

텍스트 마이닝을 활용한 매스 미디어와 소셜 미디어 의제 분석 : '마스크 5부제'를 중심으로

Mass Media and Social Media Agenda Analysis Using Text Mining : focused on '5-day Rotation Mask Distribution System'

이새미, 유승의, 안순재
동아대학교 스마트 거버넌스 연구센터

Sae-Mi Lee(emailme6@naver.com), Seung-Eui Ryu(juim0928@dau.ac.kr),
Soonjae Ahn(asj8652@gmail.com)

요약

본 연구는 코로나19 사태로 인하여 최근 이슈로 떠오르는 '마스크 5부제'에 대한 온라인 뉴스 기사와 카페글을 분석하여 언론과 대중들의 반응을 담고 있는 매스 미디어와 소셜 미디어 의제를 파악하고, 그 차이점을 알아보았다. 분석을 위해 네이버 뉴스 기사 전문 2,096건과 카페글 1,840건을 수집하고 데이터 전처리 과정과 정제과정을 거쳐 단어 빈도분석, 워드 클라우드, LDA 토픽모델링 분석을 실시하였다. 분석 결과, 매스 미디어에 비해 소셜 미디어는 '대리 구매', '개학 연기', '마스크 사용', '마스크 구입'과 같이 실생활 관련 토픽이 나타나 개인 미디어의 특성이 반영되어 정보 전달의 기능 보다는 개인의 의견, 감정, 정보를 교류하는 역할을 하는 것으로 나타났다. 본 연구에 적용된 연구방법의 적용으로 다양한 미디어 분석을 통해 사회이슈가 공중의제화 되고, 정부의제로 진화하는 정책의제설정 과정에서 참고자료로 활용될 수 있을 것이다.

■ 중심어 : | 마스크 5부제 | 매스 미디어 | 소셜 미디어 | 텍스트 마이닝 | 토픽모델링 |

Abstract

This study analyzes online news articles and cafe articles on the '5-day Rotation Mask Distribution System', which is emerging as a recent issue due to the COVID-19 incident, to identify the mass media and social media agendas containing media and public reactions. This study figured out the difference between mass media and social media. For analysis, we collected 2,096 full text articles from Naver and 1,840 posts from Naver Cafe, and conducted word frequency analysis, word cloud, and LDA topic modeling analysis through data preprocessing and refinement. As a result of analysis, social media showed real-life topics such as 'family members' purchase', 'the postponement of school opening', 'mask usage', and 'mask purchase', reflecting the characteristics of personal media. Social media was found to play a role of exchanging personal opinions, emotions, and information rather than delivering information. With the application of the research method applied to this study, social issues can be publicized through various media analysis and used as a reference in the process of establishing a policy agenda that evolves into a government agenda.

■ keyword : | 5-day Rotation Mask Distribution System | Mass Media | Social Media | Text Mining | Topic Modeling |

* 이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018S1A3A2075240)

접수일자 : 2020년 05월 06일

심사완료일 : 2020년 06월 05일

수정일자 : 2020년 06월 05일

교신저자 : 이새미, e-mail : emailme6@naver.com

I. 서론

디지털 미디어 발전에 따른 인공지능 기술의 활용 확대와 더불어 개인화 서비스 기반의 지능적, 감성적 웹이라고 일컬어지는 웹 5.0 시대의 도래로 소셜 미디어 서비스의 발달과 진화가 지속되고 있다. 이로 인해 전통적인 여론형성의 기능이 매스 미디어에서 소셜 미디어로 옮겨가고 있다. 우리나라에서도 SNS 전체 이용률은 해마다 꾸준히 증가하고 있다[1]. 비즈니스적 관점에서 도 소셜 미디어에서 발생하는 소셜 네트워킹 서비스(SNS, Social Networking Service)는 현재 전 세계 소비자들 사이에서 가장 인기 있는 온라인 활동이며 오늘날 소비자의 온라인 참여 및 참여를 위한 채널로서 점점 중요해지고 있다[2]. 이러한 현상은 계속 확대되고 있어 비즈니스 환경뿐만 아니라 정부 정책에 대한 소셜 미디어 기반의 여론이 관심을 받고 있다.

정책의제설정 과정에는 다양한 이해관계자들이 존재함에 따라 이러한 이해관계자들의 각기 다른 의견을 효과적으로 수렴하고 정책에 반영하기 위하여 대중의 여론 즉 공중의제를 파악하는 것은 필수적이다. 사회적 이슈가 여론에 의해서 정책에 반영되는 과정을 살펴보면 언론에 보도된 이슈가 대중의 관심을 이끌어내고, 이러한 대중의 관심은 미디어 의제화 되어 공중의제로 발전된다. 여기서 대중의 관심이 이슈에 계속 집중되면 그 이슈는 공중의제에서 정부의제로 한 단계 더 진화한다[3]. 이 때 사회적 이슈에 대한 대중의 관심은 소셜 미디어를 활용하여 참여자들의 의견을 분석함으로써 파악할 수 있다. 새로운 정책을 수립하는 경우와 정책에 대한 평가 및 모니터링의 과정에서 여론 파악을 위한 소셜 미디어의 활용 가능성을 탐색할 필요가 있다. 소셜 미디어는 대중의 의견수렴 및 발산이 가능한 도구이므로 정책의제설정, 정책집행, 정책평가 등의 정책과정 단계에서 활용이 가능하다[4][5]. 정책의제설정에서는 소셜 미디어 상에서 다양하게 제시되는 의견의 수집 및 분석을 통해 의제발굴에 반영될 수 있다[6]

본 연구는 마스크 5부제에 대한 온라인 뉴스와 소셜 미디어 텍스트의 실증 분석을 통해 두 매체 간 의제 차이를 비교하고자 한다. 전통적 미디어인 뉴스와 인터넷 발달에 따른 온라인 사회참여로 활성화된 소셜 미디어

를 통한 공중의 의제설정을 살펴본다. 기존의 연구에서는 전통적 미디어인 언론 즉, 매스 미디어만을 대상으로 미디어의제를 파악하였으며 정책의제설정의 관점에서 매스 미디어와 소셜 미디어 역할의 차이를 살펴본 연구는 찾아보기 힘들다[7][8]. 한편, 본 연구는 이러한 기존의 매스 미디어 의제뿐만 아니라 더 나아가 디지털 트랜스포메이션 시대라는 사회적 현상에 따라 온라인 상에서의 활발해진 공중의 사회참여를 기반으로 새로운 의제설정의 기능을 가지게 된 소셜 미디어 의제도 분석하였다.

마스크 5부제에 대한 공중의제 파악을 위해 소셜 미디어 텍스트 분석을 통해 코로나19 바이러스 사태에서 마스크 5부제 관련 키워드를 선정하고, 소셜 미디어로써 네이버와 다음 카페를 대상으로 선정한 키워드가 포함된 문서를 수집하였다. 수집된 문서의 데이터 전처리 과정을 거쳐 주요 키워드를 알아보고 이를 기반으로 문서의 주제를 분류하여 마스크 5부제에 대한 공중의제를 알아보았다.

II. 이론적 배경

1. 매스 미디어와 소셜 미디어 의제

매스 미디어는 사회 전반에서 발생하는 다양한 이슈들을 반영하고 있으며, 대중들은 매스 미디어를 통해 이슈를 접하고 이해한다[7]. 이러한 매스 미디어는 신문, TV, 라디오 등의 주요 매체를 포함하며 정부에 대한 감시자의 기능을 하고 있으므로 객관적인 사실을 보도하고 정보를 전달하는 기능에 초점을 두며 정책의제설정 과정에서 주요 역할을 수행하고 있다[9].

한편, 매스 미디어가 사회적 커뮤니케이션의 주요 채널을 점유하고 일방적인 정보 제공자의 역할을 수행하는 전통적 미디어 환경으로부터 다양한 채널을 통해 정보와 의견을 전달 및 교환할 수 있는 온라인 미디어 환경으로 변모하면서, 새로운 온라인 기반의 소셜 미디어가 의제설정의 기능을 수행하고 있음이 연구되고 있다[10-12].

페이스북, 트위터, 온라인 커뮤니티 등과 같은 소셜 미디어는 참여자들이 자유롭게 의견을 교류하는 실시

간 상호작용이 이루지는 플랫폼의 역할을 통해 여론을 형성하는 공론장의 기능을 수행하고 있으며, 그 영향력은 점차 커져가고 있다. 온라인 네트워크를 통해 새로운 의제가 생성되기도 하고 기존 이슈가 확대 재생산되기도 한다 이러한 과정에서 거대 담론이 형성되거나 [13] 사회운동이 조직화된다[14]. 즉 소셜 미디어의 의제 형성력이나 파급력은 전통 미디어에 비견하거나 이를 뛰어넘는 것으로 보인다.

소셜 미디어 상에서의 공중의제는 특정 시점에 형성된 순간적 여론이기 때문에 시간에 따른 여론의 연속성 파악이 어렵고 여론조사와 달리 대표성을 보장할 수 없다[15]. 또한, 인터넷을 통한 여론형성이나 의견교환은 내용의 사실 관계 확인이 없이 급속도로 파급되는 특성으로 인해 마녀사냥이나 사이버테러와 같은 프라이버시 침해의 경우를 종종 야기하기도 한다[16].

그러나 이러한 소셜 미디어의 부정적 특성 외에도 소셜 미디어는 온라인 사회참여의 대표적 수단으로서의 순기능을 하고 있다. 소셜 미디어에 드러나는 주요 이슈는 일반 공중의 의견을 반영할 수 있다는 점에서 중요한 의미를 지닌다. 소셜 미디어를 통한 여론조사에서는 감지할 수 없는 개개인의 의견을 실시간으로 알아볼 수 있으며, 소셜 미디어에 상에 나타난 개인들의 의견은 자발적, 쌍방향적, 비개입적, 비반작용적인 특성이 있다[17]. 이에 따라 소셜 미디어는 간단한 메시지로 개인의 의사를 자유롭게 표현할 수 있어 보다 일상적인 내용이 중요한 가치로 작용하며 개인적 관심사 위주의 정보를 선택하고 공유하려는 경향이 있다. 반면, 매스 미디어는 객관적인 사실을 보도하고 정보를 전달하는 기능에 초점을 맞추고 있다[7].

정치 참여적 관점에서 보면 투표를 전형적인 시민참여의 형태로 간주하는 것은 시민들이 미디어를 사용하여 사회문제에 대해 보다 신속하고 쉽게 감정을 표현하고 참여하는 현실을 제대로 보지 못하는 진부한 것이다 [18]. 스마트 미디어의 등장과 소셜 미디어의 대중화는 대중들에게 새로운 속도와 도달 범위로 사회 현안과 관련된 정보를 제공하며 이를 바탕으로 사회참여를 확대·고취하고 있다. 대중들은 소셜 미디어를 통해 국내외 주요 정치·사회 이슈에 개입하는 모니터 시민(monitorial citizen)의 역할을 수행한다[19]. 또한 소

셜 미디어의 확산은 대중의 참여를 확대시켜 여론의 폭발현상을 야기하였다[20]. 한편, 대중의 적극적 참여가 늘면서 이러한 대중들의 역할이 확대됨에 따라 소셜 미디어에서 생성된 데이터가 언론의 의제설정(agenda-setting)에 반영될 수 있으며 이는 역의제설정(reversed agenda-setting)으로 설명되기도 한다[21].

미디어 의제에 대한 연구로 홍유정·황주성[7]은 광주 인화학교사건에 대하여 전통 미디어와 소셜 미디어가 정책의제설정 과정에서 어떠한 차별적 역할을 수행하는지를 분석하기 위해 신문과 TV, 트위터를 대상으로 하여 두 미디어 간의 차이를 분석하였다. 강성남[20]은 소셜 미디어의 확산으로 인한 정책 환경의 변화가 기존의 정책 거버넌스에 변동을 가져오고 있음을 지적하고 어떤 의미변화를 초래하고 있는지 알아보았으며, 사례 분석을 통해 정책의제설정이 집단 주도적인 성격에서 개인 주도적으로 바뀌고 있음을 파악하였다.

2. Co-creation 기반의 온라인 사회참여

Co-creation은 경영학에서 출발한 개념으로서 공급자만에 의하여 제품이나 서비스의 가치가 창출되는 것이 아닌, 소비자도 가치를 함께 만들어 가는 즉, 공동가치의 창출을 의미한다[22]. 다양한 이해관계자들의 직접적인 참여와 협력을 가능하게 하는 Co-creation 개념을 토대로 현 시대의 온라인 사회참여 현상을 설명할 수 있다. Co-creation은 독립성과 자율성을 가진 시민 및 사회 조직 간의 네트워크를 강조하는 것으로서 현존하는 거버넌스 개념에서 나아가 시민과 조직 등 참여자 모두의 의지, 참여, 책임, 협력을 더욱 중요시하는 개념이다[23]. 특히, 정보기술의 발전과 더불어 Co-creation 개념을 토대로 과거에 비해 더욱 활발해진 온라인 참여는 실사용자 또는 수요자의 직접적이고 실질적인 의견을 바탕으로 특성을 파악할 수 있다.

Co-creation은 인터넷, 스마트 기기, SNS 등 정보기술을 기반으로 한 사회 구성원들의 적극적 참여를 통해 참여자 중심의 사회문제 해결과 가치 창출을 도모한다는 면에서 실용적이고 효과적이다. 대중들은 단순한 민원 제기에서 벗어나 생활에 직접적으로 영향을 미치는 소규모의 정책 제안을 아주 활발히 하고 있고 토론도 활성화 되고 있다. 정책의제가 형성되기 전 사회

전반의 이슈를 공중의제로 발전시키는 것이 중요하며, 대중의 사회참여를 통해 사회문제가 쟁점화 되고 쟁점화 된 사회문제는 다시 공론화 되어 정책의제화 된다.

이러한 Co-creation 개념을 바탕으로 시·공간의 제약이 없는 온라인 사회참여 현상이 증가하고 있다. 그중 소셜 미디어를 통한 대중의 사회참여는 스마트폰 사용자와 SNS 가입자 수의 증가와 더불어 사회적 의제의 선택과 트렌드 발생 방식의 변화를 가져왔다. 소셜 미디어는 기존 매스 미디어를 통해 이루어지던 단방향 의제설정에서 벗어나 실시간 소통이 가능한 새로운 개인 미디어로서의 영향력을 가지고 있다[24]. 또한 소셜 미디어의 활성화로 인한 정책 환경의 변화를 통해 소셜 미디어 사용자들의 의제 제안이 늘어나고 개인의 문제가 정책의제로 연결되기도 한다[20].

3. 텍스트 마이닝을 통한 미디어 의제 분석

텍스트 마이닝은 비정형 데이터의 한 종류인 언어로 작성된 문서에서 유의미한 정보를 추출해 내는 새로운 텍스트 분석 방법을 뜻한다. 텍스트 마이닝에서는 단어를 기본 분석 단위로 정하여 문서 내 단어의 출현 빈도, 단어들 사이의 문서 내 동시 출현 확률 등을 계산하여 정보를 파악한다. 텍스트 마이닝은 온·오프라인 상에서 의견을 나타내고 정보를 교환하는 대표적인 방식인 텍스트를 직접적으로 분석한다는 점에서 이에 대한 관심과 활용이 늘어나고 있다. 최근 인터넷과 모바일 기기의 발달과 확산으로 인해 대중이 직접 참여하여 자신의 의견이나 정보를 온라인상에 표현하는 일이 매우 쉽고 빈번해져, 블로그, SNS, 온라인 뉴스, 온라인 게시판, 인터넷 커뮤니티 등에서 발생하는 텍스트 데이터의 양이 엄청나게 증가하고 있다. 이러한 비정형 텍스트 데이터를 분석하여 기존의 데이터 마이닝에서는 다루지 못하였던 사회 여러 현상을 설명하기 위한 시도가 활발하게 이루어지고 있다. 텍스트 마이닝을 활용한 미디어 의제 분석에 대한 연구로 정지원 외[25]는 신문기사를 수집하여 장애인 노동의제를 분석하였으며, 이종혁·길우영[11]은 토픽모델링 방법을 활용하여 대통령 신년 기자회견 관련 뉴스기사를 분석하고 의제 다양성과 미디어 다양성에 대한 연구를 수행하였다.

본 연구는 Co-creation 기반 온라인 사회참여 현상

과 매스 미디어와 소셜 미디어의 특성을 이론적 근거로 하여 '매스 미디어와 소셜 미디어 의제에 차이가 있을 것이다'라는 연구문제를 설정하였다. 따라서 텍스트 마이닝을 활용하여 마스크 5부제에 대한 매스 미디어와 소셜 미디어 의제 분석을 위해 온라인 뉴스 기사와 카페글의 주요 토픽은 무엇인지 알아보고 그 차이를 살펴본다.

III. 연구 방법

본 연구에서 활용하는 매스 미디어와 소셜 미디어는 텍스트로 이루어진 비정형 데이터이므로 이를 분석하기 위해서는 텍스트 마이닝을 이용하여 유용한 의미를 추출하여야 한다. 본 연구에서는 홍순구 외[3]가 연구한 정책의제설정 자동화 단계를 수정 및 보완하여, 이슈 제기 및 촉발사건 탐지, 데이터 수집 및 정제/전처리, 빅데이터 분석 및 시각화의 단계를 거쳐 매스 미디어와 소셜 미디어 의제를 분석하였다[그림 1].

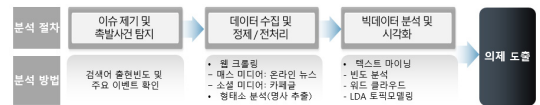


그림 1. 분석절차 및 방법

1. 이슈 제기 및 촉발사건 탐지

의제분석의 첫 번째 절차로는 사회 이슈를 알아보는 것이다. 사회 이슈를 파악하기 위해서 매스 미디어와 소셜 미디어의 검색어 출현 빈도의 변화를 확인하여 의제 발굴을 효율적으로 수행할 수 있다. '코로나'와 '마스크'를 키워드로 하여 포털사이트 다음(daum)에서 운영하는 썬트렌드를 사용하여 키워드 출현 추세를 파악하고, 검색량이 많아진 시점의 주요 이벤트를 정리하였다 [그림 2]. 다음 썬트렌드는 두 가지 키워드의 비교분석이 가능하고 뉴스, 트위터, 블로그, 인스타그램을 포함한 미디어에서 언급량 추이를 비교함으로써 사회적 이슈를 알아보기에 적합하다. 검색량을 살펴보면 국내 최초 확진자 발생 시점에서부터 신규 확진자가 급증한 이후 검색량이 대폭 늘어난 것을 알 수 있다. 이는 확진자의 수가 갑자기 증가함에 따라 마스크 구입에 대한 사람들의 불안감이 높아져 마스크 5부제 시행 전까지 검

색량이 높은 것으로 판단된다. 반면 마스크 5부제 시행 후에는 마스크 구입에 대한 혼란이 감소하여 검색량도 낮아졌으며, 국내 사망자가 증가한 시점에 다시 검색량이 높아졌다. 이러한 검색량의 변화에서 보듯이 코로나 19 사태로 인한 마스크에 대한 사람들의 관심이 급속도로 증가하여 사회 이슈화 된 것으로 판단할 수 있다.

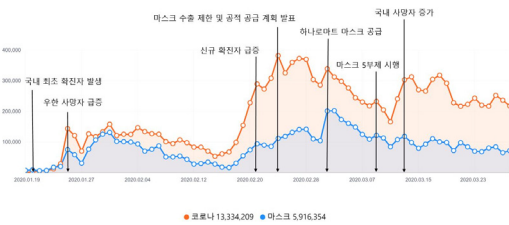


그림 2. 검색어 출현빈도 및 주요 이벤트

2. 데이터 수집

앞 단계에서 확인한 이슈에 대한 구체적 사건을 탐지하기 위하여 자료 수집의 대상을 선정해야 한다[3]. 본 연구는 매스 미디어와 소셜 미디어의 의제설정 경향을 파악·비교하기 위하여 국내 가장 규모가 큰 포털사이트인 네이버(Naver)를 대상으로 마스크 5부제 시행일인 2020년 3월 9일부터 24일까지 약 2주 간의 뉴스 2096건과 카페글 1840건을 수집하여 분석에 활용하였다. 소셜 미디어 분석을 위해 당초 네이버와 다음카페를 대상으로 데이터를 수집하였으나 다음카페에서 수집한 데이터를 살펴본 결과 데이터의 양이 네이버 카페에 비해 현저히 적고 관련성이 낮은 데이터가 많아 네이버 카페글만을 대상으로 분석하였다. 또한 카페글 외에 트위터와 페이스북의 데이터도 수집하였으나 마스크 광고글의 수가 매우 많아 본 연구에서 알아보고자 하는 마스크 5부제 관련 글이 적어 분석에서 제외하였다.

데이터 수집 방법은 웹 크롤링(web crawling) 기법을 사용하여 웹 사이트에서 뉴스 기사와 카페글을 자동으로 수집하여 데이터를 추출하였다. 본 연구의 목적인 마스크 5부제와 연관되는 뉴스와 카페글을 수집하기 위하여 검색 키워드를 '코로나', '공적마스크'로 선정하고, '코로나' or '우한' and '공적마스크'를 검색식으로 하여 네이버뉴스와 카페글을 검색하여 수집하였다. 수집 후 중복되는 뉴스와 카페글을 제거하고 마스크 5부

제와 관련성이 없는 데이터는 삭제하였다.

3. 빅데이터 분석

3.1 데이터 전처리

텍스트, 영상, 음성과 같은 비정형 데이터는 분석을 위해 계량화하는 과정이 필요한데, 본 연구에서는 비정형 데이터의 정형화를 위하여 텍스트 데이터 처리 소프트웨어인 Netminer 4를 이용하여 텍스트 분석을 수행하였다. 먼저, 수집된 뉴스 기사와 카페글의 분석을 위해 전처리 과정을 거쳐야 한다. 전처리 과정은 텍스트 데이터를 분석에 적합한 형태로 정제하는 과정으로, 텍스트 마이닝 중 가장 많은 시간이 소요된다. 수집된 데이터에서 형태소 분석을 통해 명사를 추출한 후 결과를 확인하고, 데이터를 분석에 용이하도록 정제하기 위하여 전처리 과정을 반복 수행하였다. 이러한 전처리 과정을 거치지 않은 데이터는 띄어쓰기, 유의어 처리의 문제로 인해 데이터 분석 결과에 영향을 미칠 수 있다. 첫째, 넷마이너 소프트웨어의 사용자 사전 기능을 활용하여 '앱'과 '어플', '코로나19', '코로나', '우한폐렴'과 같은 단어는 동의어 처리하였으며, '공적 마스크'와 '공적 마스크', '마스크 5부제'와 '마스크5부제' 등 같은 단어지만 띄어쓰기의 차이가 있는 단어는 모두 한 단어로 정리하였다. 둘째, 언론사명, 기사명 등 분석에 불필요한 단어도 제외어 처리하였다. 마지막으로 본 연구의 특성에 비추어 볼 때 특정 단어인 '코로나19'와 '공적 마스크'는 모든 문서가 공통으로 포함하고 있으므로 토픽을 추출하는 데 차별적 정보를 제공하지 않는다고 판단하여 제거하였다. 이와 같은 전처리 과정을 마친 후, 추출된 명사에 대한 단어 빈도(Term Frequency)분석을 실시하였다. 빈도분석 후 워드 클라우드로 결과를 시각화하였으며, 단어 빈도를 기반으로 LDA(Latent Dirichlet Allocation) 토픽모델링 분석을 수행하였다.

3.2 토픽모델링

텍스트 마이닝은 언어로 작성된 많은 양의 문서에서 유의미한 정보를 추출하는 방법이다. 토픽모델링은 텍스트 마이닝 기법 중 하나로 단어의 동시출현 정보를 바탕으로 텍스트 집합을 가장 잘 표현하는 주제를 추출하여 복잡한 문서를 토픽별로 파악할 수 있는 방법이다

[26]. 다시 말해, 토픽모델링은 문서를 구성하고 있는 단어들로부터 의미를 추출하여 문서의 주제를 자동으로 파악하는 방법을 의미한다[27]. 토픽모델링 중 대표적인 알고리즘인 LDA 토픽모델링을 적용하여 분석을 수행하였다. LDA 토픽모델링은 각 문서에서 어떤 주제가 존재하는 지를 확률적으로 나타내는 확률통계의 일종이다[28]. 하나의 문서에 다종의 주제가 존재하며 주제는 각 단어의 집합으로 나타낼 수 있다. Jacobi et al.[29]은 LDA 토픽모델링을 활용한 대용량 언론 텍스트 분석을 통해 LDA가 뉴스 콘텐츠 트렌드와 패턴을 비교적 신속하게 분석할 수 있는 유용한 도구임을 보여 주었다.

본 연구에서는 온라인 뉴스 기사와 카페글을 수집하고 LDA 토픽모델링 알고리즘을 적용하여 뉴스와 카페글에 나타나는 단어의 분포를 기반으로 뉴스와 카페글의 주제가 무엇인지를 파악하고, 이를 주제별로 분류하였다.

IV. 연구 결과

1. 단어 빈도분석

매스 미디어와 소셜 미디어 의제 분석을 위하여 네이버 뉴스 기사와 카페글을 수집하여 분석에 활용하였다. 전처리 과정을 거치고 형태소 분석을 통해 추출된 명사로 전체 문서에서의 단어 빈도를 알아보았다. [표 1]은 빈도수 상위 30개의 단어와 빈도수, 비율을 나타낸 것이다. 뉴스 기사의 형태소분석을 통해 추출된 명사 17544의 총 빈도수는 500568이며 비율은 총 빈도수 대비 명사의 빈도수를 %로 나타낸 것이다. 카페글은 뉴스 기사보다 데이터의 양이 적어 추출된 명사는 11862개였으며, 총 빈도수는 129199였다.

[그림 3][그림 4]는 뉴스와 카페글의 상위100 단어의 빈도를 바탕으로 이를 시각화 한 것이다. 워드 클라우드드는 단어의 빈도 수에 따라 문자의 크기를 결정함으로써 텍스트에서 해당 키워드의 빈도를 직관적으로 보여 줄 수 있도록 한 것이 특징이다. 뉴스 기사와 카페글에서 공통적으로 많이 언급된 단어는 '약국', '구매', '판매'이다. 뉴스 기사 단어 빈도를 살펴보면, '정부', '확진',

'지역', '상황', '공급', '마스크5부제'와 같은 단어의 빈도가 높게 나타났음을 알 수 있다. 반면, 카페글 빈도분석 결과를 보면 '사람', '집', '생각', '시간', '아이'와 같이 뉴스 기사에서 나타난 단어는 정보성 것에 비해 개인적 차원의 단어들이 주로 나타났다.

표 1. 단어 빈도분석

뉴스 기사			카페글		
단어	빈도수	비율 (%)	단어	빈도수	비율 (%)
약국	5,966	1.19	약국	1,907	1.48
구매	4,186	0.84	구매	1,588	1.23
정부	3,938	0.79	판매	984	0.76
판매	3,610	0.72	사람	843	0.65
지원	2,695	0.54	집	708	0.55
확진	2,416	0.48	생각	700	0.54
지역	2,398	0.48	구입	667	0.52
상황	2,384	0.48	시간	644	0.50
공급	2,312	0.46	가능	582	0.45
마스크 5부제	2,300	0.46	정부	573	0.44
서울	2,189	0.44	필요	536	0.41
관련	1,897	0.38	아이	534	0.41
시민	1,887	0.38	일	504	0.39
국민	1,862	0.37	사용	496	0.38
정보	1,854	0.37	정보	493	0.38
확인	1,775	0.35	앱	484	0.37
사람	1,747	0.35	확진	459	0.36
시행	1,738	0.35	상황	453	0.35
주민	1,729	0.35	정도	448	0.35
제공	1,716	0.34	확인	440	0.34
확산	1,684	0.34	가격	435	0.34
감염	1,682	0.34	재고	433	0.34
방역	1,652	0.33	바이러스	433	0.34
업체	1,649	0.33	오늘	419	0.32
바이러스	1,639	0.33	출	418	0.32
경제	1,627	0.33	마스크 5부제	416	0.32
서비스	1,590	0.32	곳	399	0.31
필요	1,545	0.31	필터	374	0.29
시간	1,512	0.30	사태	368	0.28
센터	1,447	0.29	날	361	0.28
총 빈도수	500568		총 빈도수	129199	
총 단어수	17544		총 단어수	11862	



그림 3. 뉴스 기사 워드 클라우드



그림 4. 카페글 워드 클라우드

2. 토픽모델링 결과

마스크 5부제에 대한 매스 미디어 의제를 파악하기 위하여 네이버 뉴스 2096건을 분석하였다. 연구 방법에서 설명한 바와 같이 데이터 수집과 전처리 과정을 거쳐 LDA 토픽모델링 기법으로 뉴스 기사의 주제를 추출하였다[표 2]. 토픽의 수는 연구의 목표를 고려하여 연구자가 해석에 가장 용이한 수로 결정할 수 있다[30]. 뉴스 기사의 토픽 수를 정하기 위하여 5에서 15사이의 수를 대입하여 테스트 후, 토픽의 수를 9개로 지정하고 Zhao et al.[31]과 Lu et al.[32]의 연구를 바탕으로 반복횟수(iteration)=1000, $\alpha=0.1$, $\beta=0.01$ 으로 설정하여 토픽모델링을 수행하였다. 또한 뉴스 기사 토픽과의 비교를 위하여 카페글에서도 9개의 토픽을 추출하였다[표 3].

[표 2][표 3]의 토픽 번호는 토픽의 수를 나타내는 것이며 문서 수는 각 토픽이 속하는 문서의 개수를 나타낸 것이다. 토픽 레이블은 연구자가 각 토픽 내에 속한 키워드들을 검토하여 지정할 수 있다. 토픽 레이블을 정할 시에 주의할 점은 같은 키워드라도 여러 토픽에 속할 수 있고 이 경우 각 토픽별로 키워드의 중요도는 서로 다를 수 있으므로, 토픽 내의 다른 키워드들과의 조합을 통해 토픽의 의미를 정하여야 한다. 키워드 1에서 10은 가장 확률값이 높은 키워드가 상단에 위치하고 있다. 예를 들어 토픽 7은 '정보', '서비스', '제공' 등의 순으로 토픽 7의 주제를 가장 잘 나타내는 단어라고 할 수 있다.

[표 2]의 뉴스 토픽을 살펴보면 9개의 토픽 중 토픽 3: 마스크 지원 토픽이 전체 문서에서 가장 많은 비중을 차지하고 있음을 알 수 있다. 토픽 3에 이어서 토픽 5: 마스크 5부제에 관련 내용이 많았으며, 다음으로 토

픽 7: 마스크 재고정보 순으로 나타났다. 토픽 1: 감염, 토픽 2: 마스크 공급, 토픽 9: 긴급재난대책은 비슷한 비율로 분포되어 있다. 뉴스 기사 토픽의 특징으로 추출된 9개 토픽은 마스크 5부제 관련 사회 현상을 반영하고 있다. 코로나19 사태에 대한 현재 감염 상황, 사회 운동, 정부 정책, 마스크 유통 과정에서 일어난 마스크 사기와 같은 내용을 통해 마스크 5부제에 대한 정보 전달의 기능을 주로 하고 있는 것으로 나타났다.

표 2. 뉴스 기사 토픽

토픽 번호	1	2	3	4	5	6	7	8	9
문서 수	209	207	416	155	376	168	229	131	205
토픽 레이블	감염	마스크 공급	마스크 지원	마스크 사기	마스크 5부제	사회운동	마스크 재고 정보	정부	긴급재난 대책
키워드 1	확진	업체	지원	판매	약국	사람	정보	국민	지원
키워드 2	감염	정부	지역	경찰	구매	사용	서비스	정부	학교
키워드 3	병원	공급	배부	서울	마스크 5부제	필요	제공	민주당	개혁
키워드 4	환자	생산	센터	신고	판매	손	업	보도	대책
키워드 5	대구	유통	주민	사기	시행	상황	재고	대표	정부
키워드 6	방역	물량	확진 계층	유통	주민	온라인	데이터	경제	추경
키워드 7	집단	지오영	면	수사	출생	정도	현황	정책	방역
키워드 8	센터	필터	제작	관련	시간	사회	약국	정치	장관
키워드 9	지역	가격	진달	관련	약사	바이러스	개발	통합	지급
키워드 10	교회	가짜	시민	구매	연도	운동	판매	후보	상황

표 3. 카페글 토픽

토픽 번호	1	2	3	4	5	6	7	8	9
문서 수	162	86	95	70	319	193	612	198	105
토픽 레이블	마스크 사용	마스크 지원	집단 감염 (교회)	개혁 연기	마스크 구입	마스크 재고 정보	대리 구매	마스크 5부제	마스크 공급
키워드 1	사용	지원	사람	개혁	약국	업	집	구매	정부
키워드 2	필터	지역	교회	회교	시간	정보	아이	마스크 5부제	가격
키워드 3	면	사회	나라	법	줄	재고	애	구입	판매
키워드 4	바이러스	병원	한국	연기	사람	약국	얼마	약국	공급
키워드 5	소속	의료	예배	방법	약사	판매	구입	가능	업체
키워드 6	손	필요	생각	공무원	알	확인	신항	주민	국민
키워드 7	작중	센터	확진	공부	판매	서비스	날	판매	유통
키워드 8	정도	복지	감염	무료	생각	사이트	납편	대리	물량
키워드 9	생각	경제	상황	상당	일	현황	소형	굴	생산
키워드 10	제품	기부	세계	자녀	정도	네이버	가족	자리	부족

[표 3]의 카페글 토픽을 살펴보면 토픽 7: 대리 구매에 대한 내용이 압도적으로 많았으며 토픽 5: 마스크 구입, 토픽 8: 마스크 5부제, 토픽 6: 마스크 재고정보의 순으로 비중이 높았다. 카페글은 앞서 분석한 뉴스 기사에 비해 실생활과 연관되어 있는 내용이 주를 이루었다. 마스크 대리구매, 마스크 구입과 같이 마스크 실구입에 관한 내용이 압도적으로 많았다. 또한 뉴스 기사 토픽에서는 볼 수 없었던 개혁 연기, 마스크 사용에 대한 내용이 나타나 카페글도 정보 전달의 기능을 수행

하고 있으나 1인 미디어라는 소셜 미디어의 성격이 반영되어 개인의 의견, 감정, 정보 교류의 측면이 강화되었다고 볼 수 있다.

V. 결론

본 연구는 코로나19 사태로 인하여 이슈로 떠오른 마스크 5부제에 대한 매스 미디어와 소셜 미디어의 의제를 분석하였다. 두 미디어의 분석을 위해 뉴스 기사와 카페 글을 수집하였고, 텍스트 마이닝을 활용하여 단어 빈도 분석, 워드 클라우드, LDA 토픽모델링 분석을 수행하였다. 텍스트 마이닝 결과를 통해 매스 미디어와 소셜 미디어 의제는 상이한 특성을 지니고 있음을 알 수 있었다.

연구결과를 바탕으로 다음과 같은 시사점을 도출할 수 있다. 대용량 텍스트 문서 분석을 위한 자동화된 의제도출 방법을 적용하여 분석을 보다 효율적이고 객관적으로 수행하였다. 다양한 텍스트 마이닝 방법을 적용하여 연구자의 주관에 배제하고 보다 객관적인 연구 결과를 제공하였다. 본 연구에서 사용된 연구방법의 적용으로 비정형 텍스트 데이터로 구성된 언론과 소셜 미디어의 범위를 확대하여 더욱 다양한 데이터 수집을 통해 소셜 미디어 매체별 의제 분석 등 다양한 분석을 수행할 수 있다.

또한 전통적 미디어인 매스 미디어를 대표하는 뉴스 기사와 소셜 미디어인 카페글의 분석을 통해 언론과 대중이 코로나19 사태로 인한 마스크 5부제에 대하여 어떻게 반응하는지를 살펴보고 그 차이점을 알아봄으로써 매스 미디어와 소셜 미디어를 비교한 연구를 진행하였다. 다양한 미디어 분석을 통해 사회이슈가 공중의제화 되고, 정부의제로 진화하는 정책의제설정 과정에서 참고자료로 활용될 수 있다

향후 연구방향으로 본 연구에서는 토픽을 추출하여 의제를 분석하였으나 여기에서 더 나아가 비중이 높은 토픽의 데이터를 별도로 추출하고 추가적인 토픽모델링 분석이나 네트워크 분석 등을 적용하여 높은 비중의 토픽에 대한 좀 더 세부적 내용을 파악할 수 있을 것이다. 코로나19 사태가 지속되는 시간을 고려하여 충분한 데이터 확보가 가능해지면 시계열 토픽 분석을 통해 주

요 이벤트 시점별로 구간을 나누어 토픽을 분석하여 구간별 특성을 알아보고 그 변화를 파악해봄으로써 다양한 시사점을 도출할 수 있을 것이라 예상된다.

참고 문헌

- [1] 김윤화, "SNS (소셜네트워크서비스) 이용추이 및 이용행태 분석," KISDI Stat Report, 제16권, 제7호, pp.1-9, 2019.
- [2] F. Kujur and S. Singh, "Engaging customers through online participation in social networking sites," Asia Pacific Management Review, Vol.22, No.1, pp.16-24, 2017.
- [3] 홍순구, 유승의, 김나량, 이태현, 이새미, 안순재, *스마트 거버넌스 정책과정의 혁신*, 유원북스, 2020
- [4] 이새미, 홍순구, "특허 및 뉴스 기사 텍스트 마이닝을 활용한 정책의제 제안," 디지털융복합연구, 제18권, 제3호, pp.1-12, 2020.
- [5] 안순재, 유승의, 홍순구, "전이학습을 활용한 비정형 정책데이터 감성분석 모델제안," 한국데이터정보과학회지, 제31권, 제2호, pp.405-414, 2020.
- [6] 정진명, 유기영, 구찬동, "교육정책관련 여론탐색을 위한 소셜 미디어 감정분석 연구," 정보화정책, 제24권, 제4호, pp.3-16, 2017.
- [7] 홍유정, 황주성, "정책의제설정에서 소셜미디어와 매스미디어의 역할에 관한 비교연구: 광주 인화학교사건(도가니) 을 사례로," 방송과 커뮤니케이션, 제16권, 제1호, pp.115-151, 2015.
- [8] 이수련, 최은정, "텍스트 마이닝을 이용한 SNS 와 언론의 이슈에 대한 반응 비교-"한일군사정보보호협정(GSOMIA) 종료" 를 중심으로," 디지털융복합연구, 제18권, 제2호, pp.277-284, 2020.
- [9] 남궁근, *정책학(제3판)*, 법문사, 2017.
- [10] 이동근, "일간신문과 블로그의 '미디어 간 의제설정': '최순실 게이트'사건 보도 비교 분석," 정치정보연구, 제22권, 제2호, pp.53-90, 2019.
- [11] 이종혁, 김우영, "토픽모델링을 이용한 뉴스 의제 분류와 미디어 다양성 분석: 대통령 신년 기자회견 관련 뉴스 분석을 통해," 한국방송학보, 제33권, 제1호, pp.161-196, 2019.
- [12] 최가희, 권상희, "영화뉴스와 소셜구전 (WoM) 의 상

- 관관계 연구: 영화뉴스의 매체 간 의제설정 이론을 중심으로,” 한국방송학보, 제34권, 제1호, pp.289-326, 2020.
- [13] 김정래, 나인섭, “SNS(Social Network Service)와 정책선거,” 인문사회과학연구, 제36권, pp.5-30, 2012.
- [14] 채영길, “네트워크 사회운동과 SNS: Save Jeju Island (SJI) 운동 사례,” 인터넷정보학회논문지, 제15권, 제1호, pp.89-102, 2014.
- [15] 홍주현, “소셜 네트워크 서비스 (SNS: Social Network Service) 상의 담론 분석을 통한 인지적, 정서적 측면의 여론 변화 연구,” 커뮤니케이션학 연구, 제19권, 제3호, pp.5-29, 2011.
- [16] 김성태, 김혜령, “인터넷을 통한 의제파급과 정보탐색에 관한 연구: 포털사이트를 중심으로,” 평화연구, 제19권, 제2호, pp.297-331, 2011.
- [17] 이승희, 송진, “재난보도에 나타난 소셜 미디어와 방송 뉴스의 매체 간 의제설정: 세월호 관련 보도를 중심으로,” 한국언론학보, 제58권, 제6호, pp.7-39, 2014.
- [18] M. E. Grabe and J. G. Myrick, “Informed citizenship in a media-centric way of life,” *Journal of Communication*, Vol.66, No.2, pp.215-235, 2016.
- [19] M. Deuze, “The changing context of news work: Liquid journalism for a monitorial citizenry,” *International journal of Communication*, Vol.2, pp.1-18, 2008.
- [20] 강성남, “소셜 미디어의 확산과 정책 거버넌스의 변동,” *한국사회와 행정연구*, 제24권, 제4호, pp.261-287, 2014.
- [21] 김성태, 이영환, “인터넷을 통한 새로운 의제 설정 모델의 적용: 의제 파급 (Agenda-Rippling) 과 역의제 설정 (Reversed Agenda-Setting) 을 중심으로,” *한국언론학보*, 제50권, 제3호, pp.175-204, 2006.
- [22] C. K. Prahalad and V. Ramaswamy, “Co-creation experiences: The next practice in value creation,” *Journal of interactive marketing*, Vol.18, No.3, pp.5-14, 2004.
- [23] 홍순구, 이현미, 한세익, 김종원, “청년층 일자리 창출문제에서 Co-creation 적 해결방안에 관한 연구: 부산광역시를 중심으로,” *한국산업정보학회논문지*, 제20권, 제1호, pp.91-102, 2015.
- [24] 조희순, 김정연, “소셜 미디어의 매체 특성과 참여의 커뮤니케이션: 반값등록금 관련 블로그와 트위터 내용 분석,” *사이버커뮤니케이션학보*, 제29권, 제2호, pp.95-130, 2012.
- [25] 정지원, 이재민, 최소연, “텍스트마이닝 기법을 통한 언론에서의 장애인 노동의제 분석,” *장애의 재해석*, 제11권, pp.48-100, 2018.
- [26] R. Alghamdi and K. Alfalqi, “A Survey of Topic Modeling in Text Mining,” *International Journal of Advanced Computer Science and Applications*, Vol.6, No.1, pp.147-153, 2015.
- [27] 이새미, 유승의, “온라인 리뷰 빅데이터 분석을 통한 현여울문화마을 관광 활성화 방안 연구,” *호텔리조트 연구*, Vol.19, No.1, pp.115-130, 2020.
- [28] D. M. Blei, “Probabilistic topic models,” *Communications of the ACM*, Vol.55, No.4, pp.77-84, 2012.
- [29] C. Jacobi, W. V. Atteveldt, and K. Welbers, “Quantitative analysis of large amounts of journalistic texts using topic modelling,” *Digital Journalism*, Vol.4, No.1, pp.89-106, 2016.
- [30] 백영민, *R를 이용한 텍스트 마이닝*, 한울아카데미, 2020.
- [31] W. Zhao, J. Chen, and W. Zen, “Best practices in building topic models with LDA for mining regulatory textual documents,” *CDER 9TH NOVEMBER*, 2015.
- [32] Y. Lu, M. Qiaozhu, and Z. ChengXiang, “Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA,” *Information Retrieval*, Vol.14, No.2, pp.178-203, 2011.

저 자 소 개

이 새 미(Sae-Mi Lee)

정회원



- 2017년 2월 : 세종대학교 박사(경영학)
- 2018년 11월 ~ 현재 : 동아대학교 스마트거버넌스연구센터 전임연구원

〈관심분야〉 : 서비스 마케팅, 리더십, ICT, 텍스트마이닝

유 승 의(Seung-Eui Ryu)

정회원



- 2014년 5월 : University of Texas at El Paso 박사(Information Systems)
- 2018년 9월 ~ 현재 : 동아대학교 스마트거버넌스연구센터 전임연구원

〈관심분야〉 : 국제경영학, 빅데이터, 인공지능, 텍스트마이닝

안 순 재(Soonjae Ahn)

정회원



- 2018년 8월 : 연세대학교 박사(공학)
- 2019년 5월 ~ 현재 : 동아대학교 스마트거버넌스연구센터 전임연구원

〈관심분야〉 : 빅데이터, 인공지능, 텍스트마이닝