

언어와 비언어 표현의 통합 분석을 통한 리뷰 동영상의 신뢰성 분석 시스템 설계

신희원[○], 이소정, 손규진, 김혜린, 곽서현, 김영민, 김윤희
숙명여자대학교 소프트웨어학부

gmldnjs9803@sookmyung.ac.kr, sojung17@sookmyung.ac.kr,
gj.son.19.98@gmail.com, 97grapefruit614@gmail.com, seohyun0221@sookmyung.ac.kr,
youngmin6816@gmail.com, yulan@sookmyung.ac.kr

A Design of Reliability Analysis System for Review Videos using the Integrated Analysis of Verbal and Nonverbal Sentiment

Hee-Won Shin[○], So-Jeong Lee, Gyu-Jin Son, Hye-Rin Kim,
Seo-Hyun Gwak, Yeong-Min Kim, Yoonhee Kim
Dept. of Software, Sookmyung Women's University

요 약

영상 콘텐츠 생산 간편화와 방송 채널 운영의 편리화에 따른 '영상의 시대'가 도래함에 따라 여러 제품에 대한 리뷰 영상이 관심을 받고 있다. 본 연구에서는 리뷰 영상의 언어와 비언어적 감성 분석을 토대로 통합 신뢰도 분석 시스템을 제안한다. 이를 위해, 영상 속 음성의 언어 감성 분석과 리뷰어의 표정 분석을 통해 얻은 각 감성값을 추출하고 정량화한다. 이후 표준화된 언어, 비언어적 감성값에 대한 통합 신뢰도 분석을 진행한다. 결과적으로, 리뷰 영상에 대한 신뢰도를 객관화된 지표로써 평가할 수 있다.

1. 서론

'모바일 온리'를 넘어 지금은 '유튜브 온리'라는 말이 콘텐츠 생태계 안에 돌고 있다. 그야말로 지금은 '영상의 시대'이며, 새로운 트렌드의 영상 콘텐츠가 끊임없이 생산되고 있다. 따라서, 최근 제품이나 서비스에 대한 개인적 경험과 평가를 보여주는 리뷰 동영상이 인기가 많다. 이에 따라 리뷰 분석을 위한 오피니언 마이닝에 대한 연구가 활발하게 진행되고 있다. 기존의 리뷰 분석은 리뷰 댓글을 통한 텍스트 분석 방법이 지배적이었다.[1] 그러나 영상 속 리뷰어의 정확한 감성도출을 위해서는 댓글의 분석으로는 충분하지 않고 리뷰 자체의 분석이 필요하다. 따라서 리뷰 동영상의 음성과 영상을 두루 이용하여 분석할 수 있는 시스템이 필요하다.

본 연구에서는 유튜브 리뷰에 대해 오픈소스 API와 오피니언 마이닝, 머신러닝을 이용하여 언어와 비언어 감성 분석 후, 두 가지 감성 분석 결과 일치 여부에 맞게 가중치를 주어 통합 신뢰도를 도출한다.

본 논문에서는 리뷰 영상의 언어와 비언어 감성 분석 결과에 대한 통합 신뢰도 분석 시스템(Reliability Analysis System using Integrated

Analysis of Verbal and Nonverbal Sentiment : RASIA)을 제안한다. 시스템은 영상 내 실제 리뷰 부분을 추출하고, 문장 별 timestamp에 따라 구간 분리, 음성 텍스트 변환 후 감성 분석을 실행하여 언어 감성 추출, 리뷰어의 표정을 통해 감성을 도출 후, 최종적으로 통합 신뢰도 분석기가 언어와 비언어 감성값 추출 결과를 이용하여 통합 신뢰도를 추출한다.

본 논문은 다음과 같이 구성된다. 2장은 관련 연구, 3장은 본 논문에서 제안하는 RASIA의 개요와 기능 설계, 4장에서는 실험 분석 및 결과를 설명하고 5장으로 결론을 맺는다.

2. 관련 연구

[1] 논문에서는 SNS 등에서 사용되는 인터넷 용어와 표준어의 유사성을 분석으로 비속어 감성 분석 오류를 보완하는 감성 분석 시스템을 제안했다. 하지만 복잡문이 아닌 단순한 댓글 분석에 치중했다는 한계가 있다.

[2] 논문에서는 콘텐츠에 대한 반응 동영상의 시청자 감성 분석을 통해 사용된 오브젝트에 대한 감정 파악을 연구했다. 반응 동영상의 구간별 시청자

감성 변화량이 유의미한 프레임에 등장하는 오브젝트를 식별함으로써 인물 감성과 오브젝트에 대한 관계를 분석했다.

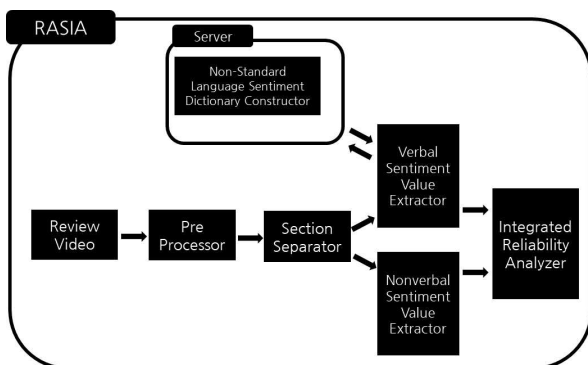
[3] 논문에서는 복합적인 감정을 표현하는 방법을 제안했다. 긍정 감성과 부정 감정은 독립적이고 공존할 수 있으며, 복합 감정은 단일 감정 단어들의 합으로 표현할 수 있음을 분석하였다.

3. Reliability Analysis System using Integrated Analysis of Verbal and Nonverbal Sentiment (RASIA)

3.1 시스템 구조

RASIA는 사용자가 입력한 리뷰 동영상의 언어와 비언어 감성값을 분석하여 통합 신뢰도를 도출하는 시스템이다. 그림 1은 본 시스템의 다이어그램이다.

사용자로부터 분석하고자 하는 리뷰 동영상을 입력 받으면 전처리기(Preprocessor)를 통해 리뷰어가 리뷰를 진행하는 부분만을 추출한다. 구간 분리기(Section Separator)에서는 리뷰어 음성의 문장 별 timestamp를 추출하여 구간을 분리한다. 언어 감성값 추출기(Verbal Sentiment Value Extractor)에서는 리뷰어의 음성을 텍스트로 변환하여 감성 분석을 실시하며 미리 구축된 감성 사전에 없는 단어가 있을 경우 비표준어 감성 사전 구축기(Non-Standard Language Sentiment Dictionary Constructor)를 통해 감성 값을 계산해 추가한다. 비언어 감성값 추출기(Nonverbal Sentiment Value Extractor)에서는 리뷰어의 표정을 통해 감성값을 추출한다. 언어의 감성값과 비언어 감성값의 통합 분석을 위하여 두 값을 긍정, 부정, 중립의 3가지 감성 표현으로 통일한다. 통합 신뢰도 분석기(Integrated Reliability Analyzer)에서는 언어와 비언어의 감성값의 일치 여부를 토대로 리뷰 동영상의 신뢰도를 계산하여 도출한다.



(그림 1) RASIA 구조도

3.2 언어와 비언어 감성값 추출과 통합 신뢰도 분석

리뷰 동영상의 감성값 분석을 위해 Google Cloud의 Speech-to-Text(STT) API[4]를 이용하여 음성 데이터를 텍스트로 변환한다. 변환한 텍스트는 감성값 분석을 위해 Java기반의 형태소 분석기인 KOMORAN[5]을 이용하여 형태소를 분석한다. 분석한 문장에 대해 감성 사전[6]과 자체 구축 비표준어 감성사전에서 긍정, 부정, 중립의 감성 값을 가져와 언어의 감성 값 도출 알고리즘을 통해 언어의 감성 값을 도출한다.

언어의 감성 값 도출 알고리즘은 다음과 같다. 모든 문장에 대해서 용언이 2개 이상일 때 용언을 기준으로 문장을 나누고 용언의 감성 값을 가져와 0인 감성이 있으면 해당 문장의 모든 형태소의 해당 감성을 0으로 정한다. 이후 문장별로 감성 값을 각각 합산한다. 문장의 단어 수에 따라 감성 값이 달라지기 때문에, 표준화를 한다. 표준화된 감성 값 중, 가장 큰 값을 가진 감성을 해당 문장의 감성 값으로 인식, 감성에 극성을 부여한다. 결과적으로 모든 문장에 대해 ‘매우 긍정(+2)’, ‘긍정(+1)’, ‘중립(0)’, ‘부정(-1)’, ‘매우 부정(-2)’의 값을 갖는다.

비언어 감성 값 도출은 이전 논문[2]에 제시된 방법과 같으며, 유의미한 구간 도출은 구간의 단위를 STT를 이용한 문장 단위로 설정한다. 리뷰어의 표정에서 얻은 8가지 감성값을 토대로 이후 언어와의 통합 분석을 위하여 긍정, 부정, 중립 3가지로 감성을 분류한다[3]. FaceAPI 결과 감성값 중, ‘Happiness’와 ‘Surprise’는 긍정으로, ‘Anger’, ‘Contempt’, ‘Disgust’, ‘Fear’, ‘Sadness’는 부정으로, ‘Neutral’은 중립으로 처리한다. 문장별 3가지로 압축된 감성 값의 정도에 따라 ‘매우 긍정(+2)’, ‘긍정(+1)’, ‘중립(0)’, ‘부정(-1)’, ‘매우 부정(-2)’의 극성을 부여한다.

언어와 비언어의 감성 분석 결과를 통합하여 리뷰 동영상의 신뢰도를 분석한다. 각각의 분석 결과는 매우 부정부터 매우 긍정까지 총 5개의 극성을 나타낸다. 이때, 리뷰 동영상의 문장별 언어와 비언어 감성 점수의 차이가 0일 경우는 매우 일치, 1은 일치, 2는 불일치, 3 이상은 매우 불일치로 4가지의 정도를 부여한다. 예를 들어, 언어와 비언어의 감성 분석 결과가 각각 ‘매우 긍정’과 ‘부정’일 경우, 차이가 3으로 매우 불일치의 결과를 갖는다.

리뷰 동영상의 신뢰도는 각각의 결과 값에 따라 가중치를 주어 모두 합산한 다음, 전체 문장의 수로

평균을 내어 구한다. 본 연구에서는 ‘매우 일치’는 1의 가중치, ‘일치’는 0.75의 가중치, ‘불일치’는 0.25의 가중치, ‘매우 불일치’는 0의 가중치를 부여하여 실험하였다. 4가지의 지표의 합의 평균은 0.5로 50% 이상의 신뢰도를 가졌을 시, 신뢰도가 높은 것으로 반대의 경우, 신뢰도가 낮은 것으로 판단한다.

3.3 비표준어 감성사전 구축

그림 2는 비표준어 감성사전 구축 방법이다. 구축 알고리즘은 논문[1]을 참고 후 수정한다. 신조어와 인터넷 용어의 유사 관계 단어를 찾아내는 Word2Vec[7] 기법을 사용한다.

1. 같은 도메인의 유튜브 리뷰 영상 50개 무작위 추출
2. KOMORAN 형태소 분석기를 통해 영상 스크립트 내 문장 분석 후 Word2Vec Skip-Gram 모델 생성
3. 코사인 유사도 상위 1000개의 유사어를 추출
4. 품사가 ‘고유명사(NNP)’, ‘일반명사(NNG)’, ‘형용사(VA)’, ‘동사(VV)’, ‘용언추정(NA)’, ‘분석불가능(NF)’인 형태소 저장
5. 각 형태소가 ‘KNU 감성사전’ 또는 ‘현재까지 구축된 비표준어 감성사전’에 등록되어 있다면 그 감성값에 해당 형태소의 코사인 유사도를 가중치로 부여하여 최종 감성값을 계산
6. 계산된 감성값을 비표준어 감성사전에 업데이트 반복

(그림 2) 비표준어 감성사전 구축방법

4. 실험 환경 및 결과 분석

4.1 실험 환경

본 연구에서는 Google Colab[8] 가상머신을 활용하였다. Ubuntu 18.04.3 LTS, Intel(r) Xeon(R) 듀얼코어, 속도는 2.20GHz이고, 캐시는 56MB이다. 소프트웨어는 Python 3.6.9를 사용한다.

4.2 언어의 감성 분석

본 연구는 ‘뷰티’와 ‘IT’ 2개의 분야에 대해 각각 11개씩 총 22개의 리뷰 동영상을 대상으로 진행하였다. 표1은 IT분야인 ‘가전주부 GJJJ’[9]의 에어팟 프로 리뷰 영상과 뷰티 분야인 ‘스완SWAN_현실남자뷰티’[10]의 파운데이션 리뷰 영상에 대한 언어 감성 분석 결과 예시이다.

<표1> 언어의 감성 분석 결과 예시

분류	문장	감성
IT	고음을 조금 덜 잡아주는 느낌 그런 느낌도 들거든요	부정
	애는 노이즈 캔슬링일 때도 화이트노이즈가 없어요	긍정
	아까 비행기랑 지하철에 비해서 유독 버스 소리는 이것 좀 노캔 커진 거 봤나 싶게끔 조금 별로예요	부정

	근데 에어팟프로를 썼을 때는 방송소리가 또 잘 들리거든요	긍정
뷰티	굉장히 얇고 자연스럽게 내 피부처럼 밀착이 되고	긍정
	뵤가 좋은 점이라면 퍼프를 쓰면 얇게 표현할 수 있고요	긍정
	땀이나 유분이 있어도 메이크업이 그대로 제 피부에 붙어 있기 때문에	긍정
	커버력이 막 높은 제품은 아니다	부정

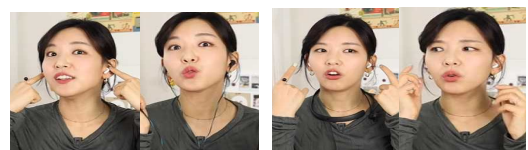
‘고음을 조금 덜 잡아주는 느낌 그런 느낌도 들거든요’는 ‘잡-’, ‘들-’이라는 2개의 용언으로 이루어진 문장이다. ‘잡-’은 부정 0.5, 중립 0, 긍정 0, ‘들-’은 부정 0.53, 중립 0.06, 긍정 0.4의 값을 가진다. 따라서 ‘잡-’ 이전에 등장하는 형태소에 대해 중립과 긍정의 감성 값은 0으로 설정 후 합산되어 문장 감성 값이 부정으로 부여된 것을 확인할 수 있다. 또한 ‘커버력이 막 높은 제품은 아니다’라는 문장에서 ‘아니-’의 부정 용언이 등장함으로써 부정의 감성 값이 부여된 것을 확인할 수 있다.

4.3 비언어의 감성 분석

도출된 영상의 비언어 값을 5가지 극성으로 최종 분류할 때의 감성 값 기준은 긍정과 중립의 경우 평균을 통해 도출하였으며 편차가 큰 부정의 경우에만 최댓값과 평균값의 중간값을 통해 도출하였다.

그림 3, 4는 ‘가전주부 GJJJ’ 영상의 긍정과 부정 표정 구간의 사진이다. 그림3 긍정 구간 표정의 극성별 감성 값은 긍정 값이 7.44 중립 값이 7.20 부정 값이 12.394로 나타났다. 해당 영상의 감성 값 기준은 긍정이 0.243, 중립 1.843, 부정 31.020이다. 이러한 기준에 따라 해당 구간의 감성별 극성 값은 긍정(2), 약한 중립(1), 약한 부정(-2)으로 나타나 구간의 감성 값 총합이 1로 긍정으로 분류되었다.

그림 5, 6은 ‘스완SWAN_현실남자뷰티’ 영상의 긍정과 부정 표정 구간의 사진이다. 그림6 부정 구간 표정의 극성별 감성 값은 긍정 값 17.995, 중립 값 0, 부정 값 31.201로 나타났다. 또한, 해당 영상의 감성 값 기준은 긍정 0.029, 중립 0.745, 부정 29.060이다. 기준에 따라 해당 구간의 감성별 극성 값은 매우 긍정(3), 강한 중립(-1), 강한 부정(-3)으로 총합 -1로 부정으로 최종 분류되었다.



(그림 3) ‘가전주부 GJJJ’ 긍정 표정 구간 (그림 4) ‘가전주부 GJJJ’ 부정 표정 구간



(그림 5)

‘스완SWAN_현실남자
뷰티’ 긍정 표정 구간



(그림 6)

‘스완SWAN_현실남자
뷰티’ 부정 표정 구간

4.4 통합 신뢰도 분석

표3은 리뷰 동영상별 신뢰도 목록이다.

<표3> 리뷰 동영상별 신뢰도

분야	유튜버	제목	신뢰도
IT	가전주 부 GJJJB	30만원 넘는 에어팟 프로 과연 살 만 할까? 노캔중독녀의 에어팟 프 로 사용기![9]	84.82
	nyanji난 지	sub) □ 내가 나에게 주는 선물4 최고의 인생템 에어팟 프로 ●● [노이즈 캔슬링 신세계] □♥[11]	31.42
뷰티	스완 SWAN_ 현실남 자뷰티	(eng) 비레디 파운데이션 사용법 & 컬러립밤 리뷰□ 손&퍼프로 바르는 법 BE READY Men's Foundation[10]	32.38
	You need 윤 이든	인생파데 발견 톰엔 제로레이어 파운데이션 품평리뷰 / 여를 파운 데이션 추천 / 커버력 좋은 파운 데이션 / 지속력 좋은 파운데이션 / 톰엔 / 윤이든[12]	75.48

84.8%로 가장 신뢰도가 높게 나온 IT분야의 ‘가전주부 GJJJB’의 리뷰 동영상의 경우 표 1의 ‘애는 노이즈 캔슬링일 때도 화이트노이즈가 없어요’, ‘근데 에어팟프로를 껐을 때는 방송소리가 또 잘 들리거든요’의 문장은 각각 ‘긍정(+1)’을 나타내는데, 이 때의 비언어 감성값(그림 3)은 모두 ‘긍정(+1)’으로 0의 차이를 가지며 ‘매우 일치’이다. 또한 ‘아 이게 엔진소리는 잡는데 바람소리는 조금 들려요’, ‘아까 비행기랑 지하철에 비해서 유독 버스 소리는 이것 좀 노캔 켜진 거 봤나 싶게끔 조금 별로예요’의 문장은 각각 ‘부정(-1)’을 나타내는데, 이 때 비언어 감성값(그림 4)은 모두 ‘부정(-1)’으로 이 역시 0의 차이로 ‘매우 일치’이며 이렇게 언어와 비언어적 표현이 일치하는 경우가 빈번함에 따라 신뢰도가 높게 평가된다.

반면에 32.4%로 신뢰도가 낮게 나온 뷰티 분야의 ‘스완SWAN_현실남자뷰티’의 경우 표 1의 ‘굉장히 얇고 자연스럽게 내 피부처럼 밀착이 되고’는 ‘긍정(+1)’의 언어 감성값을 나타내지만, 이 때 비언어 감성값(그림 6)은 ‘부정(-1)’을 나타냄에 따라 감성값의 차이가 2로 ‘불일치’이다. 또한 ‘커버력이 높은 제품은 아니다’는 ‘부정(-1)’의 값을 나타내는데 이 때의 비언어 감성값(그림 5)은 ‘매우 긍정(+2)’으로 그 차이가 3이 되어 ‘매우 불일치’이다. 이렇게 언어와 비언어적 표현 불일치 경우가 빈번함에 따라 신뢰도가 낮게 평가된다.

5. 결론

본 연구에서는 IT, 뷰티 2가지 분야에 대해 리뷰 동영상을 선정하여 리뷰어의 음성과 얼굴 표정을 통해 언어와 비언어 감성 값을 추출하였다. 추출된 언어, 비언어 감성 값의 일치 여부를 통해 통합 신뢰도 분석을 진행하였다. 분석 결과 신뢰도가 높은 경우는 70~80%대, 낮은 경우는 30%대로 극명한 차이를 보였다. 이를 통해 사용자가 정보 획득을 위한 리뷰 동영상 선택에 있어 명확한 가이드라인을 제시할 수 있을 것으로 기대된다. 향후 분야를 다양화하고 리뷰 동영상의 개수를 늘려 리뷰 동영상 신뢰도 데이터베이스를 구축할 예정이다.

이 논문(저서)은 2020년도 정부(교육부)의 재원으로 한국과학창의재단(2020년도 학부생 연구프로그램)의 지원을 받아 수행된 연구임.

참고 문헌

- [1] 김봄, 류소정, 오현주, 정윤영, 김세진, 김윤희. "SNS 빅데이터의 상대 유사도를 반영한 감성 분석 시스템 설계" 한국정보과학회 학술발표논문집 VOL.2019 NO.6 (2019):1581-1583
- [2] 정윤영, 김봄, 류소정, 권나현, 손규진, 이소정, 김윤희. "이미지 딥러닝을 통한 콘텐츠와 관련 반응 동영상의 감성 분석 시스템 설계" 한국정보과학회 학술발표논문집 VOL.2019 NO.12 (2019):1239-1241
- [3] 한의환, 차형태. "복합적 감정, mixed feelings에 대한 감정차원 연구." 감성과학 16.4 (2013): 469-480.
- [4] Google Cloud의 Speech-to-Text(STT) API <https://cloud.google.com/speech-to-text?hl=ko>
- [5] JAVA 기반 형태소 분석기 KOMORAN <https://github.com/shineware/KOMORAN>
- [6] KNU 감성 사전 <https://github.com/park1200656/KnuSentiLex>
- [7] Gensim, Word2Vec <https://radimrehurek.com/gensim/>
- [8] Google Colaboratory <https://colab.research.google.com/>
- [9] https://youtu.be/_0jft-cboDg
- [10] <https://youtu.be/PWlbsl7y1Fs>
- [11] https://youtu.be/LcGfixuPS_M
- [12] <https://youtu.be/xr58PA5UvBM>