

시청자 반응을 통한 영상 콘텐츠 객체의 감정 영향력 측정과 평가

이소정^{*1} 손규진¹ 신희원¹ 김혜린¹곽서현¹ 김영민¹ 남은비¹ 김다혜² 김윤희¹

숙명여자대학교 소프트웨어학부¹, 홍보광고학과²

{sojeong17; cnbe5494; gmlndjns9803}@sookmyung.ac.kr 97grapefruit614@gmail.com

seohyun0221@sookmyung.ac.kr {youngmin6816; eunbi2241}@gmail.com yulan@sookmyung.ac.kr

Effect Evaluation on Sentiment objects of Video content and their corresponding response

Sojeong Lee^{*1} Gyujin Son¹ HeeWon Shin¹ Hyerin Kim¹ SeoHyun Kwag¹ YoungMin Kim¹ Eunbi Nam¹

DaHye Kim² YoonHee Kim¹

Dept. of Software¹, Dept. of Public Relations and Advertising², Sookmyung Women's University

요 약

최근 다양한 영상 콘텐츠가 늘어남에 따라 단순히 텍스트로 반응을 표시하는 방법에서 반응 동영상을 통해 자신의 감정을 표현하는 등 반응이 다양해졌다. 따라서 반응 동영상 속 시청자의 감정을 분석할 필요가 있다. 본 연구에서는 영상 콘텐츠 속 다양한 등장 오브젝트를 기계학습을 이용하여 추출한다. 관련된 반응 동영상 속 시청자의 안면 감정을 측정하고 분석하여 영상 콘텐츠의 영향력이 있는 감정 오브젝트를 추출하고 사전화한다. 구축된 감정 오브젝트 사전의 영향력 효과를 검증한 결과 주요 감정 영향력에서 일치함을 보였다. 이는 동영상 제작 시에 특정 감정에 효과적인 오브젝트를 추천할 수 있어 의도된 영상 콘텐츠의 감정 효과를 기대할 수 있다.

1. 서 론

현대인의 영상 콘텐츠 시청이 증가함에 따라 자신의 감정을 영상으로 표현하는 반응 동영상의 수가 증가하고 있다. 따라서 영상 콘텐츠 속 오브젝트에 따른 시청자의 반응을 분석하는 시스템이 필요하다. 이에 따라, 영상 콘텐츠로부터 인물의 감정을 분석하기 위해 감정을 도출하는 연구[1]가 진행되고 있다.

이전 연구[4]에서 제안한 V-SAS(Video Sentiment Analysis System with Deep Learning)는 사람의 표정 이미지에 대한 감정 분석 연구[2]와 이미지 딥러닝을 사용해 영상의 오브젝트를 추적하고 인식하는 연구[3]를 활용하여 영상 콘텐츠의 유의미한 오브젝트를 특정하였다. 하지만 특정 구간의 유의미한 오브젝트들이 존재함을 분석하였으나, 다중 오브젝트에 대한 분석이 미비하고 시청자의 감정에 영향을 미치는 오브젝트별 감정을 분석하고 이에 대한 효과를 입증하지는 못하였다.

본 연구에서는 V-SAS를 확장한 SOISA(Sentimental Object Identification System using Sensitivity Analysis)를 구축하여 감정별 오브젝트 사전의 효과를 검증하는 것을 목적으로 한다. SOISA는 오픈소스 API를 이용하여 영상 콘텐츠의 구간별로 관련 반응 동영상 속 시청자의 감정을 분석하고 이를 오브젝트별 감정값을 측정한다. 감정값을 분석하기 위해 유의미한 감정값

을 가지는 오브젝트를 파악한다. 나아가 영상 콘텐츠 속 다양한 감정 오브젝트를 추출하여 이를 토대로 감정별 오브젝트 사전을 구축한다. 시청자들의 설문조사를 바탕으로 개별 오브젝트의 영향력을 비교하여 시스템의 효과를 검증하고자 한다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구, 3장은 SOISA의 개요와 기능 설계, 4장에서는 기능 및 실험 결과를 설명하며 5장으로 결론을 맺는다.

2. 관련 연구

다양한 영상 콘텐츠에 대한 시청자들의 감정과 관련한 연구가 이뤄지고 있다. [1] 논문에서는 표정뿐만 아니라 얼굴의 비언어적 특징 22개를 고려하여 6가지 감정을 도출하는 방법을 제시한다. 하지만 6가지 감정의 수는 감정을 분류하기에 적은 가지 수이고 해당 감정의 원인을 분석하지는 않았다.

[2] 논문에서는 영상에서 딥러닝을 통해 특정 오브젝트를 인식하고 검출하는 기술들에 관해 설명하고 있다. 그러나 오브젝트를 특정 목적 없이 모두 검출하기에 오브젝트의 등장으로 인한 효과를 정확하게 파악할 수 없다.

[3] 논문에서는 시청자의 표정 변화가 녹화된 동영상으로부터 감정을 인식하는 방법을 제시한다. 영상 전체 프레임을 서브 샘플링(Sub sampling)하여 감정별 프레임을 수집해 얼굴의 기하

* 이 논문은 2020년도 정부재원(과학기술정보통신부 여대학원생 공학연구팀제 지원 사업)으로 과학기술정보통신부와 한국여성과학기술인지원센터의 지원을 받아 연구되었습니다.

학적 특징과 관련된 3차원 좌표 점을 구하고 모든 쌍의 유클리드 거리를 구해 모델 학습에 필요한 특징으로 사용하였다.

이전연구인 [4] 논문에서는 콘텐츠와 관련 반응 동영상의 감정 분석 방법을 제안하였다. 이미지 딥러닝을 통해 대상 콘텐츠 속 오브젝트를 찾고 해당 오브젝트에 어떤 감정을 표현하는지 파악하여 정량화하였다. 하지만 다양한 콘텐츠의 오브젝트별 감정값을 분석하여 사전화하지 않았고, 추출된 오브젝트의 감정 영향력을 분석에 관한 연구는 부족하다.

3. SOISA(Sentimental Object Identification System using Sensitivity Analysis)

3.1 시스템 구조

SOISA는 사용자가 입력한 URL의 콘텐츠 속 오브젝트에 대한 시청자의 감정값을 분석하여 감정별 오브젝트를 추출하여 감정별 오브젝트 사전을 구축하는 시스템이다.

이전 연구[4]에서 제안한 V-SAS의 시스템 기능을 향상해 설계하였다. SOISA의 추가된 기능은 다음과 같다. 오브젝트 감정값 추출기(Object Sentiment Extractor)를 통하여 유의미한 구간에 등장한 모든 오브젝트에 대해서 오브젝트의 감정 영향력 분석 알고리즘을 통하여 감정값을 계산한다. 추출된 감정 영향력 오브젝트를 감정별 오브젝트 사전 구축기(Sentimental Object Dictionary Builder)에서 감정별로 정렬하여 사전을 구축한다.

3.2 오브젝트 감정값 분석

본 연구에서는 ‘빌리 아일리시 - bury a friend’, ‘marshmello - happier’ 뮤직비디오 2개를 대상 콘텐츠로 설정하여 오브젝트 감정값을 분석하였다. 오브젝트 감정 분석 방법은 이전 논문[4]에 기술된 방법으로 분석한다.

대상 콘텐츠에 대한 10개의 반응 동영상의 8가지 감정을 추출하고, 이에 대한 시간순서의 평균을 낸다. 이에 대한 감정값 시계열 그래프는 그림 1, 2와 같다. 그래프에 나타난 감정별 색상은 표1과 같다. 표1 중 neutral 감정은 분석에서 제외한다.

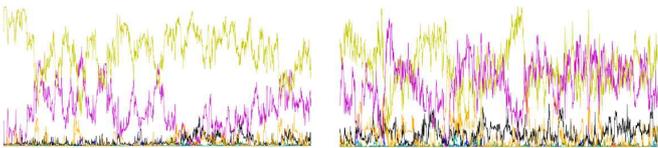


그림1 Marshmello - Happier 감정값 시계열 그래프
그림2 빌리 아일리시 - bury a friend 감정값 시계열 그래프

anger	contempt	disgust	fear
파랑	초록	빨강	하늘
happiness	neutral	sadness	surprise
보라	연두	검정	주황

표1 시계열 그래프 감정 색상표

3.3 오브젝트의 감정 영향력 분석 알고리즘

유의미한 구간에 등장하는 오브젝트의 감정 영향력을 분석하

기 위해서는 앞뒤 구간의 오브젝트와 시청자 감정과의 변화율을 분석해야 한다. 감정 영향력 분석 알고리즘은 앞뒤 구간의 오브젝트와 비교하여 앞뒤에 존재, 앞이나 뒤 중 하나만 존재, 둘 다 존재하지 않음의 3가지로 분류하여 오브젝트의 감정값 추출기를 적용하였다. 이를 분석하는 알고리즘은 다음과 같다.

유의미한 구간의 오브젝트가 존재할 때, 연속된 앞뒤 구간의 오브젝트를 비교한다. 모든 구간에서 등장하는 오브젝트의 경우 감정값에 영향을 주지 않는다고 판단하여 제외한다. 한 구간에서만 존재하는 오브젝트의 경우 존재하지 않는 구간에서의 감정 변화율 값을 계산한다. 모든 구간에서 존재 않을 시에는 앞뒤 구간의 감정 변화율 값을 평균 계산한다. 유의미한 구간의 오브젝트가 2개 이상 존재할 시에는 앞서 설명한 3가지의 판단기준을 토대로 순위를 매긴다.

4. 실험 환경 및 결과 분석

4.1 실험 환경

본 논문에서 제안한 시스템은 딥러닝 서버 1대, Intel Xeon E5-2630V3, 옥타코어, 캐시는 20MB이며 속도는 2.4GHz이다. 소프트웨어는 Python 3.7, 딥러닝 학습은 YOLO v2, CUDA 8.0, cuDNN v5.1을 사용하고 동영상 구간 분할은 OpenCV 3.2버전을 사용하였다.

4.2 대상 콘텐츠의 유의미한 구간

모든 대상 콘텐츠에 대해서 감정 변화량 값을 통합하여 평균 변화량보다 변화량이 큰 구간들을 유의미한 구간으로 추출하였다. ‘빌리 아일리시 - bury a friend’의 유의미한 구간은 91개, ‘marshmello - happier’의 유의미한 구간은 90개이다. 본 논문에서는 ‘marshmello - happier’의 유의미한 구간의 전체적 감정 변화 결과에 대해서만 예시로 설명하겠다.

그림 3은 ‘marshmello - happier’ 뮤직 비디오의 유의미한 구간 막대그래프이다. 이 그래프의 유의미한 구간의 수가 많은 것을 보아 시청자의 감정 변화가 큰 것을 알 수 있다. 전체적으로는 happiness, sadness, surprise의 감정이 큰 것을 알 수 있다. 특히 초반 부분의 happiness의 큰 값은 해당 구간이 생일 파티 장면이기 때문이라고 추정된다.

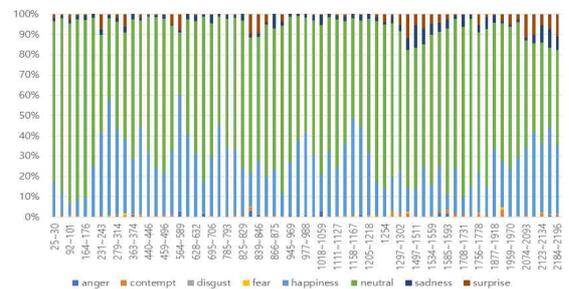


그림 3 marshmello - happier 유의미한 구간

4.3 감정별 오브젝트 사전 구축

대상 콘텐츠에 대해서 7가지 감정값에 따라 변화율 값이 큰 순서대로 정렬하여 감정별 오브젝트 사전을 구축하였다. 유의

미한 구간에 등장하는 오브젝트를 추출하여 오브젝트의 감정 영향력 분석 알고리즘을 통해 순위를 도출하였다. 같은 오브젝트가 여러 구간에서 발견되었을 시, 변화율의 평균을 구하였다.

표 2는 위에서 선정한 2가지 뮤직비디오 중 ‘marshmello-happier’에 등장하는 오브젝트들에 대해서 감정 영향력을 분석한 감정별 오브젝트 사전의 일부이다.

감정별 오브젝트 사전에서 happiness 중 11.65%로 감정변화율이 가장 큰 picture는 강아지와 여자아이가 함께 끌어안고 있는 사진이기 때문에 happiness가 크게 측정된 것으로 추정된다.

감정	오브젝트	감정 변화율(%)	순위
anger	orthodontics	0.13	1/1
happiness	picture	11.65	1/3
	cake	11.30	1/3
	cone_hat	9.99	1/3
sadness	soccer_ball	1.96	1/2
	dog	1.15	1/1
	balloon	0.99	1/3
surprise	soccer_ball	5.69	1/2
	dog	4.18	1/1
	ribbon	2.46	2/2

표 2 marshmello - happier 감정별 오브젝트 사전

4.4 감정별 오브젝트 사전 분석 및 비교 검증

본 연구에서는 20대 여성 30명을 대상으로 ‘marshmello - happier’의 오브젝트 감정에 대한 설문조사를 진행하였다. 대상 뮤직비디오에 등장하는 오브젝트들에 관해 anger, contempt, disgust, fear, happiness, neutral, sadness, surprise의 8가지 감정의 다중 투표를 허용하였다. 그림 4는 설문조사를 통해 얻은 등장 오브젝트별 감정 확률값을 나타내는 막대그래프이다.

위의 감정별 오브젝트 사전과 설문조사의 결과를 비교하기 위해 neutral을 포함하여 막대그래프를 도출하였다. 그림 5는 대상 뮤직비디오에 등장하는 모든 오브젝트에 대한 감정별 오브젝트 사전의 감정 변화율 값을 정규화한 막대그래프이다.

시청자 설문조사 결과인 그림 4에서 happiness 0.62의 확률값을 보인 picture의 경우 감정 오브젝트 사전 결과인 그림 5에서 happiness 0.77의 확률값을 보인다. 이를 보아 happiness 같은 주요한 감정값에 대해서는 시청자 설문조사 결과와 감정 오브젝트 사전이 비슷한 결과를 보임을 알 수 있다.

반대의 경우로 그림 4의 syringe의 경우 contempt 0.03, disgust 0.03, fear 0.12, neutral 0.35, sadness 0.43, surprise 0.03의 확률값을 보이지만 그림 5에서는 neutral 1.0의 확률값을 보인다. 이는 감정별 오브젝트 사전의 경우 감정 변화율 값을 계산하여 특정 수치 이하의 변화율을 가진 감정에 대해서는 유의미하지 않아 무시하기 때문이다.

결론적으로 유의미한 감정 변화율과 오브젝트의 감정값이 거의 일치한다. 그러나 영상 콘텐츠의 감정변화율이 무의미한 경우 대상 오브젝트는 추출되지 않아 무시되었으나 설문조사 결과는 감정이 표시되어 감정별 오브젝트 사전과 차이가 존재한다. 이는 설문조사 대상자가 오브젝트에 대한 의식적인 감정에

집중하는 현상으로 neutral보다 적극적으로 감정표현을 하는 것으로 판단할 수 있다. 따라서 영상 콘텐츠에서 유의미한 감정 변화율을 기준으로 추출된 오브젝트들의 감정별 오브젝트 사전은 시청자들의 감정을 직감적으로 추출하므로 보다 정확한 감정 분석하는 것에 있어 도움이 될 수 있을 것으로 판단한다.

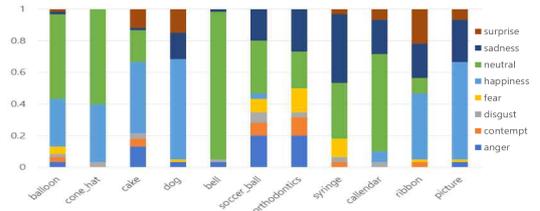


그림 4 marshmello - happier 등장 오브젝트별 감정 확률 값

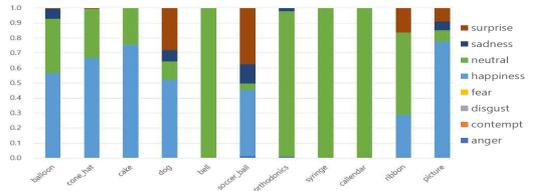


그림 5 marshmello - happier 등장 오브젝트별 감정 변화율

5. 결론

본 연구에서는 영상 콘텐츠 속 다양한 오브젝트를 기계 학습하여 추출하였다. 관련된 반응 동영상 속 시청자의 안면 감정을 측정, 분석하여 영상 콘텐츠의 영향력 있는 감정 오브젝트를 추출하고 사전화하였다. 감정 오브젝트 사전의 영향력 효과를 검증한 결과 주요 감정 영향력에서 일치함을 보였다. 이는 동영상 제작 시 특정 감정에 효과적인 오브젝트를 추천할 수 있어 의도된 영상 콘텐츠의 감정 효과를 기대할 수 있다.

참고 문헌

- [1] Philipp Michel, Rana El Kaliouby, “Real Time Facial Expression Recognition in Video using Support Vector Machines”, Proc. of the ICMI International Conference on Multimodal Interfaces 2003, pp. 258-264, 2003.
- [2] S. Ko, Y. Kim, “딥러닝을 이용한 오브젝트분류 및 검출 기술” Proc. of the IEIE Institute of Electronics and Information Engineers 2017, pp. 26-33, 2017. (in Korean)
- [3] K. Park, J. Hong, “Emotion Recognition for Spontaneous Facial Expression Video based on 3D Euclidean Distance” Proc. of the KIISE Korea Computer Congress 2017, pp. 938-940, 2017. (in Korean)
- [4] Y. Jeong, B. Kim, S. Ryu, N. Kwon, K. Son, S. Lee, “A Design of Sentiment Analysis System for Video Contents with Image Deep Learning”, Proc. of the KIISE Korea Software Congress 2019, pp. 1239 ~ 1241, 2019. (in Korean)