

Analysis of YouTube's role as a new platform between media and consumers

Tai-Sung Hur*, Jung-ju Im*, Da-hye Song*

*Professor, Dept. of Computer Science, Inha Technical College, Incheon, Korea

*Student, Dept. of Computer Science, Inha Technical College, Incheon, Korea

*Student, Dept. of Computer Science, Inha Technical College, Incheon, Korea

[Abstract]

YouTube realistically shows fake news and biased content based on facts that have not been verified due to low entry barriers and ambiguity in video regulation standards. Therefore, this study aims to analyze the influence of the media and YouTube on individual behavior and their relationship. Data from YouTube and Twitter are randomly imported with selenium, beautiful soup, and Twitter APIs to classify the 31 most frequently mentioned keywords. Based on 31 keywords classified, data were collected from YouTube, Twitter, and Naver News, and positive, negative, and neutral emotions were classified and quantified with NLTK's Natural Language Toolkit (NLTK) Vader model and used as analysis data. As a result of analyzing the correlation of data, it was confirmed that the higher the negative value of news, the more positive content on YouTube, and the positive index of YouTube content is proportional to the positive and negative values on Twitter. As a result of this study, YouTube is not consistent with the emotion index shown in the news due to its secondary processing and affected characteristics. In other words, processed YouTube content intuitively affects Twitter's positive and negative figures, which are channels of communication. The results of this study analyzed that YouTube plays a role in assisting individual discrimination in the current situation where accurate judgment of information has become difficult due to the emergence of yellow media that stimulates people's interests and instincts.

▶ **Key words:** Youtube, Twitter, Data Analysis, Vader, Sentiment Analysis, Yellow Journalism

[요 약]

Youtube는 낮은 진입장벽과 영상물 규제 기준의 모호함으로 인하여 검증되지 않은 사실을 기반으로 한 가짜 뉴스, 편파적 콘텐츠 등이 사실적으로 나타난다. 따라서 본 연구에서는 언론과 Youtube가 개인의 행동에 미치는 영향과 이들의 관계성을 분석하고자 한다. selenium, beautiful soup, Twitter API로 Youtube와 Twitter의 데이터를 무작위로 가져와 가장 자주 언급되는 키워드 31개를 분류한다. 분류된 31개의 키워드를 기반으로 Youtube, Twitter, 네이버 뉴스에서 데이터를 수집 후, NLTK(Natural Language Toolkit)의 Vader 모델로 긍정, 부정, 중립 감정을 분류 및 수치화하여 분석 데이터로 사용했다. 데이터들의 상관성을 분석한 결과, 뉴스의 부정수치가 높아 질수록 Youtube에서는 긍정적인 콘텐츠가 많아지는 것으로 분석되었다. 본 연구결과로, Youtube는 2차로 가공하여 전달되는 특성으로 인해 뉴스에서 나타나는 감정 지수와 일치하지는 않는다. 즉, 가공된 Youtube 콘텐츠는 소통의 창구인 Twitter의 긍정, 부정수치에도 직관적으로 영향을 미치게 된다. 본 연구결과는 사람들의 흥미와 본능을 자극하여 시선을 끄는 황색언론의 등장으로 정보의 정확한 판단이 어려워진 현 상황에서, 자극적이고 부정적인 영상으로 사회에 악영향을 끼치는 것으로 인식되어있는 Youtube가 도리어 개인의 식별력을 보조하는 역할을 하는 것으로 분석되었다.

▶ **주제어:** Youtube, Twitter, 데이터 분석, 베이더, 감성 분석, 황색언론

- First Author: Tai-Sung Hur, Corresponding Author: Da-hye Song
- *Tai-Sung Hur (tshur@inhac.ac.kr), Dept. of Computer Science, Inha Technical College
- *Jung-ju Im (dlawjdwn12@naver.com), Dept. of Computer Science, Inha Technical College
- *Da-hye Song (thdekgp99@naver.com), Dept. of Computer Science, Inha Technical College
- Received: 2021. 11. 12, Revised: 2022. 01. 28, Accepted: 2022. 01. 28.

I. Introduction

Youtube(Youtube)는 개방된 무료 동영상 공유 사이트이다. 이는 연령이나 성별에 상관없이 이용률이 90% 이상으로 나타나는데, 특히 4차 산업혁명에 따라 발전한 뉴미디어로 인해 'Z세대'라고 불리는 10대는 99.2%, 20대는 98.0%의 이용률을 기록했다[1]. Fig 1.의 차트에서 보는 바와 같이 30대 이상에서도 Youtube는 90% 이상의 이용률을 보인다.

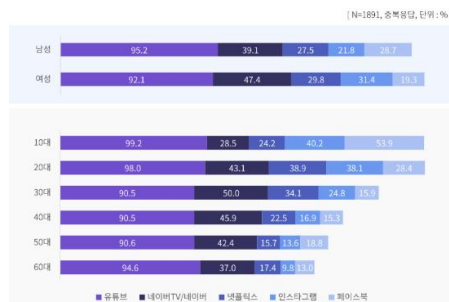


Fig. 1. Video viewing channel statistics[2]

그러나 낮은 진입장벽과 영상물 규제 기준의 모호함으로 인하여 검증되지 않은 사실을 기반으로 한 가짜뉴스, 종교 및 사상, 공격적이고 자극적인 정치 채널, 편파적 콘텐츠 등이 적나라하게 나타난다. 이에 광금주 서울대 심리학과 교수는 "일부 유튜버들이 조회 수를 이유로 위험한 콘텐츠를 올리는 실정, 이러한 콘텐츠에 자주 노출될 수록 위험성이 있다는 인식을 못 하게 될뿐더러 모방 위험성도 커진다."라고 지적했다[3]. 또한 옐로 저널리즘 관련, 가짜뉴스 전파량의 비약적 상승, 사실 확인의 어려움, 뉴스 유통 환경의 진화에 따른 탈진실 시대로 이동함에 따라 가짜뉴스로 인한 고소·고발이 이어지는 등 여러 사회 문제가 발생하고 있다.

현대경제연구원은 무분별하게 퍼져나간 가짜뉴스가 초래하는 경제적 비용이 당사자 피해액 22조 7천 700억 원, 사회적 피해액 7조 3천 200억 원 등을 포함하여 연간 30조 900억 원에 달한다고 추산하기도 하였다. 가짜뉴스는 단순히 소문과 비교하면 그 신뢰도나 영향력이 더 강력할 수 있다. 가짜뉴스가 생산되고 유통되는 과정에서 사회 전반에 걸친 다양한 분야에 영향력을 미치는 방향으로 발전할 가능성도 있다. 따라서 진짜 뉴스와 가짜뉴스의 노출 범위가 커짐에 따라 사회관계망 서비스는 어떤 영향을 받는지 체계적으로 연구할 필요가 있다[4].

연구에 앞서 데이터 수집을 위해 selenium[5], beautiful soup[6], Twitter API[7]를 사용하였다.

Twitter(Twitter)나 Youtube는 최근 많이 시청된 또는 사용된 주제들을 먼저 보여주기 때문에 수집된 데이터들을 분류하여 가장 많이 사용된 키워드 150개 중 상위 50% 데이터들 중 지역 간 분란을 조장하는 키워드 등을 제외하고 남은 31개의 키워드를 기준으로 다시 데이터를 확보하였다. 미디어별 데이터를 분석하기 위해 부정, 긍정, 중립을 기준으로 감정 지수 수치화해주는 Sentiment Analysis의 VADER 모델을 사용하였다. 분석된 감정은 부정, 중립, 긍정 지수로 나누어지게 되며 이 지수를 통해 뉴스와 소셜미디어 간의 관계성을 분석하였다[8]. 각각의 지수들은 정문(正文)일 경우 정확하게 판단이 되지만, 은어와 약어 사용이 잦은 소셜미디어의 특성을 고려하여 부정과 긍정의 척도가 되는 단어들의 가중치를 조절하였다. 전처리가 된 데이터들을 기반으로 뉴스, Youtube, Twitter 간의 상관관계 분석을 하였다. 예상과는 다르게 같은 키워드여도 뉴스는 부정지수가 높게 나왔지만, Youtube는 그와 반대로 긍정지수가 높게 나왔다. 이는 사람들이 소셜미디어를 받아들이는 모습의 차이에서 확인하였다.

사람들의 정향(동기와 가치)은 특정한 모습(양식)으로 표현된다. 광고에서 정향이라고 하면 메시지가 지향하는 것을 의미하며 이슈나 이미지에 대한 정향으로 표현할 수 있다[9]. 스포츠에 있어서 어떤 사람은 정향을 감성을 통해 보여줄 수도 있고[10], 어떤 사람은 똑같은 정향을 보여주더라도 사회적인 동기가 작용하기도 한다[11]. 정향을 정리하면 대상에 대한 가치 판단과 동기와 같이 사람들이 특정한 생각을 가지고 평가하는 것으로 볼 수 있으며 정향은 특정한 행태로 표현되는 것이다[6]. 또한, 카츠와 동료들은 다음과 같은 원칙을 제시하였다: "(a)이용자들은 목표 지향적이다, (b)이용자들은 적극적 미디어 이용자들이다, (c)이용자들은 자신의 욕구를 알고 있고 그 욕구들을 충족시키기 위해 미디어를 선택한다"[12]. 라는 관점에서 이용자는 능동적으로 정보를 선택하고 그에 대해 사고한다. 또한, 이용과 충족 이론(Uses and gratifications theory)은 미디어에 따라 다른 기능을 수행한다고 여긴다[13]. Youtube는 기존의 미디어가 각기 다른 방식으로 수행하던 기능들에서 벗어나 Youtube만의 방식으로 이용자들에게 정보추구 욕구와 오락적 기능의 복합적인 기능을 도와주는 것이다.

본 연구에서는 Youtube가 매스 미디어의 역할을 하는 동시에 사람들의 오락적인 역할도 함께 수행하는 것을 분석된 데이터를 통해 알 수 있다. 기존의 문화를 소비하는 대중들이 옐로 저널리즘이 도래한 언론의 혼란 속에서 Youtube가 새로운 방식의 플랫폼을 통하여 정보 선택을

기존의 방식보다 쉽고 능동적으로 할 수 있는 환경을 제공해 준다는 것이다. 즉, 수없이 쏟아지는 정보의 파도에서 Youtube라는 플랫폼은 개인의 목표 지향적인 정보 소비에서 가짜뉴스나 옐로 저널리즘이 생산해내는 정보 공해를 능동적으로 분별할 수 있는 능력을 키워준다. 이 능력들은 개개인이 소비하고 있는 미디어가 사실 검증에 필요한 정보를 내포하고 있는지 분별할 수 있는 시야와 안목을 넓히며, 4차 산업혁명 시대의 미디어에서 핵심적인 역할을 하게 된다.

II. Related works

1. Data collection

1.1 Selenium

Selenium은 웹 애플리케이션 테스트를 위한 포터블 프레임워크(Portable Framework)로, 브라우저 동작을 자동화시키고, 프로그래밍으로 제어하여 마치 사람이 이용하는 것과 같이 웹페이지를 요청하고 응답을 받아올 수 있다[5].

BeautifulSoup은 HTML과 XML에서 데이터를 처리할 수 있는 라이브러리로, HTML Parser를 사용하여 원하는 정보를 추출한다[7]. 다만, BeautifulSoup을 단독으로 사용하게 된다면 대량의 데이터를 동적으로 수집하기에 적절하지 못하며, 데이터 수집의 속도가 느리진다는 단점 또한 존재한다. 동적인 데이터 수집이 어렵다는 BeautifulSoup의 단점을 보완하고자 이를 Selenium으로 대체하여 사용하면, 웹 브라우저를 동적으로 사용하며 크롤링하는 것은 가능하지만, 데이터 수집에 대한 속도가 느리진다.

본 연구에서는 BeautifulSoup과 Selenium을 병행하여 사용하여 보다 빠르고 동적으로 웹 브라우저의 정보를 수집할 수 있도록 진행하였다. Chrome driver로 Youtube 브라우저를 호출한 뒤, 검색, 필터 기능을 자동화시켜 연구에 필요한 데이터가 포함된 콘텐츠의 리스트를 불러온 후, BeautifulSoup으로 수집한다.

1.2 Twitter API

Twitter API(애플리케이션 프로그래밍 인터페이스)로 Twitter 데이터에 액세스하여 엔드포인트를 호출하고, 사용자가 공개적으로 공유하기로 선택한 정보를 1회에 15분간 대략 200개씩 수집한다[7]. 이는 Selenium이나 BeautifulSoup과 같은 웹 애플리케이션을 통한 데이터 수집과 다르게 수집하고자 하는 데이터만을 가지고 올 수 있으며, Twitter 이용약관 및 개인정보 취급방침에 사용자가 게시한 데이터가 공개되거나 다른 곳에 사용될 수 있다는

Twitter 약관이 존재하기 때문에, 수집 및 연구 자료로의 활용에는 법적 이슈가 없으며, 법률적으로 검토해도 개인 정보 보호법에 위반하는 개인정보 수집에 포함되지 않는다. Twitter API는 수집하고자 하는 날짜부터 7일 안의 데이터만 얻을 수 있는 단점이 있으나, 위의 여러 장점을 가지고 있어 본 연구에서 SNS 데이터로 활용하였다.

1.3 Konlpy Open Korean Text

Konlpy는 파이썬에서 제공하는 한국어 정보처리 패키지, KAIST semantic Web Reaserch Ceter에서 개발한 Hannaum, 서울대학교 IDS 연구실에서 개발한 Kkam, Shineware에서 개발한 Komoran, 일본어용 형태소 분석기를 한국어를 사용할 수 있도록 수정한 Mecab, 과거 Twitter 형태소 분석기, 오픈소스 한국어 분석기인 Okt(Open Korean Text) 라이브러리가 내재되어 있다[14].

이중 Okt는 Konlpy에 내장된 다른 형태소 분석기와 달리 Twitter라는 SNS에 기재된 문장을 분석하는 라이브러리기 때문에 SNS에서 크롤링한 문장을 분석하는 본 연구에 가장 적합하다. 또한, Okt는 형태소의 분석뿐만 아니라 normalize 기능도 제공하는데, ‘사랑해’를 ‘사랑해’로 변환시키는 등, 표준국어대사전에 등재되어있지 않은 단어를 정규화한다. 단, 볼수록 매력 있음을 뜻하는 ‘볼매’, 심정이 몹시 어이없음을 의미하는 ‘얼렁없네’ 등과 같은 신조어 및 줄임말은 정규화되지 않는다는 단점이 있다.

1.4 NLTK VADER

NLTK(National Language Toolkit)은 파이썬에서 제공하는 NLP 패키지로, 텍스트 데이터의 특정 기호를 제거하고, 토큰화 및 불용어를 필터링하는 등 다양한 기능을 내포한다. 이중 VADER는 Sentiment Analysis를 위한 사전기반 감정분석 툴이기 때문에, Okt로도 변환되지 않은 신조어 및 줄임말을 사전에 직접 업로드 할 수 있기 때문에 본 연구의 데이터에 적절하다고 판단되어 사용하였다. VADER는 fig. 2에서 보이는 것과 같이 탑재하고 있는 텍스트별 감정평가 척도 사전을 바탕으로 문장의 감정을 긍정, 부정, 중립 3가지의 항목으로 수치화시킨다. 한 글자로 이루어진 단어는 제외시킨 후, 부정적 동사가 있으면 -0.74의 가중치를 곱하며, 영문의 경우 대문자는 강조의 의미로 보고 0.733을 긍정, 부정 여부에 따라 조정한다. 명사에 의미를 더하는 수식어는 0.293만큼 가중치를 조절하며, 수식어의 위치에 따라 정도가 조정된다. ‘그러나’와 같은 반전의 뜻을 나타내는 접속사의 경우 앞에 위치할 시 0.5를, 문장의 끝에 위치할 시 1.5를 곱한

다. 느낌표와 같이 감정이 담긴 기호에서도 가중치가 조정된다[15]. 그러나, VADER는 기존 모듈 내에 정의되어 있는 lexicon 사전을 기반으로 움직이기 때문에, 연구 데이터에 따라 추가적으로 가중치 조정이 요구된다.

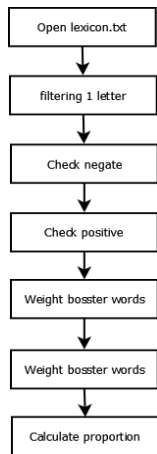


Fig. 2. Vader Flow Diagram

1.5 Collecting Data

감성 수치를 추출할 기준점이 되는 주제어들을 분류하기 위해, Twitter API를 활용해 무작위로 개인이 피드에 게시한 텍스트 데이터를 대량 추출한다. 추출한 데이터는 Konlpy의 Okt를 통해 이모지 및 불용어 제거와 정규화 단계를 진행하고, Python의 collections 모듈로 빈도수가 가장 높은 키워드 31개를 분류했다. 분류된 키워드로는 게임, 교회, 다이어트, 백신, 코로나 등이 있으며, 키워드를 중심으로 뽑은 Youtube, Twitter, 언론데이터는 각 주제어와 분류별로 1000~1500개다.

추출한 키워드가 언급된 텍스트 데이터들을 Twitter에서 추가로 수집하여 SNS 데이터를 확보하고, 마찬가지로 Youtube, naver 기사에서 추출한 데이터는 각각 Youtube, 언론데이터로 분류했다.

1.6 Emotional measurement

앞서 추출한 SNS, Youtube, 언론데이터들을 VADER를 통해 감정분류로 수치화하기 전, 정규화를 시킨 후에도 다분한 맞춤법 및 띄어쓰기, 줄임말, 신조어 등을 sentiment module이 인식할 수 있도록 기존에 존재하는 렉시온 파일에 가중치를 0.1~0.5 값으로 조정하였다. '죽어라', '빡친다', '얼탱없네'와 같이 부정적 의미를 극명하게 띄는 비속어들은 음의 값을 기재했다. 반면, '기쁘다', '존잘', '극락'과 같은 기존의 VADER 사전에 등재되어 있지 않던 긍정적 어휘들은 양의 값으로 정의했다.

1.7 Controversy rating

본 연구에 사용된 데이터에는 주제어별 Youtube의 영상길이, 조회 수별 긍정 및 부정수치, 전체 평균 부정수치와 SNS 및 언론의 긍정, 부정, 중립수치가 있다. 이중 SNS와 언론의 긍정 수치와 부정수치의 차를 구한 후 절댓값을 씌워 논쟁 수치라고 정의하였는데, 긍정과 부정의 차이가 적을수록 논란성이 높다고 판단했다. 연구데이터의 feature들간 공분산 수치를 구했을 때, Youtube의 감정이 중립에서 벗어나 긍정 및 부정수치가 높아질수록 언론상에서의 논쟁 수치 또한 높아지는 것을 확인하며 논쟁 수치에 대한 신뢰성을 확립했다.

2. Data Analysis

2.1 Collected Data

Table 1은 Youtube의 제목, SNS(instagram, Twitter)의 피드, 뉴스 제목의 긍정, 중립, 부정수치를 상관계수로 나타낸 표이며, '+'는 긍정, '*'는 중립, '-'는 부정 수치를 뜻한다. Table 1.에 대한 상세한 분석은 2장의 3절부터 5절에서 서술한다.

Table 1. Correlation coefficient of Youtube, SNS and News

	Yout ube-	Yout ube*	Yout ube+	SNS-	SNS*	SNS+	News-	News *	News +
Yout ube-	1								
Yout ube*	0.996	1							
Yout ube+	0.902	0.895	1						
SNS-	-0.030	-0.018	0.004	1					
SNS*	-0.297	-0.326	-0.398	-0.284	1				
SNS+	0.262	0.288	0.384	0.001	-0.948	1			
News-	0.417	0.397	0.644	-0.055	-0.129	0.165	1		
News *	-0.291	-0.311	-0.509	0.066	0.497	-0.526	-0.655	1	
News +	0.172	0.207	0.340	-0.057	-0.556	0.575	0.361	-0.941	1

Table 2는 수집한 31개의 주제 중 Youtube 제목의 검색 상위 5개와 하위 5개를 나타낸 표이며, 또한 주제 외에 21개의 주제에 관해 데이터를 수집하였다.

Table 2. Top 5 topics with average views and bottom 5 topics on youtube.

주제	평균	주제	평균 조회 수
방탄	52,919,828	네이버	99,842
Youtube	12,471,935	교회	109,897
마스크	4,108,699	백신	146,330
아이돌	2,816,005	정치	179,033
일본	2,024,638	언론	187,749

수집한 31개의 주제를 통해 뉴스, Twitter, Youtube에서 긍정, 부정, 중립 수치를 추출하였다. Python 3.7 환경에서 VADER 라이브러리를 사용하였다. 추출된 수치는 Excel을 통하여 상관계수, 공분산을 분석하였다.

2.2 The Effects of Youtube on User Activity

본 논문에서는 Youtube가 이용자에게 미치는 영향을 분석하였다. Youtube 이용 동기와 영화 콘텐츠 시청 의도에 미치는 선해 요인의 역할을 알아보기 위한 가설을 검증하기 위해 수도권 종합대학교에 설문조사를 실시한 대[10]. 광고 관련 수업시간 중 수업이 시작되기 전에 설문지를 나누어주고 설문 응답을 완료한 후 현장에서 수거하였다. 설문참여에 대한 금전적인 대가는 지급되지 않았다. 설문결과 총 230명의 학생이 설문문에 참여하였으며 이 중에서 남성은 70명(30.4%), 여성은 159명(69.1%)이 참여하였다. 참가자들의 나이는 19세부터 27세까지였고 평균 나이는 23.1세(표준편차 1.67)였다.

또한, 이용 동기의 경우 Youtube의 이용과 충족 차원을 알아보고 이용 그리고 충족과 세부적인 콘텐츠 장르인 영화 콘텐츠에 대한 정향의 역할에 초점을 두고 연구를 진행하였다. 연구를 진행함에 있어서 '정향'이라는 개념에 대해 논하였는데, 이를 특정한 방향성을 가진 사람들의 성향으로, 개인적인 가치와 중요성을 기반으로 한 동기 요인으로 볼 수 있다고 정의하였다. 같은 콘텐츠를 보더라도 개인이 소유한 가치와 중요성에 따라 이용 동기 및 방향성이 달라진 다는 것이다. 이 연구에 이어 본 논문에서는 다양한 콘텐츠에 대한 사용자의 감정 지수에 따른 Youtube의 새로운 역할에 대하여 연구하고자 한다.

2.3 Media-Youtube

나날이 발전하는 뉴미디어 시장에서 가장 대중적으로 사용되는 영상 플랫폼인 Youtube가 언론과 어떠한 상관관계를 가지며, 이러한 관계성이 개인에게 어떠한 영향을 미치는지에 대하여 알기 위해서는, 먼저 언론과 Youtube 사이의 관계를 파악해야 한다. 분석된 데이터들 중 언론과 Youtube의 감정 지수가 연관성이 있는지 확인하기 위하여 선형적 관계를 파악하기 위해 상관분석을 사용했다.

Fig. 3에서 보논바와 같이, Youtube와 언론 각각의 긍정, 부정수치 간 상관계수를 구한 결과, 언론의 부정수치와 Youtube의 긍정수치간의 상관계수가 가장 높았다.

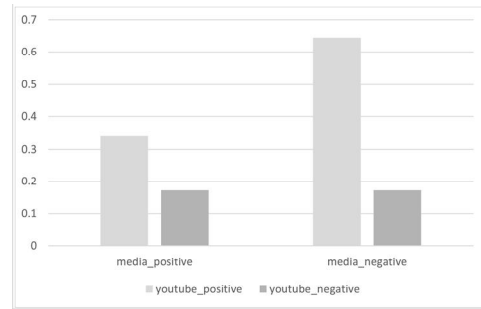


Fig. 3. The correlation coefficient between Youtube and the media's emotional figures.

언론의 부정적인 감정과 Youtube의 긍정적 감정이 어떠한 관계성을 띠고 있는지는 Fig. 4와 같이 공분산으로 확인하였다. Youtube의 긍정 수치는 언론의 부정수치와 양의 공분산 값을 가지고 있었다. 즉, 동일한 주제에서 언론이 부정적일수록 Youtube에서는 긍정적인 콘텐츠가 더욱 게시된다는 것이다.

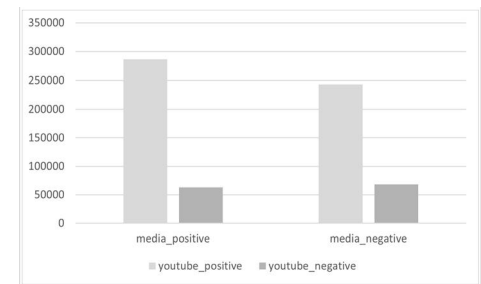


Fig. 4. The covariance between Youtube and the media's emotional figures.

Fig. 3에서 나타나는 언론과 Youtube의 방향성을 파악하기 위한 공분산 값이 Fig. 4에 나와있다. Fig. 3의 상관계수가 두 번째로 큰 항목은 긍정적인 Youtube와 긍정적 언론인데, Fig. 4에서처럼, 언론과 Youtube의 긍정 지수가 양의 공분산 값을 가지고 있다. Youtube는 기본적으로 언론의 부정, 긍정 모든 감정에 긍정적으로 반응하지만, 부정적 여론에는 높은 긍정적인 수치로 답한다는 분석이 나온다. Youtube가 언론의 부정적인 감정을 극대화하는 것이 아닌, 도리어 긍정적 콘텐츠로 균형을 맞추며 개인에게 다양한 정보의 선택권을 제공한다는 것을 알 수 있다.

2.4 Media-Twitter

앞서 도출한 결론과 연결 짓기 위해 개인이 언론에게 받는 영향을 분석하고자 언론의 controversy rating과 SNS의 긍정, 부정수치 간의 상관계수를 확인하였다. 그

결과, Fig. 5와 같이 언론의 논쟁 수치가 개인의 부정적 감정에 비해 긍정적 감정과 높은 관계성을 띠는 것을 확인했다. 반면, SNS에서 표현되는 부정적 감정은 언론의 논쟁성과 관계성이 현저히 떨어진다.

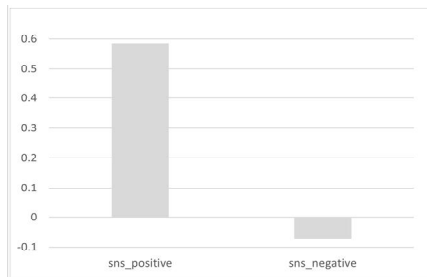


Fig. 5. The correlation coefficient between SNS and the media's controversy rating.

언론에서의 논란이 많아질수록, 개인의 긍정적 감정은 어떠한 영향을 가지는지 확인하기 위해 Fig. 6에서와 같이 언론과 SNS의 controversy rating의 공분산을 도출해 분석하였는데, 이들은 양의 공분산 값을 가지고 있었다. 언론에서 긍정적 감정과 부정적 감정이 비등하게 대립하며 나타나는 논쟁성과 SNS에 게시되는 긍정적 의견이 비례하게 나타난다는 것이다. 또한, Fig. 5에서 알 수 있듯이 개인은 언론의 논쟁에 영향받아 부정적인 태도를 취하지 않는다고 분석된다. Youtube가 대중화된 근래에는 개인이 언론의 논쟁에 휩쓸리며 그저 부정적 감정을 가지는 것이 아닌, 폭넓은 시야로 긍정적, 혹은 중립적 시선을 가지고 논점을 파악한다는 것이다.

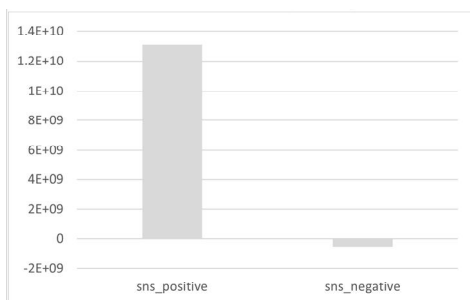


Fig. 6. The covariance between SNS and the media's controversy rating.

2.5 Youtube-Twitter

Youtube는 언론의 부정적 요소를 보충하는 것이 아닌, 감정의 균형을 맞추어 정보의 폭을 넓히고, 개인은 더 이상 언론의 논쟁에 휩쓸려 부정적 태도만을 취하지 않는다. 이 현상에 대한 원인을 Youtube와 개인의 감정간 관계에서 찾아보고자 각 언론에서 비롯된 주제가 Youtube

와 Twitter 사이의 관계에서는 어떤 영향을 미치는지에 대해 분석하였다. Fig. 7에서 보는 바와 같이, 부정지수에서는 큰 차이를 보이지 않지만, 긍정지수에서는 Youtube와 Twitter간의 상관관계수가 상대적으로 높게 확인됐다.

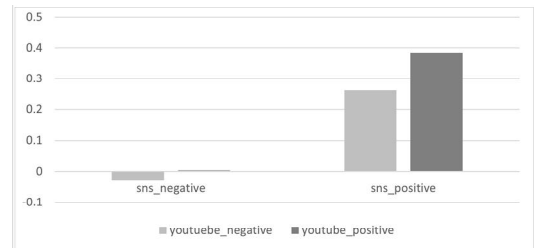


Fig. 7. The Correlation coefficient between SNS and Youtube.

SNS의 긍정 수치는 Youtube의 긍정, 부정수치 모두와 높은 상관계수를 가지고 있었고, Fig. 8에서 보이는 SNS와 Youtube의 감정 지수 공분산 값 또한 양의 관계를 띄고 있어 Youtube가 개인의 부정적 감정보다는 긍정적 감정에 더욱 영향을 끼친다는 것으로 분석된다.

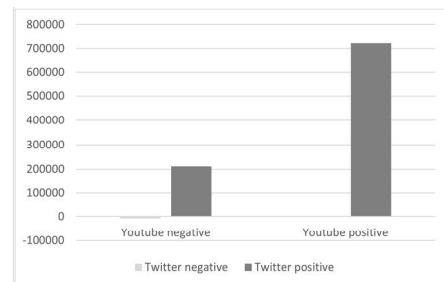


Fig. 8. The covariance between SNS and Youtube.

Youtube 데이터 중 상위 50%의 조회수를 가지는 콘텐츠와 SNS의 감정 지수를 대상으로 분석한 결과는 Fig. 9에서 확인할 수 있다.

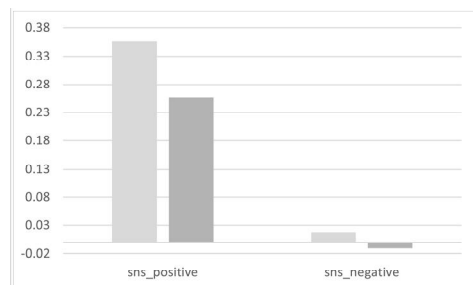


Fig. 9. Correlation coefficient between Youtube content above average views and SNS.

상위 조회수를 가지는 콘텐츠들은 그 영향력이 평균 혹은 그 이하인 조회수의 콘텐츠보다 높기에 Fig. 7에서

보여지는 평균 데이터들과의 결과가 유사한 것으로 보아, Youtube가 SNS에서 드러나는 개인의 긍정적인 감정에 영향을 더욱 크게 미친다는 것이 확인되었다. 또한 상위 조회수의 콘텐츠와 SNS의 공·부정 지수에 따른 공분산을 구한 결과, Fig. 10과 같이 SNS 긍정 수치와 양의 관계를 가진다. 이처럼 상위 조회수에서 또한 다른 콘텐츠들과 유사한 결과를 가진다는 것이 증명되었다.

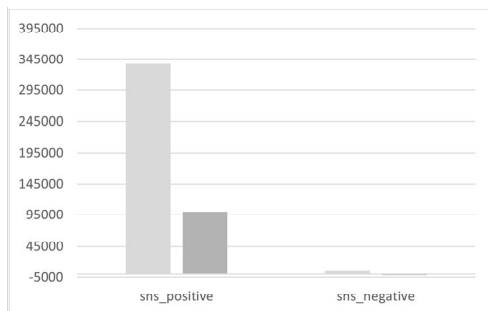


Fig. 10. Covariance between Youtube content above average views and SNS.

Youtube와 SNS간 상관계수와 공분산 지수로 보아 같은 키워드에 대한 긍정지수의 값이 상대적으로 큰 것으로 확인된다. Youtube는 가짜, 선동뉴스나 자극적인 콘텐츠의 생산 및 배포가 되는 소셜미디어의 플랫폼이기도 하지만, 언론과 SNS에서 이뤄지는 사회 상호작용에서 긍정적인 콘텐츠는 더욱 긍정적으로 인식되며 부정적인 콘텐츠는 더욱 적은 영향을 받는 소셜미디어 플랫폼으로 판단된다.

III. Result and Discussion

Youtube는 2020년 10월 8일자 KBS 뉴스에 따르면, 한 달 동안 Youtube를 사용한 우리나라 국민은 약 4천319만 명으로, 이는 전체 국민의 83%에 달하는 수치이다[17].

TV를 시작으로 이후 새로운 미디어가 등장할 때마다 이용과 충족이론은 기존 동기를 재확인하는 동시에 새로운 미디어에 맞는 동기를 추가하는 방식으로 확장되어 왔다[18]. Youtube도 TV의 동기 확장 방식과 유사하게 정보탐색, 제공, 사회적 상호작용, 자기 존재 탐색 등 기존의 동기에서 벗어나 가짜뉴스에 대한 팩트체크의 동기, 사회적 불안요소에 관한 정보의 공유 등을 통하여 본 연구결과인 언론의 부정지수가 높은 주제에 관해서도 Youtube의 긍정지수는 높은 상관 계수(0.644)를 띄며 Youtube의 부정지수와 언론의 부정지수는 0.417의 상관 계수를 가진다. 이 수치는 언론의 부정적인 키워드가

Youtube에서 재해석되어 미디어를 이용하는 사람들에게 새로운 관점을 제시한다고 분석된다. Youtube의 긍정지수가 높은 영상과 언론의 중립지수가 높은 키워드는 (-0.5)의 상관계수를 가지는 것으로 보아 상반되는 지표를 보이며, SNS에서 긍정적인 키워드와 언론에서의 긍정적인 키워드는 0.5의 상관 계수를 가지며 SNS와 언론은 같은 키워드 내에서는 긍정적인 지수가 양의 상관을 보이는 것으로 분석된다.

미디어의 이용 동기에 따라 미디어 내에서 사용되는 긍정-부정적인 뉘앙스가 달라지는 기존연구와 비교하여 [19], 본 연구에서는 미디어 내에서 사용되는 실제 게시글 및 제목에서 사용되는 단어들의 공·부정수치를 분석하여, 뉴스, Twitter와 Youtube를 사용하는 사용자들이 상호적으로 어떤 영향을 미치는지를 분석하여 새로운 관점을 제공했다.

본 연구의 모집단은 기존의 연구에서 설문조사를 통한 표본과는 다르게 실제 미디어를 이용하는 사용자들인 점에서도 기존의 연구와도 차별점이 있다. 하지만 본 연구는 데이터 수집에 있어 한 키워드당 약 600~900개의 문장으로 이루어져 있어 좀 더 많은 데이터를 확보할 필요성이 있다. 또한 중복 데이터를 제거하는 과정에서 많은 분량의 데이터가 제거되었고, Twitter나 Youtube에서 비문(悲門)의 사용이 많아 문장 단위로 감정 지수를 추출하는 과정에서 데이터 손실이 발생하였다. 본 연구에서 사용된 데이터들은 수집 시에 기간을 고려하지 않았기에, VADER로 측정된 감정 수치들이 해당 주제가 즉 동기간 대의 데이터들인지 확인이 불가하다. Twitter API가 최근의 게시글부터 탐색을 시작하며, 언론에 관한 데이터는 날짜와 상관없이 앞서 추출한 키워드와 긴밀하게 일치하는 문장이 수집되기 때문에 Youtube, Twitter, 언론 세 가지의 데이터가 동일한 주제어를 내포한 다른 문제점을 거론한 내용일 수도 있다는 것이다. 따라서, 향후 연구에서는 주제어와 감정 수치뿐만 아니라 이슈화되었던 기간까지 포함하여 분석되어야 할 것이다.

REFERENCES

- [1] EDAILY, "93% of video service users "watch YouTube"...99.2% of teenagers", <https://www.edaily.co.kr/news/read?newsId=02335366625733496&mediaCodeNo=257>
- [2] Nasmedia, "Internet User Survey NPR", <https://www.nasmedia.co.kr/NPR/2020/>
- [3] Yonhap News, "[sns world] "What if you copy the kids?" Dangerous

- Million' YouTube videos." <https://www.yna.co.kr/view/AKR20210610154300505>
- [4] Yum, Jung-Yoon & Jeong, Se-Hoon, "Predictors of fake news exposure and sharing : personality, new media literacy, and motives," Korean Journal of Journalism & Communication Studies 63(1), 7-45, (2019). 10.20879/kjcs.2019.63.1.001
- [5] Selenium, Selenium, accessed Dec 30, 2020, <https://www.selenium.dev/>
- [6] Beautiful soup, Beautiful soup, accessed Dec 30, 2020, <https://beautiful-soup-4.readthedocs.io/en/latest/#>
- [7] Twitter API, Twitter developer platform, accessed Dec 30, 2020, <https://developer.twitter.com/en/docs/twitter-api>
- [8] Hutto, Clayton, and Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." Proceedings of the International AAAI Conference on Web and Social Media. Vol. 8. No. 1, (2014).
- [9] Jun, Jong-Woo, "Effects of Uses and Gratifications of YouTube and Content Orientation on Youtube Uses," Journal of Outdoor Advertising Research 18(2), 5-21, (2021). <https://doi.org/10.3998/83336451.0011.305>
- [10] Wann, Daniel L. "Preliminary validation of the sport fan motivation scale." Journal of Sport and Social issues 19.4 (1995): 377-396. <https://doi.org/10.1177/019372395019004004>
- [11] Swanson, Scott R., et al. "Motivations of college student game attendance and word-of-mouth behavior: The impact of gender differences." Sport Marketing Quarterly 12.3 (2003).
- [12] C. Lin, "Uses and gratifications," Clarifying Communication Theories: A Hands-on Approach, pp.199-208, 1999
- [13] Katz, E., Blumler, J., & Gurevitch, M. "Utilization of mass communication by the individual," The uses of mass communications: Current perspectives on gratifications research. Beverly Hills, CA: Sage pp. 19-32, (1974)
- [14] "KoNLPy Korean Language Processing Package", Dataschool, accessed Dec 30, 2020, <https://datascienceschool.net/03%20machine%20learning/03.01.02%20KoNLPy%20%ED%95%9C%EA%B5%AD%EC%96%B4%20%EC%B2%98%EB%A6%AC%20%ED%8C%A8%ED%82%A4%EC%A7%80.html>
- [15] VADER Sentimental Analysis Code Analysis, To you who travel between stars, accessed Dec 30, 2020, <https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=ossiriand&logNo=220607426789>
- [16] Jihye Jung, Geunyoung Kim, "Excessive YouTube Usage of Middle School Boys and Internalizing and Externalizing Problems: Moderating Effects of Social Support," Korea Academy Industrial Cooperation Society's thesis paper, 22(2), 676-684, (2021). <https://doi.org/10.5762/KAIS.2021.22.2.676>
- [17] Rhee, June Woong et al. "Exploring Dispositional Media Use Motives : An Extension of the 'Uses and Gratification' Theory in a Multimedia Environment." (2006).
- [18] Kim, Wi-Geun, and Min-Jae Choi. "The effect of SNS users' use motivations on using SNS and recognizing characteristics of SNS messages: Focused on the comparison among Facebook, Twitter, Cyworld, and Me2day." Korean journal of communication and information 60, 150-171, (2012)

Authors



Tai-Sung Hur received the B.S degree in Dept. of Computer Science from Inha University in 1984, and M.S degree in Dept. of Computer engineering from Soongsil University in 1987, and Ph. D.

degree in Dept. of Computer engineering from Inha University in 1992. Dr. Hur has over 30 years of computer education. He is currently a Professor in the Dept. of Computer Science, Inha Technical College. He is interested in Data Science, Big data, Database and Internet of Things.



Jung-ju Im received B.S. degree in 2022 from the Department of Computer Science Inha Technical College, Incheon Korea. He is currently a student in Hanyang Graduate School.

His research interests include Mobile Computing, Image Recognition and Machine Learning Network.



Da-hye Song received B.S. degree in 2022 from the Department of Computer Science Inha Technical College, Incheon Korea. She is currently a student in Hanyang Graduate School.

Her research interests include Health Care, Deep Learning Network.