

인터넷 용어의 감성 분석을 통한 동영상 광고 효과 분석 시스템 설계

(Design of Video Advertisement Analysis via Analysis of
Internet Term Sensitivity)

김 세 진 [†] 김 지 은 ^{**} 성 원 영 ^{***} 김 윤 희 ^{****}
(Sejin Kim) (Jieun Kim) (Wonyoung Seong) (Yoonhee Kim)

요약 스마트폰 사용의 증가로 인해 SNS를 통한 동영상 광고의 영향력은 커지고 있으며, 이를 분석하는 것 또한 중요해지고 있다. 하지만 전통적인 설문 조사 방식의 분석 방법으로는 빠르게 유통되는 SNS 광고 효과를 측정하기에 적합하지 않으며, 기존의 시스템은 인터넷 댓글 특성인 비속어, 줄임말, 이모티콘 등을 고려하지 않으므로 이들을 고려한 분석 방법이 필요하다. 따라서 본 연구에서는 한국어 단문의 특성을 반영하여 동영상 광고의 효과를 분석할 수 있는 자동화된 시스템을 제안한다.

본 시스템은 인터넷 용어의 해석을 위해 기계학습을 사용하여 비속어 감성 사전을 구축한다. 또한, 사용자가 감성을 강조하기 위해 사용하는 이모티콘을 한국어 구문 규칙에 적용하여 감성 분석에 사용하고, 감성의 대상이 되는 명사에 대해 감성 값을 계산하여 세밀한 감성 분석이 이루어지도록 시스템을 설계 및 구현하였다. 이를 통해 동영상 콘텐츠 이용자의 만족도를 도출해 낼 수 있다.

키워드: 기계학습, 오피니언 마이닝, 광고 효과 분석 시스템, 동영상, 유튜브, 감성 분석

Abstract Analysis of the increasing influence of video advertisements via Social Networking Service (SNS) is important in identifying their effects. However, the traditional methods of survey-based analysis are not suitable for measurement of the effectiveness of SNS advertisements that are distributed rapidly via smartphone use and the current system does not consider the sensitivity of users expressed in various forms, such as slang, and emoticons. This study proposes an automated system for the analysis of the effects of video ads via video comments, reflecting the characteristics of short Korean sentences.

This system uses machine learning for the interpretation of Internet terms and compilation of a sentiment dictionary specializing in SNS short sentences. Emoticon, which is used to emphasize the sensitivity of users in comments, is used for sentiment analysis when applied to Korean syntax rules, and the system is designed and implemented for more sophisticated emotional analysis by calculating the emotional values of nouns that are subject to sentiment.

Keywords: machine learning, opinion mining, advertisement analysis service, video, youtube

· 이 논문은 2019년도 정부재원(과학기술정보통신부 여대학원생 공학연구팀제 지원사업)으로 과학기술정보통신부, 한국연구재단과 한국여성과학기술인지원센터의 지원을 받아 연구되었습니다.

· 실험에 도움을 준 김봄, 류소경, 이은아, 이지윤, 정운영에게 감사를 표하고자 합니다.

[†] 학생회원 : 숙명여자대학교 소프트웨어학부
wonder960702@gmail.com

^{**} 비 회 원 : 숙명여자대학교 법학부
amykim0116@naver.com

^{***} 비 회 원 : 숙명여자대학교 소프트웨어학부
dnjsdud9608@naver.com

^{****} 종신회원 : 숙명여자대학교 소프트웨어학부 교수
(Sookmyung Women's Univ.)
yulan@sookmyung.ac.kr
(Corresponding author)

논문접수 : 2019년 1월 31일

(Received 31 January 2019)

논문수정 : 2019년 7월 1일

(Revised 1 July 2019)

심사완료 : 2019년 7월 4일

(Accepted 4 July 2019)

Copyright©2019 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지 제46권 제9호(2019. 9)

1. 서론

광고 효과의 분석은 영향력 있는 광고 제작에 있어 필수적인 과정이다. 이와 관련한 연구는 꾸준히 진행됐지만, 이용자들의 광고에 대한 실제 반응과 태도는 매출액 자체만 놓고 그 광고에 대한 효과를 파악하는 방법으로는 측정하기 힘들다.

현대 사회에서 스마트폰 보급률이 높아짐에 따라, 스마트폰을 통한 활동은 현대인들의 삶으로부터 떼어 놓을 수 없다. 스마트폰으로 하는 다양한 활동 중 광고의 직/간접적인 영향을 받는 Social Network Service(SNS)는 많은 부분을 차지한다. 이와 관련해, SNS 광고의 효과를 분석하는 연구들이 행해지고 있고[9,10], 특히 광고에 대한 반응을 댓글을 통해 분석하려는 연구들이 진행되고 있다[11].

이러한 분석을 위해 비속어, 은어, 줄임말을 많이 포함하고 있는 인터넷 댓글에 대한 이해가 필요하다. 현재 이 댓글들의 감성을 분석하는 연구는 있었지만, 자동화된 시스템으로 이를 측정하는 방식은 아니었다[3].

본 연구에서는 줄임말, 은어, 비속어를 기계학습을 통해 해석할 수 있고, 댓글과 같은 단문에 특화된 비속어 사전을 구축한다. 또한, 인터넷상 댓글은 이용자의 감성을 강조해주는 표현의 방법으로 이모티콘 및 이모지가 빈번하게 사용되는데, 댓글 데이터의 만족도를 분석하는 기존 연구에서는 이를 반영하지 않은 상태로 분석이 이루어져 섬세한 감정평가에는 한계가 있다. 따라서 인터넷 댓글의 주된 특징인 비속어, 이모티콘 등을 고려한 새로운 알고리즘과 광고 분석 결과를 더 빠르게 보여 줄 수 있는 자동화된 동영상 효과 분석 시스템을 구현 및 구축할 필요가 있다.

본 논문에서 오피니언 마이닝과 기계학습을 통한 SNS 댓글 기반 동영상 광고 효과 분석 시스템 설계(A Video Advertisement Analysis System based on Comments using Machine Learning and Opinion Mining: VAAS)를 제안한다. 이는 기계학습을 통해 자동화된 방식으로 비속어, 은어, 줄임말 등의 인터넷 용어를 기존 감성 사전에 추가하고 해석하게 한다. 또한, 감성 언어와 이모티콘이 같이 쓰였을 때 한국어 구문 규칙을 적용하고, 댓글에 포함된 모든 명사에 대해 감성 값을 수치화하여 계산함으로써 더욱 세밀한 감성 분석을 가능하게 한다. 시스템의 주요 기능은 동영상 댓글 데이터를 수집하는 크롤러와 기계학습을 통해 줄임말, 은어, 비속어를 분석하고 이를 통해 사전을 구축하는 감성 사전 구축기, 이모지 및 이모티콘들에 감성 확률을 부여하여 분류해 놓은 이모티콘 리스트, 댓글 속 명사를 중심으로 감성 분석하는 감성 분석기이다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구, 3장은 본 논문에서 제안한 VAAS의 개요와 기능 설계, 4장에서는 기능 실험 결과를 설명하고 5장으로 결론을 맺는다.

2. 관련 연구

최근 인터넷 댓글에 관한 활발한 연구가 이뤄지고 있다. [1] 논문에서는 오피니언 마이닝을 통해 온라인 리뷰의 콘텐츠 만족도 분석 시스템을 개발하고 감성 분석 기법을 제안한다. 긍정 어휘는 +1, 부정 어휘는 -1의 값을 주는 동시에 부사에 가중치를 두어 표현 단어에 의미 강도를 부여할 수 있도록 부사에 규칙과 부정어 규칙을 적용하여 만족도를 세밀히 측정하였다.

[2] 논문에서는 설문 조사를 통해 SNS 텍스트를 대상으로 장소 선호도에 대해 측정하였다. 또한, 이모티콘의 단순한 의미를 해석하는 것뿐만 아니라 이모티콘이 함께 쓰인 문장의 감성을 토대로 이모티콘의 긍정/부정 여부를 해석하였다. 하지만 분석한 이모티콘의 개수가 현재까지 부족하고 인터넷 댓글에서 많이 쓰이는 Emoji(이모지)에 관한 연구가 부족하다.

[3] 논문에서는 인터넷 댓글, 소셜 데이터 등으로부터 비속어/은어, 이모티콘 등을 감성 분석하는 방법을 제안하였다. IRLBA(Implicitly Restarted Lanczos Bidiagonalization Algorithm)를 활용하여 메시지 분석, 감성평가 등을 하고 이를 시각화하는 모습을 보였다. 위 논문에서 비속어/은어, 이모티콘을 분석하기 위해 품사 태깅이 되지 않는 이 어휘들을 감성 평가부로 전달한 뒤 이 어휘들을 표준어로 번역하기 위해 피실험자들에게 의미를 평가하도록 하였다.

[17] 논문은 지니 인덱스 기반 특징 선택 방법과 Support Vector Machine(SVM)을 사용하여 영화 평론 데이터에 대해 감성 분류를 진행하였다. 이를 통해 정확도와 오류를 줄일 수 있었다. [18] 논문은 비정형적인 트위터 데이터에 대하여 각 트윗이 긍정인지, 부정인지, 중립인지 분석하기 위해 다양한 기계학습 알고리즘을 적용하는 논문에 대해 조사를 진행하였다. SVM과 naive Bayes 방법이 가장 높은 정확도를 보였으며, 기계학습을 사용하여 다양한 영역과 다른 언어에 대해 감성 분석을 진행할 수 있었다. [19] 논문은 짧은 텍스트에 대해 감성 분석을 적용하기 위해 긴 문장에 적용하는 신경망 언어 모델과는 차별화된 구조를 제안하였다. 이 논문은 미리 훈련된 단어 벡터 위에 Convolutional Neural Network(CNN)과 long short-term memory를 적용한 신경망 구조를 제안하였다. 이 연구들은 문장 전체의 감성에 대해 효과적으로 긍정, 부정, 중립으로 분류하였다. 하지만 두 문장이 같은 긍정적 의미를 표현하더라도, 사용한 부사어 및 이모티콘 등에 따라 감성의 정도가 달라질 수

있다. 하지만 이 방법들은 단순히 문장을 세 가지 중 하나로 분류하였기에 문장에 대한 감성을 수치화하여 만족도를 세밀하게 측정하는 방법이 아니었다.

3. 동영상 광고 효과 분석 시스템

3.1 시스템 설계

전체 시스템은 그림 1과 같이 동영상 댓글 데이터를 수집하기 위한 데이터 크롤러(Data Crawler), 쓸모없는 불용어를 제거하는 전처리기(Preprocessor-Filtering), 수집된 데이터를 형태소로 분류하는 형태소 분석기(Lexical Analyzer)와 기계학습을 통해 비속어, 은어, 줄임말을 분석하여 사전을 구축하는 감성 사전 구축기(Sentiment Dictionary Builder), 댓글에서 많이 쓰이는 이모티콘 및 이모지들에 감성 확률을 주어 분류한 이모티콘 리스트(Emoticon List)와 댓글 데이터에 대해서 오피니언 마이닝(Opinion Mining)을 적용하여 동영상에 대한 감성을 적용하는 감성 분석기(Sentiment Analyzer)로 나뉜다. 분석된 결과는 사용자에게 표와 그래프로 전달된다. 시스템의 각 기능에 대한 설명은 다음과 같다.

먼저 데이터 크롤러에서는 분석의 대상이 될 콘텐츠를 선정하기 위해서 광고명 또는 사용자의 키워드를 입력받아 웹 사이트로부터 댓글 데이터를 크롤링해 온다. 이렇게 크롤링 된 댓글 데이터들은 형태소 분석기를 거쳐 하둡 분산 파일 시스템에 텍스트 파일 형태로 자동 저장된다. 저장된 데이터들은 먼저 기계학습 기반의 사전 구축기로 비속어, 은어, 줄임말이 해석되고 극성이 분류된다. 이에 대해 자세한 설명은 3.2절에서 한다. 다음 단계에서는 오피니언 마이닝 기반의 감성 분석기로 이모티콘을 고려하며, 한국어 단문 특성을 반영하고, 명사를 중심으로 동영상 댓글 데이터의 만족도가 분석되는데, 자세한 설명은 3.3절에서 한다. 마지막으로 댓글 데이터의 만족도 분석 결과를 그림 또는 표로 시각화하여 사용자에게 보여주며 그 결과는 4절에서 설명한다.

3.2 비속어 해석을 통한 감성 사전 설계

인터넷 댓글은 그 특성상 비속어 및 은어 등이 많이 사용된다[3]. 따라서 댓글의 감성을 분석하기 위해서 기존 감성 사전에 정의가 안 된 비속어의 감성을 분석하는 것이 필요하다. 이를 위해 본 연구에서는 비속어 학습 모델을 전체 시스템에 추가하여 구성하였다. 비속어를 해석하기 위해 댓글들을 학습하여, 문장 내에서 유사한 관계에 있는 단어들을 찾아내는 word2vec 알고리즘[4]을 사용하였다. 사용한 word2vec 모델은 window 크기를 5로 하였다. 따라서 댓글에서 모든 단어는 자신의 왼쪽 5개 단어, 오른쪽 5개 단어를 유의어 후보로 갖는 방식으로 학습된다. 비속어를 해석하기 위한 모델은 다음과 같다.

- 1) 감성 사전[5]에서 단어, 품사 궁정확률, 부정확률, 중립확률만 선별하여 저장한다.
- 2) 댓글 데이터를 수집하고 수집한 댓글들을 word2vec 모델에 넣어 학습시킨다.
- 3) 댓글 데이터를 수집하고 댓글의 단어 품사를 분석하여 저장한다.
- 4) 댓글의 단어 중 비속어를 학습된 모델에 넣어 유의어를 분석한다. 이 결과로 나온 단어 중 유사도가 높은 상위 500개의 단어를 저장한다.
- 5) 저장된 단어 중 감성이 들어있는 품사인 명사, 동사, 형용사, 부사 중 하나인 단어만 필터링한다.
- 6) 필터링 된 단어 중 감성 사전[5]에 있는 단어의 감성 값을 적용하여 비속어의 감성을 분석한다. 각각 단어의 긍정일 확률, 부정일 확률, 중립일 확률의 평균을 구해 비속어의 감성을 분석한다.
- 7) 분석된 비속어는 비속어 사전에 추가한다.

이 과정을 통해 비속어를 해석하고 비속어 사전을 구축한다. 구축된 사전을 콘텐츠 만족도 분석하는 단계에서 활용하기 위해 전달한다면 비속어 해석단계가 완료된다.

3.3 콘텐츠 만족도 분석

SNS 영상 콘텐츠의 만족도를 영상의 댓글을 통해서

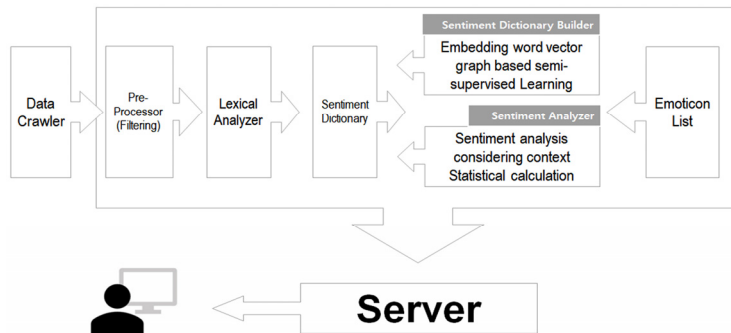


그림 1 시스템 구조도

Fig. 1 System architecture

분석하기 위해서는 세밀한 감성 분석이 이루어져야 한다. 본 연구의 목적에 맞게끔 [1,2]의 오피니언 마이닝 알고리즘들을 확장 및 수정하여 콘텐츠 만족도 도출을 위한 감성 분석 알고리즘을 설계하였다(그림 2). 설계한 알고리즘에서는 감성표현 단어의 의미 강도를 부여하기 위해 이모티콘과 감성어가 같이 쓰이는 경우를 4가지로 분류하여[2] 한국어 구문 규칙을 적용하였으며, 부사어의 경우와 유사하게 이모티콘 및 이모지에 가중치를 주어 감성 분석을 세밀화하였다.

또한, 도출된 감성 결과값을 통계적 기법에 적용하여 만족도를 계산하였다. 만족도의 계산은 특정 대상에 대한 개인의 태도를 측정하는 통계적 기법인 리커트 5점 척도[7]를 사용하였다. 알고리즘의 결과로 나온 감성 값들은 리커트 5점 척도에 적용되어 매우 만족(5) > 만족(4) > 보통(3) > 불만족(2) > 매우 불만족(1)의 결과로 도출된다.

먼저, 알고리즘의 입력 데이터는 동영상의 전체 댓글 데이터이다. 댓글 텍스트 파일에 있는 각 댓글에 대해서 형태소 분석을 하고, (단어, 품사) 형태로 배열에 집어넣는다(02-03), 그 후, 각 문장의 단어에 대해서 단어의 품사가 명사라면, 앞뒤 3어절 이내에 용언이 있는지 확인한다. 그런 뒤, 이 용언이 감성 사전에 분류된 감성단어라면, 감성 값을 받아온다(04-09). 이때의 감성 값은

3.2절에서 구축한 감성 사전과 매칭되어 가져오게 된다. 3어절 이내에 감성단어가 있다면, 앞뒤 3어절 이내에 가중치를 주는 부사어와 이모티콘 및 이모지가 있는지 확인한다. 부사어가 있다면, 부사어의 가중치를 계산하여 곱해 준다. 이모티콘 및 이모지가 있다면, 이모티콘 리스트로부터 이모티콘의 감성 확률값[6]을 이모티콘 리스트로부터 가져온다(15-16). 그 후, 이모티콘의 한국어 문법 규칙에 따라 감성 확률값[6]과 감성단어의 감성 값 조합의 긍정, 부정을 파악하여 가중치를 계산하여 곱해 준다. 명사에 대한 감성 값 계산이 끝났기 때문에, 명사와 감성 값을 벡터에 넣어주고(25), 전체 댓글에 대한 감성 값을 계산하기 위해 명사의 감성 값에 따라 0 이상일 경우에는 SLpositives에, 0 미만일 경우에는 SLnegatives에 넣는다(26-29). 그런 뒤, 영상의 전체 댓글에 대해 리커트 5점 척도로 구해진 만족도를 평균 낸다(30).

4. 실험 환경 및 결과 분석

4.1 실험 환경

본 논문에서 제안한 시스템은 master 서버 1대와 slave 서버 1대로 클러스터를 구성하고 실험을 진행하였다. 실험 환경은 표 1과 같다. 또한, 댓글 데이터의 형태소 분석은 KoNLPy의 Okt 파이썬 패키지를 사용하였다.

4.2 비속어 해석

비속어 해석에서 사용한 데이터는 유튜브버 ‘이사배’[8]의 동영상의 댓글을 2018년 5월 10일부터 11월 19일까지 수집한 것이다. 총 사용한 댓글은 1.27MB이고 단어 수는 114,337개다. 이와 더불어 표준어 데이터와 함께 학습시키기 위해 KoNLPy corpus Package[16]에서 제공하는 대한민국 국회 의안 말뭉치 91.7KB를 사용하였다.

표 2는 수집한 댓글을 토대로 비속어 “헐”, “미친”, “걸크”, “꿀팁”의 감성을 분석한 결과이다. “헐”과 “미친”은 인터넷상에서 긍정의 감탄을 나타낼 때 쓰며 “미친”의 경우 사전적으로 정의된 부정적인 의미로 해석되는 것이 아닌 긍정의 뜻으로 해석되면서 옳게 분류되었다. “걸크”와 “꿀팁”의 경우 신조어로 걸크는 ‘멋진 여

```

Algorithm: Opinion_Mining Algorithm
Input: String text
Output: Vector<Nounsenti> nsvec, double LikertScale
01: Words[][] words;
02: for each sentence in text do
03: words[num] ← ReduceForWord(sentence);
04: for each words do
05: for each words[k] do
06: if isNoun(words[k][i]) then
07: for j = -3 ... 3 do
08: if isSentimentWord(words[k][i+j]) then
09: senti_score = ml.sentiscore(words[k][i+j].word);
10: if senti_score != 0 then
11: noun_senti = senti_score;
12: for j = -3 ... 3 do
13: if isAdverb(words[k][i+j]) then
14: noun_senti *= adjscore(words[k][i+j]);
15: if isEmoticon(words[k][i+j]) then
16: emo_senti = emoscore(words[k][i+j]);
17: if (senti_score > 0 && emo_senti > 0)
18: noun_senti *= 1 + emo_senti;
19: else if (senti_score > 0 && emo_senti < 0)
20: noun_senti *= 1 - emo_senti;
21: else if(senti_score < 0 && emo_senti > 0)
22: noun_senti *= -1 - emo_senti;
23: else if(senti_score < 0 && emo_senti < 0)
24: noun_senti *= -1 + emo_senti;
25: nsvec.add(words[k][i], noun_senti);
26: if(noun_senti>0)
27: SLpositives += noun_senti;
28: else
29: SLnegatives += noun_senti;
30: LikertScale = (SLpositives * 5) / (SLpositives + SLnegatives);
    
```

그림 2 만족도 도출을 위한 감성 분석 알고리즘
 Fig. 2 Algorithm : Sentiment analysis for the calculating of satisfaction

표 1 하드웨어 명세 및 소프트웨어 리스트
 Table 1 Hardware specification and software list

| | CPU Name | CPU Speed | CPU Cores | Cache Size |
|------------------|----------------------|-----------|-----------|------------|
| master | Intel Core i7-7820X | 3.6GHz | Octa Core | 11MB |
| slave | Intel Xeon E5-2630V3 | 2.4GHz | Octa Core | 20MB |
| | Java | Hadoop | Spark | Scala |
| Software Version | 1.8.0_181. | 2.7.7 | 2.3.1 | 2.11.8 |

표 2 비속어 해석 결과 예시
Table 2 Example of slang translation

| | Positive probability | Negative probability | Neutral probability | Result |
|----|----------------------|----------------------|---------------------|----------|
| 헐 | 0.472 | 0.349 | 0.07 | Positive |
| 미친 | 0.457 | 0.389 | 0.075 | Positive |
| 걸크 | 0.463 | 0.388 | 0.072 | Positive |
| 꿀팁 | 0.407 | 0.069 | 0.043 | Positive |

성'을 의미하고 “꿀팁”은 유익한 비법 및 기술을 의미한다. 위 결과를 보면 두 단어 모두 긍정으로 분류되었으며, 이는 비속어 해석모델이 신조어도 잘 분류하는 모습이라고 할 수 있다.

그림 3은 비속어 사전의 유무에 따라 word2vec모델이 비속어를 해석하는 정확도를 비교한 것이다. 총 댓글 중 비속어 181개를 찾았고, 이를 Korean Sentiment Analyzer의 감성 사전[5]만 적용한 word2vec모델에 넣었을 경우 옳은 결과는 158개, 틀린 결과 23개로 87.2%의 정확도가 나타났다. Korean Sentiment Analyzer의 감성 사전[5]과 더불어 비속어 사전을 추가한 word2vec모델에 비속어를 넣었을 때 옳은 결과 174개 틀린 결과 7개로 96.1%의 정확도가 나타났다. 이를 통해 비속어 해석모델이 비속어의 감성 값을 유의미하게 분석해 낼 수 있다.

4.3 만족도 분석

표 3은 크롤링한 데이터에 그림 2의 감성 분석 알고리즘을 적용한 예시이다. 1번 댓글에서 ‘영상’이라는 명사가 나왔으므로 앞뒤 3어절 이내에 용언이 있는지 확인한다. 이때 ‘좋아요’라는 용언이 감성 사전에 분류되어 있으므로 감성 값을 받아온다(+1). 또한, 부사어 ‘너무’의 가중치(+2)를 ‘좋아요’의 감성 값에 곱해 준다. 그리고

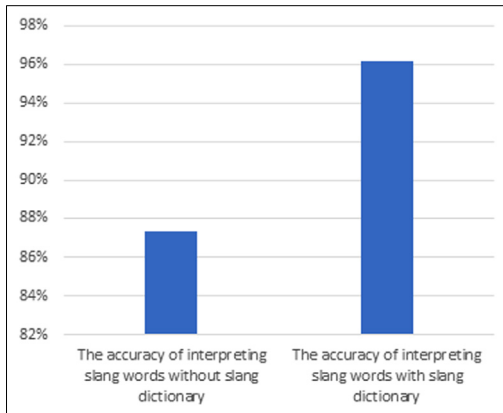


그림 3 비속어 사전 유무에 따른 word2vec모델의 비속어 해석 정확도 비교

Fig. 3 Accuracy comparison of word2vec model depending on the contents of slang dictionary

문장 마지막에 이모지가 나왔으므로 이모티콘 리스트[2]로부터 감성 확률값(+0.632)을 가져와 가중치를 곱해 주는 과정을 수행해 1번 댓글의 감성 값을 도출한다.

댓글의 만족도를 문장 단위로 분석한 뒤에 전체 감성 점수를 구하고, 긍정의 값을 가지는 단어는 SLpositives에(+9.931), 부정의 값을 가지는 단어는 SLnegatives에 저장(-2)하여 최종적으로 감성점수를 리커트 5점 척도로 환산하여 보여준다.

$$\text{LikertScale}[7] = (9.931 * 5) / (9.931 + |-2|) = 4.162$$

표 3은 모든 댓글에 대해 해당 알고리즘을 적용하여 댓글들의 만족도를 구한 뒤, 감성의 대상 명사들의 빈도수를 오름차순으로 정렬하여 사용자들이 긍정적으로 평가한 상위 5개 단어를 보여주는 결과 예시이다. 이는 만족도 리커트 5점 척도에서 4.162로 나타나 만족도 수준이 높은 것으로 평가되었다.

그림 4는 ‘이사배’의 ‘키스미’ 동영상[8]의 댓글 631개, ‘써니’의 ‘조르지오 아르마니’ 동영상[11]의 댓글 971개, ‘이사배’의 ‘맥’ 동영상[12]의 댓글 541개, ‘썸남’의 ‘에스쁘아’ 동영상[13]의 댓글 1,010개를 사용하여 감성 분석 알고리즘을 적용할 때에, 이모티콘을 적용한 경우와 적용하지 않은 경우의 만족도 점수의 총합을 비교한 것이다. 이모티콘을 감성 분석 알고리즘을 적용했을 때의 평균

표 3 만족도 분석 알고리즘 적용 예시

Table 3 Example of satisfaction analysis algorithm and its application

| Comments of Risabee - How to choose mascara well and how to curl your lashes the best way (Feat. Brand <Kiss me>) | | Sentiment Score |
|-------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------|-----------------|
| 1 | 오늘 영상 너무 좋아요 ♡ | 1*2*(1+0.632) |
| 2 | 오늘 팁 잘보고 잘알아갑니당 🙏 | 1+1*(1+0.521) |
| 3 | 마스카라 할때 한번씩 번지고 잘안지워질때 많은데 역시 키스미 짱짱합니다 | -1-1 +1 |
| 4 | 사배언니 사랑하옵니가 😊 | 1*(1+0.678) |
| 5 | 브라운 마스카라 사러가야겠다 필요했는데 🙏 | 1*(1+0.468) |

| Nouns that subject to emotional analysis | Sentiment Score |
|------------------------------------------|-----------------|
| 1 영상 | 3.264 |
| 2 팁 | 2.521 |
| 3 사배언니 | 1.678 |
| 4 키스미 | 1 |
| 5 마스카라 | -0.537 |
| SLpositives(positive emotional value) | 9.931 |
| SLnegatives(negative emotional value) | -2 |

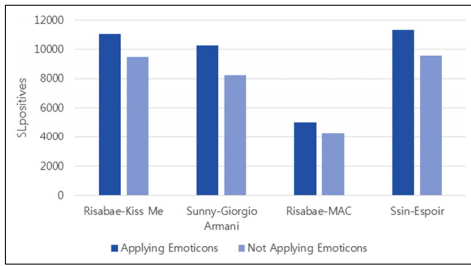


그림 4 만족도 점수 비교 결과

Fig. 4 Results of comparative satisfaction score

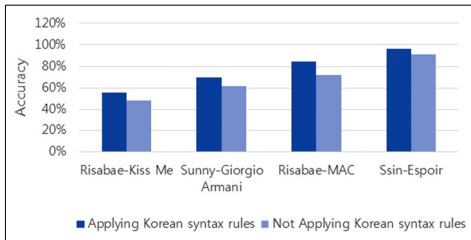


그림 5 정확도 비교 결과

Fig. 5 Result of comparative Accuracy

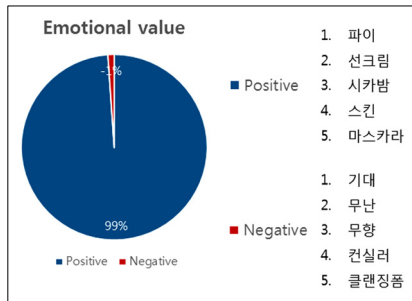


그림 6 키워드 분석 결과

Fig. 6 Result of keyword analysis

만족도 점수는 8797.54로 적용하지 않았을 때의 7341.67보다 약 19.8% 증가하였다. 이는 본 논문의 감성 분석이 기존 연구와 비교하여 더욱 감성을 세밀하게 측정할 수 있음을 알 수 있다.

본 연구의 만족도 분석 알고리즘은 감성어와 이모티콘이 함께 쓰였을 때 이의 조합의 긍정, 부정을 파악하기 위해 한국어 구문 규칙을 적용하였다. 그림 5는 위 알고리즘에 한국어 구문 규칙을 적용하였을 때와 적용하지 않았을 때의 정확도를 나타낸 것이다. 그림 4에서의 동영상 댓글 데이터를 대상으로 하여 약 6천 개씩의 명사 감성 값의 정확도를 평가하였다. 한국어 구문 규칙을 적용한 결과 적용하지 않았을 때와 비교하여 정확도가 각각 17.1%, 12.0%, 15.3%, 5.9% 증가하였다. 한국어 구문 규칙을 적용함으로써 더욱 정확한 감성 분석이 가능한 것을 알 수 있다.

그림 6은 '디렉터 파이'의 '이니스프리' 동영상[14]의 댓글 1,119개에서 계산된 감성 값과 관련해 사용자들이 SNS에 실제로 긍정적 또는 부정적으로 평가한 감성 대상 명사들을 보여준다. 왼쪽의 도표는 만족도 분석이 수행될 때 전체 댓글에 대해서 긍정을 나타내는 점수와 부정을 나타내는 점수의 총합의 비율을 계산한 결과를 나타낸다. 동영상을 긍정적으로 표현한 댓글의 점수는 9898.752점, 부정적으로 표현한 댓글의 점수는 -109.493점으로 점수의 비율은 긍정 99%, 부정 1%가 되어 왼쪽 도표와 같이 나타난다. 한편, 오른쪽 표는 만족도를 계산할 때 각 명사에 대해서 감성 값을 계산하고, 스파크[15]의 맵리듀스를 기반으로 하는 감성 분석 알고리즘을 통해 명사를 key로 명사의 감성 값을 value로 하는 분산적인 분석을 수행하여 사용자가 긍정적으로 또는 부정적으로 평가한 상위 다섯 개 명사를 보여준다. 오른쪽 표를 보았을 때 본 동영상의 가장 큰 강점은 제작자인 '파이'이며, 이 외에도 콘텐츠 이용자들은 '선크림', '시카밤', '스킨', '마스크라'에 긍정적인 반응을 보인 것을 알 수 있다. 한편, 약점으로써는 '기대'라는 단어를 통해 제품이 기대 이하였으며, '무난'하다는 것을 알 수 있다. 또한, 사용자는 '무향'이라는 요인에 대해서 부정적인 의견을 밝히며, '컨실러', '클렌징폼' 또한 이 브랜드의 단점이라고 파악할 수 있다. 결과적으로 본 시스템의 이용자들은 더 세밀하고 정확한 만족도 분석 결과와 함께 동영상 콘텐츠 혹은 제품의 강점과 약점을 파악할 수 있다.

5. 결론

본 논문에서는 광고 시장에서 지배적 지위를 가지며 그 영향력이 더 강력해지고 있는 SNS상의 광고 동영상에 빠르고 효과적으로 분석할 수 있는 자동화된 콘텐츠 만족도 분석 시스템을 설계 및 구현하였다. 또한, 인터넷 단문 특징상 자주 사용되는 비속어, 줄임말, 은어를 해석하기 위한 감성 사전 구축 방법과 댓글 데이터를 통해 동영상의 만족도를 도출하는 감성 분석 알고리즘을 구체적으로 제시하였다. 그리고 본 논문에서 설계한 시스템을 실험하기 위해 동영상 광고 콘텐츠를 선정하여 약 천 개의 댓글 데이터들을 기반으로 인터넷 단문에 특화된 감성 사전을 구축하고, 약 4천 개의 댓글 데이터들을 제안한 알고리즘에 적용해 특정 동영상들에 대한 감성 분석 결과를 도출해 냈다.

본 연구의 동영상 광고 콘텐츠 만족도 분석 시스템은 댓글에 내재하여 있는 작성자의 감성 정보를 세밀하고 정확하게 추출해 분석할 수 있으며, 대용량의 댓글 데이터들을 빠르고 효율적으로 처리할 수 있다. 본 시스템을 통해 광고주들은 동영상 광고에 대한 전체적인 의견을 빨리 정확하게 분석할 수 있으므로 향후의 광고 콘텐츠

제작과 제품의 품질 개선, 그리고 마케팅 전략을 세우는 데에서 많은 도움이 될 것으로 기대한다.

마지막으로 본 연구를 확장하기 위한 향후 연구로는 인터넷 용어 사전의 확장과 만족도 분석 시스템의 정확도 향상이 있다. 이를 통해 좀 더 지능적인 만족도 분석 시스템을 구축할 것이다.

References

[1] MoonJi Kim, EunJeong Song and YoonHee Kim, "A Design of Satisfaction Analysis System for Content Using Opinion Mining of Onlie Review Data," *Journal of KOREA INFORMATION SCIENCE SOCIETY*, Vol. 17, No. 3, pp. 107-113, Jun. 2016. (in Korean)

[2] InYeong Chae, "A Method for Analysis of Preferences of Places based on Sentimental Analysis using SNS Data : Case Study on Theme Parks in Seoul, South Korea," Aug. 2017. (in Korean)

[3] Phil-Sik Jang, "Study on Principal Sentiment Analysis of Social Data," *Journal of KOREA INFORMATION SCIENCE SOCIETY*, Vol. 19, No. 12, pp. 49-56, Dec. 2014. (in Korean)

[4] Gensim, [Online], Available, <https://radimrehurek.com/gensim/>

[5] Sentiment-dictionary from Korean Sentiment Analyzer, [Online], Available, <https://github.com/mrlee23/KoreanSentimentAnalyzer>

[6] Petra Kralj Novak, Jasmina Smalović, Borut Sluban, Igor Mozetič, "Sentiment of Emojis," Dec. 2015. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144296>

[7] LikertScale, [Online], Available, <http://ko.wikipedia.org/wiki/%EB%A6%AC%EC%BB%A4%ED%8A%B8%EC%B2%99%EB%8F%84>

[8] Risabae - Kissme Youtube video, [Online], Available, https://www.youtube.com/channel/UC9kmlDcqs aOnCkC_qzGacA

[9] KyooHoon Han, "An Investigation on Potential Determinants of Social Network Service's Promotional Effects: Based on the Analysis of Facebook Users," *Korean Academic Society For Public Relations*, Vol. 16, No. 3, pp. 132-168, 2012.

[10] Won-Bae Ji, Woon-Han Kim, "A Content Analysis of Digital Audience Replies to Video Advertising Types : Focused on Viral Video and Cable Broadcasting Advertisement"

[11] Sunny, Giorgio Armani, Youtube video, [Online], Available, https://youtu.be/nq_YGdpX3Dk

[12] Risabae, Mac, Youtube video, [Online], Available, <https://www.youtube.com/watch?v=fEegFucMCZw>

[13] Ssin, Espoir, Youtube video, [Online], Available, <https://www.youtube.com/watch?v=5oOeYzbOmmQ>

[14] Director Pi, Innisfree, Youtube video, [Online], Available, <https://www.youtube.com/watch?v=Uv06fB5ELx8&v=ko>

[15] Apache Spark, [Online], Available, <https://spark.apache.org/>

[16] KoNLPy, [Online], Available, <https://konlpy-ko.readthedocs.io/ko/v0.4.3/data/>

[17] Manek, Asha S., P. Deepa Shenoy, M. Chandra Mohan, and K. R. Venugopal, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," *World wide web 20*, No. 2 (2017): 135-154.

[18] Kharde, Vishal, and Prof Sonawane. "Sentiment analysis of twitter data: a survey of techniques," arXiv preprint arXiv:1601.06971 (2016).

[19] Manek, Asha S., P. Deepa Shenoy, M. Chandra Mohan, and K. R. Venugopal, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," *World wide web 20*, No. 2 (2017): 135-154.



김 세 진
2019년 숙명여자대학교 소프트웨어학부 졸업(학사). 2019년~현재 숙명여자대학교 컴퓨터과학과 석사과정. 관심분야는 분산컴퓨팅, 고성능 컴퓨팅, 데이터 사이언스



김 지 은
2015년~현재 숙명여자대학교 법학부 학사과정. 관심분야는 머신러닝, 딥러닝, 빅데이터



성 원 영
2015년~현재 숙명여자대학교 소프트웨어학부 학사과정. 관심분야는 머신러닝, 빅데이터, 오픈니언마이닝



김 윤 희
1991년 숙명여자대학교 전산학과 졸업(학사) 1996년 Syracuse University 전산학과 졸업(석사). 2000년 Syracuse University 전산학과 졸업(박사). 1991년~1994년 한국 전자통신연구원 연구원. 2000년~2001년 Rochester Institute of Technology 컴퓨터공학과 조교수. 2001년~2016년 숙명여자대학교 컴퓨터과학부 교수. 2017년~현재 숙명여자대학교 소프트웨어학부 교수. 관심분야는 클라우드 컴퓨팅, 워크플로우 제어, 그리드/클라우드 관리