

얼굴 감정을 이용한 시청자 감정 패턴 분석 및 흥미도 예측 연구

조인구[†], 공연우^{**}, 전소이^{***}, 조서영^{****}, 이도훈^{*****}

A Study on Sentiment Pattern Analysis of Video Viewers and Predicting Interest in Video using Facial Emotion Recognition

In Gu Jo[†], Younwoo Kong^{**}, Soyi Jeon^{***}, Seoyeong Cho^{****}, DoHoon Lee^{*****}

ABSTRACT

Emotion recognition is one of the most important and challenging areas of computer vision. Nowadays, many studies on emotion recognition were conducted and the performance of models is also improving. but, more research is needed on emotion recognition and sentiment analysis of video viewers. In this paper, we propose an emotion analysis system the includes a sentiment analysis model and an interest prediction model. We analyzed the emotional patterns of people watching popular and unpopular videos and predicted the level of interest using the emotion analysis system. Experimental results showed that certain emotions were strongly related to the popularity of videos and the interest prediction model had high accuracy in predicting the level of interest.

Key words: Emotion Recognition, Deep Neural Network, Facial Image Classification, Sentiment Analysis

1. 서 론

Youtube, Netflix와 같은 각종 Over-the-top media service(OTT)를 이용하지 않는 사람을 찾아보기 힘든 시대가 되었다. 특히 Youtube와 같은 플랫폼에서 시청자들은 해당 영상이 재미있으면 ‘좋아요’를 누르는 방식으로 영상에 대한 호감을 표현하기도 하고, 재미가 없으면 금방 꺼버리고 다른 영상을 찾기도 한다. 굉장히 높은 조회 수와 좋아요 수를 보면서, 사람들은 ‘많은 사람이 시청하는 영상의 공통점, 선

호의 근거는 무엇일까?’와 같은 궁금증을 가지게 된다. 동시에 개인이 쉽게 영상을 만들고 게시할 수 있는 시대에서, 많은 영상 제작자들은 시청자 반응 분석이 필요함과 동시에 이에 어려움을 느끼고 있다.

본 논문에서는 시청자 반응 중에서 시청자의 표정을 통한 감정분석과 이를 기반으로 한 인기 영상의 공통점을 분석하고자 한다. 이를 위하여 비디오 영상물 시청에 따른 시청자 얼굴 감정을 인식한다. 감정 인식의 로그를 이용하여 흥미 여부에 따른 감정 변이를 분석한다.

* Corresponding Author : DoHoon Lee, Address: (46241) Busandaehak-ro 63beon-gil 2, Geumjeong-gu, Busan, Korea, TEL : +82-51-510-2491, E-mail : dohoon@pusan.ac.kr

Receipt date : Feb. 12, 2022, Approval date : Feb. 23, 2022

[†] Dept. of Information Convergence Eng., Graduate School, Pusan National University (E-mail : koddakggy@pusan.ac.kr)

^{**} School of Computer Science and Eng., Pusan National University (E-mail : gyeonu99@naver.com)

^{***} School of Computer Science and Eng., Pusan National University (E-mail : jeonsy@pusan.ac.kr)

^{****} School of Computer Science and Eng., Pusan National University (E-mail : 201824589@pusan.ac.kr)

^{*****} School of Computer Science and Eng., Pusan National University

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A2A030 45884)

본 논문의 구성은 다음과 같다. 2장에서 딥러닝 기반의 감정인식 및 분류에 대해 간단히 소개한 후, 3장에서는 실험에 사용된 감정분석 시스템과 모델에 관해서 설명한다. 4장에서는 실험 방식과 실험 결과를 설명한다. 마지막으로 5장에서는 결론과 향후 활용 방안을 얘기한다.

2. 관련 연구

얼굴 표정 이미지를 통한 감정인식은 초기에 Action Unit 분석을 통해 이루어졌다. Ekman과 Friesen[10]은 1978년에 인간의 얼굴에서 나타날 수 있는 근육의 움직임들을 나타내는 FACS(Facial Action Coding System)를 발표했다. FACS에서는 46가지의 얼굴 근육 움직임들의 조합을 통해 감정을 판별한다.

이미지 분류에서 CNN이 높은 성능을 가지면서 VGGNet[6]이나 AlexNet[7]등 많은 layer를 가지는 모델들이 얼굴 이미지의 감정분석에 이용되고 있다. Mehendale[1]은 사람의 얼굴이 포함된 이미지에서 배경을 먼저 제거한 후 얼굴 이미지에서 Expressional vector extraction을 통해 감정분류를 진행하였고, 기존의 VGGNet, GoogleNet[8], Resnet[9]에 비해 높은 정확도를 가졌다. 배경을 먼저 제거한 이미지를 CNN network에 적용했기 때문에 계산 복잡도 또한 AlexNet과는 유사했으며 다른 모델보다는 훨씬 낮았다.

CNN 이후 자연어 처리에서 사용되던 transformer 개념을 이미지 처리에 적용한 ViT(Visual

Transformers) 모델이 감정분석에서 높은 성능을 나타내고 있다. Fuyan 등[2]은 이미지를 곧바로 Transformer Encoder에 적용하는 Visual Transformers 모델이 아닌 CNN을 통과한 feature maps를 Multi-layer Transformer Encoder에 적용하는 CVT 모델을 제안하였다. Global-local attention과 global self-attention을 통해 LBP feature[12]와 CNN feature를 통합시켜서 Transformer 모듈에 적용했다. RAF-DB, FERPlus 그리고 AffectNet 데이터셋을 이용하여 정확도를 계산하였으며 기존의 모델들에 비해 훨씬 우수한 성능을 나타냈다.

3. 감정분석 시스템

3.1 시청자 감정분석 시스템

먼저 시청자 감정분석을 위해 영상을 시청하는 시청자의 표정을 수집하고 그 표정을 분석하는 시스템을 제안한다. 제안하는 시스템은 영상물을 시청할 때의 감정을 샘플링 한다. 샘플링된 얼굴은 학습된 감정분류 모델에 의해 7가지 기본 감정(Neutral, Happiness, Anger, Disgust, Sad, Surprise, Fear)으로 분류된다. 분류된 감정은 시간의 흐름에 따라 그 로그가 시계열 데이터로 축적된다. 축적된 로그 데이터는 감정 변이의 패턴을 생성하고 생성된 패턴의 변화에 따라 흥미도 예측 모델에 의해 흥미도를 예측한다. 전체 흐름을 Fig. 1에서 보이고 있다.

3.2 감정분류 및 흥미도 예측

얼굴 이미지와 앞에서 언급한 7가지 감정 레이블,

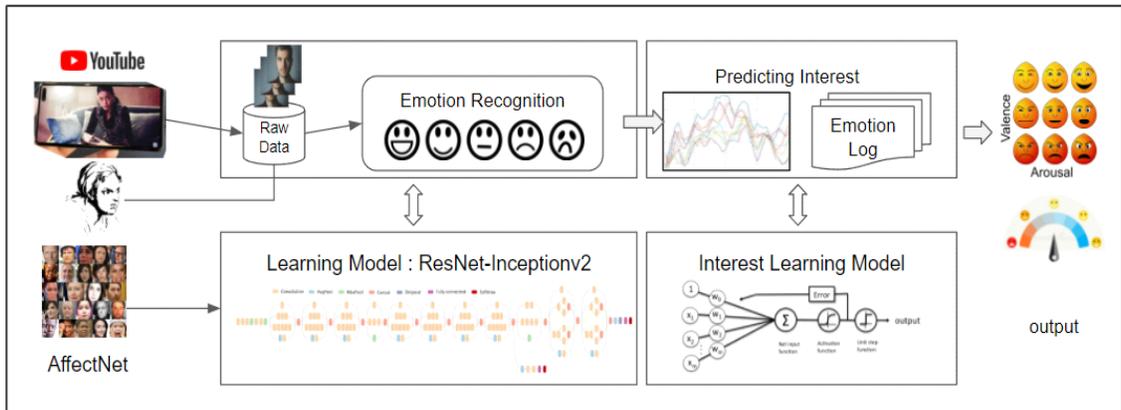


Fig. 1. Sentiment analysis system.

Valence, Arousal 값을 입력으로 하는 데이터셋을 이용하여 CNN 모델을 학습시킨다. 학습된 모델을 이용해 입력 이미지의 표정을 분석하여 7가지 감정에 대한 확률값과 Valence, Arousal 값을 도출한다.

감정분류 모델에 의해 영상물을 시청하는 동안에 시청자의 감정 변화를 알 수 있다. 영상물에 대한 감정 로그 데이터를 통해 계산한 감정 누적 비율을 기반으로 흥미도를 예측한다. 흥미도가 낮은 영상과 높은 영상을 흥미 있음과 없음으로 레이블링한 후 감정 누적 비율과 흥미도 레이블을 이용해 Logistic Regression 모델을 학습시킨다.

4. 실험 및 분석

4.1 데이터 수집

감정 누적 비율 비교와 흥미도 예측 실험을 위해 영상 선정 기준을 정한 후 영상을 선정하였다. 5명의 시청자가 10개의 영상을 시청하였으며 카메라를 이용하여 1초마다 얼굴 이미지를 캡처했다. 감정분류 모델을 이용해 감정 누적 비율을 계산한 후 흥미도와 어떤 관계가 있는지 확인했다.

4.1.1 영상 선정

소셜 미디어 통계 및 분석 웹사이트인 ‘Social Blade’와 국내 유튜브 관련 데이터 및 인사이트 제공 사이트 ‘vling’을 참고하여 동물, 코미디, 여행 카테고리의 인기 채널 5~6개를 선정하였다. 두 사이트에서 순위가 높은 채널 순으로 선정되되 외국 채널, 여러 카테고리의 영상을 다루는 채널, 영상 수가 현저히 적은 채널은 선정 대상에서 제외하였다.

선정된 채널들의 영상을 인기순으로 정렬하여 상위, 하위 영상 각각 2개씩을 선정하였다. 최근의 영상은 상대적으로 조회 수가 낮을 수 있음을 고려하여 3개월 이내에 업로드된 영상은 선정 대상에서 제외하였다. 효율적인 데이터 분석을 위해 영상의 길이가 비슷한 10분 내외의 영상들을 선정하였다. 그중 실험에서는 가장 다양한 종류의 콘텐츠가 존재하는 코미디 카테고리의 영상 10개를 대상으로 진행하였다.

4.1.2 데이터 수집 내용

코미디 카테고리에서 실험 대상으로 선정된 10개의 영상으로 데이터를 수집하였고 총 5명의 영상 시

청자가 참여하였다. 5명 중 3명의 시청자는 10개의 영상을 모두 시청하였고 1명의 시청자는 6개의 영상, 나머지 1명의 시청자는 4개의 영상을 시청하였다.

본 논문에서는 감정분석 실험을 위해 카메라를 이용하여 영상 시청자의 표정을 1초 간격으로 캡처한 후 감정분류 모델 서버로 전송하였다. 양질의 시청자 표정 데이터를 수집하기 위해 모든 시청자에게 영상을 시청하기 전에 몇 가지 주의사항을 안내하여 표정이 잘 보이도록 환경을 갖춘 후 영상을 시청하도록 하였다.

영상 시청이 완료된 후에는 방금 시청한 영상에 대한 설문조사를 진행하였다. 설문조사는 총 2가지 질문으로 구성하였다. 첫 번째 질문은 방금 시청한 영상을 이전에 시청한 경험이 있는지를 묻는 질문이고, 두 번째는 방금 시청한 영상에 대한 흥미도를 1(전혀 흥미 없음)부터 5(매우 흥미 있음)까지의 선택지를 두어 묻는 질문이었다.

4.1.3 데이터 전처리

시청자 감정 변이 수집은 카메라를 이용하여 표정을 1초 간격으로 샘플링 한다. 감정분류 모델을 이용하여 샘플링된 이미지의 표정을 분석하여 Arousal, Valence 값과 Neutral, Happiness, Sad, Surprise, Fear, Disgust, Anger 감정에 대한 확률값을 도출한다. 주어진 영상물에 대한 시청자 표정 변이가 확률값의 시계열 데이터로 변경된다.

한 영상을 시청하는 도중 시청자가 어떤 감정을 가장 많이 느끼는지 계산하기 위해 감정별 누적 E_i 의 비율 $r(E_i)$ 을 계산한다. 추출된 프레임 이미지마다 확률값이 가장 큰 감정을 E_{max} 라고 하고 어떤 감정 E_i 가 한 영상에 나타나는 횟수 $E_{max}(E_i)$ 를 계산했다. Fig. 2는 실험에서 동영상에서 추출된 프레임의 감정별 인식확률에 의해서 결정되는 감정 변이의 예를 보이고 있다. 주어진 영상물을 시청하는 동안의 감정별 누적 비율 $r(E_i)$ 는 식 (1)로 나타낸다.

$$r(E_i) = \frac{no.of E_{max}(E_i)}{|F_s|} \times 100 \tag{1}$$

이때, $E_i = \{Neutral, Happiness, Sad, Surprise, Fear, Disgust, Anger\}$ 이고 $no.of E_{max}(E_i)$ 는 주어진 감정이 최대 확률을 가지는 프레임의 빈도이다. 그리고 $|F_s|$ 은 영상물에서 샘플링된 프레임의 총수이다.

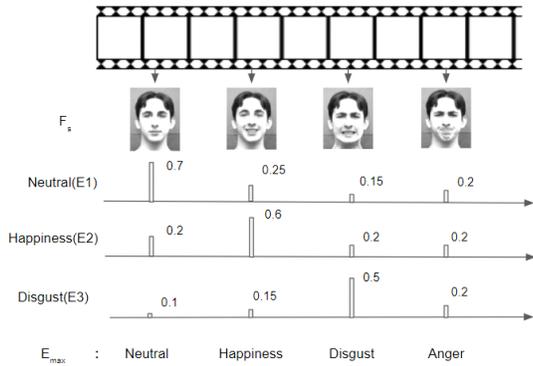


Fig. 2. Example of sampling and calculating E_{max} .

Table 1은 한 영상에 대한 4명의 시청자 각각의 감정 누적 비율을 보여주고 있다. 전체 영상, 흥미도가 낮은 영상, 흥미도가 높은 영상을 대상으로 감정 누적 비율을 계산하고 평균, 최소, 최대를 구한 후 유의미한 공통점을 보이는지 확인해 보았다.

4.2 실험 결과

4.2.1 감정 누적 비율 비교

흥미도가 다른 데이터 집합끼리 시청자 감정의 누적 비율을 비교해보았다. 가장 특징적인 감정은 Neutral과 Happiness였다. 전체 데이터를 통해 구한 감정 누적 비율 평균은 Neutral의 경우 약 51.11%. Happiness의 경우 12.79%로 나타났다. 흥미도가 낮은 데이터 집합의 감정 누적 비율 평균은 Neutral이 약 66.94%. Happiness가 약 7.54%로 전체 데이터 집합과 비교했을 때 Neutral은 약 15.8% 높게, Happiness는 약 5.2% 낮게 나타났다. 반면 흥미도가 높은 데이터 집합에서는 감정 누적 비율 평균이 Neutral의

경우 약 35.18%, Happiness가 약 17.33%로 전체 데이터 집합과 비교했을 때, Neutral은 14.9% 낮고, Happiness는 4.5% 높게 나타났다. 따라서 영상의 흥미도가 낮을 때는 Neutral의 빈도가 높고, Happiness의 빈도는 낮으며, 흥미도가 높을 때는 Neutral의 빈도는 낮고, Happiness의 빈도는 높은 경향이 있음을 확인할 수 있었다. 흥미도에 따른 감정 누적 비율 평균을 Table 2와 3에서 보이고 있다.

4.2.2 흥미도 예측 실험

총 40회의 영상 시청 중 흥미도가 3인 데이터 7개를 제거하고 흥미도가 1~2인 데이터는 0(흥미 없음)으로, 흥미도가 4~5인 데이터는 1(흥미 있음)으로 치환하였다. 그 결과 데이터 개수는 0(흥미 없음)이 12개, 1(흥미 있음)이 21개였다. 33개의 영상 시청 데이터를 train_test_split 함수를 사용하여 train data와 test data로 나눴다. test_size 파라미터는 0.25를 사용하여 9개의 데이터가 test data로, 24개의 데이터가 train data로 분류되었다. 감정 누적 비율 데이터를 logistic regression 모델에 적용하기 위해 평균 0, 표준편차 1로 정규화시킨 후 학습시켰다.

학습 데이터의 수가 적어서 발생하는 문제를 최소화하기 위해 k-fold cross validation을 사용하였다. Cross validation 이후 처음에 분류했던 test data 9개와 새로운 test data 3개(흥미 없음 1개, 흥미 있음 2개)를 추가하여 총 12개의 test data로 모델의 정확도를 측정했다. 예측한 결과 약 0.9167의 정확도를 보였다.

영상에 대한 흥미도와와의 비례 정도를 나타내는 감정 속성의 계수는 Neutral, Fear, Happiness, Anger, Surprise, Sad, Disgust 순으로 -0.41, 0.71, 0.87,

Table 1. Accumulated ratio by emotion (Video ID : eVGq8v-bXxY).

Interest level	4		4		5		5	
User	A		B		C		D	
Accumulated ratio by emotion (%)	Neutral	44.40	Neutral	50.18	Neutral	2.30	Neutral	4.32
	Fear	0.18	Fear	0	Fear	0	Fear	0
	Happiness	39.17	Happiness	30.63	Happiness	4.08	Happiness	41.91
	Anger	2.71	Anger	6.16	Anger	7.27	Anger	0
	Surprise	7.94	Surprise	1.06	Surprise	0.53	Surprise	0
	Sad	4.51	Sad	7.57	Sad	82.10	Sad	53.78
	Disgust	1.08	Disgust	4.40	Disgust	3.72	Disgust	0

Table 2. Accumulated ratio by emotion of 1~2 interest level videos.

Emotion	Average ratio (%)
Neutral	66.94
Fear	0.08
Happiness	7.54
Anger	10.37
Surprise	1.47
Sad	13.22
Disgust	0.36

Table 3. Accumulated ratio by emotion of 4~5 interest level videos.

Emotion	Average ratio (%)
Neutral	36.18
Fear	0.89
Happiness	17.33
Anger	3.86
Surprise	2.73
Sad	36.71
Disgust	2.28

-0.33, 0.35, -0.08, 0.51로 나타났다. 계수가 1에 가까울수록 흥미도와 비례하는 감정 상태라고 말할 수 있고 -1에 가까울수록 흥미도와 반비례한다. Happiness, Fear, Disgust, Surprise 순으로 흥미도와 비례하며 Neutral, Anger, Sad 순으로 흥미도와 반비례하는 결과를 얻었다.

4.2.3 토의

실험 결과 영상에 대한 흥미도에 따라 가장 편차가 큰 감정은 Neutral, Happiness임을 알 수 있었고, 흥미도 예측 실험을 통해 어떤 감정 속성이 영상에 대한 흥미도에 영향을 주는지 확인했다. 딥러닝을 통한 얼굴 감정인식의 특성상 Neutral과 Happiness에 대한 분류 정확도는 높지만, Fear, Disgust, Anger와 같은 감정은 분류 정확도가 낮다. 그 결과 Fear, Disgust, Anger와 같은 감정은 낮은 인식률로 인해서 누적 감정 비율이 비교적 낮게 계산된 것으로 보인다. 그리고 감정분석 딥러닝 모델을 학습시킬 때 주로 서양인 얼굴로 이루어진 AffectNet 데이터셋 대신, 실험에 참가한 시청자들과 같은 동양인들의 데이

터셋이나 조금 더 심리학적으로 검증된 데이터를 사용해 학습을 진행한다면 감정분석을 진행할 때 더 높은 정확도를 가질 수 있을 것이다. 또한, 시청 데이터를 수집할 때 더 다양한 영상 카테고리과 많은 시청자의 데이터를 확보하여 추가 연구를 진행한다면 더 정확한 결과를 얻을 수 있을 것으로 기대된다.

5. 결 론

본 논문에서는 인기 영상과 비인기 영상의 감정 누적 비율과의 관계를 분석하고 감정 누적 비율을 활용해 영상의 흥미도를 예측하고 확인했다. 감정분석 시스템을 통해 영상을 시청하는 시청자들의 표정 이미지를 수집하고 학습된 딥러닝 모델을 이용해서 감정을 분석했다. 흥미도에 따른 감정 누적 비율 평균을 비교해본 결과, 총 7개의 감정 중 가장 특징적인 결과를 가지는 감정은 Neutral과 Happiness였다. 또한, 7가지 감정과 흥미도 간의 관계를 알아보기 위해 회귀 분석을 진행했다. 가장 정확도가 높은 데이터 집합을 사용하여 모델을 학습시킨 결과 흥미도 예측에서 약 92%의 높은 정확도를 보였으며 각 감정 속성과 흥미도가 어떤 관계를 갖는지 확인했다.

REFERENCE

[1] N. Mehendale, "Facial Emotion Recognition using Convolutional Neural Networks (FERC)," *SN Applied Science*, Vol. 2, pp. 446, 2020.

[2] F. Ma, B. Sun, and S. Li, "Robust Facial Expression Recognition with Convolutional Visual Transformers," *arXiv Preprint*, arXiv: 2103.16854, 2021.

[3] W. Mellouk and W. Handouzi, "Facial Emotion Recognition using Deep Learning: Review and Insights." *Procedia Computer Science*, Vol. 175, pp. 689-694, 2020.

[4] J. Jeong, J. Lee, and C. Leem, "An Analysis of Characteristics and User Reactivity by Video Categories on YouTube," *Journal of Digital Contents Society*, Vol. 20, No. 12, pp. 2573-2582, 2019.

[5] C. Ryu and S. Kim, "A Research on the Design of the Youtube Video Contents Based

on the Peak-End Rule and Users' Responses," *Journal of the Korea Industrial Information Systems Research*, Vol. 26, pp. 43-56, 2021

[6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv Preprint*, arXiv:1409.1556, 2015.

[7] K. Alex, S. Ilya, and E. Geoffrey, "ImageNet Classification with Deep Convolutional Neural Networks," *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Vol. 1, pp. 1097-1105, 2012.

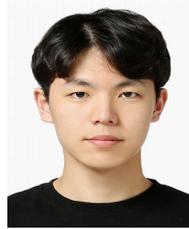
[8] S. Christian, L. Wei, J. Yangqing, S. Pierre, R. Scott, A. Dragomir et al., "Going Deeper with Convolutions," *arXiv Preprint*, arXiv:1409.4842, 2014.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv Preprint*, arXiv:1512.03385, 2015.

[10] P. Ekman, W.V. Friesen, and P. Ellsworth, *Emotion in the Human Face*, Pergamon Press, 1972.

[11] D.R. Cox, "The Regression Analysis of Binary Sequences," *Journal of the Royal Statistical Society*, Vol. 20, No. 2, pp. 215-242, 1958.

[12] C. Won, "Recognition of Facial Emotion Using Multi-scale LBP," *Journal of Korea Multi-media Society*, Vol. 17, No. 12, pp. 1383-1392, 2014.



조 인 구

2021년 부산대학교 기계공학부
학사
2021년~현재 부산대학교 정보융
합공학과 석사 과정
관심분야: 컴퓨터비전, 딥러닝,
감정인식



공 연 우

2022년 부산대학교 정보컴퓨터
공학부 학사
관심분야: 자연어처리, 빅데이터
분석



전 소 이

2018년~현재 부산대학교 정보컴
퓨터공학부 학사과정
관심분야: 웹 개발, 머신러닝



조 서 영

2022년 부산대학교 정보컴퓨터
공학부 학사
관심분야: 컴퓨터비전, 임베디드
시스템



이 도 훈

1986년 부산대학교 계산통계학과
학사
1992년 부산대학교 전자계산학과
석사
1998년 부산대학교 전자계산학과
박사

1995년~2005년 밀양대학교 컴퓨터공학과 교수
2004년~2005년 Indiana University 방문연구원
2006년~현재 부산대학교 정보컴퓨터공학부 교수
관심분야: 컴퓨터비전, 데이터 분석