

## 코로나 취약계층에 대한 한국사회의 인식: 언론 및 트위터 담론을 중심으로

김세현\*

### 국문초록

이 연구는 코로나19의 확산 이후 한국 사회에서 진행된 감염 취약계층에 대한 논의들을 경험적으로 검증하는 것을 목적으로 한다. 이를 위해 2020년 1월부터 4월까지 코로나 취약계층에 대한 언론 기사 및 트위터 게시물을 수집하였으며, 텍스트 마이닝과 의미 연결망 분석, 토픽 모델링 기법을 활용하여 분석을 진행하였다. 분석 결과를 요약하면 다음과 같다. 첫째, 언론 영역에서 코로나19로 발생한 취약계층을 소상공인과 실업자, 구직자들로 상정하는 반면, 트위터에서는 소상공인 이외에도 아동, 노인, 청년 등 연령 집단을 중심으로 어려움을 논의하고 있었다. 둘째, 논의의 흐름을 살펴보면 감염병 발생 초기에는 바이러스 전파를 막기 위한 대책이 중요하게 논의되었으나, 이후에는 경제적 어려움에 대한 논의로 전환되었다. 셋째, 언론 및 뉴미디어 모두 취약계층에 대한 기부 활동을 중요한 이슈로 다루고 있었다.

주제어: 코로나19, 감염 취약계층, 텍스트 마이닝, 의미 연결망, 토픽 모델링

---

\* 김세현\_한양대학교 SSK다문화연구센터 전임연구원(legperde@gmail.com), 단독저자

© Copyright 2020 Social Integration Research Center, Kangwon National University.  
This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## I. 들어가는 말

2019년 말 중국 우한에서 코로나19 바이러스가 처음 발생하였을 당시에는 누구도 이 전염병이 전 세계인의 일상을 뒤바꿔 놓을 것이라 상상하지 못하였다. 하지만 바이러스가 전 지구적 네트워크를 통해 급격히 확산되는 동안 이동 금지 및 격리, 주요 공공기관의 폐쇄, 의료체계의 붕괴, 경기 침체와 같은 혼돈적 상황에 놓이게 되었으며, 6개월 남짓 지난 지금 우리 사회는 전염병의 세계적 대 유행이 바뀌놓은 우리의 일상이 발병 이전으로 돌아가지 못할 것이라는 비관적 전망에 휩싸여 있다.

이 같은 코로나19의 확산은 우리 사회의 취약계층에게는 더욱 큰 위협이 되고 있다. 현재 정부에서는 코로나19와 같은 감염병 관련 취약계층을 ‘사회복지시설을 이용하는 어린이, 노인 등’<sup>1)</sup>으로 정의하고 있으나, 코로나19로 인해 발생하는 위험은 단순히 병원균 감염 문제를 넘어 사회경제적 어려움으로 확장되는 실정이다. 국내의 고용 동향을 살펴보면 2020년 3월의 경우 취업자가 19만 5천 명이 감소한 반면, 비경제활동인구는 51만 6천 명이 증가한 것으로 나타났다(통계청, 2020). 또한 국제통화기금(IMF)에서도 우리나라의 경제 전망을 전년대비 1.2%로 감소할 것으로 예상하였다(IMF, 2020). 특히 코로나19로 발생한 경제적 어려움이 임시·일용직, 프리랜서, 특수고용노동자 등 근로빈곤층 및 근로취약계층에 더욱 큰 위협이 될 것이라는 전망은 코로나19로 인한 새로운 사회적 양극화 가능성을 점치게 한다(김태완, 2020).

또한 감염병 확산에 대한 공포는 단순히 생명에 대한 위협뿐만이 아니라, 이러한 상황을 발생시키게 된 집단에 대한 원망이나 분노로 이어지기도 한다. 2003년 유행한 사스(SARS)나 2009년의 신종플루(H1N1), 2015년 메르스(MERS) 등의 사태에서도 일부의 사람들이나 집단을 비난

---

1) 감염병 예방법 49조.

하는 여론이 조성되기도 하였다. 또한 이번 코로나19의 확산 가운데에서도 일부 미디어에서 다루어진 반중(反中) 정서가 확산되거나, 특정 집단을 희생양으로 상정하고, 이들에 대한 언론 보도가 이어지면서 이들에 대한 혐오 감정이 표출되는 등 다양한 사회문화적 현상을 야기하고 있다(김수경, 2020).

코로나19 확산 이후 발생한 새로운 사회문제들의 영향력이 급격하고 광범위하게 확산되고 있다. 이러한 가운데 우리는 사회가 직면하는 문제들을 파악하고, 문제들을 해결하기 위한 정책적 방향을 결정하여 새로운 위협에 효과적으로 대비하여야 할 것이다. 따라서 이번 연구에서는 전염병 최초 발병 이후 현재까지 뉴스 미디어 및 뉴미디어 공간에서 진행된 논의들의 흐름을 분석하여 우리 사회에서 다루어지는 코로나19 관련 사회이슈를 구체적으로 파악하고자 한다.

## 1. 디지털 자료를 활용한 사회현상 분석

2010년 이후 우리 사회에서 소위 빅데이터로 통칭되는 대규모 비정형 자료들이 축적되고, 컴퓨터를 활용한 계산능력이 비약적으로 향상됨에 따라 기계학습 알고리즘을 활용한 연구 방법론이 학계에서 다시금 주목받게 되었다. 이러한 흐름에 발맞추어 메타분석 영역에서도 기계학습을 기반으로 한 메타분석이 다양하게 시도되고 있다(DiMaggio, Nag, & Blei, 2013; Chen et al., 2016; Guo et al., 2016; 신동훈·김세현, 2020). 특히 새롭게 등장한 메타분석 기법은 복잡한 동시에 매우 큰 규모의 비정형 데이터 분석에서 접근성과 신뢰성 모두를 확보할 수 있게 도움을 준다는 점에서 주목받고 있다(Mützel, 2015).

텍스트 마이닝 기법을 활용하여 전염병에 대한 미디어의 관점을 다룬 연구들 역시 다양하게 시도되어 왔다 Towers와 그의 동료들(2015)은 에볼라 바이러스와 관련된 미국 내의 TV 뉴스와 트윗, 인터넷 검색 빈도

간의 관계를 분석하면서 언론 매체가 인터넷 공간 여론 형성에 영향을 미치고 있음을 경험적으로 밝히고 있다. 또한 Househ(2015) 역시 에볼라, 바이러스에 대한 정보가 뉴스와 트위터를 통해 확산되는 관계에 주목하면서 뉴스 매체에서 제공된 정보가 사회관계망 서비스(SNS)를 통해 전달되고 있음을 밝히고 있다. 국내에서도 이와 유사하게 질병 정보와 미디어 간의 관계를 설명하려는 다양한 연구들이 진행되었다(안주영 외, 2016; 황교상 외, 2014).

텍스트 분석 기법을 이용하여 질병과 언론, 그리고 뉴미디어 간의 관계를 분석하는 것은 단순하게 의료 정보의 유통과정을 분석하는 것만이 아니라 질병이 사회적 현상에 미치는 영향력을 알아볼 수 있는 단초가 될 수 있다(Towers et al., 2015; 안주영 외, 2016). 따라서 본 연구는 코로나19가 국내에서 발생한 1월 말 이후 2020년 4월 30일까지 약 4개월(14주) 간 언론 및 뉴미디어 영역에서 진행된 코로나 취약계층에 대한 논의들을 텍스트 마이닝 및 의미 연결망 분석, 토픽 모델링 기법을 이용하여 살펴보고자 한다. 특히 코로나19의 확산 이후 우리 사회에서 관심을 가지는 코로나 취약계층의 특성을 살펴보고, 이들에 대한 사회적 지원 형태를 가늠하기 위해 텍스트 마이닝(text mining)과 의미 연결망 분석(semantic network analysis), 토픽 모델링(topic modeling) 기법을 활용하여 언론 기사와 트위터 자료를 분석하였다.

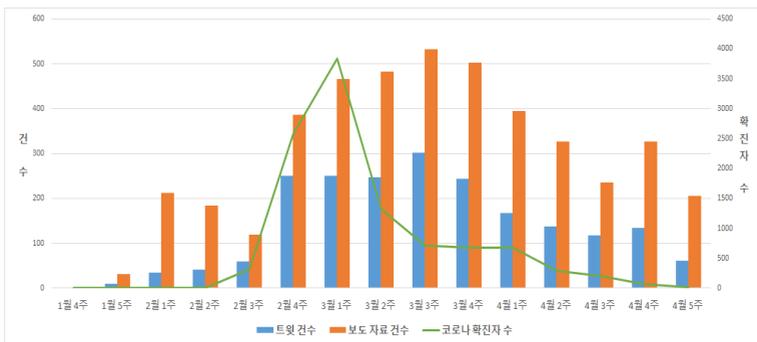
이번 연구를 통해 코로나19 감염자 확산 이후 한국 사회에서 진행된 취약계층 지원 정책 의제를 확인하고, 정책의 방향성 및 효과를 검토할 수 있는 성찰적인 토대를 마련함과 동시에 앞으로 다가올 포스트 코로나 시대에서 취약계층 지원을 위해 필요한 경험적인 자원을 제공할 수 있을 것으로 기대된다.

## II. 연구대상 및 방법론

### 1. 연구자료

이번 연구의 분석 자료는 국내 코로나19 확진자 발견 이후 2020년 4월 30일까지 작성된 언론 기사 및 트위터 게시물 중 ‘코로나’와 ‘취약계층’ 단어를 포함하는 텍스트를 대상으로 한다. 이를 위해 언론 기사는 한국언론진흥재단의 뉴스 DB인 빅카인즈<sup>2)</sup>를 활용하였으며, 트윗 자료는 파이썬(python)을 이용한 웹 크롤러를 작성하여 자료를 수집하였다. 이후 중복 자료 및 분석에 적절하지 않은 자료들을 제외하여 언론 보도자료는 4,407건, 트윗은 2,058건의 자료가 최종 분석에 이용되었다.

분석 자료를 살펴보면 2020년 1월 20일 국내에 첫 코로나19 확진자가 발생한 이후 3일 뒤인 1월 23일에 트위터에서 코로나 취약계층에 대한 게시물이 처음 등장하였으며, 언론매체에서는 이보다 조금 늦은 1월 28일에 취약계층에 대한 논의가 진행되고 있음을 알 수 있다. 다음의 <그림 1>은 국내에서 최초 확진자가 발생한 1월 넷째 주 이후 주별 확진



<그림 1> 시기별 코로나 확진자 수 및 뉴스 기사, 트윗 건수

2) <http://bigkinds.or.kr>

자 수와 뉴스 기사 건수, 트윗 건수를 제시한 것이다. 내용을 살펴보면 발병 초기에는 언론 영역에서 먼저 취약계층에 대한 논의가 활발하게 진행되었으나, 확진자 수가 급격히 증가한 2월 3주 차 이후에는 SNS 공간에서도 이들에 대한 논의가 급격히 증가하였음을 알 수 있다.

## 2. 연구방법

연구에서 수행된 자료의 전처리 및 토픽 모델링 분석은 R의 tm, LDA 패키지를 이용하였으며, 의미 연결망 구성 및 시각화를 위해서는 네트워크 분석 및 시각화 프로그램인 Netminer와 Gephi가 이용되었다.

### 1) 자료수집 및 전 처리

언론 기사들의 텍스트 정보를 제공하는 빅카인즈에서는 신문기사를 바탕으로 하는 정형화된 텍스트 정보를 제공하고 있다. 하지만 이번 연구에서 사용된 트위터 자료는 일반적 분석에 적합하지 않은 비정형 자료의 형태를 갖추고 있다. 따라서 이번 연구에서는 정형화 과정을 거친 빅카인즈 데이터를 기반으로 단어 사전을 구성한 뒤, 트위터 게시글을 정형화하기 위한 텍스트 마이닝 과정을 수행하였다.

한편, 텍스트 분석에서 단어의 출현 빈도를 확인할 경우, 검색에 사용되는 키워드가 가장 높은 빈도로 나타나게 된다. 이번 연구의 자료 수집을 위해 ‘코로나’와 ‘취약계층’이라는 단어를 기준으로 검색을 수행하였기 때문에 이후 분석에서는 ‘코로나’와 ‘취약계층’ 단어를 제외하였다. 또한 의미를 파악하기 어려운 한 글자 단어 역시 분석에서 제외하였다. 최종적으로 언론 기사 분석에서는 총 45,213개의 단어를 대상으로 분석을 진행하였으며, 트위터 분석에서는 5,976개의 단어가 분석에 포함되었다.

### 2) 의미 연결망 분석

의미 연결망 분석은 사회연결망 분석 기법을 텍스트 분석에 적용하여

단어 간의 의미 연결망을 구성하고 분석하는 기법을 의미한다(김용학, 2015). 텍스트의 공출현(co-occurrence) 관계를 통해 구성된 의미 연결망은 문서나 문서군이 담고 있는 심층적 개념들을 파악하고, 핵심 개념들 간 관계를 구성하는데 용이하기 때문에 다양한 연구 영역에서 활용되고 있다(Kleinnijenhuis, et al., 1997; 김세현, 2018).

이번 연구에서는 전처리 된 텍스트 자료를 바탕으로 출현 빈도 기준 상위 100개 키워드를 추출하고, 이들 간의 공출현 연결망을 구성하였다. 이후에 분석의 편의성 및 시각화를 위해 최소 신장 트리(minimum spanning tree) 기법을 이용하여 의미 연결망의 핵심 구조를 추출하였다. 최소 신장 트리 기법은 복잡한 구조를 형성하고 있는 연결망을 분석하여 모든 노드 간의 연결이 존재하는 동시에 전체 연결거리가 최소가 되는 최적의 연결 구조를 탐색하는 기법이다(Quirin, Cordón, Guerrero-Bote, Vargas-Quesada, & Moya-Anegón, 2008; 김용학, 2015).

마지막으로는 최소 신장트리 기법으로 구성된 의미 연결망에서 개념 간 관계 구조를 파악하기 위해 군집분석을 수행하였다. 이번 연구에서는 단어 간 응집성을 기반으로 최적의 적합도를 갖춘 군집을 추출하기 위해 CNM 알고리즘을 이용하여 의미 연결망의 군집분석을 수행하였다(Clauset et al., 2004; Theodoridis & Koutroumbas, 2006; 김세현, 2018).

### 3) LDA기법을 이용한 토픽 모델링

일반적으로 토픽 모델링이란 문서에서 나타나는 단어들을 정량적으로 분류하여 문서가 담고 있는 주제를 분류하는 다양한 기법들을 의미한다(DiMaggio, Nag, & Blei, 2013; Grimmer & Stewart, 2013). 이번 연구에서는 다양한 토픽 모델링 기법 중 생성 확률 모델(generative probabilistic model)을 기반으로 하는 잠재 디리클레 할당(Latent Dirichlet Allocation, 이하 LDA) 기법을 적용하였다. LDA 분석은 탐색적 요인분석의 한 형태로 확률 모델을 이용하여 문서에서의 단어 출현

확률을 분석하고, 소수의 토픽을 추출하는 기법이다(Steyvers & Griffiths, 2007). 이 같이 정량적인 분석 기법을 이용한 토픽 모델링은 개관적이고 재현 가능성이 높으며, 분석결과가 비교적 명료하다는 점에서 개별연구자의 통찰력에 크게 의존하는 기존 메타연구 방법에 비해 각광 받고 있다. 또한 분석대상에 대해 연구자의 사전지식이 충분하지 않은 경우에도 분석을 수행하고, 결과를 해석할 수 있다는 점에서 디지털 텍스트 자료 연구에 다양하게 활용된다.

기본적으로 LDA 분석은 문서에 몇 개의 세부 토픽이 존재한다고 가정 한 이후, 각 토픽에 대한 단어들의 출현 확률을 계산하여 개별 문서가 개별 토픽들에 속할 확률을 추정하게 된다(Blei, 2012). 따라서 LDA 기법을 이용한 토픽 모델링을 수행하기 위해서는 먼저 분석 대상에서 몇 개의 토픽을 추출할 것인지를 연구자가 결정해야 한다. 현재 LDA 분석에서 토픽 수 결정을 위한 다양한 방법들이 제시되고 있으나, 최적의 방법론에 대한 합의가 이루어지지 않았으며, 몇몇의 연구자의 경우에는 연구 대상의 특성 및 연구 목적에 따라 토픽 수를 결정하는 방법을 제안하기도 한다(Arun et al., 2010; Cao et al., 2009; Deveaud et al., 2014; Griffiths & Steyvers, 2004). 이번 연구는 코로나19와 관련에서 취약계층에 대한 논의가 언론 매체 및 SNS 매체에서 어떻게 다루어지는지를 살펴보고, 매체 간 차이를 비교하는 것이 주요한 연구 목적임을 감안하여 각 매체에서 6개의 토픽을 추출하고 분석을 진행하였다.

### III. ‘코로나’ 및 ‘취약계층’에 대한 언론기사 분석

#### 1. 주요 키워드 분석

이번 연구는 코로나 취약계층에 대해 어떠한 논의가 이루어지고 있는지를 분석하는 것을 주요 목적으로 한다. 다음 제시되는 <표 1>은 본격적인 분석에 앞서 국내 언론 기사에서 등장한 주요 단어들 중 출현 빈도를 기준 상위 50개의 단어들을 제시한 것이다.

분석 결과, 가장 높은 출현 빈도를 보인 단어는 ‘지원’(12,665회), ‘마

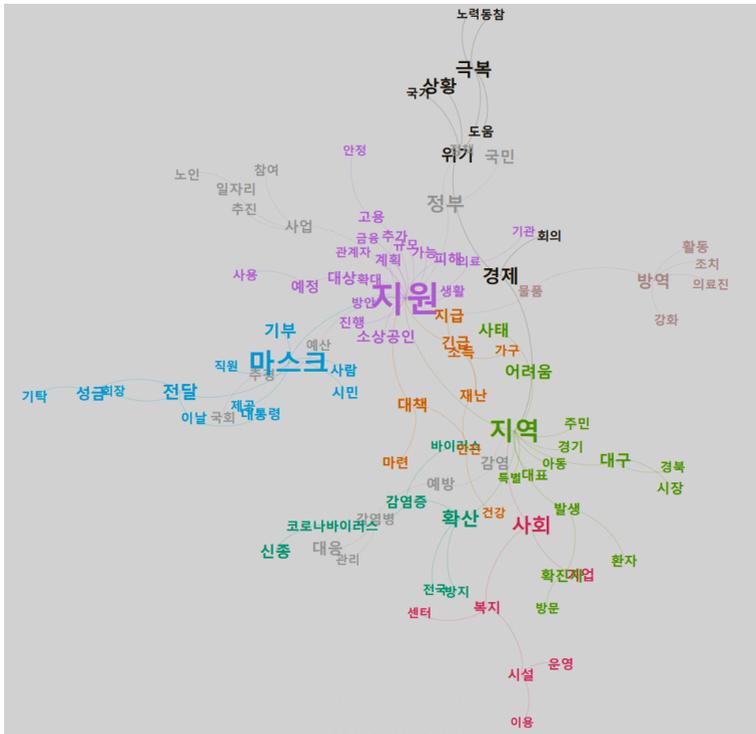
<표 1> 뉴스 기사에서의 주요단어 출현 빈도 (상위 50개 단어 기준)

순위	단어	빈도	순위	단어	빈도	순위	단어	빈도
1	지원	12,665	18	사태	2,840	35	기업	2,036
2	마스크	8,602	19	대응	2,816	36	일자리	1,998
3	지역	7,603	20	성금	2,787	37	감염증	1,997
4	정부	4,836	21	대책	2,766	38	복지	1,951
5	확산	4,599	22	지급	2,644	38	계획	1,951
6	사회	4,538	23	긴급	2,553	40	시장	1,884
7	극복	4,272	24	감염	2,489	41	고용	1,859
8	경제	4,256	25	대상	2,464	42	시설	1,855
9	전달	3,960	26	소상공인	2,368	43	대통령	1,838
10	상황	3,654	27	예정	2,311	44	마련	1,829
11	방역	3,452	28	국민	2,259	45	추경	1,800
12	기부	3,380	29	사업	2,254	46	발생	1,780
13	위기	3,299	30	피해	2,225	47	활동	1,778
14	바이러스	3,198	31	예방	2,223	48	대표	1,770
15	대구	3,050	32	재난	2,175	49	추진	1,703
16	어려움	3,028	33	확진자	2,143	50	감염병	1,674
17	신종	2,967	34	소득	2,136			

스크'(8,602회)로 나타났다. 이러한 결과는 코로나19 바이러스 확산 이후 진행된 마스크 대란 사태가 취약계층에 대한 마스크 지원에 대한 논의 이어지면서 나타난 결과로 볼 수 있다. 또한 '지역'(7,603회), '정부'(4,836회), '사회'(4,538회) 역시 주요 언론 기사에서 자주 출현 단어로 나타났다. 이러한 단어 사용 빈도를 볼 때 언론 영역에서는 통해 취약계층에 대한 지원을 담당해야 하는 핵심 행위자로 지역사회와 정부를 지목하고 있음을 간접적으로 유추할 수 있다.

## 2. 의미 연결망 분석

다음 <그림 2>는 출현 빈도 기준 상위 100개 단어들을 중심으로 공출



<그림 2> 뉴스 기사 의미 연결망 (상위 100개 단어 기준)

현 관계를 확인하여 구성된 의미 연결망을 제시한 것이다. CNM 알고리즘을 이용하여 군집분석을 수행한 결과, 총 12개의 군집이 관찰되었으며, 모듈성(modularity) 값은 0.758로 집단 간 구분이 비교적 잘 이루어지고 있음을 알 수 있다.

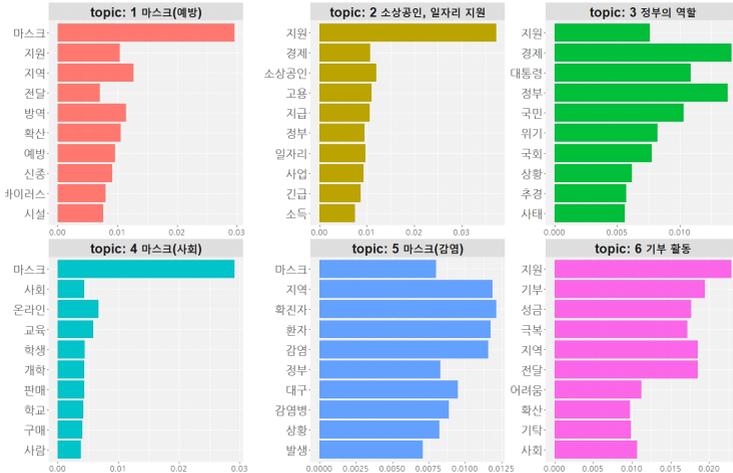
의미 연결망을 살펴보면 먼저 가장 눈에 띄는 키워드는 ‘지원’, ‘경제’, ‘위기’이다. 이 단어들 간의 관계를 살펴볼 때 코로나 바이러스 확산 이후 언론 영역에서 취약계층의 피해 및 지원 방안에 대한 논의가 진행되었음을 알 수 있다. 또한 주요 키워드인 ‘지원’ 과 함께 지원 대상을 지칭하는 ‘소상공인’ 그리고 지원 방법을 ‘금융’, ‘고용’, ‘계획’ 이같이 관찰되고 있음을 볼 때 사회문제를 해결하는 방안들에 대한 여러 논의들이 진행되고 있음을 유추할 수 있다. 마지막으로 ‘재난’ 키워드는 코로나19로 인해 발생한 현재의 상황을 재난으로 규정하고 위기 상황에 놓인 사회취약계층에 대한 경제적 지원의 필요성을 제시하는 것으로 볼 수 있다.

한편, ‘지역’, ‘사회’, ‘확산’ 키워드는 대구, 경북, 경기와 같이 지역사회 감염의 확산에 대한 기사들이 관찰된 결과로 볼 수 있다. 또한 코로나 바이러스 확산 초기 사회복지센터 시설을 중심으로 한 집단 감염 현상에 대한 뉴스 역시 의미 연결망에서 중요한 위치를 차지하고 있음을 알 수 있다. 마지막으로 ‘마스크’, ‘방역’ 키워드는 언론 영역에서 코로나 바이러스 확산 초기 마스크 부족 사태와 이후 진행된 마스크 기부 현상에 대한 내용이 중요하게 다루어진 결과로 볼 수 있다.

### 3. 토픽 모델링 결과

<그림 3>은 LDA 기법을 활용한 토픽 모델링 분석 결과를 바탕으로 도출된 6개의 토픽들과, 각 토픽들을 설명할 확률이 높은 상위 10개의 단어를 제시한 결과이다<sup>3)</sup>. 도출된 각 토픽들의 내용적 특성을 살펴보면

3) 그림에서 제시된 토픽 제목은 토픽을 구성한 내용들을 참고하여 연구자가 부여한 것이다.



〈그림 3〉 뉴스 기사 LDA 분석결과 (토픽별 상위 10개 단어)

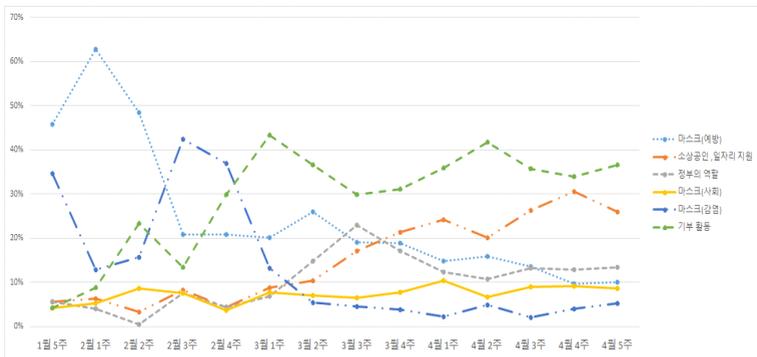
다음과 같다. 먼저 토픽 1과 토픽 4, 토픽 5는 ‘마스크’가 핵심 키워드로 제시된 토픽들이다. 먼저 토픽 1을 살펴보면 감염병 예방의 관점에서 마스크 사용에 필요성을 강조하는 동시에 마스크 지원에 대한 기사들이 분류되었으며, 토픽 4에서는 발병 이후 마스크 부족으로 인해 나타나는 다양한 사회현상에 대한 기사들이 나타나고 있었다. 또한 토픽 5에서는 특정 지역에서의 확진자 증가 현상 가운데 중요 원인으로 마스크 미착용 이 언급되고 있음을 간접적으로 보여준다.

한편, 토픽 2와 토픽 3의 핵심 키워드는 ‘경제’와 ‘지원’으로 나타났다. 토픽 2는 코로나19 사태 이후 경제적으로 어려움을 겪고 있는 소상공인 이나 실업자, 구직자들을 대상으로 한 긴급 지원 대책에 대한 논의들이 주를 이루고 있었다. 토픽 3에서는 코로나19의 확산 이후 나타난 경제 위기 상황을 극복하기 위한 주요 행위자인 ‘대통령’, ‘정부’, ‘국민’, ‘국회’가 등장함과 동시에 ‘추경’에 대한 기사가 같이 분류되었다. 마지막으로 토픽 6은 코로나19 사태 이후 나타난 ‘기부’ 및 ‘성금’ 모금, ‘지원’ 활동에 대한 기사들로 코로나 사태 극복을 위한 사회 구성원들의 노력이

기부의 형태로 나타나고 있었음을 알 수 있다.

〈그림 4〉는 코로나 취약계층에 관련한 언론 기사들 중 어떠한 내용들이 중요하게 다루어졌는지를 시간의 흐름에 따라 살펴보기 위해 토픽 모델링 결과를 토대로 재구성한 것이다. 그림을 살펴보면 코로나19의 확산 초기에는 신종 바이러스의 확산 현상과 예방, 그리고 대응 방안에 대한 논의가 주로 진행되었음을 알 수 있다. 하지만 2월 2주 차 이후에는 방역 마스크 공급 부족과 지역 사회의 감염이 심각해짐에 따라 코로나19 예방을 위한 마스크 착용 대한 논의가 증가하고 있음을 알 수 있다.

한편, 3월 이후 가장 큰 변화는 기부활동 기사의 비중이다. 기부활동 기사는 2월 3주 차 이후 증가하여 3월 이후에는 가장 많은 비중으로 보도되고 있음을 알 수 있다. 이러한 결과를 통해 코로나 취약계층에 대한 지원 보도가 제도나 국가적 지원의 필요성을 강조하기보다는 사회 구성원의 기부 활동에 초점을 맞추고 있음을 보여준다. 또한 3월 3주 차 이후에는 소상공인 및 실업자, 구직자들을 위한 일자리 지원 정책에 있어 정부의 역할을 강조하는 기사 역시 활발하게 작성되었음을 확인할 수 있었다.



〈그림 4〉 시기별 뉴스 기사 토픽 비중 변화

## IV. ‘코로나’ 및 ‘취약계층’에 대한 트위터 분석

### 1. 핵심 키워드 분석

다음 제시되는 <표 2>는 지난 2020년 1월부터 4월까지 작성된 트윗들을 텍스트 마이닝 기법을 이용하여 정형화한 뒤, 출현 빈도를 기준 상위 50개의 단어들을 정리한 것이다.

결과를 살펴보면 트위터 공간에서 가장 높은 출현 빈도를 보인 단어는

**<표 2> 트위터 공간에서 주요 단어 출현 빈도 (상위 50개 단어 기준)**

순위	단어	빈도	순위	단어	빈도	순위	단어	빈도
1	지원	671	17	경기	114	34	재난	85
2	기부	561	19	피해	112	36	나눔	84
3	마스크	494	19	성금	112	36	지역	84
4	확산	247	21	대구	108	38	도움	83
5	극복	238	22	긴급	107	39	네이버	80
6	전달	225	23	사람	103	39	바이러스	80
7	예방	203	24	신종	99	39	천만	80
8	뉴스	192	25	경제	96	42	기탁	76
9	어려움	186	25	밀알복지재단	96	43	관내	73
10	만원	166	27	아동	95	43	추경	73
11	정부	155	27	의료진	95	45	노인	72
12	소상공인	151	27	방역	95	46	장기화	71
13	사회	146	30	방지	93	46	건강	71
14	대응	135	31	필요	91	48	다음	69
15	사태	133	32	지급	90	49	기부금	65
16	국민	115	33	대상	89	49	연예	65
17	감염	114	34	위기	85			

‘지원’(671회), ‘기부’(561 회)로 나타났다. 이는 지난 코로나19 확산 이후 트위터 공간에서 많이 언급되는 콘텐츠가 취약계층을 대상으로 한 기부 미담 사례이기 때문인 것으로 풀이된다. 또한 ‘마스크’(494회), ‘확산’(247회), ‘극복’(238회), ‘예방’(203회)과 같은 단어들이 상위권에 나타난다는 사실을 미루어 볼 때 사이버 공간에서 코로나 바이러스 확산 및 예방에 관한 정보 공유가 활발하게 진행되고 있음을 알 수 있다.

## 2. 의미 연결망 분석

〈그림 5〉는 트위터 공간에서 논의된 코로나 취약계층에 대한 논의 중 출현 빈도 기준 상위 100개 단어를 중심으로 구성한 의미 연결망 결과이다. CNM 알고리즘을 이용하여 군집분석을 수행한 결과, 총 10개



〈그림 5〉 트윗 의미 연결망 (상위 100개 단어 기준)

의 군집이 관찰되었으며, 모듈성 값은 0.673으로 군집 분석 결과에 이상이 없음을 알 수 있었다.

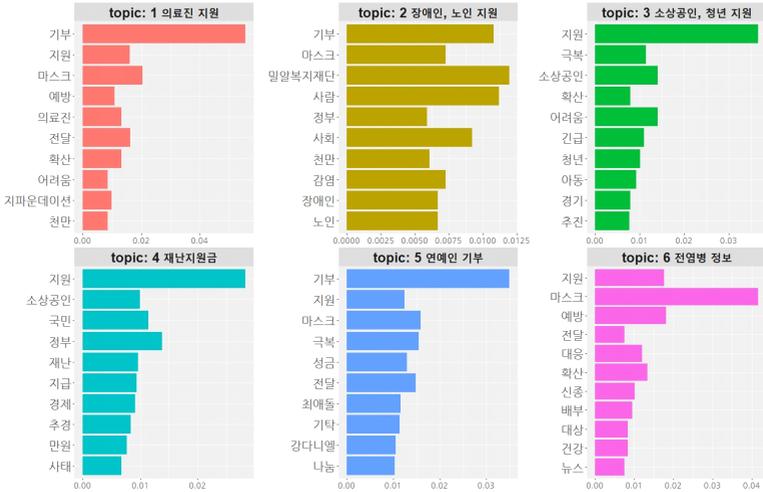
제시된 의미 연결망을 살펴보면 가장 중요한 위치를 차지하고 있는 키워드는 ‘기부’, ‘마스크’, ‘지원’임을 알 수 있다. 이러한 결과는 트위터 공간에서 취약계층에 대한 연예인과 유명인의 마스크 기부 사례에 대한 소식이 중요하게 다루어진 결과로 해석된다. 또한 ‘어려움’, ‘극복’ 키워드 역시 각 군집을 대표하는 단어로 등장하였는데, 이는 코로나 확산으로 인해 발생하는 다양한 사회경제적 어려움을 일종의 사회적 재난으로 바라보면서 취약계층이 처한 현실적 어려움을 전달하는 동시에 극복 방안에 대한 논의 결과로 볼 수 있다. 또한 의미 연결망에서 등장하는 ‘확산’, ‘지원’ 키워드는 코로나 바이러스의 확산 과정에 대한 다양한 정보들, 그리고 전염병의 급속한 확산으로 인해 발생하는 의료진 및 취약계층에 대한 지원 방안에 대해 논의된 결과가 반영된 것으로 풀이된다.

### 3. 토픽 모델링 결과

〈그림 6〉은 LDA 기법을 활용한 토픽 모델링 분석 결과를 바탕으로 도출된 6개의 토픽들과, 각 토픽들을 설명할 확률이 높은 상위 10개의 단어를 제시한 결과이다<sup>4)</sup>. 도출된 각 토픽들의 내용적 특성을 살펴보면 다음과 같다. 먼저 토픽 1, 토픽 2, 토픽 5의 핵심 키워드는 ‘기부’로 나타났다. 토픽 1을 살펴보면 의료진이 겪는 어려움과 물품 부족사태, 그리고 이들에 대한 ‘마스크’ ‘기부’에 대한 트윗들이 관찰되고 있었으며, 토픽 2에서는 ‘장애인’과 ‘노인’에 대한 ‘기부’ 및 ‘지원’에 대한 내용이 다루어지고 있었다. 마지막으로 토픽 5를 통해서는 트위터 공간에서 유명 연예인들의 기부 활동이 중요한 이슈로 다루어지고 있음을 알 수 있었다.

---

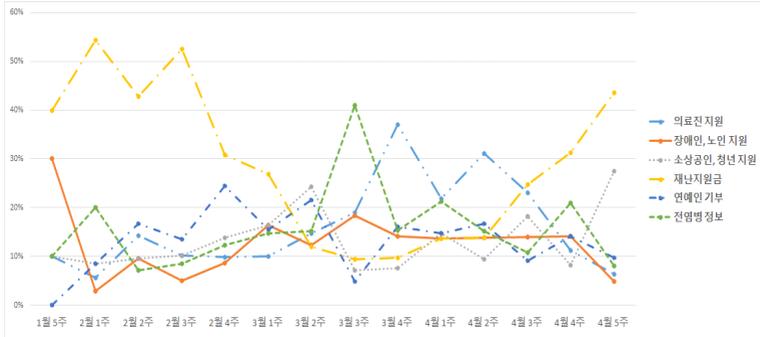
4) 그림에서 제시된 토픽 제목은 토픽을 구성한 내용들을 참고하여 연구자가 부여한 것이다.



〈그림 6〉 트위터 LDA 분석결과 (토픽별 상위 10개 단어)

한편, 토픽 3과 토픽 4의 핵심 키워드는 ‘지원’이었다. 토픽 3에서는 코로나19 사태 이후 ‘소상공인’ 및 ‘청년’, ‘아동’들이 겪는 ‘어려움’과 함께 이들에 대한 ‘지원’ 방안에 대한 논의들이 나타났으며, 토픽 4에서는 ‘정부’의 ‘재난지원금’ ‘지급’이 ‘소상공인’의 어려움 해소에 도움이 될 것이라는 논의가 관찰되었다. 마지막으로 토픽 6을 통해서도 코로나 19의 확산 현상과 이에 대한 ‘대응’, 그리고 전염병 ‘예방’을 위한 다양한 정보들을 담은 ‘뉴스’들이 트위터 공간에서 유통되고 있음을 확인할 수 있다.

〈그림 7〉은 트위터 공간에서 다루어진 코로나 취약계층에 대한 논의들 중 어떠한 내용이 중요하게 다루어졌는지를 시간 순서로 살펴보기 위해 토픽 모델링 결과를 중심으로 구성한 것이다. 먼저 가장 눈에 띄는 사실은 재난지원금에 대한 비중 변화이다. 코로나19 확산 초기에는 지원 정책에 대한 논의가 활발하게 제기되었으나, 2월 3주 이후에는 실질적인 논의가 급격히 감소하였다. 하지만 재난지원금에 대한 내용이 확정되고 기대감이 증가한 4월 이후에는 이에 대한 논의가 다시 증가하고 있음을



〈그림 7〉 시기별 트윗 토픽 비중 변화

알 수 있다. 한편, 주요 취약계층인 장애인 및 노인, 그리고 소상공인과 청년에 대한 지원은 시기에 관계없이 꾸준히 논의가 지속되고 있음을 확인할 수 있었다.

## V. 결론

코로나19에 대한 사회적 염려에도 불구하고, 한국 사회에서는 확산세가 유지·감소하는 실정이다. 물론 현재의 상황이 더욱 악화될 수 있다는 염려 역시 다수 존재하고 있지만, 많은 사람들이 일상을 발병 이전으로 되돌리려는 노력들이 진행되고 있다. 이러한 상황에서 본 연구는 코로나 19로 인해 발생한 사회취약계층의 어려움과 이들을 위한 지원 논의들이 어떻게 진행되었는지를 확인하고자 하였다. 분석 결과를 정리하면 다음과 같다.

우선 언론 영역에서의 코로나 취약계층에 대한 논의를 살펴보면 감염병 확산 방지를 위한 마스크 지원과 코로나19 확산 이후 당면한 사회적 위기를 해소하기 위해 경제적 지원 대책에 대한 논의가 핵심 이슈로 자리 잡고 있었다. 언론매체에서는 코로나19로 발생한 취약계층을 소상공인과 실업자, 구직자들로 상정하고, 이들에 대한 지역사회 및 정부의 지원

정책 필요성을 강조하고 있었다. 반면, 트위터에서는 코로나19로 발생한 취약계층으로 소상공인 이외에도 아동, 노인, 청년 등 연령집단을 중심으로 어려움을 살펴보는 경향이 나타났다. 이러한 차이를 통해 취약계층에 대한 논의가 뉴스 미디어에서 경제적인 측면을 강조하는 반면, 뉴미디어 공간에서는 세대의 문제로 접근하고 있음을 알 수 있다.

이 같은 뉴스 미디어와 뉴미디어 공간에서의 논의 구조 차이는 텍스트의 생산자의 차이에서 발현된다고 볼 수 있다. 뉴스 미디어의 경우, 전통적인 뉴스 생산자들로부터 생산되는 사회구조에 대한 거시적 담론에서부터 뉴스 소비자를 위한 내용에 이르기까지 다양한 층위의 정보들이 나타나고 있었다. 반면, 뉴미디어의 경우, 미디어를 사용하는 특정한 집단이나 텍스트 생산자의 주관이 더욱 적극적으로 반영되었기 때문에 특정 연예인이나 기관, 단체들에 대한 직접적인 언급이 더욱 빈번히 등장하고 있었다. 하지만 이 같은 행위자와 미디어 속성의 차이에도 불구하고, ‘마스크’나 ‘기부’, ‘지원’과 같은 주요 키워드가 각각의 미디어에서 중요하게 다루어진다는 점은 코로나19의 영향력이 사회영역 전반에 미치고 있다는 주장에 대한 근거로 볼 수 있을 것이다.

시기적 특징을 살펴보면 감염병 발생 초기에는 바이러스 전파를 막기 위한 대책으로 마스크에 대한 기부가 강조되었으나, 이후 사태가 지속됨에 따라 경제적 지원에 대한 논의로 이동하고 있음을 알 수 있다. 이러한 결과는 초기에 코로나19의 영향력이 보건 의료 영역에 머물러 있었으나, 시간이 지남에 따라 사회경제 전반으로 확대되었음을 간접적으로 시사한다.

한편, 언론 미디어와 뉴미디어 모두 취약계층에 대한 기부활동을 중요하게 다루고 있었다. 뉴스에서는 사회구성원의 기부 및 성금 모금, 지원에 대해 강조하는 기사가 작성되었으며, 트위터 공간에서도 유명 연예인들의 기부 활동이 중요하게 다루어지고 있었다. 물론 코로나19로 인해 발생한 위기 상황에서 사회구성원들의 다양한 기부 활동들이 소개되는 상황은 사회통합에 긍정적 영향을 미칠 것으로 보인다. 하지만 이러한

답론들이 미담의 소개의 형태에서 머물고 있으며, 취약계층을 위한 국가·제도적 지원의 필요성을 제기하지 못한다는 점에서 아쉬움이 남는다.

이 연구는 텍스트 마이닝과 의미 연결망 분석, 그리고 토픽 모델링 기법을 적용하여 뉴스와 트위터 공간에서 나타난 코로나 취약계층에 대한 논의를 제시하였다. 이를 통해 뉴스 미디어 및 뉴미디어 공간에서 사회적 약자에 대한 논의들이 어떻게 전개되는지를 탐색함에 있어 입체적인 동향 분석이 가능함은 물론, 객관적이고, 심층적 이해를 충족시키는 연구가 될 수 있을 것이다. 특히 이번 연구에서는 코로나 사태 이후 진행된 취약계층에 대한 언론 및 트위터 공간에서의 담론 구조의 변화를 정량적이고, 재현 가능성이 높은 방법을 동원하여 분석을 진행하였다는 점, 그리고 논의 구조와 변화의 양상을 시각화하여 제시하였다는 점에서 일반적 메타분석과는 차별성을 지닌다.

그럼에도 불구하고 연구에서 사용된 분석 기법들의 적합성이 학계에서 충분히 검증되지 않았다는 점, 그리고 많은 부분 정량적인 분석의 형태를 가지고 있음에도 불구하고 토픽 수 결정과 같은 부분에서 아직까지 연구자의 판단이 요구된다는 사실은 연구 결과를 해석하고 일반화하는데 있어 주의를 기울여야 하는 지점이다. 또한 토픽모델링 기법과 의미 연결망 분석 결과를 해석하는데 있어 내용 분석적 측면이 결여되어 있다는 점은 이 연구 결과를 일반화 하는데 있어 중요한 한계점으로 존재한다. 따라서 본 연구의 결과를 과도하게 일반화 하거나 절대적인 진리로 받아들이기 보다는 한국사회에서 진행된 코로나 취약계층에 대한 논의의 틀을 대략적이지만 객관적이고 빠르게 파악하는 하나의 사례로 받아들여야 할 것이다.

마지막으로 현재 코로나19로 발생한 일련의 사건들이 종식되지 않았다는 사실을 고려할 때 취약계층에 대한 사회적 논의가 더욱 확대될 가능성이 존재한다. 따라서 본 연구의 결과가 감염 취약계층에 대한 인식 전환 및 제도 개선으로 이어져 사회적 갈등을 해소하는 단초가 되기를 바란다.

## 참고문헌

- 김세현. 2018. “비정형자료분석을 통해 살펴본 한국의 다문화 연구.” 『한국인구학』 41(1): 1-27.
- 김수경. 2020. “감염병, 이념, 제노포비아: ‘코로나 19’의 정치화와 반중(反中) 현상.” 『다문화와 평화』 14(1): 22-43.
- 김용학. 2015. “한국 대중가요의 의미 연결망: 1960년대부터 2000년대까지의 변화를 중심으로.” 『대중서사연구』 21(1): 145-171.
- 김태완. 2020. “코로나 19로 인한 영향 및 사회정책 대응 방안.” 『보건복지 Issue & Focus』 385:1-8.
- 신동훈 · 김세현. 2020. “텍스트 마이닝 기법을 활용한 국내 외국인 유학생 연구 동향 분석: 시기별 연구주제 변화 탐색.” 『교육학연구』 58: 333-381.
- 안주영 · 안규빈 · 송민. 2016. “텍스트 마이닝을 이용한 매체별 에볼라 주제 분석: 바이오 분야 연구논문과 뉴스 텍스트 데이터를 이용하여.” 『한국문헌정보학회지』 50(2): 289-307.
- 통계청. 2020. “2020년 3월 고용동향.”
- 황교상 · 이태식 · 이현록. 2014. “센서스 데이터를 기반으로 만든 전염병 전파 시뮬레이션 모델.” 『대한산업공학회지』 40(2): 163-171.
- Arun, R., Suresh, V., Madhavan, C. V., & Murthy, M. N. 2010. “On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations.” *In Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Berlin, Heidelberg: Springer,
- Blei, D. M. 2012. “Probabilistic Topic Models” *Communications of the ACM* 55(4): 77-84.
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. 2009. “A Density-

- based method for adaptive LDA model selection.” *Neuro-computing*, 72(7-9): 1775-1781.
- Chen, Y., Yu, B., Zhang, X., & Yu, Y. 2016. “Topic Modeling for Evaluating Students’ Reflective Writing: A Case Study of Pre-Service Teachers’ Journals.” *In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*.
- Clauset, A., Newman, M. E., & Moore, C. 2004. “Finding community structure in very large networks.” *Physical Review E* 70(6): 066111.
- Deveaud, R., SanJuan, E., & Bellot, P. 2014. “Accurate and Effective Latent Concept Modeling for Ad Hoc Information Retrieval.” *Document Numérique* 17(1): 61-84.
- DiMaggio, P., Nag, M., & Blei, D. 2013. “Exploiting Affinities between Topic Modeling and The Sociological Perspective on Culture: Application to Newspaper Coverage of US Government Arts Funding.” *Poetics* 41(6): 570-606.
- Griffiths, T. L., & Steyvers, M. 2004. “Finding Scientific Topics.” *Proceedings of the National Academy of Sciences* 101 (suppl 1): 5228-5235.
- Grimmer, J., & Stewart, B. M. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21(3): 267-297.
- Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. 2016. “Big Social Data Analytics in Journalism and Mass Communication: Comparing Dictionary-based Text Analysis and Unsupervised Topic Modeling.” *Journalism & Mass Com-*

- munication Quarterly* 93(2): 332-359.
- Househ, M. 2016. "Communicating Ebola through Social Media and Electronic News Media Outlets: A Cross-Sectional Study." *Health Informatics Journal* 22(3): 470-478.
- IMF. 2020. *World Economic Outlook, 2020*.
- Kleinnijenhuis, J., De Ridder, J. A., Rietberg, E. M., & Robberts, C. W. 1997. "Reasoning in Economic Discourse: An Application of the Network Approach in Economic Discourse." *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*. Mahwah: Erlbaum 191-207.
- Mützel, S. 2015. "Facing Big Data: Making Sociology Relevant." *Big Data & Society* 2(2): 2053951715599179.
- Quirin, A., Cerdón, O., Guerrero-Bote, V. P., Vargas-Quesada, B., & Moya-Anegón, F. 2008. "A Quick MST-based Algorithm to Obtain Pathfinder Networks ( $\infty$ ,  $n-1$ )." *Journal of the American Society for Information Science and Technology* 59(12): 1912-1924.
- Steyvers, M., & Griffiths, T. 2007. "Probabilistic Topic Models." *Handbook of Latent Semantic Analysis*, 427(7): 424-440.
- Theodoridis, S., & Koutroumbas, K. 2006. "Clustering: Basic Concepts" *Pattern Recognition* 483-516.
- Towers, S., Afzal, S., Bernal, G., Bliss, N., Brown, S., Espinoza, B., & Mamada, R. 2015. "Mass Media and the Contagion of Fear: The Case of Ebola in America." *PloS one* 10(6).

# **Korean Society's Perception of the Covid-19 Vulnerable Group: Focusing on the News Media and Twitter Discourse**

Sehyun Kim, Hanyang University

## **Abstract**

The study aims to empirically verify discussions about the vulnerable groups in Korean society after the spread of Covid-19. From January to April 2020, news media articles and Twitter posts about Covid-19 vulnerable groups have been collected, and text mining, semantic network analysis, and topic modeling techniques have been used for analysis. The analysis results are summarized as follows. First, in the media area, the vulnerable groups caused by Covid-19 were assumed to be small business, unemployed, and job seekers, while on Twitter, in addition to small business, difficulties were discussed focusing on age groups such as children, the elderly, and youth. Second, looking at the flow of discussion, measures to prevent the spread of the virus were discussed in the early stages of infectious disease, but later turned to a discussion of economic difficulties. Third, the news media and Twitter both deal with donations to vulnerable groups as an important issue.

**Keywords:** Covid-19, Infection Vulnerable Group, Text Mining, Meaning Network, Topic Modeling