멜로디 특징을 이용하여 장르를 분류할 수 있는 알고리즘을 연구한 것을 알 수 있다[10-11].

음악의 무드 분류(audio mood classification) 평가를 위해서는 600곡의 음악 데이터베이스를 사람들의 주관적인 평가에 의해 5개의 무드 클러스터로 분류하였다. 각 클러스터에 해당하는 속성은 <표 2>와 같다.

〈표 2〉 MIREX 무드 클러스터 속성 정의

클러스터	속성
1 (정열적이고 활기 있는)	Passionate (정열적인), rousing (활발한), confident (자신감 있는), boisterous (활기 넘치는), rowdy (소란스러운)
2 (신나고 즐거운)	Rollicking (신나는), cheerful (발랄한), fun (재미있는), sweet (달콤한), amiable/good natured (쾌활한)
3 (우울하고 쓸쓸한)	Poignant (가슴 아픈), wistful (그리워하는), bittersweet (달콤 씁쓸한), autumnal (가을 분위기의), brooding (음울한)
4 (유머러스하고 재치 있는)	Humorous (유머러스한), silly (멍청한), campy (과장된), quirky (괴바른), whimsical (영동한), witty (재치 있는), wry (비꼬는)
5 (공격적이고 강렬한)	Aggressive (공격적인), fiery (불타는 듯한), tense/anxious (긴장한), intense (강렬한), volatile (불안한), visceral (본능적인)

무드 분류에 대한 결과는 미국 팝 음악 장르 분류와 마찬가지로 Ming-Ju와 Jyh-Shing이 제안한 알고리즘[8]이 68%로 가장 좋은 성능을 보였다.

작곡가 분류(audio composer identification) 기술은 30초 길이의 클래식 음악 2772곡을 데이터베이스로 사용한다. Bach, Beethoven, Brahms, Chopin, Dvorak, Handel, Haydn, Mendelssohn, Mozart, Schubert, Vivaldi 11명의 작곡가를 자동으로 분류해 주는 알고리즘을 연구한다. 2013년도 MIREX 작곡가 분류에서는 본 연구팀에서 제안한 알고리즘[12]과 Franz와 Kirk이 제안한 알고리즘[13]이 70%로 가장 우수한 성능을 보였다.

2. Audio Tag Classification

오디오 태그 분류(audio tag classification) 기술은 음악을 듣고 이 음악의 스타일이나 악기, 템포 등의 연상되는 단어를 태그하여 검색에 이용하는

기술로서, 음악의 신호적인 특징을 분석해 각 음악에 적합한 설명을 태그 형태로 부여하는 것이다. 평가 데이터베이스로는 두 가지 데이터베이스가 사용되었다. 첫 번째 데이터베이스에서는 [14]에서 사용된 3900곡 중 랜덤으로 선택된 2300곡에 대하여 MajorMiner 게임을 이용하여 총 73000개의 태그를 모았다. 두 번째 데이터베이스에서는 last.fm으로부터 3469곡에 대하여 18개의 무드 태그 그룹을 나누어 사용하였다. 이에 대한 자세한 태그 그룹 분류표는 MIREX 홈페이지를 참조하기 바란다. 이 task에 대한 2013년도 평가 결과는 아직 발표되지 않아 2012년도 MIREX 결과로 대신한다. 첫 번째 데이터베이스에 대해서는 Philippe이 제안한 알고리즘[15]이, 두 번째 데이터베이스에 대해서는 Klaus가 제안한 알고리즘[16]이 F-measure 관점에서 각각 좋은 성능을 보였다. 두 알고리즘 모두 시간 축 temporal한 특징을 이용하여 태그 분류 알고리즘을 연구한 것을 알 수 있다.

3. Audio Music Similarity and Retrieval

방대한 양의 디지털 음악에서 비슷한 음악을 찾아내는 기술이 있다. 음악의 유사성 검색(audio music similarity and retrieval) 기술이라 불리는 이 기술은 사용자에게 듣고 있는 음악과 비슷한 음악을 추천해 주거나 사용자가 좋아하는 음악을 찾아주는데 응용될 수 있다. 평가 데이터베이스로는 blues, jazz, baroque, country/western, classical 등 총 10개 장르로 이루어진 미국 팝 음악에서 30초 길이 7000곡의 음악을 이용하였다. 평가 방법으로는 사람의 주관적인 평가와 객관적으로 측정하는 방법이 있다. 주관적인 평가 방법은 평가자들이 ‘유사하지 않음’, ‘어느 정도 유사함’, ‘매우 유사함’의 3가지 중

한가지를 선택하고 0~10 중 유사한 정도를 평가하는 방법이다. 객관적인 평가 방법은 “number of times similar curve”를 그려서 평가하는 방법이다. 이 task에 대한 2013년도 MIREX 결과는 Klaus가 제안한 알고리즘[17]이 두 가지 평가 방법 모두에 대해 가장 우수한 알고리즘으로 선정되었다.

4. Query by Singing/Humming

허밍 기반 검색(query by singing/humming) 기술은 사용자가 자신의 목소리로 흥얼거린 허밍을 질의로 사용하여 음악을 찾아내는 알고리즘을 연구하는 분야이다. 이 기술 분야에서 높은 성능을 내기 위해서는 학습 데이터베이스의 양이 많아야 할 것이다. MIREX에서는 평가 데이터베이스로 세 가지 데이터베이스를 사용하였다. 중국 음악 데이터베이스인 Roger Jang의 MIRE-QBSH corpus와 IOACAS corpus, 그리고 잡음을 포함한 Noise MIDI로 Essen collection을 사용하였다. Jang 데이터베이스에 대해서는 Jang이 제안한 알고리즘들[18-19]이, IOACAS에서는 Peng이 제안한 알고리즘[20]이 가장 우수한 성능을 나타내었다.

5. 기타 Task

음악 정보 검색 시스템의 성능을 평가하는 task들 이외에도 응용 기술의 성능 향상에 간접적인 영향을 주는 일종의 기초 기술 task들이 있다. 예를 들어서, audio onset detection, audio melody extraction, audio beat tracking 등과 같이 음표의 시작 시점, 리듬 패턴, 멜로디 형태 등 중요한 음악 정보를 추출하는 기초 기술들을 평가하는 것이다.

Audio onset detection 기술의 평가 데이터베이스

으로는 솔로 드럼 30곡, 모노포닉 피치를 갖는 악기 30곡, 폴리포닉 피치를 갖는 악기 10곡, 다양한 악기들이 동시에 연주되는 15곡으로 총 85곡으로 이루어져 있다. 이 평가 task에 대한 2013년도 MIREX 결과에서는 Sebastian이 제안한 알고리즘[21]이 F-measure 관점에서 가장 우수한 성능을 나타내었다.

오디오 멜로디 추출(audio melody extraction) 기술은 폴리포닉 음악에서 멜로디의 음높이, 즉 메인 pitch를 인식하는 기술이다. 여러 개 악기가 연주하는 폴리포닉 음악에서, 악기를 구분하고 멜로디를 추출하는 기술이나 음악의 유사성 검색 기술 등에 사용될 수 있다. 사용되는 평가 데이터베이스로는 모두 4개의 데이터베이스가 있다. 첫 번째는 가라오케에서 녹음된 374곡의 중국 노래로 구성된 MIREX09 데이터베이스가 있으며, 두 번째로는 인디아 북부 지방 노래로써 다양한 인디아 악기들을 함께 녹음한 MIREX08 데이터베이스, 세 번째로는 rock, R&B, pop, jazz, sole classical piano의 장르로 구성된 MIREX05 데이터베이스, 마지막으로 2004 Audio Description Contest(ADC)로부터 구한 ADC04 데이터베이스가 있다. 2013년도 MIREX 오디오 멜로디 추출 task의 결과는 첫 번째, 세 번째 데이터베이스에 대해서는 Liming이 제안한 알고리즘[22]이, 두 번째 데이터베이스에 대해서는 Tzu-chun이 제안한 알고리즘[23]이, 네 번째 데이터베이스에 대해서는 Chris가 제안한 알고리즘[24]이 가장 우수한 성능을 보였다.

III. 결론

2005년부터 시작된 MIREX 평가회에 참여한 알

고리즘 수는 매년 꾸준히 증가하고 있다. 그만큼 음악 정보 처리 기술에 대한 관심 또한 커지고 있다는 뜻이다. 본 논문에서는 2013년도 주요 MIREX task의 결과를 살펴보았다.

해외에서 음악 정보 처리 기술이 적용된 사례를 예로 들어보면, Shazam이라는 애플리케이션을 들 수가 있다. Shazam은 스피커에서 들리는 음악을 스마트폰 디바이스에 갖다 대면 이 음악이 어떤 음악인지에 대한 정보를 알려주는 모바일 애플리케이션이다. 또한 사용자의 태그 기록을 바탕으로 음악 추천 서비스도 제공하고 있다. Shazam의 작동 원리는 입력 음악 정보를 인식해서 잡음을 제거한 뒤, 음악 정보를 추출, 서버에 저장된 데이터들과의 비교를 통해 입력 음악의 정보를 찾아내는 방식이다.

음악 정보 검색 기술에 대한 연구는 현재 미국,

유럽, 일본 등에서 활발히 진행되고 있지만, 국내에서는 부족한 것이 현실이다. 그러나 최근 들어 음악 정보 검색 기술에 대한 필요성이 대두되면서 정부 기관과 대학 및 연구소를 중심으로 다양한 연구가 진행되고 있다.

음악 정보 검색 기술은 사용자가 원하는 음악을 찾아주는 것 뿐만 아니라 표절곡을 찾아내는 분야에도 활용할 수 있을 것이다. 음악이 비슷하다는 것은 리듬이나 템포 등이 비슷하다는 것이기 때문에 두 음악의 데이터들을 비교함으로써 자동으로 유사성을 비교해 볼 수 있다. 궁극적으로는 디지털 음악 시장이 커짐에 따라 수천, 수만 곡의 음악 중에서 ‘내가 지금 듣고 싶은 음악은?’, ‘지금 분위기와 맞는 음악은?’과 같은 질문에 답을 해줄 수 있는 음악 정보 처리 기술의 개발이 필요하다.

참고 문헌

- [1] J. S. Downie, “The music information retrieval evaluation exchange (2005-2007): a window into music information retrieval research,” *Acoustical Science and Technology*, vol. 29, no. 4, pp. 247-255, 2008.
- [2] J. S. Downie, A. F. Ehmann, M. Bay, and M. C. Jones, “The music information retrieval evaluation exchange: some observations and insights,” *Advances in Music Information Retrieval*, vol. 274, pp. 93-115, 2010.
- [3] 2012: MIREX Home, http://www.music-ir.org/mirex/wiki/2012:Main_Page
- [4] 2012: MIREX Results, http://www.music-ir.org/mirex/wiki/2012:MIREX2012_Results
- [5] 2013: MIREX Home, http://www.music-ir.org/mirex/wiki/2013:Main_Page
- [6] 2013: MIREX Results, http://www.music-ir.org/mirex/wiki/2013:MIREX2013_Results
- [7] N. Orio, *Music retrieval: A tutorial and review*, Now Publishers Inc., 2006.
- [8] Audio genre classification, <http://www.music-ir.org/mirex/abstracts/2013/JJ1.pdf>
- [9] Audio genre classification, <http://www.music-ir.org/mirex/abstracts/2013/AP1.pdf>
- [10] Audio genre classification, <http://www.music-ir.org/mirex/abstracts/2013/SSKS1.pdf>
- [11] Audio genre classification, <http://www.music-ir.org/mirex/abstracts/2013/CJ1.pdf>
- [12] Audio composer identification, <http://www.music-ir.org/mirex/abstracts/2013/BBLJK1.pdf>
- [13] Audio composer identification, <http://www.music-ir.org/mirex/abstracts/2013/DM4.pdf>
- [14] MajorMiner search page, <http://majorminer.org/search/human>
- [15] Audio tag classification, <http://www.music-ir.org/mirex/abstracts/2011/PH2.pdf>
- [16] Audio tag classification, <http://www.music-ir.org/mirex/abstracts/2011/SSKSS1.pdf>
- [17] Audio music similarity and retrieval, <http://www.music-ir.org/mirex/abstracts/2013/SS2.pdf>

- [18] Query by singing/humming, <http://www.music-ir.org/mirex/abstracts/2013/YJ1.pdf>
- [19] Query by singing/humming, <http://www.music-ir.org/mirex/abstracts/2013/YJ2.pdf>
- [20] Query by singing/humming, <http://www.music-ir.org/mirex/abstracts/2013/YJ2.pdf>
- [21] Audio onset detection, <http://www.music-ir.org/mirex/abstracts/2013/SB2.pdf>
- [22] Audio melody extraction, <http://www.music-ir.org/mirex/abstracts/2013/SL1.pdf>
- [23] Audio melody extraction, <http://www.music-ir.org/mirex/abstracts/2013/YJ4.pdf>
- [24] Audio melody extraction, <http://www.music-ir.org/mirex/abstracts/2013/CF1.pdf>

필자 소개



변가람

- 2012년 : 세종대학교 정보통신공학과 학사 졸업
- 2012년 ~ 현재 : 세종대학교 정보통신공학과 석사과정
- 주관심분야 : 음악 정보 검색, 잡음 제거, 음성 인식, 화자 인식



김무영

- 1993년 : 연세대학교 전자공학과 학사 졸업
- 1995년 : 연세대학교 전자공학과 석사 졸업
- 1995년 ~ 2000년 : 삼성종합기술원 전문연구원
- 2001년 ~ 2004년 : KTH (스웨덴) 박사 졸업
- 2005년 ~ 2006년 : Ericsson Research (스웨덴), Senior Research Engineer
- 2006년 ~ 현재 : 세종대학교 정보통신공학과, 부교수
- 주관심분야 : 음악 정보 검색, 음성/오디오 신호처리 및 코딩, 잡음 제거, 패턴 인식, 정보이론