

# 영화 장르 메타데이터 생성을 위한 오디오 활용 방법에 대한 연구

용성중\* · 박효경 · 유연휘 · 문일영

한국기술교육대학교

## A Research on the Audio Utilization Method for Generating Movie Genre Metadata

Sung-Jung Yong\* · Hyo-Gyeong Park · Yeon-Hwi You · Il-Young Moon

Korea University of Technology and Education

E-mail : zeros@koreatech.ac.kr / sjjshk@koreatech.ac.kr / wlwhs3258@koreatech.ac.kr /

iymoon@koreatech.ac.kr

### 요 약

지속적으로 인터넷 및 디지털의 발전으로 많은 양의 미디어 데이터를 저장하고 온라인을 통해 개인에게 맞춤형 서비스를 제공하는 플랫폼이 등장하고 있다. 이러한 서비스를 제공하는 업체들은 미디어의 소비를 촉진 시키기 위해 개인 취향에 맞는 영화를 추천한다. 각 업체에서는 사용자가 선호할 미디어 추천을 위해 다양한 알고리즘에 대해 많은 연구를 하고 있다. 영화는 액션, 멜로, 공포, 드라마 등으로 장르를 구분하고 있으며, 영화의 오디오(음악, 효과, 음성)는 영화를 구성하는 중요한 제작 요소로 자리 잡고 있다. 본 연구에서는 영화예고편을 바탕으로 장르별 오디오를 추출하고, 장르별 오디오의 공통점을 확인 후 인공지능의 지도학습을 통해 영화 장르를 구별하고 추후 메타데이터 생성을 위한 활용방안을 제안하고자 한다.

### ABSTRACT

With the continuous development of the Internet and digital, platforms are emerging to store large amounts of media data and provide customized services to individuals through online. Companies that provide these services recommend movies that suit their personal tastes to promote media consumption. Each company is doing a lot of research on various algorithms to recommend media that users prefer. Movies are divided into genres such as action, melodrama, horror, and drama, and the film's audio (music, sound effect, voice) is an important production element that makes up the film. In this research, based on movie trailers, we extract audio for each genre, check the commonalities of audio for each genre, distinguish movie genres through supervised learning of artificial intelligence, and propose a utilization method for generating metadata in the future.

### 키워드

영화음악, 오디오, 인공지능, 메타데이터

### 1. 서론

지속적인 인터넷 및 디지털 발전으로 인해 다양한 멀티미디어 서비스가 늘고 있으며 그중에서도 많은 양의 미디어 데이터를 저장하고 온라인을 통

해 개인에게 맞춤형 서비스를 제공하는 플랫폼이 등장하고 있다.

글로벌 OTT 시장규모는 2018년부터 2023년까지 연평균성장률이 13.8%이며 국내 OTT 시장규모는 2013년부터 2018년까지 연평균성장률이 28.1%인 것으로 나타났다. 국내 OTT 시장에서 넷플

---

\* corresponding author

릭스의 시장점유율이 점차 확대되고 있고 향후 글로벌 사업자들의 국내 시장 진출로 경쟁이 치열해질 전망이다.[1] 이러한 서비스를 제공하는 업체들은 미디어의 소비를 촉진 시키기 위해 개인 취향에 맞는 영화를 추천한다. 개인의 취향을 위해 서비스 업체들은 영상의 메타데이터를 생성해 개인의 정보와 매칭하여 개인 맞춤형 서비스를 진행하고 있는데 현재 메타데이터의 생성은 사람이 영상을 보고 직접 입력하는 형태이다.

따라서 본 연구에서는 영상 미디어의 메타데이터를 추출하기 위해 영화의 장르 구분을 오디오 요소를 사용하여 활용 가능 여부를 확인한다. 또한 인공지능을 적용하여 영상 미디어의 오디오를 통해 장르를 구분할 수 있는 과정에 관한 연구를 진행하였다.

## II. 연구방법

본 연구에서는 영화의 오디오를 통해 메타데이터 생성에 있어 활용 여부를 연구한다. 우선 영화에서 음악과 음향효과 등이 어떻게 활용되고 있는지 확인하고 영화의 오디오를 통해 장르를 구분할 수 있는 요소로 활용할 수 있는지와 인공지능에 적용 방안을 제시하고 결과를 확인하고자 한다.

또한, 영화의 용량 및 저작권 문제로 영화 데이터 확보에 대한 한계점을 극복하고자 영화 예고편을 사용하여 예고편의 오디오를 분석하고 인공지능에 적용한다.

### 2.1 영화 장르 구분을 위한 오디오 추출 및 분석

영화음악은 영화와 밀접한 관계성을 맺고 있다. 영화에서 흐르는 음악만 듣고도 그 장면의 이미지를 그릴 수 있을 만큼 음악은 영화의 특징과 장면의 정서를 표현해내는 탁월한 기능을 갖고 있다. [2] 영화음악 및 효과는 영화 속에서 특징과 정서를 나타내는 요소로 영화의 오디오를 통해 장르를 파악할 수 있는 요소로 충분하다. 영화의 장르를 액션, 공포, 코믹, 멜로 4가지로 구분하고 오디오를 추출하고 위해 STFT(Short Time Fourier Transform)을 이용하여 시간에 따른 주파수 성분의 변화에 대해 분석하여 스펙트럼 이미지를 획득하였다.

표 1. 오디오 분석 과정

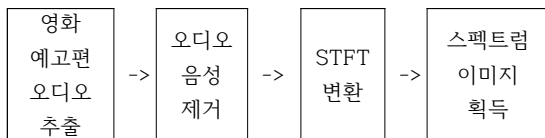


표 2. 영화 장르 구분

장르	영화예고편
코믹	극한직업
	정직한후보
	미쓰와이프
	해치지않아
액션&판타지	고질라
	백두산
	트랜스포머
공포	아이언마스크
	더 년
	어스
	곤지암
멜로	0.0MHz
	노팅힐
	개를
	원스
	뷰티인사이드

표 2와 같이 장르별 4편의 대표 영화 예고편을 선정하고 영상과 오디오를 분리하였다. 그리고 분리된 오디오에서 음성을 제거하여 STFT를 통해 각 영화 예고편의 오디오 스펙트럼을 생성하여 이미지로 저장하고 딥러닝 학습데이터로 처리하기 위해 카테고리별 분류를 하였다.

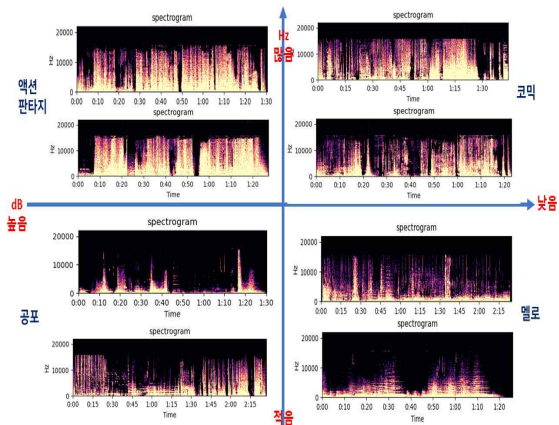


그림 1. 오디오 스펙트럼을 통한 장르 분류

그림 1과 같이 STFT를 통해 생성된 장르별 스펙트럼 이미지를 주파수와 데시벨을 기준으로 포지셔닝 맵으로 분류해보니 장르별 스펙트럼의 파형이 비슷하게 출력됨을 확인할 수 있었다. 이를 통해 스펙트럼 이미지를 인공지능에 적용한다면 오디오의 스펙트럼 이미지를 추출 후 학습된 인공지능에 입력하면 장르를 구분할 수 있을 것으로 예측하여 2단계 인공지능에 적용하여 결과를 확인하였다.

2.2 인공지능 학습 및 평가결과

오디오 분석 결과를 바탕으로 스펙트럼 이미지를 장르별로 학습 데이터로 분류하고 CNN과 로지스틱 회귀분석 모델을 적용하여 장르별 스펙트럼 분류 정확도를 확인해 보았다.

표 3. 인공지능 적용 과정

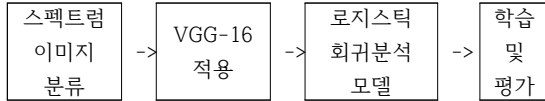


표 3과 같이 스펙트럼 이미지를 “오렌지3” 머신러닝 툴을 이용하여 학습과 평가를 진행하여 최종 결과를 확인하였다.

III. 연구결과

그림 2와 같이 Confusion Matrix 결과 공포, 멜로, 코믹 이미지는 장르별 4개의 스펙트럼이 모두 정확하게 분류된 것을 확인할 수 있었다. 그러나 액션 장르의 2개 이미지가 분류되지 못한 것을 확인하였다.

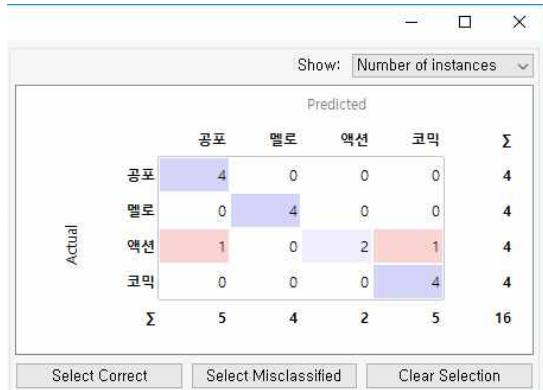


그림 2. Confusion Matrix 결과화면

그림 3은 평가 결과를 수치화 자료를 표현한 화면이다. 정확도는 97% 정도와 정밀도 90% 정도의 수치를 나타내고 있다. 평가율이 높게 나타나고 있다는 것은 오디오 스펙트럼을 바탕으로 장르 구분을 위해 의도한 학습데이터 제대로 인식하고 분류한다는 것이다. 이러한 결과는 영화 속 오디오를 이용하여 인공지능이 장르를 구분할 수 있음을 확인하고 영화의 장르를 구분한다면 장르에 대한 메타데이터를 자동생성이 가능하다는 것을 알 수 있다.

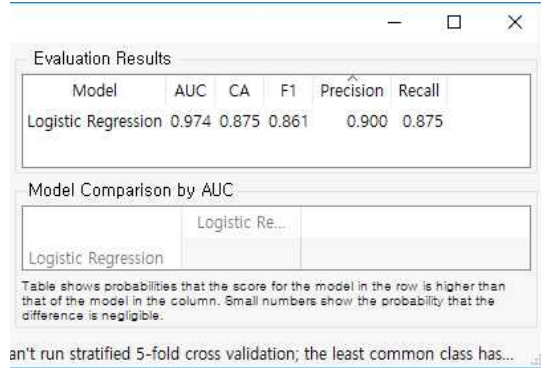


그림 3. Evaluation Result 결과화면

IV. 결 론

디지털 및 인터넷의 발달로 다양한 멀티미디어 서비스가 생겨나고 있으며, 개인 맞춤형 영상 제공 플랫폼이 많이 생겨나고 해외 업체들이 국내 시장까지 진출하려고 한다. 각 업체에서는 영상의 소비 촉진을 위해 지속적으로 개인 맞춤형 서비스 연구를 진행하고 있으며, 영상의 메타데이터에 관한 연구도 활발히 진행 중이다. 본 연구에서는 기존 사람들이 직접 메타데이터를 입력하는 방식에서 메타데이터의 요소들을 인공지능으로 학습 후 자동으로 생성하는 방법에 대해 생각하였다. 그 중 첫 번째 요소로 장르별 오디오 스펙트럼을 추출하고 스펙트럼 이미지를 CNN과 로지스틱 회귀모델에 적용하였을 때 정확도가 높게 나타났다. 하지만 액션 장르의 실패 이미지의 정확한 학습을 위해 사전 학습 데이터의 종류와 양을 늘려 학습을 진행하고, 이미지를 분류할 수 있는 특징요소들을 찾아 추가 적용하여 인공지능의 정확도 높이고 메타데이터 자동생성 시스템을 구축한다면 개인 맞춤형 서비스 및 소비를 촉진 시킬 수 있을 것으로 기대한다.

Acknowledgement

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업 (No. 2021R111A3057800) 과제 지원에 의하여 연구되었음.

References

[1] No Dong Hwan, “The prospects and market changes of OTT-based broadcasting services”, Korean Association for Broarding & Telecommunication, p101-101, 2019.11

[2] Jung Ji Young, “The Correlation of Bach Music and the Scene as Seen in Films”, Music in Modern Media The Graduate School of Ewha Womans University p1, 2007“