동영상 추천을 위한 동영상 콘텐트 분석

이원재, 신일홍, 이준규*
한국전자통신연구원, *주니스에이아이

russell@etri.re.kr, ssi@etri.re.kr, *zzunkyu@gmail.com

Video Content Analysis for Video Recommendation

Wonjae Lee, Ilhong Shin, Junkyu Lee*
Electronics and Telecommunications Research Institute, *JunisAI

요 약

본 논문에서는 동영상 추천을 위한 동영상 콘텐트 분석 시스템에 대해 논의한다. 동영상에 등장하는 객체, 인물, 배경 음악에 대한 인식 및 분석은 동영상 추천에 사용되어 시청자 취향을 더 정확하게 반영할 수 있다. 개발된 동영상 콘텐트 분석 시스템은 딥러닝 기술을 활용하여 이러한 인식 및 분석 작업을 수행한다.

Ⅰ. 서론

OTT 서비스(Over-The-Top media service)와 같은 동영상 서비스 제공 사업자는 지속적인 구독 또는 광고 노출을 위해 시청자의 취향에 맞는 동영상을 추천할 필요가 있다. 동영상 콘텐트 추천 시스템에서는 협업 필터링(Collaborative Filtering), 내용 기반 필터링(Content-Based Filtering), 딥러닝 기반 추천 기술 등을 사용한다.

최근 딥러닝 기술의 발전으로 동영상의 시각적 특징, 청각적 특징을 효과적으로 분석 및 추출할 수 있게 되었다. 이러한 분석 결과는 동영상 추천에 사용되어 추천 정확도를 향상시킬 수 있다 [1].

본 논문에서는 동영상 추천을 위해 개발한 동영상 콘텐트 분석 시스템을 소개한다.

Ⅱ. 본론

본 동영상 콘텐트 분석 시스템에서는 주로 딥러닝 기술을 사용하여 시각적, 청각적 특징을 추출한다.

1. 등장 객체 인식

시청자가 자동차, 비행기, 반려동물 등 특정 객체에 관심이 있는 경우 동영상에 등장하는 객체를 인식하여 추천에 활용하면 시청자 취향에 맞춘 동영상 추천이 가능하다. 이를 위해 본 시스템에서는 동영상에 등장하는 객체를 인식하는 기술을 개발하였다.

등장 객체 인식 시 정확도를 높이기 위해 EfficientNet-b4 [2]모델로 실내, 실외를 구분하여 인식하였다. 실내/실외 구분 후, 단일 모델을 사용하여 검출(detection)과 분류(classification)를 모두 수행할 수도 있으나, 검출과 분류를 별도로 수행하여 더 높은 정확도를 달성하였다. 객체 검출에는 YOLO v5 large 모델을, 분류에는 EfficientNet-b4 와 ResNet-50 [3]모델을 사용하였다. Class 종류가 적을 때는 EfficientNet 이, Class 종류가 많을 때는 ResNet 이 조금 더 좋은 성능을 보여, 상황에 맞추어 성능이 더 우수한 모델을 사용하였다. EfficientNet 은 Class 종류가 많아지면 특정 Class 에 대한 정확도가 현저히 떨어지는 경우가 있었다.

자동차의 경우 관심을 가지는 시청자들이 많을 것으로 예상되어 주요 자동차 모델에 대해 제조사, 모델, 색상을 인식하는 기능을 개발하였다. 모델 분류에는 ResNet-50 모델을, 색상 분류에는 EfficientNet-b4 모델을 사용하였다.

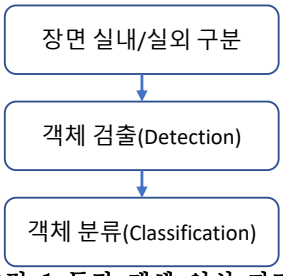

그림 1 등장 객체 인식 과정

2. 사람 얼굴 인식

시청자가 특정한 유형의 얼굴을 가진 배우들을 선호할 수 있으며, 이런 경우 동영상에 등장하는 사람 얼굴을 인식하여 분석하면 추천에 도움이 될 수 있다. 이를 위해 본 시스템에서는 동영상에 등장하는 사람 얼굴을 인식 및 분석하는 기능을 구현하였다. 동영상에 등장하는 사람 얼굴을 검출한 후, 표정을 분류하고, 유사한 얼굴들을 군집화(clustering)하였다.

얼굴 검출 시 RetinaFace [4]모델을 사용하였다. 표정 분류에는 EfficientNet-b4 모델을 사용하였다.

군집화에는 계층적 군집화(Hierarchical Clustering)를 사용하였다.

3. 음악 분류

시청자가 특정한 종류의 배경 음악을 선호하거나 특정한 분위기의 음악이 어울리는 장면들을 선호할 수 있으며, 이런 경우 음악 종류를 인식하여 분석하면 추천에 도움이 될 수 있다. 이를 위해 본 시스템에서는 동영상에 사용되는 배경 음악을 분류하는 기술을 개발하였다.

음악 분류를 위해 오디오 데이터를 10 초 단위로 분리하고, 분리된 오디오 데이터를 스펙트로그램(Spectrogram)으로 변환한다. 스펙트로그램에 대해 EfficientNet-b4 모델을 사용하여 서정적인(lyrical), 불안한(Uneasy), 신나는(Exciting), 신비로운(Dreamy) 4 종류로 분류한다.

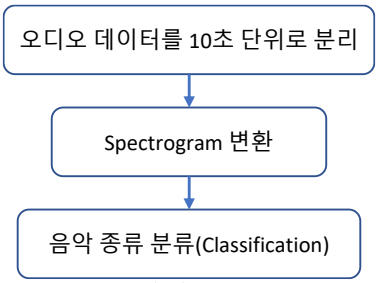

그림 2 음악 분류 과정

4. 분석 결과 관리 시스템

콘텐츠 분석 결과를 편리하게 관리할 수 있는 관리 시스템을 개발하였다. 분석결과는 MariaDB 에 저장되고, PyQt5 기반으로 개발된 사용자 인터페이스를 통해 분석결과에 대한 확인 및 수정이 가능하다.

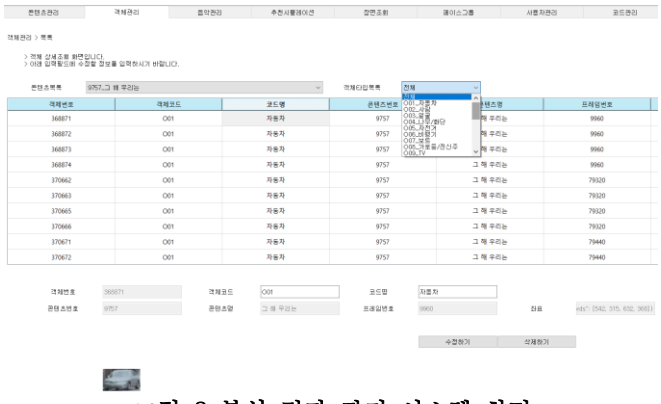

그림 3 분석 결과 관리 시스템 화면

Ⅲ. 결론

본 논문에서는 동영상 추천을 위한 동영상 콘텐트 분석 시스템에 대해 기술하였다. 분석 시스템에서는 관심 객체에 대한 선호도를 추천에 반영할 수 있도록 딥러닝 기술을 활용하여 등장 객체를 검출 및 분류한다. 선호 얼굴을 추천에 활용할 수 있도록 사람 얼굴 검출, 표정 분류, 군집화를 수행한다. 음악 취향을 고려할 수 있도록 음악 분류 기능을 구현하였다. 그리고 분석 결과를 편리하게 관리할 수 있는 관리 시스템을 개발하였다.

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임. (2022-0-00215, 다시점 자유도를 제공하는 OTT 플레이어 지능화 기술개발)

참 고 문 헌

[1] Y. Li, H. Wang, H. Liu, B. Chen. “A study on content-based video recommendation.” IEEE, Beijing (2017), pp. 4581-4585.

[2] Mingxing Tan and Quoc V Le. “EfficientNet: Rethinking model scaling for convolutional neural networks.” In Proceedings of International Conference on Machine Learning (ICML), 2019.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep residual learning for image recognition.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770-778, 2016.

[4] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, Stefanos Zafeiriou. “RetinaFace: single-stage dense face localisation in the wild.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5203-5212.