

텍스트 마이닝과 의미 네트워크 분석을 활용한 뉴스 의제 분석: 코로나 19 관련 감정을 중심으로*

유소연
한양대학교 경영대학
(soyeonyoo@hanyang.ac.kr)

임규건
한양대학교 경영대학
(gglim@hanyang.ac.kr)

전 세계적으로 퍼진 코로나 19 상황은 우리의 일상생활의 많은 부분에 영향을 끼쳤을 뿐만 아니라, 경제·사회 등 많은 부분에 걸쳐 막대한 영향력을 미치고 있다. 확진자와 사망자 수가 증가함에 따라 의료진과 대중은 불안, 우울, 스트레스 등 심리적인 문제를 겪고 있다고 한다. 장기적인 부정적인 감정은 사람들의 면역력을 감소시키고 신체적인 균형을 파괴할 수도 있으므로 코로나 19로 인한 심리적인 상태를 이해하는 것이 필수적인 상황이다. 본 연구에서는 코로나 19 감정과 관련된 뉴스 데이터를 수집하여, 텍스트 마이닝을 통해 키워드를 분류하고, 키워드 사이의 의미 네트워크 분석을 통해 단어들의 관계를 시각화하였다. 코로나 감정과 관련된 기사의 키워드에 나타난 단어들의 빈도수를 확인하고 이를 워드 클라우드 분석하였다. 키워드 빈도 분석 결과 코로나 19 감정과 관련하여 ‘중국’, ‘불안’, ‘상황’, ‘마음’, ‘사회’, ‘건강’과 같은 단어의 빈도가 높게 나타난 것을 확인할 수 있었다. 각 데이터 간 연결 중심성을 분석한 결과 키워드 중심성 네트워크에서 가장 중심적인 핵심어는 ‘심리’와 ‘코로나 19’, ‘블루’, ‘불안’이라는 단어가 높은 연결 중심성을 가지는 것을 확인할 수 있었다. 기사의 헤드라인에 나타난 주요 핵심어 사이의 동시 출현 빈도 네트워크를 그래프로 시각화한 결과, ‘코로나-블루’ 쌍이 가장 굵게 표시되었고, ‘코로나-감정’, ‘코로나-불안’ 쌍이 비교적 굵은 선으로 표시된 것을 알 수 있었다. 코로나와 관련된 ‘블루’는 우울증을 의미하는 단어로, 코로나와 우울증은 이제 관심을 가져야 할 키워드임을 확인할 수 있었다. 본 연구에서는 장기화한 코로나 19 상황에서 신체적인 방역뿐만 아니라 심리적인 방역에도 힘써야 할 이 시기에 보건 정책담당자가 빠르고 복잡한 의사결정 과정에 도움이 되고자 미디어 뉴스를 모니터링 함으로써, 더욱더 쉬운 소셜 미디어 네트워크 분석 방법을 제시하고자 한다.

주제어 : 코로나 19, 텍스트 마이닝, 의미 네트워크 분석, 동시 출현 네트워크

논문접수일 : 2020년 11월 24일 논문수정일 : 2021년 1월 20일 게재확정일 : 2021년 3월 5일
원고유형 : 학술대회용 Fast-Track 교신저자 : 임규건

1. 서론

2019년 12월에 중국에서 발생한 COVID-19 (Coronavirus Disease 2019, 이하 코로나 19)는 신종 호흡기 감염병인 코로나바이러스(SARS-CoV-2)에 의한 감염증이다. 중국, 이탈리아, 스페인, 미국, 러시아 그리고 다른 유럽 국가에서 4,670만 명의

확진자와 120만 명이 넘는 사망자가 발생했다. 우리나라는 2020년 1월 20일에 최초 확진자가 발생한 이후 10월 현재까지 26,635명의 확진자와 466명의 사망자가 발생하였다. 전 세계적으로 코로나 19가 빠르게 확산하자, 역사상 세 번째로 세계보건기구(WHO)는 가장 높은 위험등급에 해당하는 ‘팬데믹’을 선언하였다. 팬데믹은 세계적

* 이 논문은 한양대학교 교내연구지원사업으로 연구되었음(HY-20200000003396).

으로 전염병이 대유행하는 상황을 말한다. 코로나 19의 확산은 우리의 일상생활의 많은 부분에 영향을 끼쳤을 뿐만 아니라, 경제·사회 등 많은 부분에 걸쳐 막대한 영향력 미치고 있다. 우리나라는 역학조사를 통해 확진자의 동선을 파악하고 확진자와 동선이 겹치는 자가격리 대상자를 관리한다. 확진자와 자가 격리자는 코로나 19의 감염의 결과에 대한 두려움을 경험할 수 있으며, 자가 격리 중인 환자는 지루함, 우울함, 외로움 및 분노를 경험 할 수도 있다(Xiang et al, 2020). Liu et al.(2020)은 확진자와 사망자 수가 증가함에 따라 의료진과 대중은 우울, 불안, 스트레스 등 심리적 문제를 겪고 있다고 주장하였다. 전염병을 수반하는 집단 비극은 공포와 불안을 고조시키는 데, 이는 많은 사람의 행동과 심리학적 행복에 막대한 지장을 초래한다고 알려져 있다(Sivasankaran & Aleksandar, 2006). 장기적인 부정적인 감정은 사람들의 면역력을 감소시키고 신체적인 균형을 파괴할 수도 있다. 코로나 19로 인한 잠재적인 심리적 변화를 이해하는 것이 필수적인 상황이다(Li et al., 2020).

역학자들에 의해 코로나 19와 관련된 공중보건 의 감시, 신체 건강과 관련된 데이터는 체계적으로 수집되고 분석되고 있다. 하지만 확진자와 자가 격리자를 비롯한 대다수 국민의 정신 건강 문제와 심리적인 질병에 대한 역학 데이터는 제공되지 않고 있다. 특히 코로나 19에 맞서 싸우는 의료진들은 코로나 19에 대한 감염 위험과 정신 건강 문제에 모두 노출되어 있다. 이들은 전염, 격리, 낙인에 대한 두려움을 경험하고 있지만, 이에 대한 데이터가 부족한 상황이다(Xiang et al, 2020). 코로나 19가 처음 발견된 중국에서 나온 보고서에 따르면 의료진들에게서 우울증, 불안 및 불면증이 높은 유병률을 보였다(Arafa et

al., 2020).

본 연구에서는 코로나 19 상황으로 대중들이 겪는 부정적인 정서를 확인하기 위하여 검색 키워드를 ‘코로나 감정’으로 선정하였다. 코로나 19 우울증과 관련된 핵심 이슈를 파악함과 동시에 세부 이슈들과의 관계성 및 중심성을 그래프로 시각화하였다. 텍스트 마이닝(Text mining)은 비구조화(Unstructured) 또는 반구조화(Semi-structured) 된 텍스트 문서에서 정보를 추출하는 정보기술을 말하며, 토픽 모델링은 텍스트 마이닝에서 주로 사용하는 연구방법론이다(Cha et al., 2015). 텍스트 마이닝 기법을 사용하여 반복적으로 등장하는 용어의 빈도에 기초하여 내재한 주요 동향과 관계에 대해 시각화하고자 한다. 토픽모델링 기법이 소셜 미디어 데이터를 활용하여 의미 있는 결과를 도출하는데 매우 유용한 기법임을 확인할 수 있었다. 토픽모델링으로부터 추출한 토픽 사이의 관계를 분석하기 위해 문서 안에 있는 단어 간 관계를 이용한 의미 네트워크 분석을 이용할 것이다.

의미 네트워크 분석(SNA, Semantic Network Analysis)은 소셜 네트워크 분석의 발전된 형태로, 텍스트나 언어로 구성된 데이터에서 잠재된 의미구조와 맥락을 파악하기 위해 활용되는 기법이다(Hwang and Park, 2017). 토픽모델링과 의미 네트워크 분석을 함께 진행하였을 때의 장점은 쟁점이 되는 토픽 추출이 가능하고, 토픽의 중요도와 영향을 미치는 정도를 파악하고 시각화하는 데 유용하다는 점이다.

따라서 본 연구에서는 장기화한 코로나 19 상황에서 신체적인 방역뿐만 아니라 심리적인 방역에도 힘써야 할 이 시기에 보건 정책담당자가 빠르고 복잡한 의사결정 과정에 도움이 되고자 미디어 뉴스를 모니터링 함으로써, 더욱더 쉬운

소셜 미디어 네트워크 분석 방법을 제시하고자 한다. 빅데이터의 기술을 접목하여 만든 ‘빅카인즈 (Bigkinds)’를 활용하여 쉽게 뉴스 데이터를 수집하는 방법을 소개하였다. 또한 한국어 의미 네트워크 분석 도구인 KrKwic 프로그램을 사용하여 텍스트 마이닝을 수행하고 단어의 빈도수를 계산하여 키워드를 빠르게 파악하였다. 또한 소셜 네트워크 분석 전문 프로그램인 UCINET을 사용하여 연결 중심성 분석 및 군집 분석을 하고, Net Draw를 이용하여 그래프로 시각화하는 방법을 수행하였다. 본 연구에 활용된 연구 방법은 비용 절감과 동시에 사회 현상 및 변화를 빠르게 측정할 수 있다는 장점이 있다. 또한 코로나 19와 관련된 심리방역 문제에 대한 지원, 치료와 서비스 개발에 기초자료로 활용될 수 있을 것이다.

2. 이론적 배경

2.1. 선행 연구

최근 미디어 데이터를 효과적으로 의견을 수렴하고 정책에 반영하기 위해 대중의 관심을 파악하려는 주요한 도구임을 확인하려는 연구가 활발히 진행되고 있다. 신문, 라디오, TV 등의 주요 매체를 포함하고 있는 매스미디어는 사회에서 발생하는 다양한 현상들을 반영하고 있으며, 정책의제를 설정하는 과정에서 매우 중요한 역할을 담당하고 있다고 한다(Lee et al., 2020). 본 연구에서는 정책의제를 설정하는 과정에서 주요한 역할을 담당하고 있는 뉴스를 대상으로 코로나 19와 관련된 사회현상을 분석하고자 다양한 주요 매체 중 뉴스를 선택하여 연구하였다. Park

and Leydesdorff(2004)는 뉴스의 헤드라인은 언론 정보학에서 언론사가 이를 읽는 수용자에게 전달하고자 하는 내용의 핵심을 축약한 것으로 중요하게 처리된다고 주장하였다. 뉴스 데이터를 활용하여 언론을 활용하여 사회적으로 관심이 있는 주요 의제가 무엇이며 변화하는 과정을 파악하는 것과 관련된 연구는 활발히 진행되고 있다. 이러한 선행연구를 바탕으로 본 연구에서는 뉴스 헤드라인을 분석하여 코로나 19 우울증과 관련된 이슈에 대한 사람들의 감정과 인식을 파악하고, 주요 키워드들을 도출하여 분석하고자 하는 주요 의제가 무엇인지 찾아낼 수 있었다. 또한 의미 네트워크 분석 기법을 활용하여 연구 결과를 시각화하는 것이 코로나 19와 관련된 사회현상을 분석하고 유의한 결과를 도출하는데 유용한 도구임을 확인할 수 있었다.

코로나 19의 불확실성과 낮은 예측 가능성은 사람들의 신체 건강을 위협할 뿐만 아니라, 또한 많은 이론과 같이 사람들의 정신건강 특히 감정과 인지 측면에서 영향을 미친다고 한다(Li et al., 2020). 갑작스러운 전염병은 사람들에게 불안, 우울증 및 스트레스를 유발할 수 있기에 전염병 기간 대중의 심리상태를 중요하게 평가하여 정신건강과 관련된 적절한 가이드라인을 제때 제공하는 것이 중요하다고 한다(Wang et al., 2020).

행동 면역 시스템(BIS) 이론에 따르면 사람들은 자기보호를 위해 혐오와 불안과 같은 부정적인 감정을 느낄 때나 부정적인 인지평가에 대해 자체보호를 위해 잠재적인 질병 위협에 직면한 사람들은 회피행동을 하는 경향이 있다고 한다(Terrizzi et al., 2013).

스트레스 이론과 인지된 위협 이론이 주장하는 바는 공중보건 비상사태는 더 많은 부정적인 감정을 유발하고, 인지평가에도 영향을 미친다

(Noriss, 2002; Solvic, 1987).

건강을 위협하는 팬데믹과 같은 상황에서 이를 해결하기 위한 효과적인 방법의 하나는 의사소통을 하는 것이다. 소셜 미디어 게시물에서 특정 위험 인식이 존재하는지와 사용자가 게시물에 어떻게 반응하는지 위험 커뮤니케이션 이론은 설명하고 있다(Jeanine et al., 2020).

따라서 본 연구에서는 장기화로 접어든 코로나 19 상황에서 코로나 19 감정과 관련된 뉴스 빅데이터를 활용하여 언론을 통해 사회에서 발생하는 주요 의제는 무엇이고 어떻게 변화하고 있는지를 텍스트 마이닝 기법과 의미 네트워크 분석을 활용해 확인해보고자 한다.

2.2. 텍스트 마이닝(Text Mining)과 토픽 모델링(Topic Modeling)

텍스트 마이닝은 데이터 마이닝 기법의 하나로, 텍스트를 사용하여 유용한 패턴이나 관계를 추출하고 그 안에서 유의미한 정보나 가치를 추출하여 분석하는 과정을 말한다. 텍스트 마이닝은 키워드 분석을 통해 텍스트 데이터로부터 유의미한 정보를 선별하여 추출할 수 있고, 이를 활용하여 여론이나 사회적 현상, 추이를 분석할 수 있다. 텍스트 분석에 대한 수요가 증가하는 원인을 Kim et al. (2017)은 “텍스트 분석을 통하여 새로운 지식을 창출하고자 하는 수요가 증가하고, 이러한 수요 증가로 인해 텍스트 분석에 관한 관심과 연구가 활발해지는 선순환 구조가 형성된 것”이라고 분석하였다.

최근 모바일과 웹의 확산은 대중이 자신의 정보를 공유하거나 의견을 제시하는 일을 가능하게 하였다. SNS, 블로그, 뉴스, 인터넷 커뮤니티 등에서 발생하는 텍스트의 양이 점점 증가하고

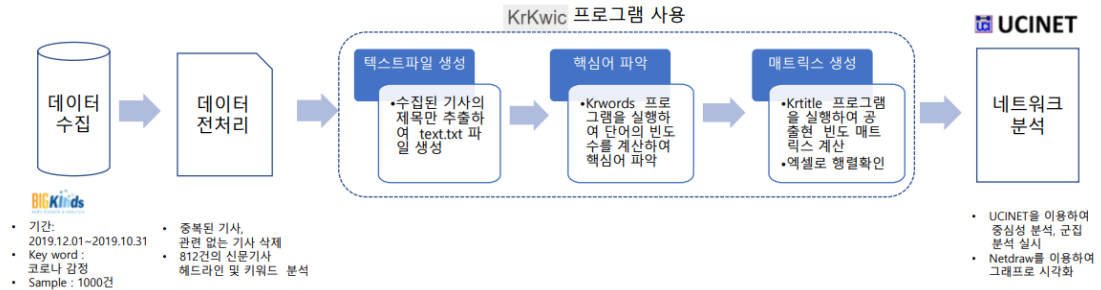
있다. 비정형 데이터인 텍스트를 분석하여 여러 사회현상을 설명하려는 시도가 활발하게 이루어지고 있다(Liu and Lim, 2019; Yu and Lim, 2015; Lee et al. 2018).

텍스트 마이닝을 활용한 뉴스 분석에 대한 연구로 Lee et al. (2020)은 텍스트 마이닝을 활용하여 매스 미디어와 소셜 미디어 의제 분석에 관한 연구를 수행하였다.

토픽모델링(Topic Modeling)은 문서나 문헌 속의 텍스트에서 확률적으로 주제들을 추출하는 텍스트 마이닝(Text mining) 기법의 하나이다(Blei et al., 2003). 토픽모델링은 단어를 근간으로 하여 토픽을 추출하고, 그에 해당하는 텍스트들을 대표하는 특정 이슈나 주제들을 자동으로 추출하는 분석기법으로 뉴스 빅데이터를 가지고 잠재적인 주제나 쟁점을 분석하는 데 유용한 기법이다.

2.3. 의미네트워크분석(SNA, Semantic Network Analysis)

소셜 네트워크 분석(SNA, Social Network Analysis)의 정의를 Sohn (2002)은 “의사소통 집단 내 개체의 상호작용에 관심을 두고, 개체 간 연결 상태의 연결 구조의 특성을 계량적으로 파악하여 시각적으로 표현하는 분석 기법”이라고 하였다. Lee et al. (2017)은 소셜 네트워크 분석은 개인, 지식, 장소, 집단 및 정보 등과 같은 특정 노드(개체)의 관계를 설명하는 이론으로, 노드들과 노드들을 연결하는 링크(link)로 표현되고, 이를 통해 중요한 위치에 있는 개체나 조직들의 도출이 가능하며 전략적으로 중요한 위치를 기반으로 정책적 함의를 할 수 있다고 하였다(Lee et al., 2017; Gim and Koo, 2020; Park and



〈Figure 1〉 Research Model

Lim, 2015). 소셜 네트워크 분석의 발전된 형태인 의미 네트워크 분석(SNA, Semantic Network Analysis)은 의미 네트워크 분석, 텍스트 네트워크 분석, 키워드 네트워크 분석 등으로 명칭이 혼재되어 사용되고 있다(Hwang and Kim, 2019). Diesner & Carley(2005)는 “텍스트 네트워크 분석은 소셜 네트워크 분석기법을 적용하여 각 단어를 노드로 정의하고 단어와 단어의 동시 출현 (co-occurrence)을 단어(노드) 간 관계로 코딩하여 텍스트 내 단어 간의 연결망을 추출할 수 있게 하는 방법”이라고 하였다. 텍스트 네트워크 분석의 핵심은 문서 전체에서 특정한 단어의 상대적인 영향력 크기를 파악하고, 그 단어들의 관계 안에서 특정단어가 하는 역할이 무엇인지를 분석하는 것이라고 설명하였다(Chung et al., 2019; Yoon and Suh, 2019). Chung et al. (2019)는 “전체 문서에서 특정 단어의 상대적 영향력의 크기를 파악하고, 단어 간의 관계 속에서 특정단어가 어떤 역할을 하는지 분석하는 것이 텍스트 네트워크 분석의 핵심”이라고 하였다. 다양한 소셜 네트워크 분석을 통해 전체 문서가 전달하려는 의도를 이해하기 위해서는 네트워크 중심성이 비교적 높게 나타난 개념을 찾아내야 한다. 네트워크를 파악하기 위해 중요한 역할을 하는 중심

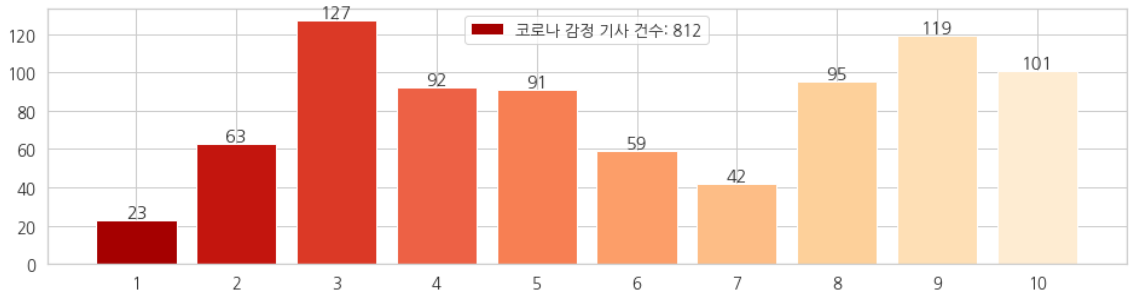
성은 개체가 전체 네트워크 구조상 중심에 위치하는지를 나타낸 지표이다.

본 연구에서는 근접 중심성(Closeness centrality), 매개 중심성(Betweenness centrality), 고유벡터 중심성(Eigenvector centrality), 연결 중심성(Degree centrality)의 네 가지 중심성을 측정하였다. 높은 수준으로 네 가지 중심성이 나타나는지를 확인하고 문헌이나 문서에서 내용이나 의미구조를 형성하는 데 어떤 역할을 하는지 확인하고자 한다.

3. 연구 방법

3.1. 분석개요

본 연구는 코로나 19 감정과 관련된 뉴스의 헤드라인을 분석하여 진행하였다. 분석 절차는 누구나 쉽게 이용할 수 있는 빅카인즈 (BigKinds)로부터 뉴스 데이터를 수집하고, 연구자에 의해 데이터 정제과정을 거친 후, 한국어 의미 네트워크 분석 도구인 KrKwic 프로그램을 사용하여 텍스트 마이닝을 수행하였다. Kwords 프로그램을 실행하여, 기사의 제목에 나타난 단어의 빈도수를 계산하여 핵심어를 쉽게 파악하였다. 소셜 네



(Figure 2) The number of COVID19 Emotion News

트위크 분석 전문 프로그램인 UCINET을 사용하여 연결 중심성 및 군집 분석 등을 실시하고, Net Draw를 이용하여 그래프로 시각화하는 방법을 수행하였다.

3.2. 분석 대상

본 연구의 목적인 코로나 19와 관련된 정신건강 이슈를 분석하기 위해 뉴스 빅데이터 분석 서비스 플랫폼인 빅카인즈 (Bigkinds)를 활용하여 코로나 19가 처음 발생한 2019년 12월부터 2020년 10월까지 보도된 1,000건의 기사를 대상으로 분석하였다. 본 연구에 필요한 핵심 키워드를 얻기 위해 ‘코로나 감정’으로 설정하고 전체 언론사를 대상으로 기사를 수집하였다. 수집된 1,000건의 기사 중 주제에 해당하지 않는 기사와 중복되는 기사를 제외하고 812건의 기사를 선별하여 분석에 활용하였다. 코로나 확진자 수가 급격히 증가한 후인 3월과 9월에 코로나 감정 관련 기사 건수가 증가한 것을 확인할 수 있었다.

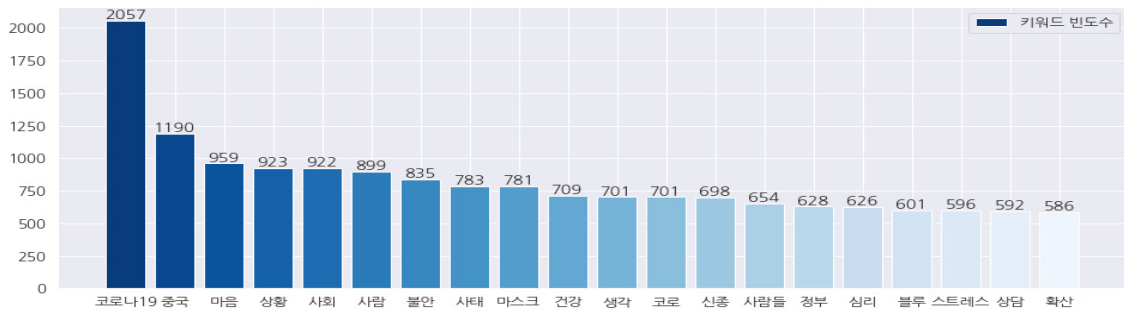
3.3. KrKwic을 이용한 텍스트 마이닝

본 연구에서는 코로나 19와 관련된 언론사별 기사의 특징을 분석하기 위해 의미 네트워크 분

석방법론을 활용하였다. 소셜 네트워크의 발전된 형태인 의미 네트워크 분석은 특정한 단어의 상대적인 영향력 크기를 파악하고, 그 단어들의 관계 안에서 특정단어가 하는 역할이 무엇인지를 분석하는 것이다. 다양한 형태의 메시지에서 핵심어를 파악하려는 목적으로 만들어진 KrKwic 프로그램은 한국어 의미 네트워크 분석 프로그램이다(Park and Leydesdorff, 2004). KrKwic의 분석 절차는 세 단계로 나누어 진행된다. 첫째, 코로나 감정이라는 키워드로 수집된 기사의 제목만 추출하여 text.txt파일을 생성하였다. 이때 주의하여야 할 점은 미국 표준 부호인 다양한 언어를 표현하지 못했던 ASCII의 한계를 표현하기 위해 ASCII의 확장판인 ANIS로 인코딩하여 저장해야 한다. 두 번째 단계로 krwords를 실행하여 wrdfreq.txt파일을 생성하여 빈도수를 계산하였다. wrdfreq.txt 파일 내 단어 중 본 연구 목적에 맞게 텍스트를 정제(cleansing)하여 핵심 키워드 19개를 도출하여 words.txt 파일을 생성하였다. KrKwic 프로그램의 특징은 동의어나 단어의 파생어들을 자동으로 처리하지 못한다. 따라서 연구자의 주관적인 판단이 개입되지 않는 범위에서 데이터의 슬라이싱(Slicing) 및 정제과정을 수행해야 한다. 셋째, krtile을 실행하여 매트릭스를

〈Table 1〉 Frequency of Keywords

Keywords	Frequency	Keywords	Frequency	Keywords	Frequency
코로나 19	2057	사태	783	정부	628
중국	1190	마스크	783	블루	601
마음	959	건강	709	스트레스	596
상황	923	생각	701	상담	592
사회	922	코로	701	확산	586
사람	899	신종	698		
불안	835	사람들	654		



〈Figure 3〉 Frequency of Keywords graph

생성하고, words.txt 파일의 핵심 키워드에 맞추어 엑셀에 표현해 준다. 이 행렬 파일은 textmining.xlsx로 저장한다. 이 데이터를 가지고 소셜 네트워크 분석 프로그램인 UCINET을 활용하여 그래프로 표현하고 매개 중심성(Betweenness centrality), 근접 중심성(Closeness centrality), 연결 중심성(Degree centrality), 고유벡터 중심성(Eigenvector centrality)과 군집 분석을 시행하였다.

4. 분석 결과

4.1. 빈도분석

〈Table 1〉는 코로나 감정과 관련된 기사에 나타난 핵심 단어들의 빈도를 나타낸 것이다. 검색

키워드를 코로나 감정으로 설정하여 기사를 수집하였다. 키워드의 빈도수는 파이썬의 matplotlib을 활용하여 그래프로 시각화하고, 워드 클라우드로 분석하였다. 핵심어의 빈도 순위는 검색어인 코로나와 감정을 제외하고 분석하였다. ‘코로나 19(2,057회)’가 가장 많이 출현하였으며, ‘중국(1,190회)’, ‘마음(959회)’, ‘상황(922회)’, ‘사회(922회)’, ‘사람(899회)’, ‘불안(835회)’, ‘사태(783회)’, ‘마스크(783회)’, ‘건강(709회)’, ‘코로(701회)’, ‘생각(701회)’, ‘신종(698회)’, ‘사람들(654회)’, ‘심리(626회)’, ‘블루(601회)’, ‘스트레스(601회)’, ‘상담(592회)’, ‘확산(568회)’ 순으로 나타났다. 〈Figure 3〉은 코로나 19와 관련된 키워드의 빈도수를 상위 20단어의 빈도수를 가지고 그래프로 시각화한 것이다. 〈Figure 4〉는 뉴스 키워드의 빈도를 채도와 크기에 따라 직관적으로 보

여줄 수 있는 워드 클라우드로 시각화하였다. 워드 클라우드 분석 시 전처리 과정에서 검색어로 설정한 ‘코로나’와 ‘감정’은 제거하고, 의미 없는 ‘코로’라는 단어도 제거해 주었다. 코로나 19 감정과 관련하여 ‘중국’, ‘불안’, ‘상황’, ‘마음’, ‘사회’, ‘건강’과 같은 단어의 빈도가 높게 나타난 것을 확인할 수 있었다.



〈Figure 4〉 COVID 19 Word Cloud

4.2. 중심성분석

Bhat & Milne(2008)은 소셜 네트워크의 개체 및 구조들의 영향력을 파악하고자 중심성 (centrality) 분석이 가장 많이 사용된다고 하였다. 중심성 분석은 네트워크에서 특정 노드가 중심에 위치하는 정도의 지표를 분석하는 것을 말하며, 직접적으로 한 노드에 연결되어 있는 노드의 중심성을 측정하여 연결 정도를 보여주는 분석 방법이다.

앞서 분석한 코로나 19 감정과 밀접한 관련이 있는 단어 19건에 대하여 중심성 분석을 하였다.

<Table 2>는 매개 중심성(Betweenness centrality) 과 연결 중심성(Degree centrality), 근접 중심성 (Closeness centrality) 그리고 고유벡터 중심성 (Eigenvector centrality)의 분석 결과를 제시하고 있다.

각 데이터 간 연결 중심성을 분석한 결과 연결 중심성은 ‘심리(31.000)’, ‘코로나 19(23.000)’, ‘블루(23.000)’, ‘불안(22.000)’, ‘마음(21.000)’ 순으로 높은 연결 중심성을 가지는 것으로 확인할 수 있다. 연결 중심성의 관점에서 볼 때 키워드 중심성 네트워크에서 가장 중심적인 핵심어는 ‘심리’, ‘코로나 19’, ‘블루’, ‘불안’이라는 단어가 높은 연결 중심성을 가지는 것을 확인할 수 있었다.

Hwang and Kim(2019)은 근접 중심성, 고유벡터 중심성, 연결 중심성, 매개 중심성 네 가지의 중심성이 전부 상위에 있다는 것은 문서나 문헌에서 의미구조와 맥락, 내용을 형성하는 데 매우 중요한 역할을 한다고 하였다. ‘심리’, ‘코로나 19’, ‘블루’, ‘불안’, ‘마음’은 코로나 감정과 관련된 주요한 역할을 하는 단어임을 확인할 수 있었다.

교차점 역할을 하는 것은 매개 중심성과 연결 중심성이 모두 높은 노드와 상대적으로 연결 중심성은 낮지만, 매개 중심성은 높은 노드일 경우 하위구조 사이의 의미를 연결해 주는 것이라고 한다. 또한, 매개 중심성은 낮지만 연결 중심성은 높은 노드들은 하위 집단에서 그 집단의 의미를 형성하는데 일반적으로 영향을 미치는 개념으로 이해할 수 있다고 하였다(Park and Chung, 2013). 분석한 결과 연결 중심성이 높게 나타난 ‘코로나 19’, ‘마음’, ‘심리’, ‘불안’이라는 단어는 직접적으로 다른 노드들과 연결 관계가 높은 것으로 나타났고, 매개 중심성도 높은 결과를 나타

<Table 2> COVID 19 Emotion related Keywords Centrality analysis

	Word	Degree	Betweenness	Closeness	Eigen vector
1	코로나 19	23.000	39.594	0.667	0.374
2	중국	11.000	1.492	0.514	0.118
3	마음	21.000	31.414	0.667	0.324
4	상황	2.000	0.000	0.474	0.046
5	사회	6.000	10.406	0.545	0.070
6	사람	6.000	0.000	0.474	0.049
7	불안	22.000	24.622	0.667	0.393
8	사태	8.000	0.000	0.462	0.157
9	마스크	8.000	4.222	0.563	0.125
10	건강	4.000	0.000	0.429	0.086
11	생각	2.000	0.000	0.409	0.041
12	신중	12.000	13.011	0.529	0.134
13	사람들	6.000	0.000	0.474	0.049
14	정부	2.000	0.000	0.400	0.011
15	심리	31.000	32.161	0.720	0.512
16	블루	23.000	8.244	0.545	0.462
17	스트레스	2.000	0.000	0.429	0.047
18	상담	7.000	0.000	0.514	0.158
19	확산	4.000	0.833	0.450	0.069

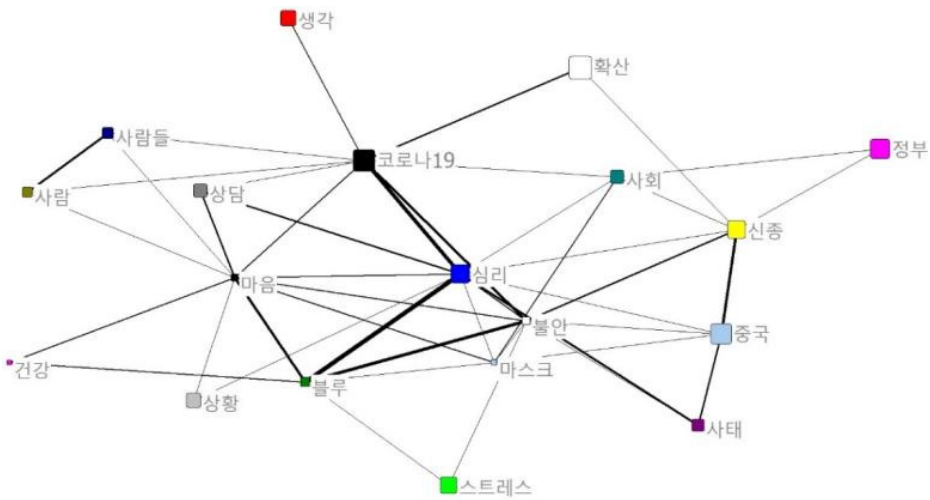
냈기 때문에 하위 노드들 사이에서 의미의 연결을 해주는 역할을 한다고 해석할 수 있다. 또한 ‘사회’ 키워드는 연결 중심성은 상대적으로 낮게 나타나 중심에서 떨어져 있는 것을 알 수 있었지만 매개 중심성은 비교적 높게 나타나 서로 다른 키워드들을 이어주는 ‘문지기’ 역할을 하는 것을 알 수 있었다.

4.3. 그래프 분석

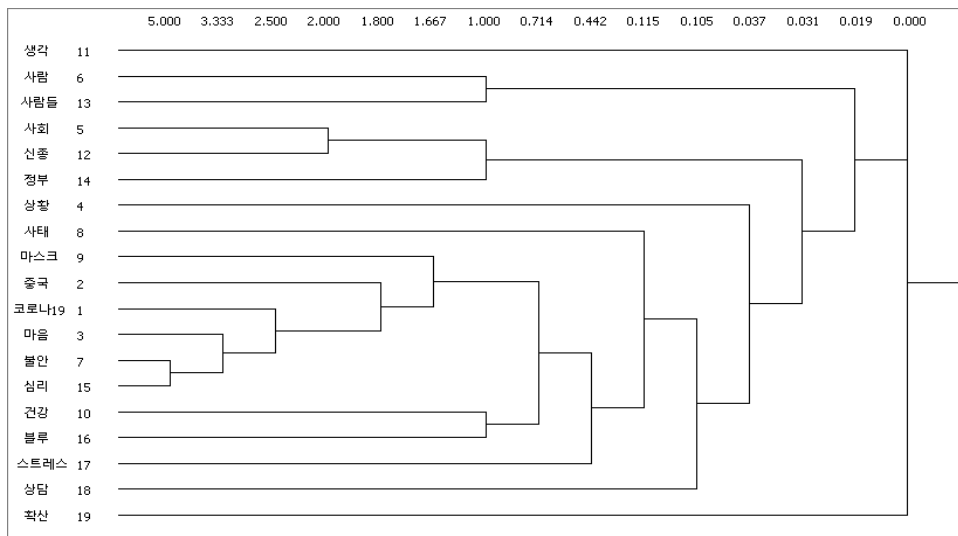
<Figure 5>는 코로나 감정으로 검색한 기사의 헤드라인에 나타난 주요 핵심어 사이의 동시 출현 빈도 네트워크를 그래프로 가시화한 것이다. UCINET은 소셜 네트워크 분석 프로그램으로 이를 사용하여 중심성 및 군집 분석을 시행하였다.

본 연구에서는 UCINET에 포함된 지도화 프로그램인 Net Draw를 이용하여 네트워크로 시각화하였다. 사각형은 핵심어를 나타내고, 선(line)은 동시에 두 단어가 출현한 빈도를 나타내는 것이다. 선(line)의 굵기는 동시 출현 빈도에 비례하는데, 두 단어가 동시에 출현한 빈도가 높으면 높을수록 선의 굵기가 굵어진다.

분석 결과는 ‘코로나-블루’ 쌍이 가장 굵게 표시되었고, ‘코로나-코로나 19’ 쌍이 그 뒤를 이어 ‘코로나-심리’ 쌍이 굵은 선으로 표시되었다. ‘코로나-감정’, ‘코로나-불안’ 쌍이 비교적 굵은 선으로 표시된 것을 알 수 있었다.



<Figure 5> COVID 19 Emotion related Keywords Network graph



<Figure 6> COVID 19 Emotion related Dendrogram for Hierarchical Clustering

4.4. 군집 분석

계층적 군집 분석은 사용하기에 매우 간단한 방법의 하나이며, 고차원 또는 다차원 데이터를 시각화하는 데 있어 기본적인 방법이다(Shin et

al., 2017). <Figure 6>은 앞서 도출한 핵심 키워드를 군집 분석에 대한 덴드로그램이다. Kwahk (2014)는 군집 분석은 개별 액터(actor)들을 서로 간의 상호 관련성을 바탕으로 동질적인 집단으로 묶어주는 통계기법이기 때문에, 군집 분석은

구조적 등위성을 분석하기 위한 가장 자연스럽고도 유용한 방법이라고 주장하였다. 계층적 군집화를 위한 덴드로그램에서 군집의 수는 크게 3개로 나누어진 것을 볼 수 있다. {생각}, {사람, 사람들, 사회, 신중, 정부, 상황, 사태, 마스크, 중국, 코로나 19, 마음, 불안, 심리, 건강, 블루, 스트레스, 상담}, {확산}으로 군집이 형성된 것을 확인할 수 있었다.

5. 결론

본 연구는 전 세계적으로 빠르게 확산하고 있는 코로나 19 감정과 관련된 뉴스 데이터를 수집하여 텍스트 마이닝 기법을 활용하여 키워드를 분류하고, 의미 네트워크 분석을 통해 단어들의 관계를 시각화하여 제시한 논문이다. 코로나 19의 확산은 우리의 일상생활뿐만 아니라, 사회-경제 등 많은 부분에 걸쳐 막대한 영향력을 미치고 있다. 전염병을 수반하는 집단 비극은 공포와 불안을 고조시키는데 이는 많은 사람의 행동과 심리학적 행복에 막대한 지장을 초래한다고 알려져 있다. 장기적인 부정적인 감정은 사람들의 면역력을 감소시키고 신체적인 균형을 파괴할 수도 있기 때문에 코로나 19로 인한 잠재적인 심리 변화를 이해하는 것이 필수적이다(Li et al, 2020). 보건 정책담당자들의 빠르고 복잡한 의사결정 과정에 도움이 되고자 급속히 전파되는 미디어 뉴스를 모니터링 함으로써, 더욱더 쉬운 소셜 네트워크 분석 방법을 제시하고자 하였다.

연구 절차는 다음과 같다. 코로나 19와 관련된 정신건강 이슈를 분석하기 위해 신문기사 빅데이터 분석 서비스 플랫폼인 빅카인즈(Bigkinds)를 활용하여 코로나 19가 처음 발생한 2019년 12

월부터 2020년 10월까지 보도된 1,000건의 기사를 대상으로 분석하였다.

본 연구에 필요한 핵심 키워드를 얻기 위해 ‘코로나 19 감정’으로 설정하고 전체 언론사를 대상으로 기사를 수집하였다. 수집된 1,000건의 기사 중 주제에 해당하지 않는 기사와 중복되는 기사를 제외하고 812건의 기사를 선별하여 분석에 활용하였다. 빅카인즈 (BigKinds)로부터 데이터를 수집하고, 연구자에 의해 데이터 정제과정을 거친 후, 키워드의 빈도수는 파이썬의 matplotlib을 활용하여 그래프로 시각화하고 워드 클라우드로 분석하였다. 한국어 의미 네트워크 분석 도구인 KrKwic 프로그램을 사용하여 텍스트 마이닝을 수행하였다. keywords프로그램을 실행하여, 기사의 제목에 나타난 단어의 빈도수를 계산하여 핵심어를 파악하였다. 소셜 네트워크 분석 전문 프로그램인 UCINET을 사용하여 연결 중심성, 매개 중심성, 고유 벡터 중심성, 근접 중심성을 분석하고, 군집 분석을 시행하고, Net Draw를 이용하여 그래프로 시각화하였다.

분석 결과는 다음 네 가지로 설명할 수 있다. 첫째, 키워드 분석 결과, 코로나 19 감정과 관련하여 ‘중국’, ‘불안’, ‘상황’, ‘마음’, ‘사회’, ‘건강’과 같은 단어의 빈도가 높게 나타난 것을 확인할 수 있었다. 둘째, 중심성 분석 결과, 연결 중심성은 ‘심리(31.000)’, ‘코로나 19(23.000)’, ‘블루(23.000)’, ‘불안(22.000)’, ‘마음(21.000)’ 순으로 높은 연결 중심성을 가지는 것으로 확인할 수 있다. 연결 중심성의 관점에서 볼 때, 가장 중심적인 핵심어는 ‘심리’, ‘코로나 19’, ‘블루’와 ‘불안’이라는 단어가 높은 중심성을 가지는 것을 확인할 수 있었다. 셋째, 뉴스의 헤드라인에 나타난 주요 키워드 사이의 동시 출현 빈도 네트워크를 그래프로 가시화하였다. 선의 굵기는 동시 출현

빈도에 비례하는데, 두 단어가 동시에 출현한 빈도가 높으면 굵은 선으로 표시된다. ‘코로나-블루’ 쌍이 가장 굵게 표시되었고, ‘코로나-코로나 19’ 쌍이 그 뒤를 이어 ‘코로나-심리’ 쌍이 굵은 선으로 표시되었다. ‘코로나-불안’, ‘코로나-심리’ 쌍이 비교적 굵은 선으로 표시된 것을 알 수 있었다. 코로나와 우울증은 이제 관심을 가져야 할 키워드임을 확인할 수 있었다. 넷째, 앞서 도출한 핵심 키워드를 가지고 군집 분석한 결과, 덴드로그램을 통해 3개의 군집을 추정할 수 있었다.

본 연구의 시사점은 다음과 같다. 첫째, 코로나가 처음 발생한 2019년 12월부터 2020년 10월까지 54개의 언론사로부터 수집한 뉴스를 대상으로 ‘코로나 19 감정’이라는 키워드를 더욱더 쉽게 분석했다는 것이다. 비용 절감과 동시에 사회 현상 및 변화를 측정할 수 있다는 장점이 있다. 둘째, 소셜 네트워크 분석 기법과 텍스트 마이닝 기법을 통합함으로써 토픽이 형성할 수 있는 네트워크 조직에 관심을 두고 네트워크 분석을 한 결과를 바탕으로 큰 영향력을 가진 토픽을 선정하여 시각화한 것이라고 할 수 있다. 셋째, 코로나 19 감정과 관련된 주제와 주요 키워드를 함께 제시하고 시각화함으로써 의료 정책 담당자들이 현상을 파악하고 연구를 할 때, 다양한 관점을 제공해 줄 수 있다는 점이다. 이를 이용하여 코로나 19와 우울증과 관련된 심리방역 문제에 대한 지원, 치료와 서비스 개발에 기초자료로 활용하는 데 도움을 주고자 한다.

연구 결과를 바탕으로 본 연구는 다음과 같은 학술적인 시사점을 가진다. 첫째, 코로나 19 감정과 관련하여 사회적으로 형성되고 있는 주요 의제가 무엇인지 텍스트 마이닝과 의미 네트워크 분석을 활용하여 실증적이고 체계적으로 분

석한 연구라는 점에서 학술적 의의가 있다. 둘째, 비정형 데이터인 텍스트를 분석하여 코로나 19 감정과 관련된 핵심 의제를 도출해 낼 수 있었다는 점이다. 비정형 데이터인 텍스트를 분석하는 기법은 다양한 분야에서 활용되고 적용될 수 있다.

본 연구는 다음과 같은 한계점을 가진다. 첫째, <Table 1>에 제시한 ‘사람’과 ‘사람들’의 빈도 분석 결과는 ‘사람’ 899회, ‘사람들’ 654회로 다르게 분석되어 별개로 집계하였으나, 중심성 분석 결과 중심성 값이 동일하게 분석되었다. 이를 바탕으로 형태 및 갈래는 다르나 본질적으로 일치하는 의미가 있는 단어들인 연구 과정을 거쳐 동일한 분석 결과를 도출할 수 있음을 확인하였다. 따라서 전처리 과정에서 텍스트 분석 전 어근 동일화(stemming) 및 더욱 세분된 단어 분류 작업을 통해 두 단어를 같은 형태로 통일해서 분석하는 방법을 제시하고자 한다. 둘째, 코로나 정신건강과 관련하여 뉴스 헤드라인만을 분석하였다는 점이다. 향후 연구에서는 학술논문, 연구 보고서, 페이스북, 트위터, 뉴스, 블로그의 리뷰 분석을 통해 좀 더 다양한 의견을 수집하여 연구해볼 필요가 있다. 셋째, 키워드를 좀 더 다양하게 하여 분석해 볼 필요가 있을 것이다. 장기화에 진입한 코로나 19에 대한 심리방역에 관심을 가져야 할 시기에, 코로나와 관련된 다양한 현상들에 대해 체계적으로 연구해 볼 필요가 있을 것이다.

참고문헌(References)

- Arafa, A., Z. Mohammed, O. Mahmoud, M. Elshazley, and A. Ewis, “Depressed, anxious,

- and stressed: What have healthcare workers on the frontlines in Egypt and Saudi Arabia experienced during the COVID-19 pandemic?," *Journal of Affective Disorders*, Vol.278, (2021), 365-371.
- Baht, S. S. and S. Milne, "Network effects on cooperation in destination website development," *Tourism Management*, Vol.29, No.6(2008), 1131-1140.
- Blei, D. M., A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, Vol.3, (2003), 993-1022.
- Cha, Y. J., J. H. Lee, J. E. Choi, and H. W. Kim, "A Topic Modeling Approach to Marketing Strategies for Smartphone Companies," *Knowledge Management Research*, Vol.16, No.4(2015), 69-87.
- Chung, P. L., H. C. Ahn, and K. Y. Kwahk, "Identification of Core Features and Values of Smartphone Design using Text Mining and Social Network Analysis," *Korean Journal of Business Administration*, Vol.32, No.1(2019), 27-47.
- Diesner, J. and K. M. Carley, "Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a Novel Method for Network Text Analysis," *Causal Mapping for Research in Information Technology*, (2005), 81-108.
- Gim, E. J. and J. H. Koo, "Analysis of Social Network Change Characteristics of Participants in Urban Regeneration Project Using Net Miner: Focused on the Urban Regeneration Leading Area in Suncheon-City," *Journal of Information Technology Services*, Vol 19, No.1(2020), 1-16.
- Guidry, J. P. D., M. Messner, S. L. Meganck, P. B. Perrin, A. Lovari, and K. E. Carlyle, "#Ebola: Tweeting and Pinning an Epidemic," *Atlantic Journal of Communication*, (2020), 1-16.
- Hwang, S. I. and M. K. Kim, "An Analysis of Artificial Intelligence(A.I.)_related Studies' Trends in Korea Focused on Topic Modeling and Semantic Network Analysis," *Journal of Digital Contents Society*, Vol.20, No.9(2019), 1847-1855.
- Hwang, S. I. and Y. W. Park, "An Analysis of Arts Management-Related Studies' Trend in Korea using Topic Modeling and Semantic Network Analysis," *Korean Association of Arts Management*, Vol.50, No.1(2019), 5-31.
- J. A. Terrizzi Jr., N. J. Shook, and M. A. McDaniel, "The behavioral immune system and social conservatism: A meta-analysis," *Evol. Hum. Behav.*, Vol.34, (2013), 99-108.
- Kim, N. G., D. H. Lee, H. C. Choi, and W. X. S. Wong, "Investigations on Techniques and Applications of Text Analytics," *The Journal of Korean Institute of Communications and Information Sciences*, Vol.42, No.2(2017), 471-492.
- Kwahk, K. Y., *Social Network Analysis*, 2nd edition, CheongRam Publishers, 2019.
- Lee, J. M., M. N. Liu, and G. G. Lim, "A study on the revitalization of tourism industry through Big Data analysis," *Journal of Intelligence and Information Systems*, Vol.24, No.2(2018), 149-169.
- Lee, S. M., S. E. Ryu, and S. J. Ahn, "Mass Media and Social Media Agenda Analysis Using Text Mining : focused on 5-day Rotation Mask Distribution System," *The Journal of the Korea Contents Association*, Vol.20, No.6(2020), 460-469.

- Lee, S. H. and H. Y. Lee, "A Data Mining and Social Network Analysis to Understand Multi-Destination Tour Behavior of Inbound Free Independent Tourists in Seoul," *Korean Journal of Business Administration Academic Conference*, (2017), 321-334.
- Li, S., Y. Wang, J. Xue, N. Zhao, and T. Zhu, "The impact of COVID-19 epidemic declaration on psychological consequences: A study on active weibo users," *International Journal of Environmental Research and Public Health*, Vol.17, No.6(2020), 2032.
- Liu, M. N. and G. G. Lim, "Word-of-Mouth Effect for Online Sales of K-Beauty Products: Centered on China SINA Weibo and Meipai," *Journal of Intelligence and Information Systems* Vol.25, No.1(2019), 197-218.
- Liu, S., L. Yang, C. Zhang, Y. Xiang, Z. Liu, S. Hu, and B. Zhang, "Online mental health services in China during the COVID-19 outbreak," *The Lancet Psychiatry*, Vol.7, No.4(2020), e17-e18.
- Norris, F. H., M. J. Friedman, and P. J. Watson, "60,000 Disaster Victims Speak: Part II. Summary and Implications of the Disaster Mental Health Research," *Psychiatry*, Vol.65, No.3(2002), 240-260.
- Park, C. S. and C. W. Chung, "Text Network Analysis: Detecting Shared Meaning through Socio-cognitive Networks of Policy Stakeholders," *Institute of Governmental Studies*, Vol.19, No.2(2013), 73-108.
- Park, B. E. and G. G. Lim, "A Study on the Impact Factors of Contents Diffusion in Youtube using Integrated Content Network Analysis," *Journal of Intelligence and Information Systems* Vol.21, No.3(2015), 19-36.
- Park, H. W. and L. Leydesdorff, "Understanding the KrKwic: A computer program for the analysis of Korean text," *Journal of The Korean Data Analysis Society(JKDAS)*, Vol.6, No.5(2004), 1377-1387.
- Sivasankaran, B. and J. Aleksandar, "Mass hysteria revisited," *Current Opinion in Psychiatry*, Vol.19, No.2(2006), 171-174.
- Sohn, D. W., *Social Network Analysis*, KyungMoon Publishers, 2002.
- Solvic, P., "Perception of risk," *Science* 1987, Vol.236, No.4799(1987), 280-285.
- Shin, M. S., M. G. Park, and S. H. Bae, "Nano Technology Trend Analysis Using Google Trend and Data Mining Method for Nano-Informatics," *Journal of Society of Korea Industrial and Systems Engineering*, Vol. 40, No. 4(2017), 237-245.
- Wang, Y., Y. Di, J. Ye, and W. Wei, "Study on the public psychological states and its related factors during the outbreak of coronavirus disease 2019 (COVID-19) in some regions of China," *Psychology, Health & Medicine*, Vol.26, No.1(2021), 13-22.
- Xiang, Y., Y. Yang, W. Li, L. Zhang, Q. Zhang, T. Cheung, and C. H. Ng, "Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed," *The Lancet Psychiatry*, Vol.7, No.3(2020), 228-229.
- Yoon, J. E. and C. J. Suh, "Research Trend Analysis by using Text-Mining Techniques on the Convergence Studies of AI and Healthcare Technologies," *Journal of Information Technology Services*, Vol 18. No.2(2019), 123-141.

Yu, D. S. and G. G. Lim, "A Study on the eWOM and Selecting Movie According to Online Media and Replies," *Journal of Information*

Technology Services, Vol.14, No.2(2015), 177-193.

Abstract

Analysis of News Agenda Using Text mining and Semantic Network Analysis: Focused on COVID-19 Emotions

So-yeon Yoo* · Gyoo-gun Lim**

The global spread of COVID-19 around the world has not only affected many parts of our daily life but also has a huge impact on many areas, including the economy and society. As the number of confirmed cases and deaths increases, medical staff and the public are said to be experiencing psychological problems such as anxiety, depression, and stress. The collective tragedy that accompanies the epidemic raises fear and anxiety, which is known to cause enormous disruptions to the behavior and psychological well-being of many. Long-term negative emotions can reduce people's immunity and destroy their physical balance, so it is essential to understand the psychological state of COVID-19. This study suggests a method of monitoring medial news reflecting current days which requires striving not only for physical but also for psychological quarantine in the prolonged COVID-19 situation. Moreover, it is presented how an easier method of analyzing social media networks applies to those cases. The aim of this study is to assist health policymakers in fast and complex decision-making processes. News plays a major role in setting the policy agenda. Among various major media, news headlines are considered important in the field of communication science as a summary of the core content that the media wants to convey to the audiences who read it. News data used in this study was easily collected using “Bigkinds” that is created by integrating big data technology. With the collected news data, keywords were classified through text mining, and the relationship between words was visualized through semantic network analysis between keywords. Using the KrKwic program, a Korean semantic network analysis tool, text mining was performed and the frequency of words was calculated to easily identify keywords. The frequency of words appearing in keywords of articles related to COVID-19 emotions was checked and visualized in word cloud ‘China’, ‘anxiety’, ‘situation’, ‘mind’, ‘social’, and ‘health’ appeared high in relation to the emotions of COVID-19.

* School of Business, Hanyang University

** Corresponding author: Gyoo-gun Lim
School of Business, Hanyang University
220 Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea
Tel: +82-2-2220-2593, E-mail: gglim@hanyang.ac.kr

In addition, UCINET, a specialized social network analysis program, was used to analyze connection centrality and cluster analysis, and a method of visualizing a graph using Net Draw was performed. As a result of analyzing the connection centrality between each data, it was found that the most central keywords in the keyword-centric network were 'psychology', 'COVID-19', 'blue', and 'anxiety'. The network of frequency of co-occurrence among the keywords appearing in the headlines of the news was visualized as a graph. The thickness of the line on the graph is proportional to the frequency of co-occurrence, and if the frequency of two words appearing at the same time is high, it is indicated by a thick line. It can be seen that the 'COVID-blue' pair is displayed in the boldest, and the 'COVID-emotion' and 'COVID-anxiety' pairs are displayed with a relatively thick line. 'Blue' related to COVID-19 is a word that means depression, and it was confirmed that COVID-19 and depression are keywords that should be of interest now. The research methodology used in this study has the convenience of being able to quickly measure social phenomena and changes while reducing costs. In this study, by analyzing news headlines, we were able to identify people's feelings and perceptions on issues related to COVID-19 depression, and identify the main agendas to be analyzed by deriving important keywords. By presenting and visualizing the subject and important keywords related to the COVID-19 emotion at a time, medical policy managers will be able to be provided a variety of perspectives when identifying and researching the regarding phenomenon. It is expected that it can help to use it as basic data for support, treatment and service development for psychological quarantine issues related to COVID-19.

Key Words : COVID-19, Text mining, Semantic Network Analysis, Co-occurrence Network

Received : November 24, 2020 Revised : January 20, 2021 Accepted : March 5, 2021

Corresponding Author : Gyoo-gun Lim

저 자 소개



유소연

한양대학교 경영대학 비즈니스 인포매틱스학과 박사과정에 재학중이다. 관심분야는 빅데이터 분석, 인공지능, 텍스트 마이닝 등이다.



임규건

한양대학교 경영대학 임규건 교수는 KAIST 전산학 학사, POSTECH 컴퓨터 석사, KAIST 경영공학 박사학위를 취득하였고, 삼성전자, KT, 국제전자상거래연구센터(ICEC) 연구위원, 세종대학교 경영학과 교수를 역임하였다. 관심분야는 혁신 비즈니스 모델, IT서비스 혁신, 인공지능과 경영, e-Business 등이며, 2018년 IT서비스 우수연구인상을, 2009년 IT Innovation 유공자 지식경제부 장관 표창과 2007년 SW산업발전 유공자 정통부 장관 표창을 수여하였다. 주요 저서로는 ‘경영을 위한 정보기술’, ‘e-비즈니스 경영’, ‘디지털경제시대의 경영정보시스템’ 등 전문서적과 다수의 논문과 특허가 있다. 또

한, 아시아최초 상용인터넷인 KORNET 상용화, 중국 Shanghai Telecom SI사업전략, 한국영화기술 로드맵, KTI 사업전략, 나라장터 (G2B) 효과평가, 행정정보화(G4C) 성과분석, 국가정보보호지수개발, 국방정보화 수준평가모형, IT혁신인력양성종합대책, 국가디지털식별체계(UCI), 저작권정품 인증제도, SW사업자신고제도 개선, SW기술자신고제도개선 등 다양한 IT혁신 분야의 프로젝트를 수행하였다 .