

코로나19 관련 키워드 분석: 토픽 모델링과 의미 연결망 네트워크 분석을 중심으로

김동욱* · 이민상* · 정재영* · 김현철**

** 상명대학교 소프트웨어학과

COVID19 Related Keyword Analysis: Based on Topic Modeling and Semantic Network Analysis

Dong-wook Kim*, Min-sang Lee*, Jae-young Jeong* and Hyun-chul Kim**

** Dept. of Software, Sangmyung University

ABSTRACT

In the era of COVID-19 pandemic, COVID related keywords, news and SNS data are pouring out. With the help of the data and LDA topic modeling, we can check out what media reports about COVID-19 and vaccines. Also, we can be clear how the public reacts to the vaccine on social media and how this is related with the increasing number of COVID-19 patients. By using sentimental analysis methodology, we can get to know about the different kinds of reports that Korea media send out and get to know what kind of emotions that each media company uses in majority. Through this procedure, we can know the difference between the Korean media and the foreign ones. Ultimately, we can find and analyze the keyword that suddenly rose during the COVID-19 period throughout this research.

Key Words : COVID19, LDA Topic Modeling, Sentiment and Emotion Analysis, CONCORAnalysis

1. Introduction

2019년 코로나 바이러스가 처음 확인되고 지금까지 전 세계적으로 큰 피해를 입히고 있다. 세계보건기구(WHO: World Health Organization)는 감염병 위험 수준에 따라 6단계로 구분해 놓았는데 2020년 3월 11일 코로나19에 대해 전염병 경보 최고 단계인 6단계(세계적 대유행), 즉 팬데믹을 선언하였다. 이런 코로나 팬데믹 시대에서 코로나와 관련된 키워드로 기사와 SNS 데이터들이 쏟아져 나오고 있다.

따라서 본 연구에서는 언론의 기사를 통하여 코로나 팬데믹 속에서 우리나라에 어떠한 일들이 발생했고 어떻게 변화되었는지 토픽 모델링으로 알아보고 해외와 우리나라의 언론사에서 코로나 관련 키워드를 어떤 관점에서

주로 보도하는지 알아보고자 감성 분석을 진행하였다. 또한, 자신의 의견을 서로 공유하는 SNS의 특징을 이용하여 코로나 관련 키워드를 통해 사람들의 여론 및 감정을 파악하고자 하였다. 마지막으로 코로나 전과 후 어떠한 단어들에 새롭게 도출되었는지 알아보고자 연구를 진행하였다. 2장에서는 관련 연구에 대하여 설명하며, 4장에서는 데이터 수집과 전처리에 관하여 기술한다. 5장에서는 각 분석 방법을 세부적으로 기술하도록 한다. 6장에서는 각 분석 방법에 따른 결과를 기술한다. 마지막 7장에서는 본 논문의 결론에 대하여 기술한다.

2. Related Works

본 장에서는 언론의 감성 분석, LDA 토픽 모델링, CONCOR분석과 관련된 연구를 소개한다.

Ki et al. (2020)에서는 농촌태양광에 대한 언론과 블로그

†E-mail: hkim@smu.ac.kr

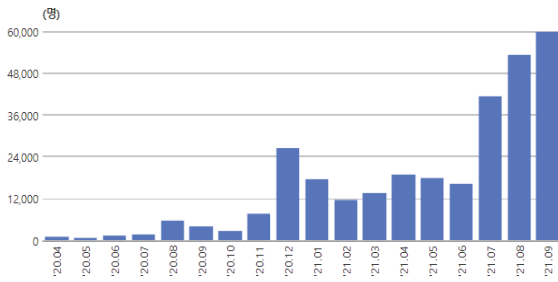


Fig. 1. Number of COVID-19 cases in 2021.

포스트의 감성을 긍정, 부정, 중립 세 가지 척도로 분류한다[1]. Park et al. (2020)에서는 전력반도체 패키징 기술 동향 분석을 위해 LDA 토픽모델링을 이용하였다[2]. Nam et al. (2018)에서는 차량용 반도체 기술 추세 분석과 기술 트렌드를 연구하기 위해 LDA 토픽모델링을 활용하였다[3]. 코로나19와 관련하여 토픽모델링을 진행한 선행연구 Suh et al. (2021)에서는 코로나19 대유행 기간 ‘마스크’ 키워드와 관련한 언론 보도 및 태도 변화에 대한 분석을 진행하였다[4]. Yun et al. (2020)에서는 코로나19 이후에 바뀐 한국인들의 여행 행동패턴에 대해 CONCOR 네트워크 연결망을 분석하였다 [5].

본 논문에서는 감정과 감성을 포함하여 총 10가지 척도로 감정을 더욱 풍부하게 분류하였다. 또한 언론 외에도 SNS의 데이터로 코로나 키워드와 관련하여 토픽모델링을 진행하였고, 선행연구와는 다르게 SNS 데이터를 코로나19 전 기간을 포함하여 CONCOR 분석을 통해 코로나19 전과 후의 변화를 알아보았다.

3. Data

3.1 Data Crawling

언론 데이터는 국내 언론 데이터와 해외 언론 데이터를 수집하였다. 수집한 국내 언론사들은 YTN, SBS, OBS, MBC, KBS, 한국일보, 한겨레, 중앙일보, 조선일보, 오마이뉴스, 연합뉴스, 세계일보, 서울신문, 문화일보, 매일경제, 동아일보, 내일신문, 국민일보가 있다. 수집한 해외 언론사들은 The Guardian, CNN, BBC, Asahi가 있다. 국내 언론 데이터들을 수집하는 방법은 네이버에서 특정 키워드(백신, 코로나)에 대하여 기간과 언론사를 설정하여 검색 후 나오는 결과들을 Selenium을 사용한 웹 크롤링을 통해 수집한다. 해외 언론 데이터들을 수집하는 방법은 각 언론사의 사이트에서 검색란에 특정 키워드들(Vaccine, Corona, Covid19)을 검색한 결과를 Selenium을 사용한 웹 크롤링을 통해 수집한다.

SNS의 데이터는 Instagram과 Twitter의 데이터를 Selenium

을 이용한 웹 크롤링 방법을 통하여 직접 수집하였다. Twitter는 고급 검색 기능을 이용해 기간 설정, 제외 단어(성인 광고 키워드), 링크, 답글 제거를 설정하여 코로나 관련 키워드(백신, 여행)의 데이터를 수집하였다.

3.1.1 Dataset

수집한 언론 데이터셋은 날짜(Datetime)과 기사 본문(Str)으로 이루어져 있고 각 언론사별로 csv파일로 저장하여 사용하였다. ‘백신’ 키워드로 수집한 데이터를 설정한 기간은 2020년 11월 1일부터 2021년 10월 31일까지이고 총 수집한 기사는 약 10만 개 정도이다.

언론과 마찬가지로 SNS의 데이터셋은 날짜(Datetime)와 본문(Str)의 형식으로 이루어져 있고 csv 파일로 저장하여 사용하였다. ‘백신’ 키워드로 2020년 11월 1일부터 2021년 10월 31일까지의 데이터 11,106개를 수집하였고 ‘여행’ 키워드로 2018년 2월 1일부터 2021년 10월 31일까지의 데이터를 61,857개 수집하였다.

3.2 Data Processing

3.2.1 Preprocessing of LDA Topic Modeling

본 연구에서는 ‘코로나’와 ‘백신’ 키워드 관련하여 토픽 모델링 진행을 위해 konlpy를 사용하여 모든 텍스트의 0부터 9까지의 숫자, 영문자, 특수문자를 제거하였고 pfizer와 같은 중요한 영어 단어는 한국말로 바꿔서 데이터 누락을 방지했다. 또한 okt를 사용해 문장 텍스트 데이터를 토큰화하고 조사, 대명사 등을 제거한 명사만 추출하여 본 연구에 사용하였다. 동시에 분석에 의미가 없는 ‘코로나’, ‘주사기’ 같은 단어들은 불용어로 처리하여 제거하였다.

3.2.2 Preprocessing of Sentiment and Emotion Analysis

감성분석을 위해 NRC감성사전을 사용하는데 NRC감성 사전에서 사용할 수 있는 언어는 영어이고 불용어를 제거하는 데 한글은 언어적 어려움이 있다. 따라서 한글로 된 국내 언론은 Google Cloud Translation을 사용하여 영어로 번역하여 진행하였다. 영어로 번역된 국내 언론 기사들을 단어 단위로 나누어 숫자와 특수 기호를 제거하여 불용어 처리를 진행하였다.

3.2.3 Preprocessing of CONCOR Analysis

‘여행’ 키워드로 알아본 단어연결망 연구에서는LDA 토픽 모델링과 마찬가지로 한국어 텍스트 문장 데이터를 토큰화 시켜 명사들만 추출하여 사용했다. 다만 단어간 의미연결망을 알아보기 위해 명사들의 위치가 중요하기 때문에 문장구조를 유지한 채 토큰화를 진행했다.

3.2.4 Period Classification

LDA 토픽모델링을 위한 기간 분류는 분석하는 키워드에 따라 다르게 진행하였다. 우선 ‘코로나’ 키워드로는 2020년 2월부터 2021년 10월까지 대한민국 통계청 코로나 19 확진자 현황 데이터 그래프에서 최고 확진자 수 날짜를 기점으로 전후 15일을 가져와 각 시기당 30일의 데이터를 분석하였다. ‘백신’ 키워드로는 ‘백신’에 대한 여론을 알아보려고 진행하였으므로 백신의 도입된 2021년을 기준으로 3개의 기간으로 나눴다.

감성분석 기간 분류는 2020년 11월부터 2021년 10월까지의 데이터를 사용하였으며, 언론 기사의 감성 변화를 알아보기 위해서 초기를 2020년 11월 부터 2021년 1월 중기를 2021년 2월 부터 2021년 6월, 말기를 2021년 7월 부터 2021년 10월로 설정하고 데이터를 분류하여 기간별 감성 분석을 진행하였다.

CONCOR 분석을 진행하기 위해 ‘여행’이라는 키워드로 코로나19 전 후 기간을 분석하기 위해 국내 코로나19 확진자 발생일 전후 1년 10개월을 사용하였다. 이로써 코로나 전은 2018년 2월부터 2019년 12월, 코로나 후 기간은 2020년 1월부터 2021년 10월로 지정하여 분석하였다.

4. Analysis Methodology

4.1 LDA Topic Modeling

LDA 토픽모델링은 각 단어나 문서의 숨겨진 주제를 찾아, 문서와 키워드를 묶어 주제를 형성하는 텍스트 마이닝 기법이다[6]. LDA 토픽모델링은 두 가지의 과정으로 이루어지는데, 우선 문서에 존재하는 토픽 수를 설정해준 다음 설정한 토픽 개수들에 해당하는 단어들을 랜덤하게 토픽들로 할당해준다. 이후 토픽의 각각 단어들마다 반복적인 검증을 거친다.

다음으로 베이지안 이론(Bayesian Theory)의 개념을 차용하여 특정 단어 하나만 토픽이 잘못 배정되어 있고, 다른 단어들은 올바르게 할당되었다는 가정을 세우고 난 뒤 디리클레 분포(Dirichlet Distribution)로 계산된 다른 단어들의 prior 값들을 참고하여 특정 단어의 토픽 위치를 결정하는 과정을 반복한다. 최종적으로 임의의 토픽에 비슷한 단어들끼리 할당하고 분류의 과정을 거친다. 토픽모델링의 시각화는 pyLDAvis를 사용해서 진행했다.

4.2 Sentiment and Emotion Analysis

각 언론사들이 작성한 기사 내용을 토대로 감성분석을 진행하여 언론사마다 기사에 어떤 감성을 담아서 작성하였는지 기사를 긍정적으로 작성하였는지 부정적으로

작성하였는지, 혐오감을 조성하였는지 등에 대해서 알아본다. 감성분석은 크게 사전 기반 감성 분석과 지도 기계학습 기반 감성 분석으로 나눌 수 있는데 본 논문에서는 사전 기반 감성 분석을 사용하였다. 전처리된 데이터들을 통해 각 단어들이 어떤 감성을 가졌는지 NRC 감성사전을 통해 가져온다. NRC 감성사전에는 긍정, 부정 및 플루릭의 수레바퀴에 기초하여 인간이 가지는 가장 기본적인 감성 8가지인 공포, 기대, 신뢰, 놀람, 슬픔, 기쁨, 역겨움, 분노로 감성 정보를 더욱 풍부하게 분류하고 14,182개의 감성어를 보유하고 있다. NRC 감성사전을 통해 언론 기사들의 감성별 비율을 분석한다.

4.3 CONCOR Analysis

CONCOR(CONvergence of iteration CORrelation) 분석은 단어 간 유사성을 중심으로 단어들의 관계패턴을 도출하고, 군집화 된 해당 클러스터들의 성격을 알아볼 수 있는 분석 방법이다. 본 연구에서는 ‘여행’ 키워드로 크롤링한 트위터 데이터에 대해 코로나19의 전후 기간에 클러스터들의 변화를 살펴보기 위해 이 방법을 사용하였다. CONCOR 분석을 위해 전후 기간 각각 토큰화된 명사들로 단어 출현 빈도 수(word frequency) 와 동시 출현 매트릭스(co-occurrence matrix)를 생성하였다. 전 과정에서 생성한 데이터들을 UCINET의 Netdraw 프로그램을 사용하여 ‘여행’ 키워드로 도출된 단어들을 그룹화 하였다.

5. Results

5.1 LDA Topic Modeling

5.1.1 Topic Analysis of Keyword ‘COVID-19’ during the COVID-19 Pandemic

5.1.1.1 1st wave

2020년 2월 25일부터 2020년 3월 31일까지 ‘코로나’ 키워드로 언론 데이터를 뽑아 도출된 토픽 결과는 Fig. 2와 같다. Fig. 2는 토픽 모델링을 통해 각 토픽 별 비중이 큰 단어들을 일부 선택해 표시한 결과다. 도출된 주제를 확인한 결과 1st wave는 코로나19가 발생한 초기 기간으로 ‘온라인’, ‘개학’, ‘연기’, ‘학교’ 등의 키워드들이 나왔다. 이는 코로나 초기 학교 개학이 연기되었고, 학교 수업이 온라인으로 진행된다는 것과 관련되어 있다. 9번 토픽에서는 주가와 관련된 키워드들이 확인되었는데, ‘증시’, ‘코스피’, ‘지수’, ‘폭락’, ‘공포’ 등과 같이 이 당시 코스피지수가 역대급으로 폭등하여 폭락에 대해 공포를 나타내는 것이다.

1st wave	Topic Num	Topic words	2nd wave	Topic Num	Topic words
2020.02.25 ~2020.03.31	3	개발, 연기, 학교, 교육부, 온라인	2020.08.20 ~2020.09.20	5	채용, 연기, 시험, 수시
	9	중시, 코스피, 지수, 폭락, 주식		6	택배, 노동, 버스, 손해
	12	요양, 대구, 집단, 부산		17	육사, 전광훈, 의혹, 소양, 비리
3rd wave	Topic Num	Topic words	4th wave	Topic Num	Topic words
2020.12.05 ~2021.01.04	15	동부, 구치소, 수용자, 양성	2021.09.13 ~2021.10.13	9	비엔날레, 사진, 예술, 전시
	21	생산, 자동차, 업계, 전기, 위기		15	학교, 학생, 대학, 교육, 축제, 종교
	23	모더나, 화이자, 사형, 계약, 승인		18	일상, 회복, 위드, 전환

Fig. 2. Topic Modeling by the keyword “COVID-19” during the pandemic with News Data.

5.1.1.2 2nd wave

2020년 8월 20일부터 2020년 9월 20일까지 도출된 2nd wave의 토픽 5에서는 ‘채용’, ‘연기’, ‘시험’, ‘수시’ 등의 키워드들이 나왔다. 이는 주요 국가 시험들이 연기됨에 따라 나타난 것이다. 6번 토픽에서는 ‘택배’, ‘노동’, ‘버스’, ‘손해’ 등의 키워드가 나왔는데, 이는 이 기간에 택배업들이 고속성장을 하여 수익의 30%가 늘고, 노동직과 버스회사는 적자를 본 것과 관련이 있다.

5.1.1.3 3rd wave

2020년 12월 5일부터 2021년 1월 4일까지 도출된 3rd wave에는 백신의 도입과 겹치는 대유행 기간이라 ‘모더나’, ‘이자’, ‘계약’, ‘승인’ 등의 백신 접종에 대한 시작에 대한 키워드들이 나왔다. 토픽15번에서는 ‘구치소’, ‘동부’, ‘수용자’, ‘양성’ 등의 키워드들이 나왔다. 이는 서울동부구치소에서 집단감염사태가 일어난 것과 관련이 있다. 토픽 21번에서는 ‘생산’, ‘자동차’, ‘업계’, ‘위기’ 등의 키워드들이 확인되었는데 이는 자동차 생산이 부진하여 자동차업계의 위기가 찾아온 것에 대한 것이다.

5.1.1.4 4th wave

2021년 9월 13일부터 2021년 10월 13일까지 도출된 토픽 3번의 ‘비엔날레’, ‘사진’, ‘예술’, ‘전시’ 등의 키워드들은 2021년 9월에 개막하는 광주 비엔날레와 더불어 서울시에서도 많은 문화예술 프로그램을 개최한 것과 관련이 있다. 토픽 15번에서는 ‘학교’, ‘학생’, ‘대학’, ‘교육’, ‘축제’ 등의 키워드들이 관찰되었다. 토픽18번에서는 ‘일상’, ‘회복’, ‘위드코로나’, ‘전환’ 등의 키워드가 관찰되었는데 이는 국내에서 나아가 세계적으로 위드코로나를 시행하고

있다는 것, 일상으로의 복귀를 꾀하고 있는 것과 관련이 있다.

5.1.2 Topic Analysis of Keyword ‘vaccine’ during the COVID-19 Period

5.1.2.1 Press

2020년 11월부터 2021년 1월까지 초기 언론에서 ‘백신’ 키워드로 도출한 토픽은 다음 Fig. 3의 상단과 같다. 토픽 6번에서는 ‘유통’, ‘보관’, ‘준비’ 등의 키워드들이 나왔는데 백신접종을 위해 백신에 대해 보관과 유통에 대한 것하고 연관이 있다. 토픽 14번에서는 ‘변이’, ‘옥스퍼드’, ‘존슨앤존슨’, ‘일본’ 등의 키워드들이 확인되었다. 이는 이 시기에 옥스퍼드 대학교와 존슨앤존슨등 제약사가 백신에 대한 승인이 임박하고 일본에서 새로운 변이 바이러스가 확인된 것 하고 연관이 있다.

다음은 2021년 2월부터 2021년 6월까지 중기 언론에서 도출된 토픽이다. 토픽 20번에서는 ‘변이’, ‘영국’, ‘효능’, ‘남아공’ 등의 키워드들이 확인되었는데, 이는 알파, 베타 바이러스가 확산함에 따라 백신에 대한 효능에 대한 의문을 가지는 것 하고 관련이 있다.

2021년 7월부터 2021년 10월까지 말기 언론에서 도출된 토픽 3번은 ‘심근염’, ‘소아’, ‘심낭염’, ‘혈전증’ 등 코로나 19에 대한 백신 접종 후 생기는 부작용들에 대한 키워드들이 도출되었다.

5.1.2.2 SNS (Social Network Service)

2020년 11월부터 2021년 1월까지 초기 SNS에서 ‘백신’ 키워드로 도출한 토픽은 Fig. 3 하단과 같다. 토픽 8번에서 ‘중국’, ‘이상’, ‘시노백’, ‘구매’ 등을 통해 중국에서 개발한 시노백 백신에 대해 이상이 있다는 것과 연관이 있다.

	초기(20.11-21.02)	중기(21.02-21.06)	말기(21.07-21.10)
언론	Topic Num: 6 Topic words: 유통, 보관, 준비, 저온	Topic Num: 5 Topic words: 신고, 사망, 조사, 의심, 노인	Topic Num: 1 Topic words: 시노백, 격리, 접촉, 사업장, 장애
	Topic Num: 9 Topic words: 중시, 코스피, 하락, 마감, 포인트	Topic Num: 11 Topic words: 생산, 노바백스, 한국, COVAX, 바이오	Topic Num: 2 Topic words: 바이든, 민주당, 존슨, 종리, 협약
	Topic Num: 14 Topic words: 변이, 옥스퍼드, 존슨앤존슨, 일본, 연구	Topic Num: 20 Topic words: 변이, 영국, 연구, 효능, 남아공	Topic Num: 3 Topic words: 심근염, 소아, 심낭염, 혈전증
SNS	Topic Num: 1 Topic words: 세계, 안전, 성공, 기업, 안정	Topic Num: 6 Topic words: 효과, 문제인, 물량, 부족	Topic Num: 2 Topic words: 후유증, 사진, 인증
	Topic Num: 2 Topic words: 독감, 언론, 희망	Topic Num: 7 Topic words: 부작용, 영향, 교차	Topic Num: 9 Topic words: 운동, 피곤, 근육통, 그림
	Topic Num: 8 Topic words: 중국, 이상, 단계, 구매, 감염증	Topic Num: 14 Topic words: 전여, 예약, 격정, 병원	Topic Num: 11 Topic words: 부작용, 이상, 다행

Fig. 3. Topic Modeling by the keyword “vaccine” during the 2021 COVID19 period with Media and SNS.

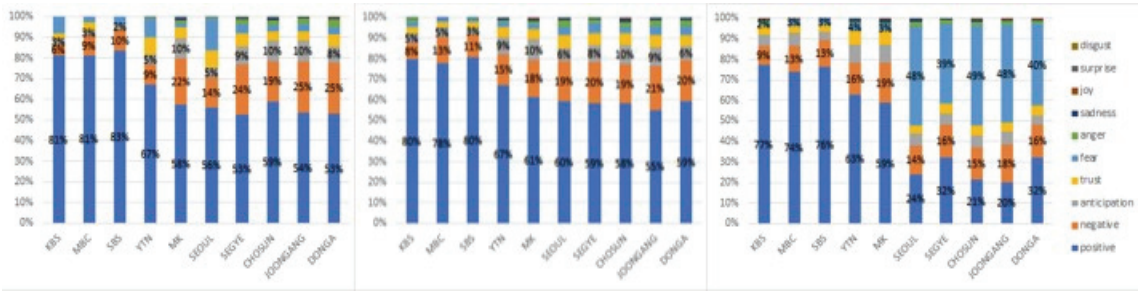


Fig. 4. Sentiment and Emotion Analysis Domestic.

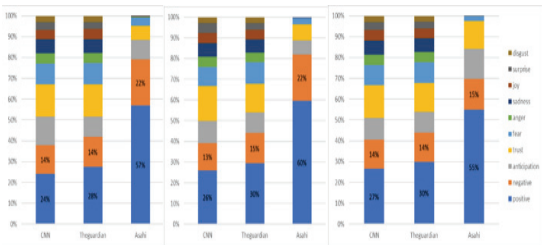


Fig. 5. Sentiment and Emotion Analysis Overseas.

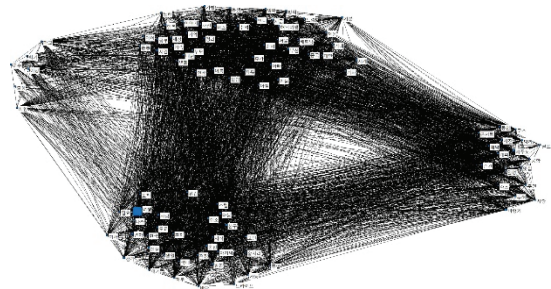


Fig. 6. CONCOR Analysis by the keyword 'Travel' Before COVID19 period.

2021년 2월부터 2021년 6월까지 중기 14번 토픽에서는 ‘잔여’, ‘예약’, ‘걱정’, ‘병원’ 등의 키워드들이 확인되었다. 이는 SNS 상에서 잔여 백신에 대한 예약에 어려움을 겪었던 것과 백신 접종에 대한 걱정을 한 것과 연결된다.

2021년 7월부터 2021년 10월까지 말기 SNS에서 도출된 토픽 9번에서는 ‘운동’, ‘피곤’, ‘근육통’, ‘그림’과 같은 키워드들이 나왔는데 이는 백신을 접종 후 운동, 그림과 같은 활동을 하거나 피로감과 근육통을 느꼈다는 것과 연관이 있다.

5.2 Sentiment and Emotion Analysis

‘백신’ 키워드에 대한 언론 별 감성분석 결과 국내 언론 전부 긍정적인 감성이 가장 많이 도출되었다. 이를 기간을 초기(20.11 ~ 21.01), 중기(21.02 ~ 21.06), 말기(21.07 ~ 21.10)로 나누어 감성 분석을 진행한 결과 초기, 중기에는 모든 언론사가 긍정적인 감성을 가장 많이 서술하였으나 말기에 몇몇 언론사들은 두려움 감성이 가장 많이 도출되었다. Fig 4는 감성분석 결과를 비율로 나타낸 것으로 차례대로 초기, 중기, 말기에 대한 그래프이다.

Fig 4를 통해 국내 언론사들은 초기, 중기에는 백신에 대해 긍정적으로 기사를 작성하였으나 말기부터 서울신문, 세계일보, 조선일보, 중앙일보, 동아일보에서는 백신에 대한 두려움을 표출하는 기사 작성 양이 늘어났다.

해외 언론 같은 경우에는 CNN, TheGuardian, Asahi 언론에 대한 감성분석을 진행하였다. Fig 3은 그 결과를 국내 언론과 마찬가지로 초기, 중기, 말기로 나눈 감성분석 결과 그래프이다.

Fig 5을 통해 CNN, The Guardian에 비해 Asahi는 백신에 대해 긍정적인 기사를 많이 작성하였고 해외 언론 같은 경우에는 기간의 변화에 따라 기사에서 볼 수 있는 감성의 변화가 국내 언론에 비해 상대적으로 미미했다. 국내 언론과 해외 언론을 비교하자면 국내 언론이 백신에 대한 긍정적인 기사 비율이 압도적으로 많고 해외 언론은 감성들이 국내 언론보다 다소 균등하게 이루어져 있음을 알 수 있다.

5.3 CONCOR Analysis

5.3.1 CONCOR Analysis on ‘Travel’ with UCINET (2018.02~2019.02)

2018년 2월부터 2019년 12월까지 ‘여행’ 키워드로 진행한 CONCOR분석의 결과는 Fig 6와 같다. Fig 6에는 총 4개의 군집으로 클러스터링 된 것을 확인할 수 있다. 본 연구에서 Cluster Depth와 Scrunch Factor 모두 4로 설정했다. 대표적으로 ‘여행’ 키워드의 클러스터는 ‘호텔’, ‘일본’, ‘드라이브’,

‘생일’ 노드 등을 확인할 수 있다. 또한 제주도 노드로 묶인 클러스터에서도 ‘탐라’, ‘공항’, ‘맛집’ 등의 노드들도 확인된다.

5.3.2 CONCOR Analysis on ‘Travel’ with UCINET (2021.01~2021.10)

2020년 1월부터 2021년 10월까지 ‘여행’ 키워드로 진행한 CONCOR 분석의 결과는 Fig. 6와 같다. Fig. 6와 다르게 Fig. 7에는 총 3개의 군집으로 클러스터링 된 것이 확인된다.

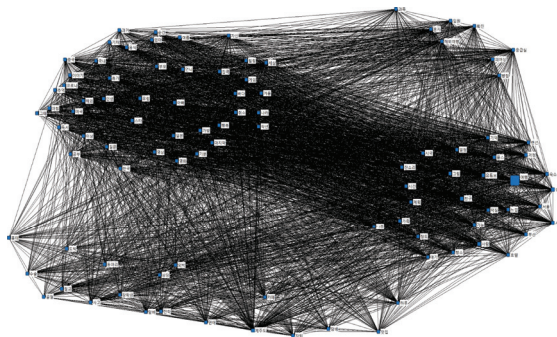


Fig. 7. CONCOR Analysis by the keyword ‘Travel’ After COVID19 period.

본 연구에서는 Cluster Depth는 4, Scrunch는 6으로 설정하여 결과를 도출했다.

5.3.1에서 진행된 코로나19 전 기간과는 다르게 본 분석에서 여행 클러스터에서 ‘여행’ 키워드 주변으로 새롭게 ‘렌선’, ‘플스’, ‘유튜브’, ‘그림’, ‘덕질’ 등의 노드들이 관찰되었다. 이를 통해 코로나 19 기간동안 SNS에서 ‘여행’에 관련한 노드들을 대신해 이런 트렌드의 변화가 있음을 알 수 있다. Fig 5하단에 클러스터링 되지 않은 노드 중 ‘제주도’ 노드가 확인되었는데, 6.3.1의 분석과는 다르게 ‘제주도’ 클러스터가 존재하지 않는 것을 보아 코로나 19 전 기간 보다 후 기간에 제주도 여행에 대한 관심이 줄어들었음을 예상할 수 있다.

6. Conclusion

코로나19에 대한 국내 백신 완전 접종률은 80%가 넘어 가는데 비해 코로나19의 확진자 수는 역대 최고치를 달성하고 있다. 단순히 백신 개발로 수그러 들 줄 알았던

팬데믹은 길어져만 가고 있고, 나아가 코로나19가 우리의 생활패턴을 바꾸고 있다. 따라서 본 논문은 코로나19 관련 키워드들로 구성된 언론과 SNS 데이터를 활용해 숨겨진 주제와 국내여론을 파악하였다. 또한, 언론에서 작성하는 기사들을 통해 백신에 대한 감성을 추출하고 언론사마다 기사에 담아내는 감성을 알고 그에 맞춰서 백신에 대한 정책을 실현하도록 정보를 제공한다. 마지막으로 코로나19의 유행 전과 비교해서 코로나19가 국내 트렌드에 어떠한 변화를 주었는지 분석해보았다. 본 논문의 결과로 사용하여 코로나19 관련 빅데이터 플랫폼을 생성하고 전국민이 트렌드를 공유할 수 있는 효과가 기대된다. 향후 연구에서는 ‘재난지원금’ 등 코로나19와 관련한 정부 정책 키워드에 관련하여 분석을 시행하거나, 해외 SNS 데이터를 사용하여 국내외 비교분석을 진행할 수 있을 것이다.

참고문헌

1. J. Ki, “Application of Sentiment Analysis and Topic Modeling on Rural Solar PV Issues : Comparison of News Articles and Blog Posts”, The Society of Digital Policy & Management, pp. 17-27, (2020).
2. K. Park and G. Choi, “A Study on Technology Trend of Power Semiconductor Packaging using Topic model”, J. Microelectron. Packag. Soc., 27(2), 53-58 (2020).
3. D. Nam and G. Choi, “Technology Trend Analysis in the Automotive Semiconductor Industry using Topic Model and Patent Analysis”, Journal of Korea Technology Innovation Society, 21(3), 1155-1178 (2018).
4. Y. Suh, K. Koh, and J. Lee, Journal of the Korea Institute of Information and Communication Engineering 25(5), 734-738(4 pages), (2021).
5. Y. Sung, K. Kim, and H. Kwon, “Big Data Analysis of Korean Travelers’ Behavior in the Post-COVID-19 Era,” Sustainability, 13, 310 (2021).
6. H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao, “Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,” Multimedia Tools Appl 78, 1 5169–15211 (2019).

접수일: 2022년 6월 9일, 심사일: 2022년 6월 20일,
게재확정일: 2022년 6월 22일