

# 통계 분석과 뉴스 기사 토픽 모델링을 통한 코로나19가 자살에 미치는 영향 분석

권민지\*, 김준철\*

\*서울기술연구원 데이터사이언스센터

mjkwon@sit.re.kr, kjc@sit.re.kr

## The Effect of Covid-19 on Suicide through Statistical Analysis and Topic Modeling of News Articles

Minji Kwon\*, Junchul Kim\*

\*Center for Data Science, Seoul Institute of Technology

### 요 약

전 세계적으로 확산된 코로나19의 장기화로 인해 국민들은 경제적, 심리적 어려움을 겪고 있으며, 이에 따른 자살 시도에 대한 우려가 높아지고 있다. 본 연구에서는 자살사망자 통계와 자살 관련 뉴스 기사의 토픽 모델링을 통해 코로나19가 자살에 미친 영향을 분석하였다. 그 결과 수치적으로는 재난 직후 자살률이 일시적으로 감소하는 ‘허니문 기간’을 보였고, 의미적으로는 자살 예방에 대한 중요성이 지속적으로 부각되었다. 또한 유명인 또는 사회적으로 이슈화된 사건에 대한 수사 및 사실관계가 언론을 통해 드러났으며, 연초를 지나도 꾸준히 유지되는 경제 관련 이슈가 도출되었다.

### 1. 서론

대한민국은 OECD 국가 중 자살률 1위<sup>1)</sup>를 기록하고 있으며, 2020년도 기준 사망원인 중 자살이 5위(4.3%)<sup>2)</sup>를 차지했다. 한편 전 세계적으로 코로나 바이러스가 확산되고 장기화되며, 국내외 경제는 생산량이 감소하고 실업률이 증가하는 등 타격을 받고 있으며,[1] 심리적으로도 코로나19 이후 42.8%가 우울을, 36.1%가 불안을 겪고 있다.[2]

이에 본 연구에서는 자살사망자 통계와 자살 관련 뉴스 기사를 수집해 수치적, 의미적 분석을 수행하여 코로나 전후의 변화를 도출하였다.

### 2. 선행연구

대량의 뉴스 기사 데이터에서 키워드를 추출하고 사회 이슈 및 트렌드 변화를 도출하는 연구들이 수행되어왔다. 47,816건의 코로나19 관련 뉴스 기사 분석을 통해 위기경보 4단계별 토픽을 도출해 시사점을 도출하였으며,[3] 8,863건의 국내 KCI 등재지에서 토픽 모델링을 통해 도시와 관련된 연구동향을 분석하였다.[4] ‘코로나 블루’ 키워드를 포함하는 국내 포털 사이트 및 소셜미디어 글 6,976건을 수집해 분석한 연구에서는 ‘우울감’, ‘극복’, ‘장기화’, ‘마음’, ‘불안’ 등이 빈번히 나타난 키워드로 도출되었다.[5]

### 3. 분석 데이터

#### 3.1 자살사망자 통계

통계청(KOSIS)에서 2019.01.01.~202108.31. 기간의 월별 자살사망자 통계 데이터를 수집하였다.<sup>3)</sup>

#### 3.2 뉴스 기사

빅카인즈<sup>4)</sup> 뉴스 검색을 통해 2019.01.01.~202108.31. 기간의 ‘자살’, ‘극단적 선택’ 키워드를 포함하는 뉴스 기사 URL을 수집한 뒤, 파이썬 newspaper 라이브러리를 활용하여 본문을 크롤링하였다. 전체기간 동안 중복된 데이터를 제거한 결과, 수집된 데이터는 총 48,016건이다.

### 4. 분석 방법론

#### 4.1 데이터 전처리

뉴스 기사 데이터 전처리는 수집된 뉴스 기사에서 중복된 데이터를 제거하고 MeCab 형태소 분석기를 활용하여 뉴스 기사 본문에서 명사를 추출하고 bigram을 활용해 빈번하게 연속해서 출현하는 단어들을 묶어주었다. 불용어(Stopword)는 검색 키워드인 ‘자살’과 의미가 없는 명사들인 ‘때문’, ‘관련’ 등의 단어를 정의하였다.

3) 2021.10.20. 통계 기준

4) 빅카인즈는 뉴스 기사의 수집, 분석, 저장 등의 서비스를 제공(<https://www.bigkinds.or.kr/>)

1) 뉴스기사(<https://www.hankyung.com/economy/article/2021071921931>)

2) 2020년 사망원인통계 결과, 통계청

4.2 토픽 모델링

LDA(Latent Dirichlet Allocation) 모델은 대량의 텍스트 데이터로부터 주제분류를 도출하는 대표적인 모델로 각 문서별 토픽 분포와 주제별 단어 분포를 추정한다.[6] LDA 모델에서 알파 값을 0.1, 샘플링 반복횟수를 2,000회로 설정하여 전체 뉴스 기사 본문으로부터 10개의 토픽을 도출하였다. 토픽 수 결정은 coherence score가 높은 5, 6, 10, 13개의 후보 중 의미 해석이 용이했던 10개로 설정하였다.

5. 분석 결과

5.1 통계 시각화

월별 ‘자살’ 관련 뉴스 기사 건수와 자살 자망자수를 시각화 한 결과는 (그림1)과 같으며, 사망자 수와 뉴스 기사 수가 연관관계를 가지는 것을 확인할 수 있다. 2019년 1월, 10월, 11월, 2020년 7월, 9월, 11월, 12월, 2021년 1월, 2월, 6월의 경우 유명인의 자살 또는 사회적으로 이슈가 된 자살 사건으로 뉴스 건수가 도드라지게 높은 것을 확인할 수 있다.

각 해의 1월부터 8월까지의 전체 자살사망자 수를 보면 2019년 9,180명, 2020년 9,054명, 2021년 8,634명으로 매년 1.4%, 4.6%씩 감소하였다. 전문가들은 이러한 현상을 ‘허니문 기간(Honeymoon period)’<sup>5)</sup>이라 언급하며 이후의 지속적인 경제적, 심리적 어려움 등으로 인한 자살 증가 가능성에 대해 우려하고 있다.[7]

5.2 키워드 분석

전체 뉴스 기사 본문에서 키워드를 추출하여 출현 빈도를 도출한 결과, 가장 빈번하게 출현한 키워드는 ‘생각’, ‘경찰’, ‘사건’, ‘문제’, ‘자신’, ‘자살 예방’, ‘수사’, ‘조사’로 자살 사건과 관련된 수사, 예방이 주요한 사항임을 알 수 있다. ‘생각’과 관련해서는 자

살을 나타내는 ‘극단적인 생각’, 자살 동기와 관련된 ‘소문날 수 있던 생각’, 주변인들의 자살 동기 추론과 관련된 ‘하는 생각이 든다.’, 고인을 그리워하는 ‘생각 많이 하고 있어요.’ 등의 문장이 도출되었다.

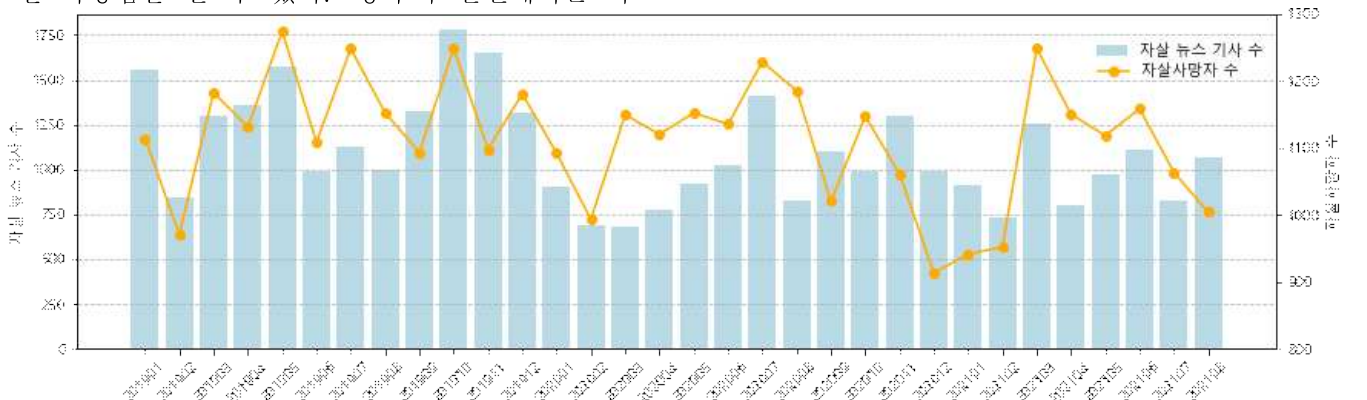


생각	36,526	경찰	36,381	사건	33,362
문제	31,328	자신	27,671	자살 예방	24,692
수사	22,869	조사	21,621	시간	20,638

<표 1> 키워드 빈도 및 워드 클라우드

5.3 토픽 모델링

LDA 모델을 통해 도출된 10개의 토픽과 해당 토픽의 주요 뉴스 기사 제목은 <표 2>와 같다. 토픽을 구성하는 주요 키워드를 통해 Topic 3은 경찰 수사, Topic 5는 자살 예방, Topic 8은 경제와 관련된 주제임을 추론해 볼 수 있다. (그림 2)의 월별 토픽 분포를 보면 뉴스 기사에서 자살 예방과 관련된 Topic 5와 마음, 심리와 관련된 Topic 10은 꾸준히 높은 비중을 보였으며, 수사와 관련된 Topic 3과 피해자, 가해자와 관련된 Topic 9는 이슈화된 사건이 발생한 시기에 높은 비중을 보였다. 경제 관련 Topic 8은 2019년도에는 연초에 집중적으로 나타났으나 코로나19 발생 후에는 경기 악화, 실업 등의 이유로 인한 극단적인 선택을 하거나 우려되는 상황으로 하반기에도 지속적으로 일정 비중을 유지하는 것으로 나타났다.

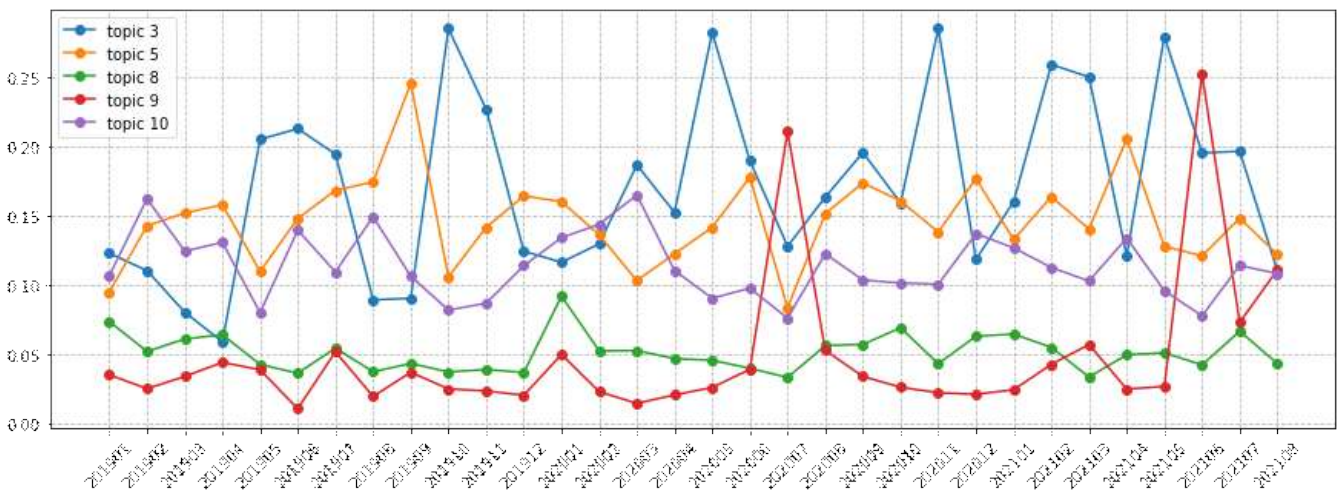


(그림 1) ‘자살’ 관련 뉴스 기사 수 및 자살사망자 수

5) 국가적 재난 발생 시 단합 사회적 긴장 등의 이유로 자살률이 감소하다 향후 자살률이 증가하는 현상

Topic	주요 키워드	주요 뉴스 기사 제목
1	미국, 일본, 중국, 북한, 정부	튀니지 주재 美 대사관 근처서 폭발 발생 "자살폭탄 테러 추정"
2	사건, 방송, 피해자, 자신, 범행	텔레그램 음란물방 운영하던 'n번방' 제보자, 방송사 인터뷰 후 극단적 선택
3	경찰, 극단 선택, 발견, 수사, 조사	서산경찰서, 자살의심자 신속히 발견 가족 품에 무사히 인계
4	안전, 지역, 사고, 차량, 시민	AI와 결합한 CCTV 택시 성추행범 잡고 극단선택 막았다
5	자살 예방, 지원, 센터, 정신 건강, 상담	정부, '코로나 블루' 예방 위한 심리 방역 나선다
6	여성, 가족, 학생, 부모, 사회	청소년 사망 원인 8년째 1위 '극단적 선택'
7	국민, 문제, 검찰, 정치, 상황	진천군보건소, 우울증 환자 치료관리비 지원
8	경제, 세계, 사회, 기업, 회사	노인 빈곤과 자살 대책, 코로나19 사태 대응처럼
9	피해자, 의원, 위원회, 가해자, 성추행	공군 성추행 피해자 사건 다음날부터 "더는 살 이유가 없다"
10	우리, 생각, 자신, 이야기, 마음	불안, 두려움, 공포는 질병보다 큰 적이다

<표 2> 토픽 모델링 결과



(그림 2) 월별 주요 토픽 분포

6. 결론

코로나19 전후로 ‘자살’과 관련된 통계와 뉴스 기사를 수집해 수치적, 의미적 분석을 수행한 결과, 수치적으로는 ‘허니문 기간’을 보이며 코로나 이후 자살률이 지속적으로 감소한 모습을 보였다. 의미적으로는 자살 관련 기사에서 주요한 키워드는 자살 사건의 수사와 자살 예방이었으며, 월별 주요 토픽에서도 자살 예방은 꾸준히 높은 비중을 보였고, 사건과 관련된 토픽들은 이벤트를 중심으로 피크값을 보였다. 또한 코로나19로 인해 경제 관련된 이슈가 연초뿐만 아니라 이후에도 일정 수준을 유지하는 모습이 도출되었다.

ACKNOWLEDGEMENT

본 논문은 서울기술연구원(2021-AH-002, 한강교량 CCTV 영상감시 통합관제체계 고도화)의 지원을 받아 수행된 연구임.

참고문헌

[1] 김경빈, 코로나19 사태에 따른 국내외 거시경제 동향과 시사점, 연금포럼 78호, 66-73, 2020.  
 [2] 장우심, 팬데믹 상황(COVID-19)에서 우울, 불안, 스트레스가 자살생각에 미치는 영향, 상담심리교육복지 8(4), 235-246, 2021.  
 [3] 김태중, 뉴스 빅데이터를 활용한 코로나19 언론보도 분석 :토픽모델링 분석을 중심으로, 한국콘텐츠학회논문지 20(5), 457-466, 2020.  
 [4] 장선영, 정승현, 토픽모델링을 이용한 도시 분야 연구동향 분석, 한국산학기술학회논문지 22(3), 661-670, 2021.  
 [5] 임유하, 빅데이터를 통해 본 ‘코로나 블루’에 대한 이슈 분석, 한국심리상담학회지 33(2), 829-852, 2021.  
 [6] Blei, D. M., Ng, A. Y., Jordan, M. I., “Latent Dirichlet Allocation”, Journal of Machine Learning Research 3, 993-1022. 2003.  
 [7] 이은환, 코로나19 세대, 정신건강 안녕한가!, 경기연구원, 이슈&진단, 1-25, 2020.