KoBERT기반 Youtube 자막 감정 분석 연구

최다은, 김효민, 이혜린, 황유림
서울여자대학교 정보보호학과
ekdrmsnl@swu.ac.kr, gyals4574@swu.ac.kr, harlee11@swu.ac.kr, yurimeeee@swu.ac.kr

Sentimental Analysis of YouTube Korean Subscripts Using KoBERT

Da-Eun Choi, Hyo-Min Kim, Hae-Rin Lee, Yu-Rim Hwang
Dept. of Information Security, Seoul Women's University

요 약

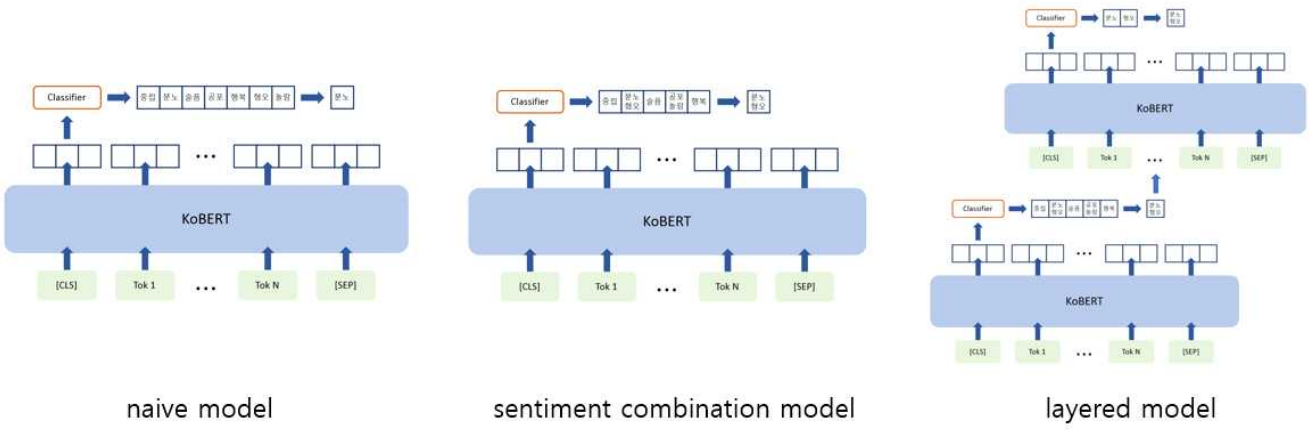
YouTube 이용자의 급증으로 많은 사람이 유튜브 알고리즘에 의해 무분별한 영상에 노출되고 있다. 이는 YouTube 이용자에게 부정적인 영향을 미칠 수 있으며 더 나아가 사회적으로 미성숙한 미디어 문화를 조장할 수 있다. 본 논문에서는 YouTube 컨텐츠에 대한 감정분석 연구를 처음으로 시도한다. 구체적으로, YouTube 컨텐츠 자막에 대해 기존의 자연어 처리 기반 감정분석 기법을 적용하여 성능을 분석한다.

1. 서론

최근 YouTube 이용자의 급증으로 많은 사람이 유튜브 알고리즘에 의해 무분별한 영상에 노출되고 있다. 사회인지 이론[1]에 의하면 직접적인 경험 없이 영상 콘텐츠로부터 노출되는 폭력행위와 그에 관련된 지식을 습득하는 것이 가능하다. 이는 무분별한 영상 시청이 이용자에게 부정적인 영향을 미칠 수 있으며 더 나아가 사회적으로 미성숙한 문화를 조장할 수 있다. 그러므로 시청자는 무분별한 영상 시청이 아닌 자신이 시청할 영상에 대한 파악이 필요하다. 이는 YouTube 영상의 감정분석을 통해 자신이 어떤 감정이 내포된 콘텐츠를 소비하는지 인지시켜 시청 영상을 파악할 수 있다. 영상의 감정분석은 자막 감정분석으로 접근 가능하다. 기존 KoBERT를 활용한 감정분석 연구에서는 영상 댓글이나 리뷰를 대상으로 하는 연구가 대부분이었으며 유튜브 영상의 자막을 감정 분석한 사례는 없었다. 댓글은 문장의 길이가 짧고 비약이나 생략이 강하지만 유튜브 영상의 자막은 상대적으로 문장의 길이가 길며 주로 장면 전환이 일어나기에 다양한 감정이 포함되는 특징이 있다. 본 논문에서는 YouTube 영상의 한국어 자막 감정에 대해 분석하여 YouTube 이용자가 시청할 영상의 주요 감정을 미리 인지할 방법을 제안한다.

자연어 처리(Natural Language Processing, NLP)는 자막 분석에 가장 바탕이 된다. 구글에서 공개한 transformer[2] 기반의 BERT (Bidirectional Encoder Representations from Transformers)가 좋은 성능을 내는 모델로 평가받고 있다. 그러나 BERT는 영어 텍스트를 바탕으로 학습되므로 한국어 데이터 셋을 학습하기에는 한계가 있다. 한국어로 된 데이터 셋을 바탕으로 사전 학습하는 언어 모델로는 Multilingual BERT가 있지만 KoBERT[3]가 네이버 영화 댓글을 대상으로 한 분류 테스트에서 BERT-base multilingual cased 대비 0.026% 높은 Accuracy를 보여 본 논문에서는 KoBERT 모델을 사용한다.

최근 KoBERT를 활용한 감정 분류 연구는 KoBERT를 fine-tuning해 감정 분석하는 방법에서 더 나아가 다양한 방법이 시도되고 있다. 해당 방법으로는 긍/부정을 분류하는 감성 분석에서 세부 감정분석 분류 층을 추가한 이중구조 방법[4]과 기존 모델에서 단어 단위의 감정 정보도 동시에 학습시키는 방법[5]이 있다. 단어 단위의 감정 정보를 학습시키는 방법은 약 0.1% 성능향상이 있었고 이중구조 방법은 기존 모델 대비 약 1% 성능이 향상되었다. 본 연구는 더 큰 정확도 향상을 보인 이중구조 방법에서 문장을 긍/부정으로 분류한 후 세부 감정으로 나누는 아이디어를 착안하여 감정분석 연구를 진행한다.



[그림 1] KoBERT 학습 모델

2. 모델 학습

2.1 학습 방법

본 논문에서 사용한 감정 분석 모델은 총 4가지 방법으로, KoBERT fine tuning을 통해 모델을 학습시켜 감정 분류를 시도한다.

방법 1은 [그림 1]의 기본 모델과 같이 7가지 감정을 학습시킨 모델로 입력 데이터를 7개의 감정 중 하나의 감정으로 분류한다.

방법 2는 7가지 감정(중립/분노/혐오/공포/행복/슬픔/놀람) 중 비슷한 감정의 양상을 가진 공포/놀람과 혐오/분노의 분류 정확도를 높이기 위해 두 가지 감정을 하나의 감정으로 결합하여 5가지 감정으로 학습하였고 [그림 1]의 sentiment combination model(sc-model)과 같이 5개의 감정으로 분류한다.

방법 3은 기본 모델에서 같은 범주 내의 감정이 나올 경우 정답 처리하였다. 따라서 입력된 7가지 감정의 데이터는 [그림 1]의 sc-model과 같이 5개의 감정 중 하나로 분류된다.

방법 4는 [그림 1]의 layered model과 같이 sc-model을 거쳐 나온 중립, 행복, 슬픔은 그대로 도출되고 두 가지 감정을 통합시킨 (공포/놀람)과 (혐오/분노)의 경우, 이중 감정 분류기를 통해 세부 감정에 대해서 분석하도록 구성하였다.

2.2 학습 데이터 셋

본 연구에서는 SNS 및 온라인 댓글을 크롤링하여 구축된 AI Hub '한국어 감정 정보가 포함된 단발성 데이터셋'을 사용한다. 이 데이터셋은 기쁨, 슬픔, 놀람, 분노, 공포, 혐오, 중립의 7개 감정으로 레이블링된 38,594문장으로 구성되어 있다. sc-model의 경우 비슷한 양상을 가지는 두 감정을 통합할 시, 타 감정의 데이터 수보다 2배가량 커져 데이터 편향이 발생할 수 있다. 따라서 sentiment combination model을 활용할 때 통합할 감정의 학습 데이터 수를 각각 절반으로 줄인 후 결합하여 각 감정별로 균등한 데이터 수를 유지한다. 기본 모델의 학습 데이터도 sc-model과 비슷한 개수로 맞춰주었다. 검증 데이터는 3,860개, 테스트데이터는 3,859개로 감정별 데이터 수는 균등하다.

2.3 실험

문장을 tokenizing한 단어의 개수가 64개 이하이기 때문에 max_len을 64로 설정하였다. batch_size와 num_epoch, learning_rate는 여러 파라미터에 대해 실험을 한 결과, 가장 높은 정확도를 보이는 값으로 [표 1]과 같이 설정하였다.

[표 1] 모델 파라미터

	sc-model	기본 모델
max_len	64	64
batch_size	64	128
num_epochs	5	4
learning_rate	5e-5	5e-5

[표 2]와 같이 방법 3이 방법 2보다 정확도가 약 1.32% 높은 정확도를 보인다.

[표 2] 5가지 감정 분류 방법별 정확도

	방법 2	방법 3
accuracy	65.14	66.46

방법 2와 방법 3의 감정별 정확도는 [표 3]과 같

다. 다른 감정과 레이블을 통합하지 않은 슬픔, 중립, 행복의 정확도는 방법 2에서 더 높았고 동일 감정으로 통합시킨 4개의 혐오, 분노, 공포 놀람의 감정은 방법 3에서 더 높은 정확도를 보였다.

[표 3] 5가지 감정 분류 방법별 감정 Recall

Recall	중립	분노 혐오	공포 놀람	행복	슬픔
방법 2	50.4	64.7	60.5	82.9	69.0
방법 3	39.4	69.6	68.9	82.1	61.5

감정 세분화를 위해 동일 감정 범주에 해당하는 (분노/혐오)와 (공포/놀람)에 대해 이중분류를 진행하는 모델을 생성하였고 각각 62.43%, 79.25%의 정확도를 보였다. 그 결과 [표 4]와 같이 방법 4가 방법 1보다 정확도가 약 3.62% 낮았다.

[표 4] 7가지 감정 분류 방법별 정확도

	방법 1	방법 4
accuracy	57.75	54.13

방법 1과 방법 4의 감정별 정확도는 [표 5]에서 확인이 가능하다. 다른 감정과 레이블을 통합하지 않은 슬픔, 중립, 행복의 정확도는 방법 4에서 더 높아졌고 동일 감정으로 통합시킨 4개의 혐오, 분노, 공포 놀람의 감정은 대체로 낮아졌음을 알 수 있다.

[표 5] 7가지 감정 분류 방법별 감정 Recall

Recall	중립	분노	혐오	공포	행복	슬픔	놀람
방법1	39.4	55.9	28.2	56.4	82.1	61.5	57.1
방법4	50.4	40.3	36.9	50.2	82.9	69.0	46.1

3. 유튜브 자막 감정 분류

3.1 유튜브 자막 데이터 셋

다양한 감정이 내포된 SBS 드라마 공식 유튜브 채널 'SBS 캐치'에서 드라마 "펜트하우스"의 4개의 영상을 선정하여 각 영상에 대하여 자막을 추출하고, 추출한 자막에 대하여 10초 단위와 맥락 단위로 자막의 길이를 조정했다. 맥락 단위는 이전 문장과 다음 문장 사이의 공백이 5초 이상이면 맥락의 전환이 이루어지는 곳으로 판단하여 자막을 자른 것이다. 수집한 약 300개의 자막 데이터는 !(느낌표), ?(물음표), .(온점), ,(쉼표)를 제외한 특수 기호를 제거하는 전처리를 진행하고 자체적으로 총 7가지의 감정으로 레이블링을 수행하였다. 본 데이터 셋은 해당 주소[6]에서 확인이 가능하다.

[표 6] Youtube 영상별 감정별 데이터 개수

	A	B	C	D
맥락				
10초				

3.2 실험 결과

7가지 감정으로 라벨링 된 4개의 영상들의 자막을 자체적으로 만든 모델 4개를 Bagging 하여 유튜브 영상의 감정을 분석했다. 결과는 [표 6], [표 7], [표 8], [표 9]과 같으며, 방법 1의 경우 영상별로 정확도 약 32.1%~51.7%를 방법 2의 경우 45.4%~64.7%, 방법 3의 경우 45.4~70.5%, 마지막으로 방법 4의 경우 30.7%~44.1%의 정확도를 나타냈다.

[표 7] Youtube 영상별 감정분석 정확도
방법 1

	A	B	C	D
맥락	38.1	32.1	36.3	35.2
10초	51.7	48.5	35.2	41.0

※ 4개의 영상을 각각 알파벳으로 지정

[표 8] Youtube 영상별 감정분석 정확도
방법 2

	A	B	C	D
맥락	50.0	46.4	45.4	64.7
10초	57.1	47.1	50.0	46.1

[표 9] Youtube 영상별 감정분석 정확도
방법 3

	A	B	C	D
맥락	57.6	46.4	45.4	70.5
10초	57.1	48.5	47.0	46.1

[표 10] Youtube 영상별 감정분석 정확도
방법 4

	A	B	C	D
맥락	30.7	32.1	36.3	41.1
10초	42.8	37.1	44.1	38.4

데이터의 개수가 가장 많은 영상 B의 감정별 정확도는 [표 10], [표 11], [표 12], [표 13]와 같다. 나머지 영상에 대한 감정별 정확도는 깃 허브 주소[6]에서 확인이 가능하다.

[표 11] 감정 Recall
방법 1

Recall	중립	분노	혐오	공포	행복	슬픔	놀람
맥락	66.6	20.0	0.0	0.0	100	50.0	33.3
10초	100	55.5	75.0	16.6	71.4	28.5	35.7

[표 12] 감정 Recall
방법 2

Recall	중립	분노 혐오	공포 놀람	행복	슬픔
맥락	50.0	40.0	33.3	100	100
10초	68.7	44.8	20.0	80.0	25.0

[표 13] 감정 Recall
방법 3

Recall	중립	분노 혐오	공포 놀람	행복	슬픔
맥락	100	46.1	27.2	100	100
10초	100	55.5	19.0	66.6	28.5

[표 14] 감정 Recall
방법 4

Recall	중립	분노	혐오	공포	행복	슬픔	놀람
맥락	75.0	0.0	7.1	50.0	100	100	25.0
10초	68.7	33.3	20.0	14.2	66.6	50.0	12.5

영상을 구성하는 각 감정별 데이터 개수의 편차가 크기 때문에 감정별 정확도 또한 편차가 큰 것으로 유추된다.

4. 결론 및 향후 연구

본 논문에서는 비슷한 감정 양상을 가진 감정들을 하나의 감정 범주로 통합하고 세부 감정을 분류하는 모델을 제안하였지만, 실험을 통해 기존의 KoBERT로 7가지 감정들을 학습시킨 모델보다 정확도가 낮음을 알 수 있었다. 데이터 셋을 구성하는 감정 개수의 편차가 크기 때문에 감정별 정확도를 비교하는 것에 어려움이 있어 향후 연구에서는 데이터 셋을 더 많이 확보하여 연구를 진행할 예정이다. 추가로 더 나은 성능의 감정 분석 모델을 연구하여 유튜브 영상 감정 분석의 정확도를 높일 예정이다.

참고문헌

[1] Bandura, “Social cognitive theory: An agentic perspective.”, Annual review of psychology, Vol. 52, pp. 1-26, 2001

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NAACL, 4171-4186, 2019.

[3] https://github.com/SKTBrain/KoBERT

[4] 이화경,“대화 패턴 인식 기반의 다중 감정분류 모델 개발 및 적용, 과학기술정보통신부 및 정보통신기획평가원, 2021, 1915-1917

[5] 이영준, 한국어 감정 분석을 위한 합동 학습 기반 KoBERT 모델, 산업통상자원부, 2020, 568 - 570

[6] https://github.com/bigcarrot0722/kobert_emotion_analysis.git