

음악 속성 태그를 이용한 유사 음악 추천 방법

서용석, 임동혁
한국전자통신연구원(ETRI)

yongseok@etri.re.kr

Similar music recommendation method using music attribute tags

Yongseok Seo, Dong-Hyuck Im
Electronics and Telecommunications Research Institute

요약

본 논문은 음악의 무드, 악기, 테마 속성을 자동 태깅하고 이러한 음악 태그에 기반하여 입력 음악과 유사한 곡을 DB 에서 검색하여 추천하는 방법에 관한 것이다. 음악 분류를 위한 주요 속성인 무드, 악기, 테마는 각 음악 속성별 태그 데이터셋을 구성하여 딥러닝 모델을 학습한다. 학습이 완료된 모델을 활용하여 태그 예측값을 추출하고 이를 기반으로 태그별 유사도가 높은 음악을 검색하여 추천한다. 또한 실험을 통해 각 음악 속성의 자동 태깅 정확성을 검증하였다.

I. 서론

디지털 음악 시장이 커짐에 따라 방대한 양의 음악을 효과적으로 관리할 필요성과 함께 대용량 음악 검색이나 저작권-free 유사 음악 추천에 관한 기술 수요가 증가하고 있다.

본 논문에서는 음악의 무드, 악기, 테마를 자동태깅하는 딥러닝 모델과 이를 이용한 유사 음악 검색 및 추천에 관한 구성과 방법을 제안하고 각 자동태깅 모델의 성능을 확인한다.

II. 본론

최근 음악 자동태깅 연구는 대부분 딥러닝 방법을 이용하고 있다[1-2]. 오디오 신호에서 멜스펙트로그램과 같은 저수준의 특징을 추출한 후 CNN(Convolutional Neural Networks) 등을 이용하여 분류하는 방식이다.

본 논문에서는 음악의 무드, 악기, 테마를 자동으로 태깅하는 딥러닝 모델을 학습하고, 이를 기반으로 유사한 음악을 검색 및 추천할 수 있는 방법을 소개한다. 딥러닝 기반 모델 학습을 위해서는 음악 데이터와 음악별 무드, 악기, 테마가 태깅된 데이터셋이 준비되어야 한다.

정제하여 45 개의 무드 태그를 갖는 약 20 만곡을 필터링하였다[3]. 악기 태깅을 위한 데이터셋은 Free Music Archive 음원 데이터를 기반으로 polyphonic 악기 태그로 구성된 레이블 데이터를 재구성한 총 2 만개의 오디오 데이터셋인 OpenMIC-2018 을 활용하였다. 악기 관련 태그는 관악기, 현악기, 타악기 등의 태그들과 함께 'voice' 태그를 포함하고 있으며, 그림 2 와 같은 총 20 개의 태그들로 구성된다.

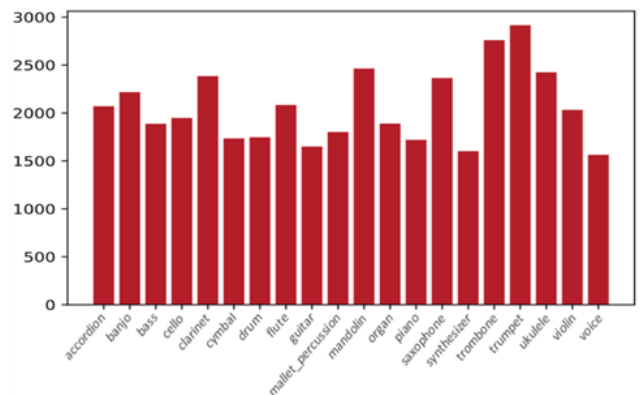


그림 2. OpenMIC-2018 데이터셋의 악기 태그 분포

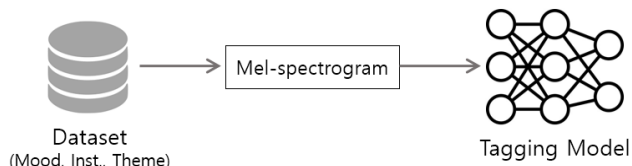


그림 1. 딥러닝 기반 음악 속성 자동태깅 모델

음악을 표현하는 속성 중 분위기나 느낌을 묘사하는 음악 무드 태깅을 위해 Million Song Dataset (MSD)을

음악을 표현하는 태그들 중 무드나 악기 태그와 다르게 주로 활용되는 사용처와 음악 자체가 표현하고자 하는 주제를 나타내는 테마 태그가 존재한다. 스페인의 UPF-MTG 에서 발표한 MTG-Jamendo 데이터셋은 음악 자동태깅 태스크에 대한 학습용 데이터셋으로 195 개의 다양한 카테고리 태그 약 55,000 곡으로 구성되어 있다. 본 논문에서는 MTG-Jamendo 데이터셋의 무드/테마 카테고리에서 14,788 곡을 선별하여 모델 학습에 사용하였으며, 그림 3 과 같이 23 개의 태그들로 구성된다.

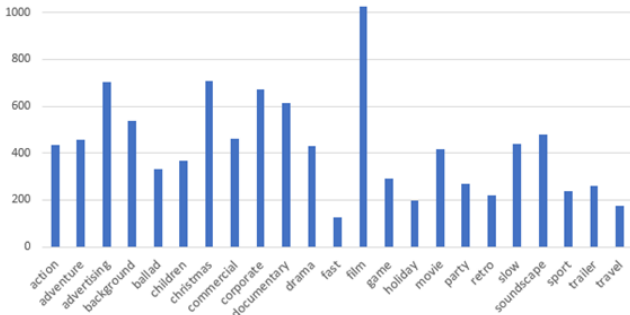


그림 3. Jamendo 데이터셋에서 선별한 테마 태그 분포

음악 자동태깅 모델은 7 개의 CNN layer 와 1 개의 FC layer, 그리고 각 태그에 대한 예측값을 출력할 수 있는 분류기를 포함한 구조로 구성된다. 음악의 무드, 악기, 테마와 같은 고차원 특징을 모델링하기 위해서 다양한 신경망 구조 모델이 활용될수 있으나, 본 논문에서는 계층적 구조를 갖는 다수의 CNN layer 연결 구조가 시간축 또는 주파수축의 제약없이 음악적 특징을 학습하기에 가장 적합한 것으로 판단하였다. 모델의 입력 데이터의 경우에도 waveform, mel-spectrogram 등의 다양한 입력 형태가 활용될 수 있는데, 10 초 길이의 음원을 CNN 모델에 적합한 mel-spectrogram 형태로 오디오 데이터를 변환하여 사용하였다.

음악의 무드, 악기, 테마 속성을 자동 태깅하는 모델의 성능 평가를 위해 각각 52,524 개, 3,533 개, 1,849 개의 테스트 데이터를 대상으로 ROC-AUC 값을 측정한 결과 45 개 무드 태깅 정확도 0.828, 20 개 악기 태깅 정확도 0.919, 23 개 테마 태깅 정확도는 0.839 로 측정되었다.

학습이 완료된 자동태깅 모델들에 대한 입력 음원의 태그 예측값을 추출하여 추천 음악 DB 의 태그 예측값들과의 코사인 유사도를 계산하여 가장 유사한 음악을 추천하는 방법을 그림 4 에 나타내었다. 또한 그림 5 와 같이 임의의 입력 음악에 대해 무드, 악기, 테마 속성이 유사한 음악을 Free Music Archive 의 2 만곡

데이터셋을 검색 대상으로 활용하여 Top-N 을 추천하는 것이 가능함을 확인하였다.

III. 결론

본 논문에서는 음악의 무드, 악기, 테마 속성을 자동태깅하는 딥러닝 모델의 구성과 모델 학습을 위한 데이터셋 구성 방법을 제안하였다. 제안한 방법으로 학습한 자동태깅 모델의 성능을 확인하고, 이를 이용하여 음악 속성 태그를 이용한 유사음악 검색 및 추천 방법과 결과를 소개하였다. 본 연구결과는 음악 정보검색 분야 뿐만 아니라 저작권 침해가 의심되는 음악을 대체하는 유사 음악 검색 및 추천에도 활용될 것으로 기대한다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024 년도 저작권기술 연구개발사업으로 수행되었음 (과제명: 교육 콘텐츠에 대한 인공지능 기반 저작권 침해 의심요소 검출 및 대체 재료 콘텐츠 추천 기술 개발, 과제번호: CR202104003)

참 고 문 헌

- [1] K. Choi, G. Fazekas, and M. Sandler, "Automatic tagging using deep convolutional neural networks," in Proc. ISMIR 2016, pp. 805-811.
- [2] T. Kim, J. Lee, J. Nam, "Sampling-level CNN architectures for music auto-tagging using raw waveforms," in Proc. ICASSP 2018, pp. 366-370
- [3] 서용석, 임동혁, "음악 분류 모델 학습을 위한 데이터셋 구성," 한국통신학회추계학술발표회, 2022.

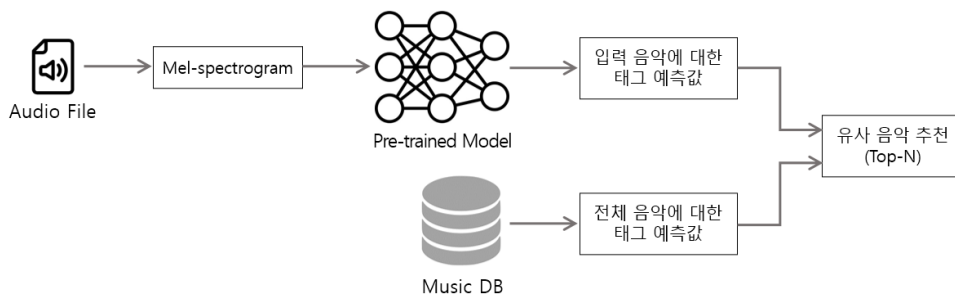


그림 4. 유사 음악 검색 및 추천 기술 구성도

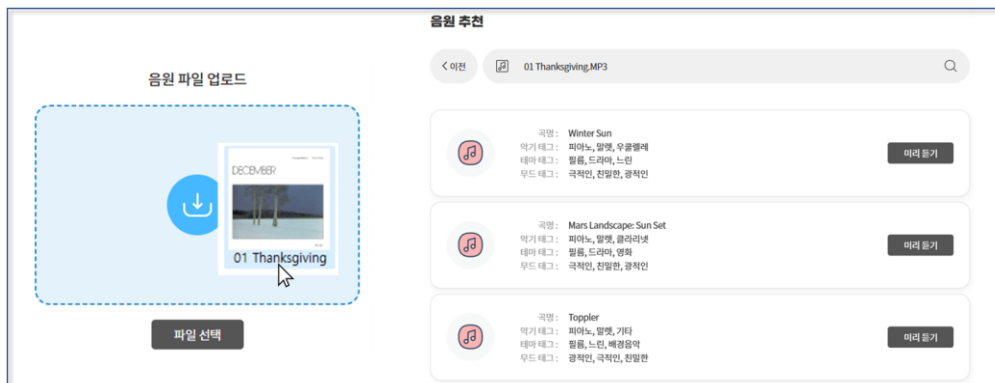


그림 5. 유사 음악 검색 및 추천 결과 예시