

토픽모델링을 활용한 사회문제 해결방안 모색 연구: 사회양극화와 다중격차의 관점에서

유명현¹ · 임세민² · 오서경² · 송지훈^{3*}

¹한양대학교 한국교육문제연구소 전임연구원

²한양대학교 교육공학과 석사과정

^{3*}한양대학교 교육공학과 교수

Solution to Social Problems Using Topic Modeling: From the Perspective of Social Polarization and Multiple Disparities

Myunghyun Yoo¹ · Se Min Yim² · Seokyoung Oh² · Ji Hoon Song^{3*}

¹Research Professor, Institute for Educational Research, Hanyang University, Seoul 04763, Korea

²Master's Course, Department of Educational Technology, Hanyang University, Seoul 04763, Korea

^{3*}Professor, Department of Educational Technology, Hanyang University, Seoul 04763, Korea

[요약]

본 연구에서는 사회양극화 및 다중격차의 개념을 통해 최근 주요하게 논의되는 사회문제들을 도출하고 관련 현상을 해석하고자 한다. 이를 위해 주요 3개 일간지의 사회면 기사를 수집하고 토픽모델링 방법으로 LDA(Latent Dirichlet Allocation)를 사용하여 20개의 주요 토픽을 도출하였으며 사회양극화와 다중격차의 관점에서 해석하였다. 또한 시간적 흐름에 따른 세부 토픽별 변화를 확인하여 주요 논의 분야의 경향성을 예측하였다. 마지막으로 시사점과 향후 연구를 위한 제언을 기술하였다.

[Abstract]

This study aimed to extract major social problems recently discussed through social polarization and multiple disparities and interpret the related phenomena. Specifically, social media articles from three major daily newspapers were collected and 20 major topics were identified using Latent Dirichlet Allocation (LDA) as a method for topic modeling and interpretation. In addition, the tendency of topics was predicted by confirming changes in topics over time. Finally, the implications and research limitations are discussed.

색인어 : 토픽모델링, 사회양극화, 다중격차, 사회문제, 사회기사

Keyword : Topic Modeling, Social Polarization, Multiple Disparities, Social Problems, Social Article

<http://dx.doi.org/10.9728/dcs.2023.24.8.1741>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 27 June 2023; **Revised** 17 July 2023

Accepted 20 July 2023

***Corresponding Author, Ji Hoon Song**

Tel: +82-2-2220-2642

E-mail: psu.jihoonsong@gmail.com

I. 서론

소득의 격차에 따른 경제적 원인에서 기인한 전통적 사회 양극화 문제는 산업 전반에서 진행되는 디지털 전환과 COVID-19에 따른 시대적 변화와 함께 복합적인 사회구조적 양극화 현상을 야기하고 있다. 양극화는 사회 불평등이 극대화되고 중간계층이 축소하면서 사회계층이 양극단으로 쏠리는 현상으로[1], 주로 경제적 측면에서 논의되는 경우가 많다[2]. 그 원인도 경제환경의 급변과 산업 고용구조의 취약성으로 보는 접근이 대다수이나, 사회 양극화는 말 그대로 사회적 문제로 논의될 필요성이 있다. 즉, 경제적 차원으로만 다루어서는 원인 파악과 해결이 불가능하며 사회 전체적인 차원의 문제로 접근해야 하는 것이다[3]. 이는 교육, 의료, 성별, 세대적 측면에서 국민의 삶의 질과 직결되는 사회구조적 양극화라는 점에서 다양한 사회문제와 직결되고 문제 해결을 위한 선제적 접근이 필요하다[4].

최근 한국사회에서 발생하고 있는 사회양극화는 다양한 형태와 양상을 보인다. 특히 교육분야에서 발생한 양극화와 여성의 돌봄 부담은 증가하였다. 그리고 COVID-19 이후 발생한 학습결손과 교육의 격차는 지역과 소득에 따라 심화 정도에 차이를 보이며 교육분야에서 장기적으로 해결해야 할 과제로 논의되고 있다. 이에 따라 COVID-19 이후 가중된 여성의 돌봄 부담을 완화하고, 경력단절 위기에 놓인 여성들을 위한 일·가정 양립과 교육복지 지원 확대 방안이 요구되고 있다. 다음으로 한국이 고령화 사회에 진입하면서 의료 이용 환경에서 나타나는 양극화는 보건복지 정책의 중대한 해결 목표가 되었다. 디지털 헬스케어의 발달은 새로운 의료 소외계층으로 디지털 취약계층을 대두시켰다. 우리나라의 경우 의료분야 정보격차의 논의는 상대적으로 늦은 편으로, 2021년 발표된 ‘국민건강증진종합계획’에서 건강 형평성 제고 전략으로 의료분야 정보격차를 다루고 있어 의료정보 및 서비스에 대한 접근격차가 지역 및 세대에 따라 다양한 양상을 보여 복합적인 해결방안이 필요하다. 또 다른 사회양극화 및 사회문제 영역으로 고용과 노동의 안전망 약화를 들 수 있다. 세계화·디지털화로 기업환경의 불확실성이 증대되며 기업 측면에서는 높은 혁신성과 유연한 고용시스템을, 근로자 측면에서는 안정성이 높은 새로운 고용시스템을 요구하며 청년층의 고용 위기가 심화되었다. 산업구조의 전환 과정에서 선제적 대응을 통해 약화된 고용·노동 안전망을 재고하고 중층적 고용안전망의 구축이 요구되는 시점이다.

이렇듯 사회양극화 및 사회문제가 발생하는 원인은 매우 복잡적이며, 최근 한국사회는 소득, 자산, 교육, 주거, 문화, 건강 등 불평등의 여러 차원이 서로 중첩되는 현상이 심화되고 있다[5]. 한 영역에서의 단일한 격차가 아니라 다중격차가 생겨나 구조화되고 고착되면, 시장의 모순과 불평등을 시정하고자 하는 복지국가의 과제 또한 복잡해질 수 있으며, 개별 격차의 논의에서 벗어나 다학제 간 논의의 필요성이 증가하

고 있는 것이다[6]. 따라서 본 연구에서는 사회양극화 및 다중격차의 개념을 통해 최근 주요하게 논의되는 사회문제들을 도출하고 관련 현상들을 해석하고자 한다.

본 연구의 목적은 사회면 기사를 수집하여 최근 한국 사회의 주요 토픽들을 도출하고, 복합적인 사회문제 해결과 선제적 대응을 위해 다학제 간 이해를 위한 기초자료를 제공하는 것이다. 이를 위해 최근 사회면 기사에서 논의된 주요 토픽들을 도출하고, 사회양극화와 다중격차를 해소하기 위한 근거 기반 실천을 제공하고자 한다. 구체적으로 토픽모델링을 통해 최근 한국 사회의 문제들과 관련된 최적의 잠재 토픽을 도출함으로써 주요 쟁점을 확인하고, 중요 논의 분야와 관련 요인을 확인한다.

본 연구를 수행하기 위한 연구 문제는 다음과 같다.

- 연구문제 1. 사회면 기사를 통한 연구 맥락상의 토픽 분포는 어떻게 이루어져 있는가?
- 연구문제 2. 도출된 토픽은 어떻게 분류하고 전략적으로 이해해야 하는가?

II. 이론적 배경

2-1 사회양극화와 다중격차

사회양극화는 사회 안에서 상층과 하층의 격차가 벌어지며 분열되는 현상을 의미한다[1]. 중간층이 소멸되고 사회에서 양극화가 심화되면 집단 간 갈등으로 이어지게 된다. 이는 사회의 경제, 교육, 임금과 같은 다양한 문제를 유발한다. 한국 사회에서 양극화의 원인은 단순한 경제적 불평등을 넘어 다차원적이고 중층적인 현상으로 바라볼 필요가 있다[7].

다중격차는 다양한 영역에서 격차가 중복되어 나타나는 것을 의미한다[8]. 다중격차는 소득, 주거, 자산과 같은 다양한 불평등의 영역들로 구성되어 있다[5]. 예를 들어 정규직과 비정규직, 대기업과 중소기업, 수도권 대학과 지방대, 특목고·자사고와 일반고 격차 문제 등이 있다[9]. 이러한 불평등의 범주들은 지속적으로 상호작용하며 다중격차를 고정시키거나 심화시킨다[10].

격차가 발생하고 있는 범주들은 각 범주별로 격차가 증가하는 것뿐만 아니라 격차의 중첩이 심화되고 있음을 보여준다[9]. 예를 들어 대기업과 중소기업 간 차이가 정규직과 비정규직 사이의 격차와 중첩되면서 고용 격차를 나타내고 있으며[9], 소득과 자산의 불평등은 가족형성 기회의 불평등으로 이어질 수 있다[11]. 특히, 결혼과 출산 문제에서 경제적 불평등에 영향을 주고 있다[11].

다중격차는 각 영역의 사회문제가 상호작용하는 특성으로 인해 문제가 심화되고 반복적으로 악순환되는 특징이 있다[6]. 즉 사회에 세습된다는 점에서 격차 완화를 위해 사회양

극화와 다중격차 문제에 주목할 필요가 있다. 그러나 복합적 원인으로 인해 발생하는 다중격차가 해소되거나 완화되는 데에는 많은 어려움이 있다. 하나의 영역에서 발생하는 격차가 아닌 다중격차로 인한 문제 발생이 구조화되고 사회에 고착된다면, 사회 불평등과 모순을 개선하고자 하는 국가 과제 및 대응책도 복잡해질 수 있다[6]. 따라서 사회문제 해결을 위해 다차원적·복합적 대응으로 접근하는 것은 다중격차 해소를 위한 중요한 시각으로 바라볼 수 있다. 그동안 한국사회에서는 다중격차 문제의 범주를 단일 차원으로 분리하여 바라보는 경향이 존재했으며, 하나의 공통분모로 병합하여 동시에 극복해야 하는 관점으로 여기고 있어[4] 다중격차와 관련한 정책적 대응에 한계가 있었다. 선제적·예방적 정책대응 방안을 구축하기 위해 다중격차의 복합적 요인을 고려하여 문제점을 파악하는 노력이 필요하다.

2-2 다중격차의 주요 논의 영역

한국사회는 자산, 소득, 주거, 교육 등 다양한 차원의 불평등 문제가 중첩되는 현상이 증가하고 있다[6]. 지속적으로 증가하는 사회문제를 토대로 다차원적인 결핍상황에 주목하여 논의되고 있는 다중격차의 주요 논의 영역을 살펴보고자 한다.

1) 고용정책의 사각지대

고용취약계층의 보호 제도가 필요하다. 실직자들은 생계를 유지하기 위해 일용직과 같은 불안정한 일자리도 마다하지 않고 일을 한다. 하지만 불안정한 일자리로 인해 실업이 반복되는 악순환에 머무르게 된다. 이러한 악순환을 벗어나기 위해 소득을 보전하는 것이 필요하다. 고용보험의 실업급여가 소득보전의 기능을 담당하게 된다. 하지만 고용보험은 보험료 납부가 어려운 청년, 경력단절여성, 저소득층 등은 급여를 받을 수 없다. 따라서 고용보험을 중심으로 하는 고용정책은 지속적인 사각지대를 형성하게 된다[12].

2) 인구감소 및 지역소멸 시대

수도권과 비수도권의 격차는 지속되고 있다. 수도권으로 인구 이동이 집중되면서 지방 지역의 인구가 감소하게 되었다. 국가는 지역 주도 균형발전에 대해 항상 강조해오고 있다. 하지만 정치, 경제, 행정 등의 분권 현실은 여전히 많은 한계에 부딪히고 있다. 그리고 비수도권에 생활 인프라 예산 편성 여력이 쉽지 않아 편의시설 설립이 어려운 상황이다. 이로 인해 지방의 인구수 감소가 지속되며 지역 소멸가능성 문제가 함께 제기되고 있다[12].

3) 농어촌지역 재생

농어촌 지역 특성을 고려한 지역발전이 어려움을 겪고 있다. 농어촌 인구는 10년간 지속적으로 감소하며 고령화 비율은 매년 증가하고 있다. 이러한 상황에서 수도권과 비수도권 격차에 관한 문제의식에만 중점을 두고 있다. ‘농촌’과 ‘어촌’

의 지역 특성을 고려한 지역 발전 정책은 논의되고 있지 않다. 따라서 ‘지역소멸’ 문제보다 세분화된 관점으로 접근하여 농어촌지역의 특색을 살려 발전시킬 수 있는 실질적이고 구체적인 방안이 제시되어야 한다[13].

4) COVID-19 이후 교육격차 심화

COVID-19 이후 기초학력 미달 학생 비율이 증가하고 있다. 원격수업이 도입되면서 대면 교육의 기회가 줄어들어 교육의 격차가 심화되고 있다. 농어촌학생, 장애학생, 다문화학생 등 사회적 취약계층은 원격수업 확대 상황에서 많은 어려움을 겪고 있다. 디지털 격차와 정보격차 등으로 인해 온라인 수업에 대한 진입장벽을 느끼며 수업 이해도에 따른 학습격차가 발생하고 있다. COVID-19 장기화로 인해 학생들의 학습격차 확대, 원격수업의 질과 학습효과 저하, 사교육 의존도 증가 등으로 인한 계층 간 교육 불평등이 더욱 심화되고 있다[12].

2-3 다중격차 문제의 접근 방법

최근 한국사회의 불평등이 심화되면서 다중격차는 심각한 사회문제로 여겨지고 있다[1]. 구체적으로 소득, 자산, 주거 등 다양한 불평등의 요소들이 서로를 강화시키는 메커니즘 역할을 하고 있으며[5], 불평등 요소들의 상호작용으로 인해 지속적으로 불평등 문제가 중첩되어 사회문제가 심화되고 있다. 따라서 다중격차의 문제가 사회에 고착화되지 않고 적절한 해결 방안을 모색하기 위해 다양한 차원에서 문제를 파악하고 논의하는 것이 필요하다.

한국의 다중격차 문제는 경제·산업, 사회, 지역·국토, 인구 등 다양한 측면으로 문제를 바라보고 파악하기 위해 노력을 기울이고 있다. 그러나 주요 논의 영역을 비롯해 다양한 카테고리 고리의 사회격차 문제를 파악하기에는 한계점이 존재한다. 따라서 다각화된 시점으로 사회문제를 바라볼 때 주요 논의되는 영역 이외에 논의가 제대로 이루어지지 못한 사회격차 문제를 파악할 수 있을 것이다.

본 연구에서는 최근 주요하게 논의되고 있는 다중격차 문제를 포함하여, 다학제 간 논의가 필요한 다양한 사회문제를 도출하여 다중격차와 사회양극화의 관점에 입각해 한국사회가 당면한 사회문제를 다루어 보고자 한다. 연구방법으로는 다양한 영역의 사회문제를 살펴보기 위해 사회기사를 활용하여 사회문제를 분석하였다. 토픽모델링을 활용하여 사회기사에서 제시된 격차 문제를 파악하여 주제별로 공통 내용을 설정하여 분류하였다.

이러한 분류를 토대로 소외받았던 사회격차 문제까지 함께 살펴보고 다양한 측면에서 사회문제를 바라보고자 한다. 현재 우리 사회에서 주요 문제와 주목받지 못한 사회문제까지 함께 고려하여 다중격차 해소를 완화시킬 수 있는 방법에 기여하고자 한다.

III. 연구 방법

본 연구는 뉴스 기사에서 다루지는 사회 이슈와 관련된 주제와 시간의 흐름에 따른 주제의 변화를 살펴보기 위해 토픽 모델링 분석을 실시하였다. 분석 절차는 첫째, 분석 대상 자료 수집, 둘째, 데이터 전처리, 셋째, LDA 토픽모델링 분석, 넷째, 시계열 분석을 통한 Hot & Cold 토픽 분류의 순서이다.

3-1 분석 대상 선정 및 자료 수집

본 연구에서는 사회 카테고리의 뉴스 기사를 분석 대상으로 선정하였다. 자료 수집은 2022년 6월부터 2022년 12월 까지 7개월 간의 조선일보, 중앙일보, 동아일보의 사회 카테고리에 등록된 40,861개의 기사를 웹크롤링을 통해 수집하였으며, 구체적으로 기사 제목, 기사 본문, 기사 등록일자를 수집하였다. 수집된 뉴스 기사 제목 및 본문을 기준으로 사회적 이슈를 담지 않을 것이라 판단한 기사들을 제외하였다. 예를 들어 단순 일기예보, 운세, 복권, 외국어 학습코너, 본문 없이 사진만 제시된 기사가 이에 해당하였다. 분석 대상인 뉴스 기사는 2,733편의 기사를 제외하여 최종적으로 총 38,128편으로 확정되었다.

3-2 데이터 전처리

수집한 기사 본문은 비정형 데이터로 분석에 앞서 데이터 전처리 과정이 필요하다. 데이터 전처리는 파이썬 3.10을 통해 진행되었으며 과정은 다음과 같다. 첫째, KoNLPy[14] 패키지를 통해 기사별 명사를 추출하여 토큰화하였다. 이때 명사를 구분하는 사전은 Okt(open-korean-text)를 사용하였으며, 뉴스 기사에 등장하는 명사 일부를 연구자가 추가하였다(예: 확진자, 증大本 등). 둘째, 문서에서 빈번하게 도출되거나 분석에 불필요한 조사 및 관용구 등의 불용어를 제외하였다(예: 가령, 때문 등). 셋째, 분석에서 의미를 도출하지 못하는 1글자 단어를 제외하였다(예: 수, 가 등). 최종적으로 60,780개의 단어로 구성된 38,128개의 말뭉치(corpus)가 분석에 사용되었다.

3-3 LDA를 통한 토픽모델링 분석

뉴스 기사의 주요 주제를 확인하기 위해 토픽모델링을 실시하였다. 토픽모델링은 문서 내 관찰된 단어를 통해 직접적으로 보이지 않는 문서의 주제를 추론하는 방법으로, 본 연구에서는 토픽모델링의 대표적 방법인 LDA(Latent Dirichlet Allocation)[15],[16]를 사용하였다. 이를 통해 전체 기사 집합에서의 토픽, 각 기사별 토픽의 비중, 각 토픽에 포함되는 단어의 분포를 도출할 수 있다.

LDA는 파이썬 3.10을 통해 수행하였으며 오픈소스로 개

발된 Gensim 패키지를 사용하였다[17]. Gensim 패키지는 대량의 문서 모음을 분석하는데 적합한 알고리즘을 사용하므로[16],[18], 많은 양의 뉴스 기사를 분석하는 본 연구에 적합하다고 판단하였다.

LDA 모델을 구성하기 위해서는 잠재된 토픽의 개수를 사전에 정의할 필요가 있으며 본 연구에서는 복잡도(perplexity)[19]와 응집도(coherence)[20]를 참고하여 토픽의 개수를 판단하였다. 이때 디리클레 분포의 하이퍼파라미터 α 와 β (Gensim 패키지에서는 eta로 표시한다)는 'auto'로 설정하였다. 토픽 개수별 복잡도와 응집도를 비교했을 때 복잡도는 낮을수록, 응집도가 높을수록 토픽 개수가 적절하다고 볼 수 있다[21]. 최대 60개까지의 토픽 개수를 가정하여 혼잡도와 응집도를 계산한 결과는 그림 1과 같다. 복잡도는 14개 토픽 이상일 때 감소폭이 줄어들고, 응집도는 18개 이상에서 일정 수준 이상으로 증가하지 않음을 확인할 수 있었다. 따라서 본 연구는 18개 이상의 토픽 개수에서 주요 키워드의 중복 및 분산이 적고 토픽 별 해석의 용이성을 고려하여 토픽 개수를 선택하였다. 최종적으로 토픽 개수를 20개로 설정하여 분석을 수행하였다.

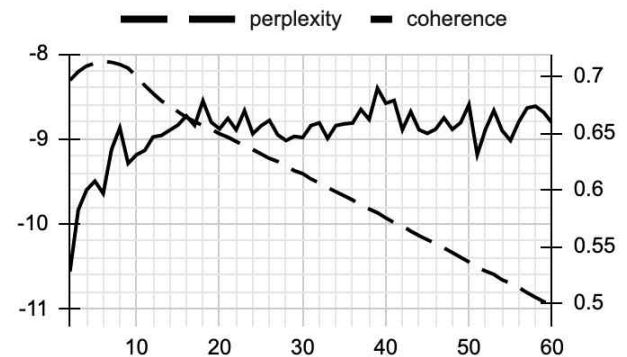


그림 1. 토픽 개수에 따른 복잡도(좌) 및 응집도(우)
 Fig. 1. Perplexity (left) and coherence (right) according to the number of topics

도출된 각 토픽의 명명을 프로그램을 통한 자동화 시도가 있었으나 부정확하고 일관성을 보장하지 않으므로[22] 토픽의 명명은 연구자가 수행할 필요가 있다. 각 토픽의 명칭이 상세 내용을 의미적으로 포괄하면서도 다른 토픽과의 변별력을 갖추기 위해 TextRank를 통해 각 토픽과 관련성이 높은 30개의 단어를 도출하였고[23], 연구자 간의 논의를 바탕으로 토픽을 명명하고 대표하는 단어 10개를 선정하였다. 이를 위해 LDA를 통한 모델을 시각화하고 관련된 단어를 도출할 수 있는 LDAvis[24]의 파이썬 패키지인 pyLDAvis을 이용하였다. 이때 각 단어의 관련성을 설정하는 λ 값은 토픽에서 출현하는 빈도가 높으면서 변별성을 확보하기 위한 .6의 값을 사용하였다[23].

이후 도출된 토픽을 관련된 주제끼리 상위군집하여 유목화

하였다. 유목화의 기준으로는 사회 이슈 키워드를 다루고 있는 국회입법조사처의 2023 올해의 이슈[12]를 활용하였다.

3-4 시계열 회귀분석을 통한 Hot & Cold 분류

시간에 따른 토픽의 추이를 분석하는 시계열 분석은 기간에 따른 토픽 비중(%) 추이에 대한 선형 회귀분석으로 수행하였다[25],[26]. 구체적으로 독립변수는 월단위로 설정하고, 종속변수는 각 문서의 토픽 비중으로 설정하였으며, 유의 수준 5%의 범위에서 통계적으로 유의하다고 판단하였다. 회귀계수 값이 양수일 경우 토픽의 추세를 Hot 토픽으로, 음수일 경우 Cold 토픽으로 분류하였다.

IV. 연구 결과

4-1 토픽 도출

LDA를 통한 토픽모델링 분석에 따라 도출된 20개 토픽과 토픽을 대표하는 10개의 단어는 표 1에 제시하였다. 토픽의 분류는 정치, 국제, 경제·산업, 사회, 지역·국토의 기준[12]을 사용하였으며, 토픽의 번호 순서는 전체 말뭉치에서 해당 토픽을 나타내는 단어가 차지하는 비율이 높은 순이다. 토픽모델링 분석을 시각화한 결과는 그림 2에 제시하였다.

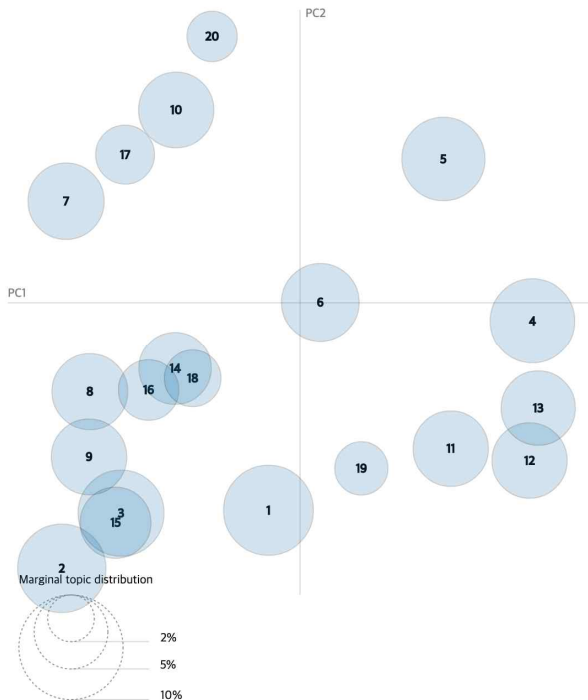


그림 2. 토픽 간 거리분포
Fig. 2. Intertopic distance map

표 1. 토픽 도출 결과
Table 1. Result topic list

Topic	Category	Main keyword	Word proportion
Topic 1: 사회문제에 대한 국가와 정부의 역할과 책임	정치	국민, 장관, 국회, 정부, 대통령, 회의, 정책, 논의, 인사, 비판	7.4%
Topic 2: 학령인구 감소에 따른 대학입시와 공교육	지역·국토	대학, 학생, 교육, 시험, 교육부, 수능, 교사, 학과, 모집, 평가	7.1%
Topic 3: 현 정부의 정책 및 계획 수립	정치	사업, 계획, 내년, 주택, 예산, 추진, 지역, 재정, 연금	6.8%
Topic 4: 성범죄 재판 및 선고	사회	혐의, 선고, 피해자, 재판, 징역, 스토킹, 집행유예, 살인, 성폭행, 성폭력	6.6%
Topic 5: 강력범죄에 대한 수사	사회	경찰, 신고, 조사, 발견, 혐의, 입건, 수사, 체포, 살해, 범행	6.3%
Topic 6: 심리적 트라우마 및 마음챙김의 필요성	사회	사람, 이태원, 생각, 마음, 희생, 유가족, 참사, 사과, 추모, 이틀	5.5%
Topic 7: 기상 이변에 따른 폭우, 태풍, 한파의 피해	사회	기온, 지역, 태풍, 기상청, 해상, 전국, 예상, 예보, 바람, 추위	5.3%
Topic 8: COVID-19 팬데믹 지속	사회	코로나, 방역, 접종, 입국, 유행, 감염, 백신, 중국, 마스크, 검사	5.3%
Topic 9: 글로벌 축제 및 행사 재개	사회	세계, 행사, 문화, 한국, 관광, 공연, 축제, 개최, 월드컵, 전시	5.2%
Topic 10: 안전사고 발생 및 대비 필요성 제기	사회	사고, 화재, 소방, 발생, 안전, 터널, 현장, 참사, 구조, 피해	5.2%
Topic 11: 법원 재판 및 판결사례에 대한 관심	사회	대법원, 소송, 판결, 법원, 징계, 처분, 재판, 손해배상, 인정, 청구	5.2%
Topic 12: 정치 비리 관련 이슈 및 논쟁	정치	대표, 의원, 검찰, 대장동, 민주당, 의혹, 변호사, 성남, 혐의	5.2%
Topic 13: 사기 및 채용비리 관련 수사 증가	사회	수사, 검찰, 장관, 검사, 법무부, 압수수색, 보고서, 사건, 혐의	5.1%
Topic 14: 쓰레기 배출 및 환경오염 논쟁 심화	사회	사용, 판매, 쓰레기, 배달, 구매, 제품, 서비스, 음식, 플라스틱, 식당	4.8%
Topic 15: 소외계층을 위한 보호정책 필요	사회	사회, 청년, 복지, 보육, 지원, 청소년, 아동, 노인, 가정, 계층	4.6%
Topic 16: 공공의료시스템 및 의료서비스 사각지대	지역·국토	병원, 의료, 환자, 치료, 건보(건강보험), 진료, 수술, 보험, 사망, 질환	3.3%
Topic 17: 장애인 예산 확보 및 권리보장 요구	사회	지하철, 서울, 열차, 집회, 시민, 시위, 철도, 지역, 장애인, 정치	3.2%
Topic 18: 노동자의 안전권리 보장제도 시급	사회	화물, 노조, 파업, 연대, 운송, 조합원, 근로자, 임금, 운임, 노동조합	2.9%
Topic 19: 투자 및 금융사기 피해사례의 증가	사회	사기, 금융, 투자, 주식, 보이스피싱, 횡령, 화폐, 거래, 자금, 재산	2.6%
Topic 20: 음주운전 관련 안전망 대비의 필요성	사회	음주운전, 차량, 고속도로, 운전자, 사고, 스쿨존, 음주, 자동차, 도로, 보행자	2.3%

*To describe the keywords derived through text analysis, Table 1 was written in Korean

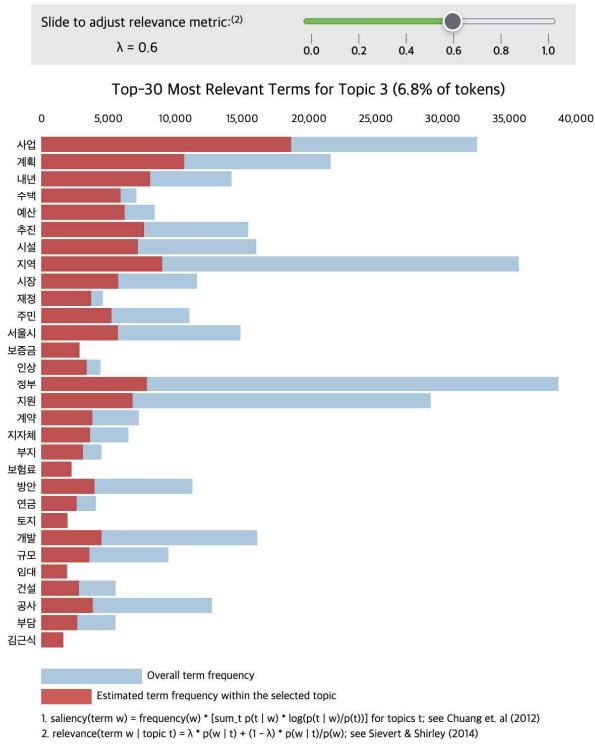


그림 3. 정치영역 토픽(토픽 1)
 Fig. 3. Politics category topics (Topic 1)

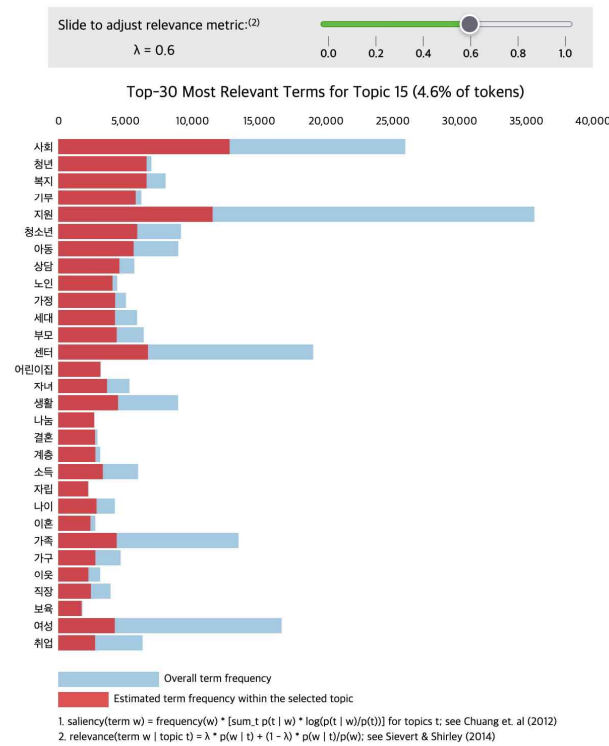


그림 4. 사회영역 토픽(토픽 15)
 Fig. 4. Social category topics (Topic 15)

정치영역에서는 토픽 1, 3, 12가 논의되었다. 그림 3은 토픽 1에 대한 단어의 목록을 나타낸다. 토픽 1은 사회문제에 대한 국가와 정부의 역할과 책임으로 국민, 장관, 국회 등의 키워드와 함께 논의되었다. 뉴스 기사에서는 다양한 사회문제가 발생한 뒤 원인 분석과 함께 책임 소재를 다뤘으며, 특히 사회문제가 다수의 국민에게 영향을 미치고 파급력이 크다는 점에서 국가와 정부의 역할과 책임이 두드러졌다.

토픽 3은 현 정부의 정책 및 계획 수립으로 사업, 내년, 예산 등의 키워드와 함께 논의되었다. 사회문제를 대응하기 위해서 정부 차원의 계획, 정책 그리고 예산을 수립할 필요성이 제시되었다. 주택, 보험료, 연금, 토지, 건설, 공사 등 다양한 영역의 계획이 수립되고 추진되는 점을 뉴스 기사에서 확인할 수 있었다.

토픽 12는 정치 비리 관련 이슈 및 논쟁으로 대장동 등의 키워드를 필두로 검찰 조사가 이뤄지고 있으며 새롭게 밝혀지는 의혹 또는 정보에 대해 뉴스 기사에서 다루고 있었다.

사회영역에서는 토픽 4-11, 13-15, 17-20까지 15개의 토픽이 논의되었다. 그림 4는 토픽 15에 대한 단어의 목록을 나타낸다. 토픽 4는 성범죄 재판 및 선고로 혐의, 재판, 스토킹 등의 키워드와 함께 논의되었다. 뉴스 기사는 스토킹, 성폭력, 성범죄 등의 범죄의 발생 및 실태와 함께 이에 대한 가해자의 처벌 및 재판 결과를 다루고 있었다. 신당역 스토킹 살인 이후 관심이 높아진 스토킹 키워드와 더불어 아동 및 청소년에 대한 그루밍 성범죄 등을 주목해 볼 수 있다.

토픽 5는 강력범죄에 대한 수사로 명명하여 경찰, 신고, 수사 등의 키워드와 함께 논의되었다. 살인 사건과 비롯하여 잔혹하면서도 엽기적인 강력범죄들이 발생하고 있으며 뉴스 기사들은 이들에 대한 수사과정, 밝혀진 사실에 대해 다루었다. 특히 아동을 대상으로 한 학대 및 살인사건은 적지 않은 사회적 파장을 일으켰다.

토픽 6은 심리적 트라우마 및 마음 챙김의 필요성으로 사람, 이태원, 추모 등의 키워드와 함께 논의되었다. 이태원 참사와 같이 안타까운 사건의 발생 이후 많은 국민들에게 마음의 상처가 생겼다. 참사를 추모하는 애도의 물결과 더불어 이태원 참사를 직간접적으로 겪은 이들을 위한 심리적 지원의 필요성이 언급되었다.

토픽 7은 기상 이변에 따른 폭우, 태풍, 한파의 피해로 기온, 지역, 예보 등의 키워드와 함께 논의되었다. 전례 없는 기상이변에 따라 올해 국내에 여름은 태풍, 집중 호우에 따라 전국적으로 피해가 발생하였다. 특히 반지하 등의 거주취약계층의 피해가 부각되었으며, 이에 대한 후속 조치가 논의되었다.

토픽 8은 COVID-19 팬데믹 지속으로 코로나, 방역, 입국 등의 키워드와 함께 논의되었다. 각종 바이러스 변이가 나타나면서 COVID-19 팬데믹이 지속되고 있다. 오랜 기간 동안 지속된 방역 정책에 따라 국민의 피로감이 증가하며 COVID-19를 대응하기 위한 정책에 어려움이 있었다. 백신 접종, 입국자 조치, 방역 조치 해제, 겨울철 독감과의 더블 팬데믹 대비 등이 주로 논의되었다.

토픽 9는 글로벌 축제 및 행사 재개로 세계, 행사, 문화 등의 키워드와 함께 논의되었다. COVID-19 이후 국제적인 축제, 관광 등이 활발하게 이루어지고 있다. 카타르 월드컵을 비롯하여 해외여행, 헬러윈, 지역축제, 해돋이 등의 소식이 뉴스 기사에 등장하였다.

토픽 10은 안전사고 발생 및 대비 필요성 제기로 사고, 화재, 현장 등의 키워드와 함께 논의되었다. 화재, 침수 등 안전 사고가 발생하여 인명 및 재산 피해가 발생하였다. 집중호우로 인한 지하 침수, 카카오 IDC 화재, 과천 방음터널 화재 등이 뉴스 기사에서 언급되었다.

토픽 11은 법원 재판 및 판결사례에 대한 관심으로 소송, 판결, 손해배상 등의 키워드와 함께 논의되었다. 과거 흔하지 않았던 범죄 판결 사례뿐만 아니라, 기존의 판결에 불복하여 항소심에서의 변화를 다루기도 하였다. 이 과정에서 피해자에 대한 제도적, 물리적 안전장치의 필요성이 드러났다.

토픽 13은 사기 및 채용비리 관련 수사 증가로 검사, 압수 수색, 혐의 등의 키워드와 함께 논의되었다. 아파트, 오피스텔 전세 사기 등으로 인한 피해가 주목받음과 더불어 기관 및 기업에서의 채용비리와 관련하여 공정성이 강조되었다. 이들을 수사하는 과정에서 압수수색 등이 이뤄졌다.

토픽 14는 쓰레기 배출 및 환경오염 논쟁 심화로 사용, 쓰레기, 배달 등의 키워드와 함께 논의되었다. 카페 및 음식점에서의 식음료 포장재료로 인한 플라스틱 배출 증가와 더불어 쓰레기 분리배출 공간 부족 및 쓰레기 처리시설 확대 필요성에 대한 언급이 있었다. 쓰레기 소각장에 대한 지역갈등과 합의의 어려움, 인구 집중화로 인한 수도권 쓰레기 포화 상태의 문제점이 지적되었다.

토픽 15는 소외계층을 위한 보호정책 필요로 사회, 복지, 지원 등의 키워드와 함께 논의되었다. 저소득층, 독거노인, 보육원 출신의 자립준비청년 등 다양한 소외계층이 처한 현실과 대응의 필요성에 대해 논의가 되었다. 이들은 경제적 어려움뿐만 아니라 사회로부터 심리적 소외감, 불안감 등과 같은 정서적 차원의 어려움도 있었다. 지원을 받을 수 있음에도 불구하고 정보가 부족하여 복지를 받지 못하는 복지신청주의에 대한 비판이 있었으며, 수원 세 모녀 사건 등에서 크게 부각되었다.

토픽 17은 장애인 예산 확보 및 권리보장 요구로 지하철, 시위, 시민 등의 키워드와 함께 논의되었다. 전국 장애인 차별철폐 연대(전장연) 장애인과 같은 교통약자를 위한 이동권 보장 강화를 요구하며 시위를 진행하였는데 이로 인해 지하철을 이용하는 시민의 불편함이 뉴스 기사에서 다뤄졌다.

토픽 18은 노동자의 안전권리 보장제도 시급으로 노조, 연대, 근로자 등의 키워드와 함께 논의되었다. 민주노총 화물연대, 대양조선해양 하청업체 파업 등 노동자의 권리에 대한 대립이 나타났다. 안전운임제, 혹은 노동자의 안전에 대한 책임으로 중대재해처벌법과 같은 제도적 측면에서의 논의가 있었다.

토픽 19는 투자 및 금융사기 피해사례의 증가로 투자, 주식, 보이콧 등의 키워드와 함께 논의되었다. 토지나 금 등

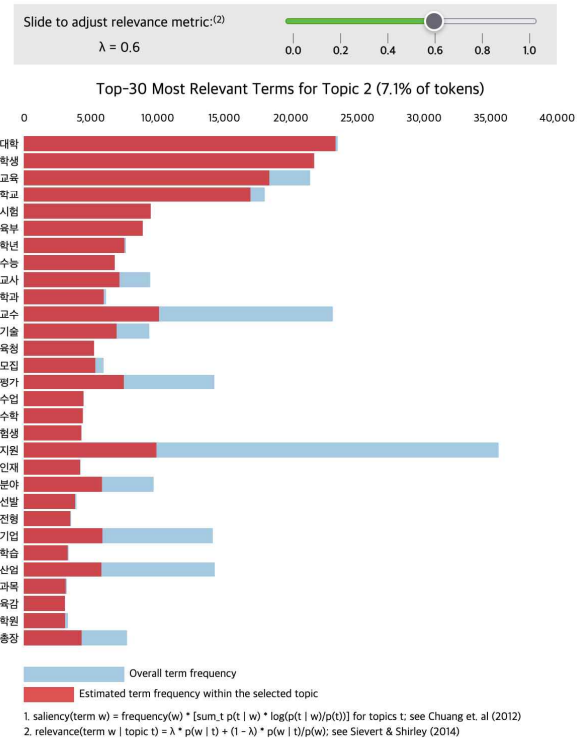


그림 5. 지역·국토영역 토픽(토픽 2)
 Fig. 5. National territorial and regional category topic (Topic 2)

특정 상품을 미끼로 투자를 유도하고 투자금을 가로채는 사례, 고수익 알바 등으로 속여 보이콧팅 가해자로 만드는 사례 등 정보의 차이로 인해 피해가 나타나는 사례가 다뤄졌다. 주식이나 가상자산 투자 실패로 인해 개인회생 채무조정을 신청하는 사례 등이 나타났다.

토픽 20은 음주운전 관련 안전망 대비의 필요성으로 고속도로, 운전자, 사고 등의 키워드와 함께 논의되었다. 음주운전으로 인한 사고가 지속적으로 발생함에 따라 이에 대한 방지 대책 제시의 필요성이 지적되었다.

지역·국토 영역에서는 토픽 2와 토픽 16이 논의되었다. 그림 5는 토픽 2에 대한 단어의 목록을 나타낸다. 토픽 2는 학령인구 감소에 따른 대학입시와 공교육으로 교육부, 수능, 학생 등의 키워드와 함께 논의되었다. 학령인구의 감소는 교육계에 영향을 미쳤다. 대학입시와 관련해서 특히 지방대학이 신입생을 모집하기 위한 노력 및 대비와 함께, 적은 수의 학생을 대상으로 공교육이 어떠한 기여와 역할을 할 수 있을지에 대한 논의가 고교학점제 등의 논의와 함께 이뤄졌다.

토픽 16은 공공의료시스템 및 의료서비스 사각지대로 병원, 환자, 건보(건강보험) 등의 키워드와 함께 논의되었다. 뉴스 기사는 소아외과를 비롯한 필수의료의 의료시스템 공백에 대해 다루었으며, 이로 인한 사망, 의료진의 업무 과부하 등의 문제가 지적되었다. 특히 지방의 경우 그 문제가 더욱 심각한 것으로 나타났다.

4-2 토픽 별 Hot & Cold 분류

2022년 6월부터 12월까지의 7개월 간 뉴스 기사의 토픽 별 비중 추이를 선형회귀분석을 통해 도출한 결과는 표 2와 같다. 시간의 흐름에 따라 토픽의 비중이 증가하는 경우 Hot 토픽으로, 시간의 흐름에 따라 토픽의 비중이 감소하는 경우 Cold 토픽으로 분류하였다.

표 2. 시계열 분석을 통한 Hot & Cold 분류
Table 2. Hot & Cold classification through time series analysis

Topic	Estimate	SE	t	p	Hot&Cold
Topic1	-.00295	.00035	-8.42	< .001	cold
Topic2	.000684	.000405	1.69	.091	-
Topic3	.00113	.000327	3.45	< .001	hot
Topic4	.000515	.000493	1.04	.296	-
Topic5	-.00204	.000469	-4.34	< .001	cold
Topic6	-.000694	.000282	-2.46	.014	cold
Topic7	-.00266	.000423	-6.29	< .001	cold
Topic8	-.00426	.000481	-8.87	< .001	cold
Topic9	-.00138	.000315	-4.39	< .001	cold
Topic10	.00417	.000383	10.9	< .001	hot
Topic11	-.00101	.000292	-3.45	< .001	cold
Topic12	.00553	.000393	14.063	< .001	hot
Topic13	-.000244	.000346	-0.706	.481	-
Topic14	-.00227	.000289	-7.86	< .001	cold
Topic15	.000326	.000263	1.24	.216	-
Topic16	-.000427	.00025	-1.71	.087	-
Topic17	.00326	.000282	11.56	< .001	hot
Topic18	.00209	.00027	7.74	< .001	hot
Topic19	.000684	.000207	3.3	< .001	hot
Topic20	-.000456	.000241	-1.89	.059	-

Hot 토픽은 시간이 흐름에도 논의가 활발한 영역으로 해석할 수 있다. 도출된 Hot 토픽은 3(현 정부의 정책 및 계획 수립), 10(안전사고 발생 및 대비 필요성 제기), 12(정치 비리 관련 이슈 및 논쟁), 17(장애인 예산 확보 및 권리보장 요구), 18(노동자의 안전권리 보장제도 시급) 그리고 19(투자 및 금융사기 피해사태의 증가)로 나타났다. 시간의 흐름에 따른 Hot 토픽의 토픽별 비중 추이를 시각화한 것은 그림 6과 같다.

토픽 3의 경우 2022년 새 정부가 구성됨과 함께 2023년을 맞이하여 차년도에 계획을 수립하는 과정에서 문서의 토픽 비중이 증가하는 것으로 예상된다. 토픽 10의 경우 침수로 인한 피해, 이태원 참사, 방음터널 화재 등 예방 가능했던 사고가 다수 발생함에 따라 후속 조치에 대한 논의가 증가하는 것으로 해석할 수 있다. 토픽 12의 경우 지속적으로 의혹이 제시되고 있으나 혐의가 확정되지 못하여 지속적으로 논의되고 있는 것으로 보인다. 토픽 17의 경우 전장연 시위가 지속

됨에 따라 관련 기사가 누적되고 있는 것으로, 토픽 18의 경우 노동 관련 파업이 산업계에 미치는 영향이 크에 따라 논의가 지속되는 것으로 볼 수 있다. 특히 화물연대 파업의 경우 안전운임제에 대한 일몰제의 연장된 상태로, 이후에도 논의가 활발해질 가능성이 있다. 토픽 19의 경우 피해 금액이 크고 범행이 교묘해짐에 따라 경각심을 알리는 차원에서 토픽 비중이 증가하는 추세인 것으로 보인다.

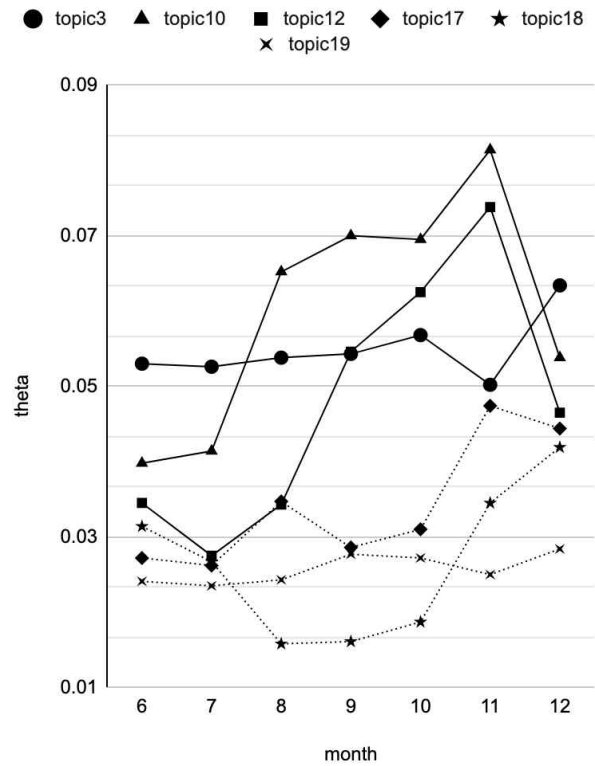


그림 6. Hot 토픽
Fig. 6. Hot topics

Cold 토픽은 시간이 지나면서 논의가 감소한 영역으로 해석할 수 있다. 도출된 Cold 토픽은 1(사회문제에 대한 국가와 정부의 역할과 책임), 5(강력범죄에 대한 수사), 6(심리적 트라우마 및 마음 챙김의 필요성), 7(기상 이변에 따른 폭우, 태풍, 한파), 8(코로나 팬데믹 지속), 9(글로벌 축제 및 행사 재개), 11(법원 재판 및 판결사태에 대한 관심) 그리고 14(쓰레기 배출 및 환경오염 논쟁 심화)로 나타났다. 시간의 흐름에 따른 Cold 토픽의 토픽별 비중 추이를 시각화한 것은 그림 7과 같다.

토픽 1과 토픽 6의 경우 이태원 참사와 같은 커다란 참사가 당시 매우 높았던 토픽의 비중이 시간의 흐름에 따라 감소한 것으로 판단된다. 토픽 5의 경우 계곡살인 등의 사건 등이 시간의 흐름에 따라 토픽의 비중이 감소한 것으로 예상된다. 토픽 7의 경우 호우, 태풍의 피해 및 논의가 8~9월에 집중되어 시간이 지남에 따라 논의가 감소하는 것으로 보인다. 다만 한파가 찾아옴에 따라 다시 증가할 가능성이 있다고 볼 수 있다. 토픽 8과 9의 경우 코로나와 관련된 방역 정책의 완화와

함께 토픽의 비중이 줄어들고 있는 것으로 예상된다. 같은 맥락에서 토픽 14의 경우 코로나 상황을 반영하여 일회용품 사용의 정책 변화에 따라 전체적인 토픽 비중이 감소하는 것으로 해석할 수 있다. 토픽 11의 경우 다른 토픽에 비해 상대적으로 토픽의 비중이 줄어든 것으로 볼 수 있다.

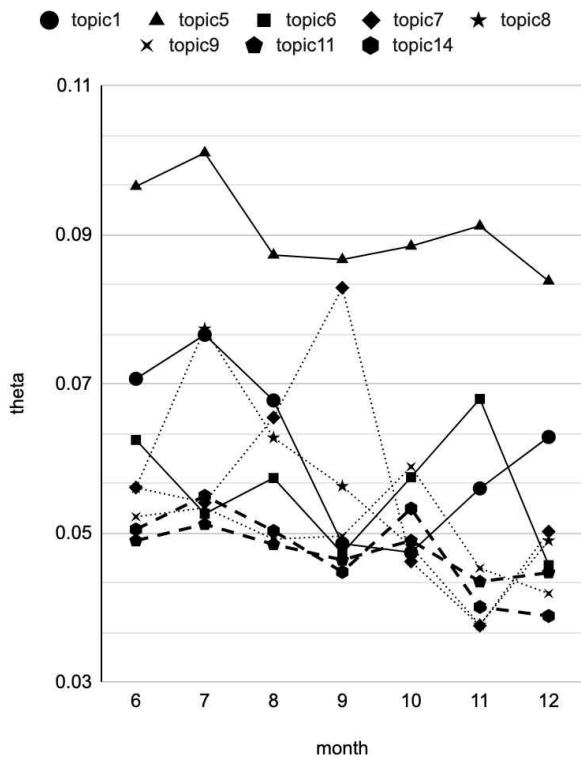


그림 7. Cold 토픽
Fig. 7. Cold topics

V. 결 론

본 연구의 결과는 다음과 같이 요약할 수 있다. 최근 한국의 사회문제 해결방안 모색을 위하여 사회양극화와 다중격차의 관점에서 주요 3개 일간지의 사회면 기사를 수집하고 토픽 모델링 분석을 통해 주요 토픽들을 도출하여 다학제 간 이해를 위한 기초자료를 제공하였다. 토픽모델링 분석 결과 사회문제와 관련해 중점적으로 논의되고 있는 20개의 주요 토픽들을 도출하였으며 정치, 국제, 경제·산업, 사회, 지역·국토의 다섯 가지 기준으로 분류하고 대표적 키워드를 통해 주요 현상들을 해석하였다. 또한 시간적 흐름에 따른 세부 주제별 변화를 확인하여 주요 논의 분야의 경향성을 예측하였다. 한국 사회는 최근 사회와 경제구조 대전환 시기에 새로운 사회위험이 지속적으로 발생하며 점차 상시화 되고 있다. 이러한 사회문제를 다양한 관점에서 바라보고 데이터 기반 문제해결 접근방식을 통해 야기되고 있는 새로운 사회문제를 살펴볼 필요가 있으며, 정량적 분석과 예측은 사회 각 영역에서 선제

적·예방적 대응을 가능케 한다. 특히 본 연구결과 시계열 분석에 따라 Hot 토픽으로 분류된 주제들은 향후 지속적으로 논의될 가능성이 있는 영역으로써, 문제해결을 위한 장기적 노력이 필요한 영역으로 해석할 수 있다. 특히 안전사고 발생 및 대비 필요성 제기(토픽 10), 투자 및 금융사기 피해사례의 증가(토픽 19) 등은 지속적으로 논의될 가능성이 높은 사회문제 영역으로 중장기적 문제해결전략이 필요함을 시사한다.

본 연구의 결과를 통한 시사점과 활용 방안은 다음과 같다. 첫째, 토픽모델링 기법을 사회문제로 분석하기 위해 적용하여 근거기반실천을 제공한다. 토픽모델링은 문서의 집합 속에서 숨겨진 주제를 찾아 비슷한 내용끼리 분류하여 문서 집합 내 숨겨진 주제를 찾아내기 위한 분석 방법으로, 도출된 다양하고 실증적인 분석 결과를 바탕으로 사회문제에 관한 기초자료와 체계적 이해를 제공할 수 있다. 또한 급변하는 사회문제 영역에서의 환경변화를 과학적 데이터에 기반한 객관적인 자료를 통해 분석함으로써 발생 원인을 진단 및 예측하여 문헌 연구, 해외 벤치마킹 등 중관적인 데이터를 통한 문제 해결 대안 제시의 한계점을 보완할 수 있다.

둘째, 사회문제 해결을 위한 정책적 함의를 제공한다. 사회양극화 및 다중격차의 영역에 대한 논의는 향후 사회문제 해소 정책의 방향과도 밀접한 관련을 맺고 있으며 본 연구의 결과로 사회문제 관련요인에 대한 기초자료를 제공할 수 있다. 또한 시간적 흐름에 따른 세부 주제별 변화를 확인함으로써, 향후 지속적으로 논의될 영역을 예측하여 선제적 대응을 위한 정책적 함의를 제공할 수 있다.

마지막으로, 사회문제와 관련된 다학제 간 논의를 제공한다. 본 연구에서는 최근 한국의 사회문제와 주요 현상을 사회양극화와 다중격차를 통해 탐색하고자 하였다. 집단 내 격차가 다차원적 영역에서 중복하여 나타나는 상태인 다중격차의 이해를 위해서는 문제가 발생하는 복합적 영역에 대한 이해가 필수적이다. 최근 한국사회에서는 세대, 지역, 성별, 인종, 계층 등에 따라 이해집단이 다양해지면서 나타나는 중첩적이고 복합적인 사회적 갈등이 증가하였고, 젠더 갈등, 세대 갈등, 지역 갈등 등 단일한 영역이 아닌 다중적 영역에서의 문제들이 상호작용하며 발생하고 있다. 따라서 국제, 경제, 경제·산업, 사회, 지역·국토 등 다학제적 관점에서의 논의가 이루어질 필요성이 있으며, 본 연구의 결과를 통해 복합적인 사회문제를 설명하는 데 기여할 수 있다.

본 연구의 한계점 및 향후 연구를 위한 제언은 다음과 같다. 먼저 본 연구는 최근 한국사회의 문제와 현상을 반영하기 위해 주요 3개 일간지의 사회면 기사를 분석 대상으로 제한하였다. 향후 연구에서는 정치, 경제 등 추가적인 영역의 기사를 수집하고 보다 긴 수집기간을 통해 장기간의 사회문제의 변화양상을 확인할 수 있을 것이다. 다음으로, 본 연구에서는 국내의 사회현상을 분석하는 것으로 분석범위를 설정하고 그에 따른 시사점을 논의하였다. 향후 연구에서는 같은 시점의 국제적 사회현상에 대한 추가적인 분석을 통해 국내의 사회현상의 비교분석이 가능할 것이다.

감사의 글

이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구로(NRF-2021S1A5C2A03088191), 관계부처에 감사드립니다.

참고문헌

- [1] Y. Lee, "The Role of Culture on Perceptions of Polarization and Class Gap," *Journal of Governmental Studies*, Vol. 26, No. 2, pp. 323-357, August 2020. <https://dx.doi.org/10.19067/jgs.2020.26.2.10>
- [2] K.-C. Nam, "Social Polarization and Social Work Practice," *Journal of Critical Social Welfare*, Vol. 22, pp. 71-116, August 2006.
- [3] H. S. Geum, S. H. Ahn, C. U. Park, S. W. Kim, Y. K. Ryu, B. Y. Kim, and H. S. Jung, "Composition and Meaning of Multi-Dimensional Distances - Multi-Dimensional Distances Index in Korean Society -," in *Proceedings of Fall Conference on Korea Association of Social Welfare Policy*, Seoul, pp. 220-250, November 2012.
- [4] J.-W. Shin, S. E. Jeong, and J. Y. Chang, "Social Cleavages and Social Policy Response in the Context of Multiple Inequalities," *Journal of Korean Social Trend and Perspective*, No. 97, pp. 81-110, June 2016.
- [5] G. S. Hwang, "In Search of Conceptualization of 'Multiple Disparities'," *Journal of Korean Social Trend and Perspective*, No. 97, pp. 9-44, June 2016.
- [6] E. Choi, "The Role of Social Allowance in an Era of Increasing Multiple Disparities," *Journal of Human Ecology*, Vol. 24, No. 1, pp. 85-92, April 2020. <https://doi.org/10.36357/johe.2020.24.1.85>
- [7] W.-S. Kim, "Social Polarization in Korea and Multi-dimensional Theory of Justice," *Social Philosophy*, No. 26, pp. 223-250, October 2013.
- [8] G. Byun and K. Kim, "Determinants of Multiple Disparities in Quality of Life among Young Adults in South Korea," *Social Welfare Policy*, Vol. 46, No. 2, pp. 257-285, June 2019. <https://doi.org/10.15855/swp.2019.46.2.257>
- [9] Y. S. Choi, "Inequality Cumulative Society: Overlap and Consequences of Gaps," in *Proceedings of Spring Conference on Korea Association of Social Welfare Policy*, Online, pp. 133-149, June 2021.
- [10] H. Park and M. Park, "On the Capital and Habitus of Provincial Youth in South Korea," *Economy and Society*, No. 131, pp. 55-87, September 2021. <http://dx.doi.org/10.18207/criso.2021..131.55>
- [11] S. Y. Kim and M. J. Park, Characteristics and Challenges of Youth Inequality in Barrier Society, Seoul Institute, Seoul, Policy Report No. 326, June, 2021.
- [12] J. A. Song, J. Y. Jeon, S. Y. Jo, J. J. Lee, S. J. Huh, S. H. Kim, ... and S. Y. Park, NARS Issue of the Year 2022, NARS(National Assembly Research Service), Seoul, 31-9735044-001928-01, December 2021.
- [13] S. W. Lee, J. J. Lee, S. J. Huh, J. M. Song, S. H. Kim, G. H. Kim, ... and M. S. Huh, NARS Issue of the Year 2023, NARS(National Assembly Research Service), Seoul, 31-9735044-001990-01, December 2022.
- [14] E. Park and S. Cho, "KoNLPy: Korean Natural Language Processing in Python," in *Proceedings of Annual Conference on Human and Language Technology*, Chuncheon, pp. 133-136, October 2014.
- [15] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, Vol. 3, No. 1, pp. 993-1022, January 2003.
- [16] B. V. Barde and A. M. Bainwad, "An Overview of Topic Modeling methods and Tools," in *Proceedings of International Conference on Intelligent Computing and Control Systems*, Madurai, India, pp. 745-750, June 2017. <http://doi.org/10.1109/ICCONS.2017.8250563>
- [17] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of LREC 2010 Workshop on New Challenges for NLP Framework*, Valletta, Malta, pp. 46-50, May 2010.
- [18] M. D. Hoffman, F. Bach, and D. M. Blei, "Online Learning for Latent Dirichlet Allocation," in *Proceedings of the International Conference on Advances in Neural Information Processing Systems 23 (NIPS 2010)*, Vancouver, Canada, pp. 856-864, December 2010.
- [19] G. Heinrich, "Parameter Estimation for Text Analysis," *Technical Report*, Vol. 1, pp. 1-32, May 2005.
- [20] M. Röder, A. Both, and A. Hinneburg, "Exploring the Space of Topic Coherence Measures," in *Proceedings of the 8th ACM International Conference on Web Search and Data Mining (WSDM '15)*, Shanghai, China, pp. 339-408, February 2015. <https://doi.org/10.1145/2684822.2685324>
- [21] D. Y. Lee and H. S. Yi "Exploring Methods for Determining the Appropriate Number of Topics in LDA: Focusing on Perplexity and Harmonic Mean Method," *Journal of Educational Evaluation*, Vol. 34, No. 1, pp. 1-30, March 2021. <http://dx.doi.org/10.31158/JEEV.2021.34.1.1>
- [22] Q. Mei, X. Shen, and C. Zhai, "Automatic Labeling of Multinomial Topic Models," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge*

Discovery and Data Mining (KDD '07), San Jose: CA, pp. 490-499, August 2007. <https://doi.org/10.1145/1281192.1281246>

- [23] E. Kim and Y. Seo, "A Method of Calculating Topic KeyWords for Topic Labeling," *Journal of the Korea Society of Digital Industry and Information Management*, Vol. 16, No. 3, pp. 25-36, September 2020. <http://dx.doi.org/10.17662/ksdim.2020.16.3.025>
- [24] C. Sievert and K. E. Shirley, "LDavis: A Method for Visualizing and Interpreting Topics," in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, Baltimore: MD, pp. 63-70, June 2014. <http://doi.org/10.3115/v1/W14-3110>
- [25] J. S. Park, S.-G. Hong, and J.-W. Kim, "A Study on Science Technology Trend and Prediction Using Topic Modeling," *Journal of the Korea Industrial Information Systems Research*, Vol. 22, No. 4, pp. 19-28, August 2017. <https://dx.doi.org/10.9723/jksiis.2017.22.4.019>
- [26] T. L. Griffiths and M. Steyvers, "Finding Scientific Topics," *PNAS*, Vol. 101, No. suppl_1, pp. 5228-5235, April 2004. <https://doi.org/10.1073/pnas.0307752101>

오서경 (Seokyoung Oh)



2016년 : 배재대학교 경영학과 (학사)

2022년~현 재: 한양대학교 교육공학과 석사과정
※ 관심분야 : 조직학습, 인적자원관리, 지식창출

송지훈 (Ji Hoon Song)

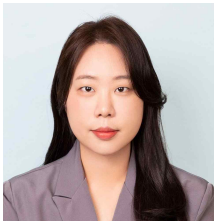


2005년 : The Pennsylvania State University (M.Sc. 인적자원과 조직개발)

2008년 : The Pennsylvania State University (Ph.D. 인적자원과 조직개발)

2008년~2012년: Oklahoma State University 조교수
2012년~2015년: University of North Texas 조교수
2015년~현 재: 한양대학교 교육공학과 교수
※ 관심분야 : 조직개발, 지식경영, 조직창의성, 지식창출

유명현 (Myunghyun Yoo)



2021년 : 한양대학교 (교육학박사)

2016년~2021년: 한양대학교 교육공학과(교수체제설계 및 기업 교육연구 전공)

2020년~2021년: 한양사이버대학교 학습데이터센터 직원

2021년~현 재: 한양대학교 한국교육문제연구소 전임연구원/연구조교수

※ 관심분야 : 경력적응성, 경력개발, 지식관리, 학습분석학

임세민 (Se Min Yim)



2019년 : 한양대학교 교육공학과 (이학사)

2021년~현 재: 한양대학교 교육공학과 석사과정

※ 관심분야 : 조직학습, 지식경영, 조직개발