JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 25, No. 4: 495~500, Apr. 2021

대규모 빅데이터 분석 기반 COVID-19 Pandemic 분석결과

김나현[1] · 오하영[2*]

Analysis of COVID-19 Pandemic based on Massive Big Data Analysis

Na-Hyeon Kim[1] · Hayoung Oh[2*]

[1]Undergraduate Student, Sungkyunkwan University, Seoul, 03063 Korea
[2*]Associative Professor, College of Computing & Informatics, Sungkyunkwan University, Seoul, 03063 Korea

요 약

본 연구의 목적은 본 논문의 목적은 최근 확산되고 있는 코로나바이러스감염증-19로 인한 위기를 국내 뉴스 빅데이터를 활용하여 규명하는 것이다. 본 논문은 COVID-19로 인한 위기와 관련된 언론기사들을 한국언론진흥재단의 뉴스 빅데이터 분석 시스템 '빅카인즈(BIGKinds)'를 활용하여 분석하였다. 본 논문에서는 약 10개월의 기간을 설정한 후, '코로나'와 '위기' 키워드를 중심으로 총 54개의 언론사의 관련기사들을 추출하였다. '코로나'와 '위기' 두 키워드 간의 상관계수를 파악하고, 연관어 분석을 통해 경제, 사회, 국제, 문화 각 대표 카테고리 별로 COVID-19로 인해 어떤 위기를 맞고 있는지 파악하고자 한다. COVID-19 사태는 경제, 사회 등 모든 부분에 큰 타격을 주고 있는 만큼 빅데이터를 활용한 본 논문은 COVID-19 사태 위기 극복을 위한 기초자료로 활용될 수 있을 것으로 기대된다.

ABSTRACT

This paper is to identify the recent growing crisis from coronavirus infections-19, using domestic news big data. This paper analyzed media articles related to the crisis caused by COVID-19 using the Korea Press Foundation's news big data analysis system 'BIGKinds'. In this paper, a total of 54 media articles were extracted around the keywords 'Corona' and 'Crisis', after a period of about 10 months. We want to understand the correlation coefficient between the two keywords "Corona" and "Crisis" and to understand what kind of crisis the COVID-19 is facing for each representative category of economy, society, international and cultural. As the COVID-19 crisis is taking a heavy toll on the economy, society and any other categories, this research using big data is expected to be used as a basic data to overcome the crisis of COVID-19.

키워드 : 코로나19, 감염병, 뉴스 빅데이터, 언론 보도 분석, 상관계수

Keywords : COVID-19, Infectious disease, News big data, News analysis, Correlation coefficient

Received 30 December 2020, **Revised** 24 February 2021, **Accepted** 12 March 2021

*** Corresponding Author** Hayoung Oh (E-mail: hyoh79@gmail.com, Tel:+82-2-583-8585)
Associative Professor, College of Computing & Informatics, Sungkyunkwan University, Seoul, 03063 Korea

Open Access http://doi.org/10.6109/jkiice.2021.25.4.495 print ISSN: 2234-4772 online ISSN: 2288-4165

ⓒThis is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/li-censes/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright Ⓒ The Korea Institute of Information and Communication Engineering.

Ⅰ. 서 론

2019년 12월 중국 우한에서 처음 발생한 코로나바이러스감염증-19(COVID-19)로 인해 2020년 1월 20일 국내에서 COVID-19에 감염된 첫 확진자가 발생했고[1], 2020년 8월 15일 0시기준 COVID-19 국내 총 누적 확진자 수는 15,039명, 누적 사망자 수는 305명에 이르렀다.[2] 세계보건기구(World Health Organization, WHO)는 전세계로 확산되는 COVID-19에 대응하여 2020년 1월 30일 '국제적 공중보건 비상사태 (PHEIC)'를 선포하였으며, 이후 3월 11일 홍콩독감, 신종플루에 이어 사상 세 번째로 '팬데믹(Pandemic)'을 선포하였다. COVID-19 사태는 단순한 공중보건의 위기를 넘어서서 정치·경제·사회·문화 등 모든 분야에 걸쳐 영향을 미치는 심각한 위기이다. 특히나 신종 감염병 중에서도 호흡기 질환인 COVID-19는 전파력이 강하고 파급효과가 크기 때문에 장기적인 국면으로 이어졌을 때 각국에 미치는 영향이 상당히 치명적이다.

COVID-19의 멈출 줄 모르는 확산으로 전 세계는 각국의 국민들에게 사회적으로 거리 두는 것을 권장하고 있다.[3] 국내에서도 COVID-19 확산에 따라 정부는 2020년 2월 23일 감염병 위기경보 단계를 '경계' 단계에서 최고 단계인 '심각' 단계로 격상했다. 정부는 중앙재난안전대책본부를 설치하여 중앙정부 및 지방자치단체의 지원 체계를 한층 강화하고, COVID-19의 해외 유입 차단, 환자 발견 및 접촉자 격리 등의 봉쇄정책, 지역사회 확산 차단 및 최소화를 위한 조치를 추진했다.[4] 초기의 사회적 거리 두기에 대한 요구 기간은 몇 주에 불과했지만, 이는 몇 달로 확대되어 어느새 COVID-19는 장기적인 싸움이 되었다.

이러한 Pandemic 상황일수록 언론의 뉴스 보도량은 크게 증가하며, 언론은 감염병 위험을 알리는 경고자이자 위험 인식의 확산자, 국민 의견의 전달자로서의 역할[5]을 수행한다. 대부분의 언론사에서 카테고리를 정치, 경제, 사회, 국제, 문화, 스포츠 등으로 나누고 있기 때문에 각 영역별 뉴스 빅데이터 분석은 COVID-19로 인한 위기를 진단하는데 효과적이다. 다만, 본 연구에서는 각 언론사들의 정치적 성향에 따라 특정 인물과 단체의 유불리한 내용을 보도하여 객관성을 잃을 수 있다는 점을 우려하여 '정치' 카테고리에 대한 분석은 제외하였다. 소수 언론사의 뉴스로는 신뢰성, 편향성 등의 문제를 일으킬 수 있지만 방대한 양의 뉴스 데이터베이스에 빅데이터 분석 기술을 접목한다면, 이는 COVID-19로 인한 위기를 진단하는 데에 객관적인 지표가 될 수 있다.[6] 본 논문의 구성은 다음과 같다. 1장 서론을 시작으로 2장에서는 뉴스 데이터 수집 및 추출 과정을, 3장에서는 각 데이터 분석 결과가 의미하는 바에 대해 설명한다.

Ⅱ. 연구 방법

2.1. 연구 대상 및 범위

본 연구는 경제, 사회, 국제, 문화 카테고리별 COVID-19로 인한 위기를 국내 주요 신문사와 방송사 뉴스를 통해 분석하고자 한다. 분석 대상 뉴스들은 한국언론진흥재단이 제공하는 뉴스 빅데이터 분석 시스템 '빅카인즈(BIGKinds)'를 활용하여 수집했다. 빅카인즈 서비스는 중앙지, 경제지, 지역종합지, 방송사, 전문지를 비롯한 총 54개 매체로부터 수집한 뉴스로 구성된 통합 데이터베이스에 빅데이터 분석 기술을 접목하여 만든 새로운 뉴스 분석 서비스이다. 분석기간은 중국 우한에서 COVID-19가 발생한 뒤 최초로 국내에서 첫 확진자가 발생한 2020년 1월부터 2020년 10월까지 총 10개월(2020.01.01-2020.10.31)이다. 데이터 수집을 위한 검색어는 '코로나', '위기'로 선정했다. 이를 바탕으로 하여 빅카인즈로부터 데이터 전처리와 형태소 추출이 완료된 정제 데이터를 제공 받아 키워드 트렌드 분석과 연관어 분석을 실시했다.

2.2. 데이터 수집 및 추출

연구에서는 경제, 사회, 국제, 문화 카테고리별로 '코로나'와 '위기'에 대한 키워드 트렌드 분석과 연관어 분석을 실시했다. 먼저, 키워드 트렌드 분석을 통해서는 2020년 1월부터 2020년 10월까지 빅카인즈로부터 제공받은 기사 건수 데이터의 월간 추이를 키워드별 살펴볼 수 있다.

또한, 경제, 사회, 국제, 문화 카테고리별로 COVID-19에 따른 위기의 정도는 다를 것이며, 이는 '코로나'와 '위기'라는 두 키워드 간의 피어슨 상관계수를 통해서 파악했다. 표 1과 같이 '코로나'와 '위기'라는 두 변수 간의 관련성을 나타내는 피어슨 상관계수는 +1에 가까울수록 강한 비례관계를 의미하며, -1에 가까울수록 강

한 반비례관계를 의미한다.

Table. 1 Pearson correlation coefficient (Pearson's r)

Range	Strength of association
0	No association
0 to ±0.25	Negligible association
±0.25 to ±0.50	Weak association
±0.50 to ±0.75	Moderate association
±0.75 to ±1	Very strong association
±1	Perfect association

언론기사에서 '위기'라는 키워드는 다른 개념들과 함께 나타나기 마련이다. 그 연관어들을 분석하면 언론에서 각 카테고리별로 어떤 위기를 보도하고 있는지 알 수 있다. 모든 연관어 분석은 1,000건의 뉴스를 대상으로 하였으며, 데이터 유형은 가중치와 키워드 빈도수를 기준으로 하였다. 가중치 정보는 사용자가 입력한 질의어에 기반하여 검색한 결과에 대해 공식적 분석과 워드 클러스터링을 하여 동적으로 상호 작용하는 시맨틱 네트워크를 생성하는 토픽랭크 알고리즘에 기반해 출력된 결과이며, 키워드 빈도수는 가중치 순으로 선정된 연관어를 해당 연관어가 등장한 횟수 기준으로 재정렬한 결과이다.

Ⅲ. 데이터 분석 결과

3.1. 경제

'경제' 카테고리 내의 '코로나', '위기' 키워드에 대한 뉴스는 총 128,057건이 추출되었다. 그 중 정확도가 높은 기사 20,000건을 대상으로 월별 추이를 분석한다. 그림 1과 같이 경제 관련 기사에서 코로나로 인한 위기를 다룬 기사는 2020년 04월에 3.017건으로 가장 많았고, 03월, 05월이 뒤를 이었다.

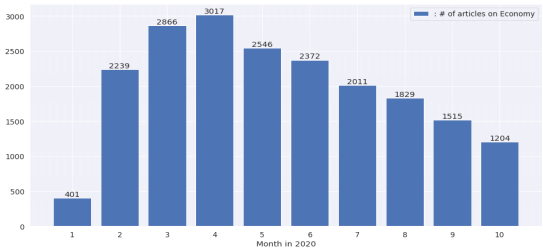

Fig. 1 Bar graph of Economic news articles

'코로나, 위기'에 대한 연관 키워드 분석 결과, 1위는 '기업들'(가중치 79.86, 키워드 빈도수 756), 2위는 '중소기업'(가중치 67.99, 키워드 빈도수 731), 3위는 '장기화'(가중치 52.67, 키워드 빈도수 283) 순으로 결과가 나타났다. 이는 다음 표 2와 같다. 그림 2는 가중치 순으로 선정된 연관어를 해당 연관어가 등장한 횟수 기준으로 재정렬한 워드클라우드를 나타낸다.

Table. 2 Associated keywords for 'COVID-19', 'Crisis' in Economy

rank	name	weight	tf
1	companies	79.86	756
2	small businesses	67.99	731
3	protracted	52.67	283
4	Europe	43.93	389
5	small business owner	40.78	352
6	pandemic	38.87	170
7	China	33.16	628
8	direct hit	32.61	126
9	virus infections	29.87	284
10	experts	28.73	164
11	car	27.3	765
12	self-employeed	25.11	149
13	manufacturing	24.83	499
14	economic slump	24.62	118
15	Samsung Electronics	23.5	325
16	world economy	23.49	135
17	semiconductor	21.56	453
18	Hyundai Motors	21.21	273
19	major company	19.66	279
20	first half	17.76	300

Fig. 2 Wordcloud of Associated keywords in Economy

연관어 분석 결과 상위 20개의 키워드 중에서 경제와

관련성이 높은 키워드로는 기업들, 중소기업, 소상공인, 직격탄, 자동차, 자영업자, 제조업 등이 있다. 분석 결과로 미루어 볼 수 있듯, 경제 카테고리 내 언론은 '기업'을 중점적으로 다루고 있다. 더불어, 중소기업, 소상공인, 자영업자들이 받는 경제적인 타격을 집중적으로 다룬다.

표 3은 각 카테고리별 '코로나'와 '위기' 두 키워드 간의 상관계수를 보여준다. 전반적으로 경제, 사회, 국제, 문화 모든 카테고리에서 상관계수는 뚜렷한 양적 선형관계를 보여주며 COVID-19로 인한 위기는 모든 영역에서 이슈가 되고 있음을 파악할 수 있다. 하지만, 경제 카테고리 내에서의 상관계수는 0.8246로 강한 양적 선형관계를 보이고 있다. 이는 언론보도에서 COVID-19로 인한 위기는 다른 영역들보다도 경제에 미치는 타격을 강조하고 있음을 시사한다.

Table. 3 Correlation coefficient between 'COVID-19' and 'Crisis'

Category	Correlation coefficient
Economy	0.8246
Society	0.6946
International	0.5652
Culture	0.6744

한국은행에 따르면 올해 2분기 국내총생산(GDP) 성장률이 전분기 대비 -3.2%로 집계됐다고 밝혔다. 3.2%의 역성장은 IMF(국제통화기금) 외환위기 당시인 1998년 4분기(-3.8%) 이래 21년 6개월 만에 가장 낮은 실적으로 기록됐다. 대부분의 나라가 그렇듯 전체 산업 중 GDP에서 차지하는 비중은 제조업과 서비스업이 대부분이다. 우리나라의 경우 국가GDP 대비 제조업의 비중은 약 28%이고 서비스업의 비중은 약 62%이다. 때문에, COVID-19로 인한 제조업과 서비스업의 경영악화는 곧 GDP성장률 악화와 직결되어 있다.

COVID-19로 인한 경기 위축으로 공장에서 생산한 제품이 팔리지 않고 재고로 쌓여가고 있다. 가파르게 상승중인 제조업 출하 대비 재고 비율은 3월 110.8%에서 5월 128.8%까지 치솟았다. 이는 곧 평균가동률의 하락으로 이어지는데 통계청 통계포털인 코시스(kosis)에 따르면 1~8월 평균 제조업 가동률은 70.0%로 IMF외환위기가 일어난 1998년(67.6%) 이후 22년만에 최저치를 나타내고 있다.[7] 특히, 제조업 중에서도 글로벌 공급망과 해외 수출에 크게 의존하는 자동차 업계는 가장 큰 타격을 맞으며 위기에 처해 있다. 연관어 분석 결과에서도 알 수 있듯 '자동차'는 COVID-19와 위기를 다룬 경제 관련 뉴스 기사에서 가장 빈번하게 언급된 키워드이기도 하다.

COVID-19가 확산되면서 위축된 소비심리로 인해 서비스업도 직격탄을 맞았다. 한국은행의 소비자동향조사에 따르면 실제로 2020년 1월까지 100을 상회하던 소비자 심리지수는 2020년 2월, 3월, 4월을 지나며 96.9, 78.4, 70.8로 하락하는 모습을 보였다. 소비자 심리지수의 하락은 국내 경제상황 및 향후 경기에 대한 부정적인 전망을 나타낸다. 특히, 사회적 거리두기로 인해 외출 횟수가 감소하고 실내 공간에서 생활하는 시간이 늘어남에 따라[8] 요식업, 도소매업, 여행레저업, 운수창고업, 숙박업계의 중소상공인과 기업은 실적 부진 뿐만 아니라 그로 인한 고용 문제까지 겪고 있다.[9]

GDP, 소비자심리지수 등 경제관련 지표와 위의 COVID-19로 인한 위기를 다룬 경제 관련 기사 건수는 유사한 추이를 보이고 있다. 최악의 수치를 나타냈던 2분기의 GDP성장률은 3분기에 들어 전년 동기 대비 4.9% 회복했다. 내수는 긴급재난지원금의 지급과 방역강도의 완화로 회복세를 보이고 있다 하더라도 수출과 제조업에서는 여전히 부진한 흐름이 이어지고 있는 실정이다.

3.2. 사회

'사회' 카테고리 내의 '코로나', '위기' 키워드에 대한 뉴스는 총 108,980건이 추출되었다. 그 중 정확도가 높은 기사 20,000건을 대상으로 월별 추이를 분석한 결과, 사회 관련 기사에서 코로나로 인한 위기를 다룬 기사는 2020년 2월에 2,991건으로 가장 많았고, 03월, 08월이 뒤를 이었다.

'코로나, 위기'에 대한 연관 키워드 분석 결과, 1위는 '확진자'(가중치 67.64, 키워드 빈도수 772), 2위는 '장기화'(가중치 41.07, 키워드 빈도수 222), 3위는 '사람들'(가중치 40.06, 키워드 빈도수 248) 순으로 결과가 나타났다. 이는 다음 표 4와 같다. 그림 3은 가중치 순으로 선정된 연관어를 해당 연관어가 등장한 횟수 기준으로 재정렬한 워드클라우드를 나타낸다.

Table. 4 Associated keywords for 'COVID-19', 'Crisis' in Society

rank	name	weight	tf
1	confirmed patient	67.64	772
2	protracted	41.07	222
3	people	40.06	248
4	virus infections	36.97	344
5	depression	32.62	186
6	social distance	30.12	157
7	medical staff	22.02	179
8	anxiety	20.06	164
9	Ministry of Health and Welfare	19.77	150
10	confirmed	19.48	178
11	quarantine authorities	19.11	30
12	Incheon	19.02	275
13	infectious disease	18.13	376
14	experts	18	103
15	job	16.96	165
16	Department of Education	16.8	237
17	KCDC	16.2	99
18	KCTU	15.56	66
19	Metropolitan area	14.87	160
20	medical world	14.05	130

Fig. 3 Wordcloud of Associated keywords in Society

연관어 분석 결과 상위 20개의 키워드 중에서 사회와 관련성이 높은 키워드로는 사람들, 우울감, 의료진, 불안감, 확진 판정, 일자리 등이 있다. 사회 카테고리 내 언론은 '사람'을 중점적으로 다루고 있다. 사회적인 측면에서 봤을 때 COVID-19의 가장 큰 위기는 국민들이 겪는 심리적 불안감이다. 최근 '코로나블루'라는 신조어까지 등장했는데, 이는 COVID-19 확산으로 일상에 큰 변화가 닥치면서 생긴 우울감이나 무기력증을 뜻한다. 경기연구원은 코로나19 장기화로 인해 우리나라 국민 47.5%가 불안과 우울감을 경험하고 있다고 밝혔다.[10] 한국건강증진개발원의 '코로나19로 인한 건강 상태' 조사 결과에 따르면 코로나블루를 경험한 응답자는 그 원인으로 '외출 및 모임 자제로 인한 사회적 고립감'(32.1%), '감염 확산에 따른 건강 염려'(30.7%), '취업 및 일자리 유지의 어려움'(14.0%)을 꼽았다. COVID-19과 같은 감염병의 특성상 사회적 거리두기, 자가격리 등으로 일상 생활이 중단될 수 있기 때문에 개인의 심리적 우울감으로 인한 위험은 더 크다. 학생, 의료진, 직장인 누구 하나 예외 없이 보이지 않는 바이러스로 인한 공포감과 그로 인한 스트레스를 함께 겪으며 심리적 탈진 상태인 번아웃(Burn out)을 경험하고 있다. 또한, 일부 확진자를 슈퍼전파자라 부르며 무차별적 공격을 일삼는 사회적 낙인 현상 역시 사회문제로 대두되고 있다. 마지막으로, 청년층의 고용문제와 실업문제 역시 심각한 사회문제로 여겨지고 있다. 앞서 언급한 경제 위기 상황 속에서 기업은 인력 감축과 신규 채용 연기의 방식으로 대응하고 있는데 이 모든 것은 청년, 비정규직, 영세영업자 등 취약집단의 몫이 되기 때문이다.

Ⅳ. 결 론

COVID-19로 인한 위기는 전반적으로 경제, 사회, 국제, 문화 모든 영역에서 이슈가 되고 있지만, 언론보도에서는 특히 경제에 미치는 타격을 강조하고 있다. 제조업의 경영악화, 위축된 소비심리로 인한 서비스업의 직격탄, 중소상공인들의 생계유지 어려움 등 경제적인 측면에서는 해결해야 하는 과제가 많다. 그 외 영역의 데이터 분석 결과, 사회 카테고리에서는 국민들이 겪는 심리적 불안감과 사회적 낙인 현상, 고용 문제와 같은 사회문제가 위기로 작용하며, 국제 카테고리에서는 각국의 위기대응능력과 국제관계가, 문화 카테고리에서는 영화, 뮤지컬, 연극, 대중음악 등 취소된 공연들로 인한 피해 등이 위기로 작용함을 발견할 수 있었다.

REFERENCES

[1] Korea Centers for Disease Control and Prevention (KCDC). Confirmation of confirmed cases of COVID-19 from abroad

during quarantine phase, raising the alert for the crisis of infectious disease to the 'caution' stage, and responding [Internet]. Available: http://www.mohw.go.kr/react/al/sal0301vw.jsp?PAR_MENU_ID=04&MENU_ID=0403&page=1&CONT_SEQ=352435.

[2] Korea Centers for Disease Control and Prevention (KCDC), Coronavirus Infection-19 Domestic Status [Internet]. Available: http://ncov.mohw.go.kr/tcmBoardView.do?contSeq=359024.

[3] J. Kent. (2020). COVID-19 data shows how social distancing impacts virus spread. *Health IT Analytics* [Internet]. Available: https://healthitanalytics.com/news/covid-19-data-shows-how-social-distancing-impacts-virus-spread.

[4] Korea Centers for Disease Control and Prevention (KCDC), Coronavirus Infection-19 Government Meeting Briefing [Internet]. Available: http://ncov.mohw.go.kr/tcmBoardView.do?brdId=&brdGubun=&dataGubun=&ncvContSeq=353064&contSeq=353064&board_id=&gubun=ALL.

[5] A. Y. Kang. (2020, Feb), The reports of infectious diseases should be more accurate than anything else... Need to discuss guidelines. *Korean Journalists Association* [Internet]. Available: http://journalist.or.kr/m/m_article.html?no=47266.

[6] R. Casero and Andreu, "Impact of Covid-19 on the media system. Communicative and democratic consequences of news consumption during the outbreak," *El profesional de la información*, vol. 29, no. 2, 2020.

[7] W. S. Jeong. (2020, Oct). The worst manufacturing cold wave in 22 years. The Korean economy is mired in a structural adjustment dilemma. *Chosunbiz* [Internet]. Available: https://biz.chosun.com/site/data/html_dir/2020/10/22/2020102200665.html.

[8] A, L. Andersen, E, T. Hansen, N. Johannesen, and A. Sheridan, "Consumer Responses to the COVID-19 Crisis: Evidence from Bank Account Transaction Data," *CEBI Working Paper Series*, 2020.

[9] H. S. Park, J. H. Pan, H. C. Jung, and S. J. Kim, "The Impact of the Covid-19 on the Seoul Economy and Countermeasures for Small Business and Tourism," Policy Report of the Seoul Institute, pp. 1-32, 2020.

[10] K. D. Ha. (2020, June). Psychological instability, social stigma, economic crisis, medical team burnout...How do you deal with Corona Blue?. *Medicatenews* [Internet]. Available: https://www.medigatenews.com/news/3163621084.

김나현(Na-Hyeon Kim)

성균관대학교 경영학과 재학

※ 관심분야 : 금융, 경제 및 데이터 분석

오하영(Hayoung Oh)

Sungkyunkwan University Professor (2019~)

Ajou University Professor (2016~2019)

Soongsil University Professor (2013~2016)

Ph.D. in computer engineering at Seoul National University (2013)