

토픽모델링을 활용한 4차 산업혁명의 이슈 분석

노설현

안양대학교 ICT융합공학부 통계데이터과학전공 조교수

Analysis of Issues Related to the Fourth Industrial Revolution Based on Topic Modeling

Seol-Hyun Noh

Assistant Professor, Department of Statistical Data Science, ICT Convergence Engineering, Anyang University, Gyeonggi, 14028, Korea

[요 약]

본 연구는 국내의 4차 산업혁명과 관련된 기사들을 LDA 알고리즘에 기반한 토픽모델링 기법으로 분석하여 4차 산업혁명 관련 주요 이슈들을 도출하고 세부적으로 분석함으로써 이미 우리에게 다가온 4차 산업혁명에 대비하기 위한 전략을 수립하는 데 유용한 정보를 생산하고자 하였다. 본 연구에서는 '4차 산업혁명'을 검색어로 하여 추출된 11개의 중앙지와 8개의 경제지, 주요 방송사의 2018년부터 2019년까지 4,389건의 기사를 대상으로 오픈 소프트웨어인 R을 활용한 토픽모델링 기법을 사용하여 토픽별 키워드들을 추출하였다. 각 토픽의 키워드 간 연관성을 나타내는 PMI (Pointwise Mutual Information) 측도를 높이도록 relevance 파라미터 λ 를 최적화하여 토픽별 키워드를 추출하였으며, 키워드들로부터 타당한 근거를 바탕으로 토픽 명을 추론하였다. 추출된 토픽들은 4차 산업혁명의 핵심기술 분야와 경제, 산업, 교육, 문화 전반에서 일어나고 있는 변화 및 정부의 대응 정책들을 폭 넓게 나타내었다.

[Abstract]

In this study I analyzed domestic articles related to the 4th industrial revolution by topic modeling method based on LDA algorithm. I derived major issues related to the 4th industrial revolution and analyzed them in detail to prepare for the 4th industrial revolution already approached to us. Keywords extracted from 4,389 articles of eleven metropolitan newspapers, eight business newspapers and major broadcasting companies; articles were selected by searching for the keyword "the 4th industrial revolution". Keywords were extracted by optimizing the relevance parameter λ to improve the measure of pointwise mutual information (PMI), which shows the correlation among the keywords of each topic, and topic names were inferred from keywords based on valid evidence. The extracted topics widely showed the core technological fields of the 4th industrial revolution, the changes occurring throughout economy, industry, education, and culture, and government response policies.

색인어 : 토픽 모델링, LDA 알고리즘, 4차 산업혁명, 이슈 분석, 빅데이터

Key word : Topic Modeling, LDA Algorithm, Fourth Industrial Revolution, Issues Analysis, Big Data

<http://dx.doi.org/10.9728/dcs.2020.21.3.551>



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 13 February 2020; **Revised** 15 March 2020
Accepted 25 March 2020

***Corresponding Author; Seol-Hyun Noh**

Tel: +82-31-463-0856

E-mail: shnoh@anyang.ac.kr

I. 서론

4차 산업혁명에 대한 화두는 2016년 1월에 스위스 다보스에서 개최된 세계경제포럼(World Economic Forum, WEF)에서 WEF의 회장 클라우스 슈밥(Klaus Schwab)이 “4차 산업혁명(The Fourth Industrial Revolution)”이란 책을 통해 4차 산업혁명의 정의, 4차 산업혁명을 이끄는 기술, 4차 산업혁명의 영향력, 4차 산업혁명의 기술적·사회적 이슈에 대한 설문조사 등을 포괄적으로 다루면서 시작되었다[1].

1차 산업혁명을 기계에 의한 생산, 2차 산업혁명을 대량생산, 3차 산업혁명을 컴퓨터 혁명 혹은 디지털 혁명으로 규정할 수 있으며, 산업혁명이 ‘혁명적’ 효과를 낼 수 있었던 이유는 기술혁신의 상호연관성과 시너지 효과 때문이었다. 슈밥은 4차 산업혁명을 3차 산업혁명의 연장선이 아닌 근거로 속도, 범위 그리고 시스템에 미치는 충격을 들고 있다. 현재와 같은 비약적인 발전 속도는 전례가 없으며, 모든 나라와 산업을 충격에 빠뜨리고 있고, 생산, 관리, 통제 전반에 걸쳐 시스템의 변화를 예고하고 있다는 것이다[2].

산업혁명의 역사를 감안하여 산업혁명이 성립하기 위한 조건을 살펴보면 ① 해당 산업혁명을 선도하는 핵심기술이 존재해야 하고, ② 핵심 기술은 다른 기술혁신과 연결되면서 포괄적인 연쇄효과를 유발해야 하며, ③ 해당 산업혁명으로 인한 경제적 구조의 변화가 이전의 시기와 구분되어야 하고, ④ 사회문화적 차원에서도 이전의 시기와 구분되는 변화가 있어야 한다[1]. 본 연구는 국내의 4차 산업혁명과 관련된 기사들을 토픽 모델링 기법으로 분석하여 4차 산업혁명 관련 주요 이슈를 도출하고 세부적으로 분석함으로써, ‘4차 산업혁명’을 또 하나의 산업혁명으로 바라볼 수 있는지 살펴보고 이미 우리에게 다가온 4차 산업혁명에 대비하기 위한 전략을 수립하는데 유용한 정보를 생산하는데 그 목적이 있다. 뉴스 미디어는 사회 구성원들의 높은 수준의 동의가 나타난 뉴스로서의 가치가 있는 의제들을 보도하므로, 뉴스 기사들을 대상으로 토픽모델링 기법을 통해 높은 빈도로 추출된 주제(의제)들은 사회 구성원들이 공통으로 갖고 있는 관심과 의견 수렴 및 합의에 의해 도출된 주제들로 볼 수 있다. 따라서 본 연구는 ‘4차 산업혁명’이 과학기술과 산업뿐 만 아니라 경제, 교육, 문화 전반에 어떠한 영향을 끼치고 변화를 일으키고 있는지 통찰할 수 있고 국민의 관심 분야 및 이슈를 알 수 있다는 점에서 중요한 연구라고 할 수 있다.

II. 본론

2-1 선행연구 및 연구자료

토픽 모델링 기법은 특히 빅데이터에 대한 동향을 파악하는데 유용하게 활용되고 있는 기법이다. 박자현, 송민(2013)[3]은 토픽 모델링을 통해 국내 문헌정보학 연구 동향을 분석하였고,

Flora(2017)[4]는 토픽 모델링을 이용하여 웹 블로그의 콘텐츠 동향을 분석하였다. Lau, Collier, Baldwin(2012)[5]은 트위터 데이터를 토픽 모델링을 통해 분석하여 온라인 트렌드를 분석하였으며, Jeong, Song(2014)[6]은 컴퓨터 사이언스와 의학 분야에서 논문, 특히, 웹 뉴스 등 다양한 매체를 통해 토픽 모델링을 이용하여 연구 동향을 분석하였다. 황서이, 김문기(2019)[7]는 토픽모델링 기법을 사용하여 1985년부터 2018년까지 총 1,691편의 학술논문에서 서명, 주제어, 초록을 대상으로 인공지능 분야의 연구동향을 파악하였다.

4차 산업혁명은 클라우스 슈밥이 2016년 1월에 스위스 다보스에서 개최된 세계경제포럼(World Economic Forum, WEF)에서 처음으로 언급하면서 알려졌다. 홍정우, 문혜정(2017)[8]은 한국의 ‘4차 산업혁명’ 관련 뉴스를 대상으로 언어 간의 의미연결성을 분석하는 오픈 소프트웨어 khcoder를 분석도구로 사용하여 사회 이슈를 도출하였다. 분석 결과 주된 분야는 비즈니스, 교육, 정책, 기술의 네 가지 분야였으며 주요 이슈는 각각 벤처투자, 융합학문, 경제성장, 일자리창출이었다. 홍정우, 문혜정(2017)[8]의 연구 결과에서는 개체와 노드 사이의 거리와 링크를 기준으로 의미망을 해석하였으며 이슈가 큰 틀에서 구분되어 세부적인 분석이 필요하다고 판단하였다. 현재까지의 연구에서는 토픽모델링 기법을 활용하여 기술동향 및 연구동향을 주로 분석하여 사회적 매체인 4차 산업혁명 관련 뉴스 기사를 대상으로 이슈 분석을 수행하여 4차 산업혁명이 사회, 경제, 산업, 문화 전반에 어떠한 영향을 끼치고 있는지, 사회 구성원들이 4차 산업혁명에 대하여 어떤 분야에 관심을 갖고 어떤 의견을 가지고 있는지에 대한 분석이 필요하다고 판단하였으며 이러한 분석을 통해 이미 우리에게 다가온 4차 산업혁명에 대비하기 위한 전략을 수립하는데 유용한 정보를 생산하고자 하였다.

본 연구에서는 ‘4차 산업혁명’을 검색어로 하여 추출된 11개의 중앙지와 8개의 경제지, 주요 방송사의 2018년부터 2019년까지 4,389건의 뉴스 기사를 대상으로 오픈 소프트웨어인 R을 활용한 토픽모델링 기법을 사용하여 토픽별 키워드들을 추출하였다. 각 토픽의 키워드 간 연관성을 나타내는 PMI(Pointwise Mutual Information) 측도를 높이도록 relevance 파라미터 λ 를 최적화하여 토픽별 키워드를 추출하였으며, 키워드들로부터 타당한 근거를 바탕으로 토픽 명을 추론하였다.

본 연구에서는 2018년 1월 1일부터 2019년 12월 31일까지를 분석기간으로 설정하였다. 분석 대상은 한국언론진흥재단의 기사데이터베이스인 빅카인즈(www.kinds.or.kr)에서 ‘4차 산업혁명’을 검색어로 하여 조선일보, 중앙일보, 동아일보를 비롯한 11개의 중앙지와 매일경제, 서울경제를 비롯한 8개의 경제지, 각종 지역종합지와 주요 방송사에서 추출된 4,389건의 기사이다. [그림 1]은 기사를 수집한 언론사를 나타낸다.

Press (54/54)						
Central (11)						
Kyungnyang Newspaper	Kookmin Daily	Dong-A Daily	Naeil Newspaper	Munhwa Daily	Seoul Newspaper	Segye Daily
Joosen Daily	JoongAng Daily	Hankyoreh	Hankook Daily			
Economy (8)						
Maeil Economy	Money Today	Seoul Economy	Asian Economy	Ajou Economy	Financial News	Korea Economy
Herald Economy						
Local(28)						
Broadcasting company (5)						
KBS	MBC	OBS	SBS	YTN		
Specialized (2)						
Digital Times	Electronic Newspaper					

그림 1. 기사를 수집한 언론사
Fig. 1. Media from which articles are collected

2-2 이론 및 연구방법론

본 연구에서는 문서들의 주제를 도출하기 위해 토픽 모델링 방법을 사용하였다. 토픽 모델링은 학문적 매체인 논문, 기술적 매체인 특허, 사회적 매체인 웹뉴스와 같이 대량의 문헌들을 연구 대상으로 삼는 다양한 분야에서 매체 별로 토픽들의 키워드를 추출하여 토픽을 정의하고, 토픽들의 동향 분석을 통해 각 매체 별로 빅데이터의 동향을 분석하는 도구로 활용되고 있다. 특히 토픽 모델링 분야에서 학계에서 표준으로 인식되고 있는 Blei, Ng, Jordan(2003)[9]이 제안한 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 알고리즘은 각 주제 별 단어 수 분포를 기반으로, 주어진 문서들이 어떤 주제들을 다루고 있는지를 예측하는 데 높은 성능을 보이는 알고리즘이다. LDA 알고리즘의 개요는 다음과 같다.

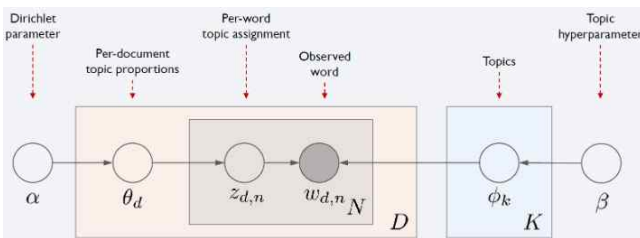


그림 2. LDA 알고리즘의 구조
Fig. 2. Architecture of LDA algorithm

- (1단계) Dirichlet parameter β 를 이용하여 전체 말뭉치(corpus)의 초기 토픽 할당 분포인 ϕ_k 를 설정한다. ($k \in \{1, 2, \dots, K\}$, K 는 토픽 수)
- (2단계) Dirichlet parameter α 를 이용하여 d 번째 문서의 초기 토픽 할당 분포인 θ_d 를 설정한다. ($d \in \{1, 2, \dots, D\}$, D 는 corpus를 구성하는 문서의 수)
- (3단계) d 번째 문서의 n 번째 단어에 할당된 토픽 $z_{d,n}$ 을 깁스샘플링(Gibbs Sampling) 방법을 통해 모든 단어에 대한 토픽 할당 정보가 수렴할 때까지 업데이트하여 ϕ_k 와 θ_d 의 사후 확률분포를 구한다.

LDA 알고리즘의 학습 결과를 시각적으로 표현하는 라이브러리인 LDAvis에서는 가중치(weight parameter) λ ($0 \leq \lambda \leq 1$)를 이용하여 각 토픽의 키워드들에 대한 relevance metric을 측정함으로써 각 토픽의 키워드들의 순위(rank)를 조정한다. 토픽 k 의 단어 w 에 대한 relevance metric $r(w, k)$ 의 정의는 다음과 같다[10].

$$r(w, k|\lambda) = \lambda \log(\phi_{kw}) + (1 - \lambda) \log\left(\frac{\phi_{kw}}{p_w}\right) \quad (1)$$

ϕ_{kw} 는 단어 w 가 토픽 k 에 등장할 확률이며 p_w 는 단어 w 가 말뭉치(corpus)에 등장할 확률이다. $\lambda = 1$ 인 경우는 토픽 별로 가장 자주 등장하는 단어들을 우선적으로 키워드로 선택하게 되며 $\lambda = 0$ 인 경우는 토픽 간에 차이가 많이 나타나는 단어를 우선적으로 키워드로 선택하게 된다. 따라서 토픽 별 키워드들로부터 타당한 토픽을 추론하기 위해서는 최적의 λ 의 값을 설정하여 토픽 별 키워드들을 도출할 필요가 있다. 본 연구에서는 각 토픽의 임의의 두 키워드 쌍에 대한 단어 간 연관성을 나타내는 척도 PMI가 커지도록 λ 의 값을 설정하여 각 토픽이 특징적으로 분명하게 추출될 수 있도록 하였다. PMI의 정의는 다음과 같다[11].

$$PMI(w_i, w_j) = \log \frac{p(w_i, w_j)}{p(w_i)p(w_j)} \quad (2)$$

$p(w_i)$ 는 단어 w_i 가 전체 말뭉치(corpus)에 등장할 확률이며 $p(w_i, w_j)$ 는 단어 w_i 와 w_j 가 한 토픽에 동시에 등장할 확률이다.

단어 w 가 토픽 k 에 특징적으로 나타나는 단어일수록 $\frac{\phi_{kw}}{p_w}$ 의 값이 커지므로 λ 를 0에 가까운 작은 수로 설정하면 토픽 간에 차이가 많이 나는 단어를 키워드들로 선택하게 된다. 이 경우 한 토픽의 키워드에 속하는 두 단어 w_i 와 w_j 를 선택하여 PMI를 측정하면 λ 를 1에 가까운 큰 수로 설정한 경우보다 PMI의 값이 커지므로 키워드들로부터 토픽을 추출하기가 용이해진다. [그림 3]은 2-1의 연구자료에 대하여 LDA 알고리즘을 사용한 토픽모델링을 통해 도출한 한 토픽에 대해 출현 빈도가 높은 상위 30개의 키워드들에 대하여 λ 에 따른 PMI의 변화를 관찰한 결과이다.

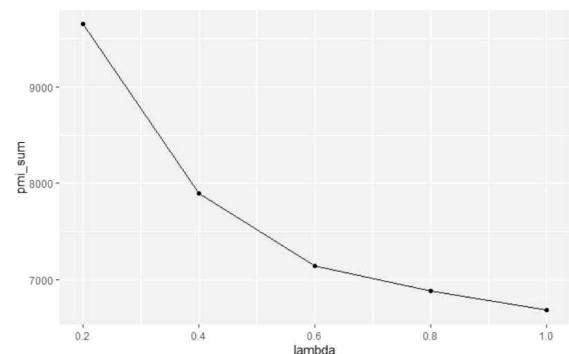


그림 3. λ 에 따른 한 토픽의 키워드 간 PMI 합의 추이
Fig. 3. Trend of PMI between keywords for a topic by λ

본 연구에서는 LDA 알고리즘을 활용한 토픽 모델링 기법을 활용하여 2-1의 연구자료에 대해 [그림 4]와 같은 연구 절차를 통해 국내의 4차 산업혁명 관련 주요 이슈를 세부적으로 분석하였으며, λ 를 0.2로 설정하여 키워드 간의 연관성이 커지게 함으로써 키워드들로부터 토픽을 추출하기가 용이해지도록 하였다.

빅카인즈(www.kinds.or.kr)에서 수집된 4,389건의 기사 문서들을 말뭉치(corpus)로 읽어 들인 후, 명사 형태소가 가장 핵심적인 의미를 지니고 있을 것으로 판단하여 명사 중심으로 데이터를 추출하고 불용어(stopwords)를 제거하는 데이터 전처리 과정을 거쳐 LDA 알고리즘을 적용할 수 있는 데이터 형태로 가공하였다. LDA 알고리즘을 적용할 때, 초기 값으로 설정하는 Dirichlet parameter α 와 β 의 값을 0.02, 0.04, 0.06, 0.08, 0.1에 대해 실험하였으며 실험 결과 추출되는 키워드들에 큰 변화가 없어 토픽 별 키워드 구분이 좀 더 명확해지도록 α 와 β 의 값을 0.02로 설정하였다. LDA 알고리즘의 입력 값인 토픽 수 K 의 최적 값을 결정하기 위해 K 의 값을 변화시키면서 perplexity를 측정하였다. perplexity는 전체 문서들의 집합인 말뭉치(corpus)의 단어 발생 확률이 커질수록 단조 감소하는 측도로서 다음과 같이 정의된다. 전체 문서들의 집합을 D 라 하고, D 를 구성하는 문서의 수를 M , d 번째 문서의 총 단어 수를 N_d , d 번째 문서의 단어 벡터를 $\mathbf{w}_d = (w_1, w_2, \dots, w_{N_d})$ 라 하면,

$$\text{perplexity}(D) = \exp \left\{ - \frac{\sum_{d=1}^M \log \{p(\mathbf{w}_d)\}}{\sum_{d=1}^M N_d} \right\} \quad (3)$$

이다[9]. K 의 값이 5, 7, 10, 15, 20, 25, 30인 경우에 대해 실험하여 perplexity를 측정하였으며, [그림 5]의 결과를 얻었다. 모든 단어의 발생 확률이 커질수록 perplexity는 작아지므로 perplexity의 감소 폭이 작아지는 $K=15$ 를 토픽 수로 설정하여 토픽 모델링 분석이 효율적으로 진행되도록 하였다. [그림 6]은 효율적인 perplexity를 나타내도록 토픽 수를 15로 설정하여 추출된 토픽들이 중복되는 부분이 작게 전체 말뭉치(corpus)에서 고르게 분포함을 보여주는 Intertopic Distance Map이다.

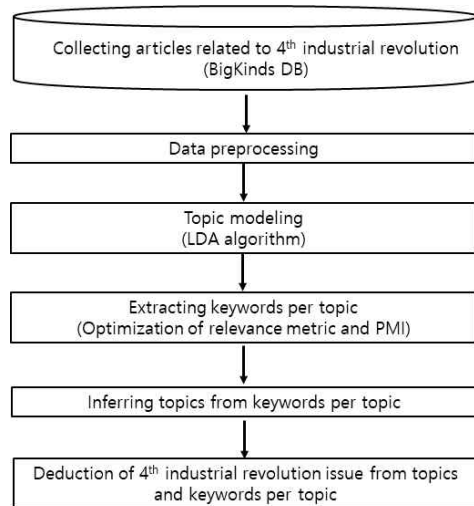


그림 4. 연구절차
Fig. 4. Research process

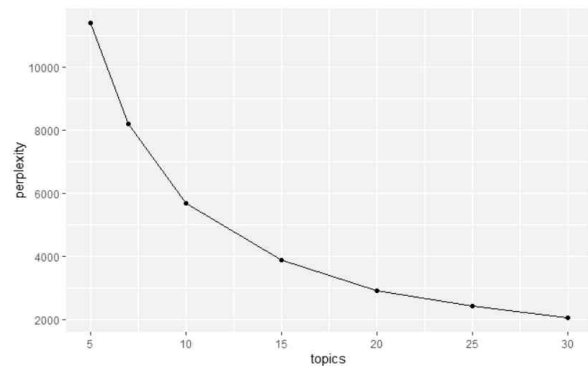


그림 5. 토픽 수에 따른 perplexity의 변화
Fig. 5. Perplexity of number of topics

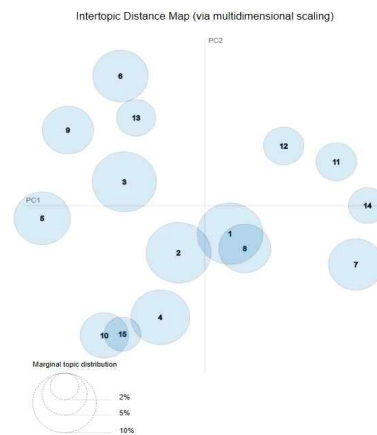


그림 6. $K=15$ 인 경우의 Intertopic Distance Map
Fig. 6. Intertopic Distance Map ($K=15$)

2-3 연구결과

표 1. 토픽모델링 분석 결과

Table 1. Result of topics modeling analysis

No	Topics	Freq. (%)	Keywords
1	The Role of Humanities in the 4th Industrial Revolution (4차 산업혁명 시대에 인문학의 역할)	10.6	human(인간), world(세상), humanity(인류), author(저자), person(사람), civilization(문명), leader(리더), technology(테크놀로지), AlphaGo(알파고), book(저서), nature(자연), emotion(감성), scientist(과학자), philosophy(철학)
2	Big Data and Deregulations of Personal Information (빅데이터와 개인정보 규제 완화)	10.3	regulation(규제), privacy(개인정보), relaxation(완화), restriction(제한), data(데이터), service(서비스), information(정보), permit(허용), government(정부), innovation(혁신), amendment(개정안), big data(빅데이터), institution(제도), protect(보호), uses(활용), rule(규정), industry(산업), enterprise(기업), fintech(핀테크), finance(금융), bill(법안), separation(분리), congress(국회)
3	Governmental and Corporate Activities for Innovation in the 4th Industrial Revolution (4차 산업혁명 시대의 혁신을 위한 정부와 기업의 활동)	10.1	industry(산업), support(지원), innovation(혁신), enterprise(기업), field(분야), small and medium enterprises(중소기업), government(정부), foundation(창업), startup(스타트업), growth(성장), business(사업), invest(투자), manufacturing(제조업), bio(바이오), creation(창출), new industry(신산업), propel(추진), patent(특허), competitiveness(경쟁력), upbringing(육성), enlargement(확대), strategy(전략), market(시장), fourth(4차), venture(벤처), healthcare(헬스케어), ecosystem(생태계)
4	Fourth Industrial Core Technology Areas (4차 산업혁명의 핵심기술 분야)	7.8	5G, AI, service(서비스), data(데이터), platform(플랫폼), KT, SK Telecom(SK텔레콤), communication(통신), autonomous driving(자율주행), information(정보), cloud(클라우드), artificial intelligence(인공지능), IoT, security(보안), VR, uses(활용), solution(솔루션), big data(빅데이터), ICT, autonomous vehicles(자율주행차), network(네트워크), commercialization(상용화), vehicle(차량), video(영상), smartphone(스마트폰), Internet of Things(사물인터넷), smart(스마트), contents(콘텐츠)

No.	Topics	Freq. (%)	Keywords
5	Smart City Creation and Industry Promotion (스마트시티 조성 및 산업진흥)	7.5	area(지역), city(도시), furtherance(조성), business(사업), propel(추진), plan(계획), support(지원), Ulsan(울산), Busan(부산), Gyeongbuk(경북), Gwangju(광주), center(센터), inducement(유치), Pohang(포항), complex(단지), smart city(스마트시티), erection(건립), municipality(지자체), move in(입주), Dae-gu(대구), Daejeon(대전), industrial complex(산업단지), industry(산업), ministry of environment(환경부)
6	Job Problem in the 4th Industrial Revolution (4차 산업혁명 시대의 일자리 문제)	6.6	job(일자리), labor(노동), wage(임금), minimum wage(최저임금), policy(정책), employ(고용), government(정부), worker(노동자), wage increase(인상), labor union(노조), income(소득), work(근로), economy(경제), employee(근로자), irregular workers(비정규직), working hours(근로시간), shorten(단축), labor market(노동시장), Korean confederation of trade unions(민주노총), society(사회), labor and management(노사), consumption(소비), youth(청년), major company(대기업), small and medium enterprises(중소기업)
7	Energy Paradigm Shift in the 4th Industrial Revolution (4차 산업혁명 시대 에너지 패러다임의 전환)	6.5	safety(안전), energy(에너지), nuclear power plant(원전), calamity(재난), accident(사고), power(전력), nuclear power(원자력), electric(전기), occur(발생), uses(활용), renewable energy(재생에너지), ocean(해양), fine dust(미세먼지), solar energy(태양광), equipment(설비), build(구축), plan(계획), hydrogen(수소), de-nuclear power(탈원전)
8	Investment in Leading Companies in the 4th Industrial Revolution (4차산업 선도기업에 대한 투자)	5.8	chairman(회장), stock price(주가), fund(펀드), subsidiary(계열사), stock(주식), flotation(상장), vice-chairman(부회장), yield(수익률), analyst(애널리스트), ETF, group(그룹), M&A(인수), holding company(지주사), investment category(종목), KOSPI(코스피), disposal(매각), securities firm(증권사), stock market(증시), listed company(상장사)

No.	Topics	Freq. (%)	Keywords
9	Changes in Education in response to the 4th Industrial Revolution (4차 산업혁명에 대응한 교육의 변화)	3.8	education(교육), student(학생), university(대학), school(학교), class(수업), scholastic ability(수학), scholastic ability test(수능), teacher(교사), coding(코딩), ministry of education(교육부), curriculum(교육과정), general student admissions(학종), subject(과목), educational superintendent(교육감), department(학과), career(진로), science and engineering(이공계), enter university(진학), entrance examination(입시), admission(입학), course(교과), education office(교육청), selection(선발), human resources(인재)
10	Blockchain Technology and Finance (블록체인 기술과 금융)	3.3	blockchain(블록체인), money(화폐), virtual(가상), deal(거래), virtual currency(가상화폐), cryptography(암호), bitcoin(비트코인), speculation trading(투기), hacking(해킹), investor(투자자), currency(통화), Ethereum(이더리움), issue(발행)
total			72.3

[표 1]은 ‘4차 산업혁명’을 검색어로 하여 수집된 4,389건의 기사들을 대상으로 토픽 모델링을 통해 도출한 주제 별 키워드들과 키워드들로부터 유추한 토픽 명을 나타낸 것이다. 각 토픽의 임의의 두 키워드 쌍에 대한 단어 간 연관성을 나타내는 척도 PMI가 커지도록 λ 의 값을 0.2로 설정하여 키워드들을 도출함으로써 각 토픽이 특징적으로 분명하게 추출될 수 있도록 하였다. Frequency는 전체 문서에서 각 토픽의 단어가 등장한 빈도를 나타낸다. Topic 11 ~ topic 15는 주제 별로 추출된 키워드들이 명확한 토픽을 나타낸다고 판단하기 어려워 분석 대상에서 제외하였다. Topic 1 ~ topic 10의 토픽 명을 주제 별 키워드들을 바탕으로 다음과 같이 유추하였으며, 각 토픽들은 4차 산업혁명 관련 국내의 주요 이슈들을 나타낸다.

Topic 1의 토픽 명을 ‘4차 산업혁명 시대 인문학의 역할’로 추론한 근거는 다음과 같다. 4차 산업혁명의 기술은 단순한 산업의 변화 뿐 아니라 경제, 사회, 문화 전반에 변화를 일으키고 있다. 아인슈타인은 과학기술이란 결국 인류 내면의 목표에 따라 방향을 잡게 된다고 말한 바 있다. 그러므로 기술의 발전에 인간을 맞추는 것이 아니라 인간의 목적에 비추어 기술 발전의 방향을 결정해 나가야 하며, 인문학적 성찰을 통해 기술 발전의 바람직한 방향을 모색하는 것이 필요하다. 기술이 사회에 미치는 영향을 분석하는 주체는 사람이며 윤리적 행동의 기본원칙(도덕철학), 존재의 본성(형이상학), 지식과 이해의 본질(인식론) 등을 아우르는 인문학적 소양을 통해 앞으로 기술이 사회

에서 일으킬 변화를 바람직한 방향으로 이끌어 나갈 수 있는 인문적 지성을 갖춘 인재를 양성하는 것 또한 중요하다.

Topic 2의 토픽 명을 ‘빅데이터와 개인정보 규제 완화’로 추론한 근거는 다음과 같다. 빅데이터는 당초엔 수만, 수억 테라바이트(Terabytes)에 달하는 거대 데이터 집합만을 지칭했으나, 점차 그 데이터의 수집 및 분석의 과정 전반에서 이용되는 관련 도구, 플랫폼, 분석기법까지 포괄하는 용어로 발전하였다. 빅데이터에서 개인정보란 빅데이터 서비스를 위해 수집·저장 및 이용되는 개인을 식별할 수 있는 개인에 대한 모든 정보를 의미한다. 4차 산업혁명을 맞아 데이터 활용이 국가와 기업의 핵심 경쟁력으로 떠오르면서, 2018년 8월에 정부는 ‘데이터 규제혁신’ 정책을 발표하였다. ‘데이터 규제혁신’은 기업과 소상공인, 소비자 모두에게 도움이 되고 혁신 성장과 직결된다며 규제혁신의 필요성을 강조하였다. 데이터 경제 활성화를 위한 규제혁신은 산업계의 요구를 만족시키는 한편 빅데이터를 활용해 일자리를 늘리고 전문 인력을 양성하고자 하는 정부의 의지가 반영된 것이다. 그동안 엄격하게 통제되었던 개인정보 활용이 가능해지면 미국보다 2년 가까이 뒤쳐진 인공지능 기술 격차도 줄일 수 있을 것으로 전망된다. 개인정보가 허락 없이 오용되거나 사생활 침해와 같은 문제점이 발생할 수 없도록 개인정보의 안전한 관리와 보호 역시 중요하다는 국회와 시민단체의 의견을 반영하여 추가 정보의 사용 및 결합이 없는 특정 개인을 알아볼 수 없도록 정보 집합물 간 결합 근거를 마련하고 개인정보처리자의 책임을 강화하는 등의 절차를 수립하는 것이 매우 중요하다.

Topic 3의 토픽 명을 ‘4차 산업혁명 시대의 혁신을 위한 정부와 기업의 활동’으로 추론한 근거는 다음과 같다. 정부에서는 4차 산업혁명 시대의 경제·사회 전반의 혁신을 위해 과학기술, 산업, 사람, 사회제도의 4대 분야의 혁신을 위한 정책을 수립하여 추진하고 있다(기획재정부 홈페이지 혁신성장 포털, <http://www.moef.go.kr/pa/archiveMain.do>). 신산업을 육성하기 위해서는 기업에 대한 정부의 정책이 중요하므로 관련 키워드들이 높은 비중으로 도출되었다. 시장, 성장, 창출, 혁신의 키워드들은 시장 환경이 변화함에 따라 새로운 가치창출과 혁신이 필요함을 의미한다. 대기업들은 ‘4차 산업혁명’ 관련 스타트업에 대한 투자를 매년 확대하고 있다. 의사결정 구조가 복잡한 대기업보다 규모가 작은 스타트업들이 시장 변화에 빠르게 대응할 수 있어 혁신에 유리하기 때문이다. 전 세계 기업들은 자국의 상황에 맞는 전략을 수립하여 4차 산업혁명에 대응하고 있다[12]. 일본은 자국이 현실 데이터(real data)에 강점이 있다는 것에 주목하고 ‘현실 데이터 플랫폼’ 구축을 목표로 5대 핵심 전략을 수립하였다. 중국은 제조업 혁신을 위한 30개년 국가 전략인 ‘중국제조 2025’를 수립하였다. 제조업 혁신의 선두주자인 독일은 첨단기술전략(High-Tech Strategy)인 ‘인더스트리 4.0’을 2014년에 핵심 전략으로 채택하였다.

Topic 4의 토픽 명을 ‘4차 산업혁명의 핵심기술 분야’로 추론한 근거는 다음과 같다. 4차 산업혁명은 topic 4의 기술들이 나타내듯이 초연결, 초지능, 초융합을 특징으로 갖는다. 자율주

행차, 스마트, 5G, VR(Virtual Reality, 가상현실), IoT(Internet of Things, 사물인터넷), AI(Artificial Intelligence, 인공지능), 클라우드, 상용화의 키워드들은 '4차 산업혁명'의 핵심기술인 자율주행자동차의 개발 및 상용화 이슈를 나타낸다. 자율주행자동차에 장착된 각종 센서에서 검출되는 데이터와 운전자의 운전 방식 데이터를 통해 인공지능 기술인 딥러닝은 자율주행자동차의 인지·판단·제어 기능에 활용되고 있다. 자율주행자동차는 인공지능과 IT 기술이 결합된 디지털 디바이스로 발전하고 있으며 클라우드 컴퓨터와 연결되어 차량 및 도로 인프라에 설치된 각종 IoT 센서들과의 통신을 통해 다른 차량, 도로 상태 등의 주변 환경과 실시간으로 정보를 주고받는다. 또한 자율주행자동차는 향후 전방 유리창의 디스플레이 구현을 통해 AR(Augmented Reality, 증강현실) 기능을 탑재할 것으로 전망된다. 2015년 미국 라스베이거스에서 개최된 세계 최대 가전전시회인 CES(Consumer Electronics Show)에서는 자율주행자동차가 상대 차량과 정보를 주고받거나, 차량 내에서 모바일 오피스를 구축하는 등 자동차와 사물인터넷이 융합되는 미래상을 통해 향후 IT 융합의 혁신은 자동차와 사물인터넷 분야로 옮겨가고 있음을 시사하였다. 사물인터넷은 스마트 가전과 스마트 홈을 넘어 자동차와 도시 인프라까지 확산되고 있으며 이러한 현상은 점차 가속화될 전망이다.

인공지능과 빅데이터 역시 '4차 산업혁명'의 핵심기술이다. 인공지능은 머신 러닝(Machine Learning) 알고리즘으로 구현되며, 머신 러닝은 훈련 데이터(Training Data)를 통해 학습된 속성을 기반으로 예측 및 분류하는 알고리즘이다. 인공지능 기술은 방대한 양의 훈련 데이터, 즉 빅데이터를 필요로 한다. 클라우드를 기반으로 한 빅데이터 기술의 발전으로 머신 러닝에 사용할 수 있는 데이터의 양과 종류가 획기적으로 늘어났고, 딥러닝의 방대한 자료를 처리할 수 있는 컴퓨팅 플랫폼도 빠른 속도로 발전하고 있다. 이에 빅데이터 기술은 단순한 데이터 분석 그 이상으로, 인공지능 기술과 결합되어 새로운 비즈니스 가치를 대규모로 제공해줄 것으로 전망된다.

Topic 5의 토픽 명을 '스마트시티 조성 및 산업진흥'으로 추론한 근거는 다음과 같다. 클라우드 슈밥은 그의 저서 '제4차 산업혁명'에서 인간과 사물, 사물과 사물간의 네트워크를 연결하여 물리적 공간에서 제약받지 않고, 사물인터넷의 기술이 미래의 도시를 혁신적으로 변화시킬 것으로 전망하였다. 이러한 혁신적 변화의 도시를 스마트시티(Smart City)라고 한다. 스마트시티는 4차 산업혁명의 플랫폼으로서 도시기반시설은 물론 주거, 의료, 물류, 교통, 문화, 복지 등의 생활밀착형 스마트 도시행정 서비스 시행으로 국민의 삶의 질을 증대시킬 수 있으며 노후화된 도시 인프라 시스템을 지능형 관리시스템으로 전환해 국가 전체 유지관리 비용을 획기적으로 절감하고 많은 새로운 일자리를 창출할 수 있다. 또한 국제 스마트시티 시장의 진출로 국가 성장동력을 창출할 수 있고, 4차 산업혁명의 기술 진화에 선도적으로 대응할 수 있다. 2017년에 조직된 대통령 직속 기구인 4차산업혁명위원회에서는 2018년 공기기업 사업지인 '세종시 5-1 생활권'과 '부산 에코델타시티'를 스마트시티

국가 시범단지 사업지로 선정하여 융복합 신기술의 테스트베드로 활용하여 산업 생태계를 조성하고 국내 스마트시티 발전 방향을 제시하는 한편 해외진출을 도모하고 있다. 그 외에도 서울, 인천, 경기도, 전라도, 강원도, 경상도에 스마트시티 시범도시를 추진하고 있으며, 충청도, 세종, 대전, 대구에서 스마트시티 건설사업을 추진하고 있다[13].

Topic 6의 토픽 명을 '4차 산업혁명 시대의 일자리 문제'로 추론한 근거는 다음과 같다. 한국경제는 장기화되는 저성장, 고령화, 청년 실업과 비정규직 확산으로 고용이 불안정한 상황이다. 이 상황에서 4차 산업혁명으로 인간의 일자리가 빼앗길 수 있다는 담론은 미래의 고용에 대한 부정적 인식이 확산될 수 있는 조건이 된다. 기술변화의 일자리에 대한 영향은 대체효과, 보완효과, 생산효과의 종합이다. 즉, 신기술로 더 저렴한 기계가 개발되어 도입되면서 노동력의 대체로 이어지는 대체효과, 신기술 기반 기계의 도입으로 해당 기계를 다룰 수 있는 노동력에 대한 수요 증가인 보완효과, 신기술이 생산성을 향상시켜 생산량이 증가하여 노동수요가 증가하는 생산효과로 구성된다[14]. 4차 산업혁명이 야기할 일자리에 대한 양적 영향에 대해서는 연구 결과 사이에 상당한 괴리가 존재하고 있다. 하나의 일자리 안에서도 직무 구성의 차이에 따라 자동화의 영향이 다를 수 있다. 일자리를 유지한 상태에서 ICT 기반의 자동화는 일부 직무를 대체하는 형태이기 때문에, 비판적 전망처럼 급격한 일자리 소멸이 현실화되기는 쉽지 않다. 더구나 한 분야에서의 기술 발전은 또 다른 분야의 혁신으로 연결돼 새로운 산업과 직업이 생기면서 소멸되는 일자리를 상쇄할 것이라는 전망도 있다. 직무구성의 변화는 고용구조의 변화, 일자리의 성격 변화, 일자리의 질에 상당한 영향을 미칠 것을 예상할 수 있다. 따라서 노사는 혁신적 변화에 대해 어떻게 적응할지를 논의할 기구를 마련하는데 공동의 노력을 기울여야 한다.

기술혁신에 따른 일자리 총량에 대한 충격의 크기와 무관하게 공정혁신, 제품혁신에 따라 직업의 구성, 기존 직업수행에 필요한 직무구성의 변화는 불가피하며, 이는 기존 근로자는 물론 청년 구직자들에게 새로운 숙련습득을 위한 평생학습 과제를 제기한다. 즉 변화하는 숙련수요에 대응한 교육훈련과 적응능력의 향상이 중요하다. 지금까지는 "교육-노동생산-여가"의 연령분절형 생애과정(각각 청년기-중년기-노년기에 배타적으로 경험)이 지배적이었다면 앞으로는 "교육노동생산여가"의 연령통합형 생애과정(연령과 관계없이 평생에 걸쳐 교육을 받고 노동과 생산을 하고 여가를 즐기는 생애과정을 경험)으로 전환될 것으로 전망된다[15]. 새로운 아이디어를 생성하고 계획하는 창조성이 요구되는 시대에서 사람들은 일과 개인의 삶 사이의 균형을 의미하는 '일생활균형(Work and Life Balance)'을 직장을 구할 때 중요한 가치로 여기고 있으며, 기업에서도 우수한 인재들이 창의적으로 일할 수 있는 업무 환경을 조성하고 업무 효율성을 극대화하기 위해서 선택적 근로시간제, 재택근무제, 근로시간 단축 근무 등의 제도를 통해 '일과 가정의 양립'이 가능한 일터를 조성하고 있다.

Topic 7의 토픽 명을 '4차 산업혁명 시대 에너지 패러다임의

전환'으로 추론한 근거는 다음과 같다. 4차 산업혁명 기술은 에너지 산업의 근본적인 변화를 유도하고 새로운 서비스 창출에 기여하고 있다. 에너지전환은 원전과 화력발전을 줄이고 신재생에너지를 확대하는 것은 물론 에너지의 효율적인 소비를 포함한다. IoE(Internet of Energy)를 통해 에너지는 언제, 어디서든 양방향으로 전달될 수 있고, 에너지 소비 모니터링이 개별 기기단위에서 지역, 국가, 글로벌 단위로 가능하게 된다. 소비자는 신뢰성 높고 유연하며 효율적이면서도 경제적인 에너지 공급네트워크를 통해 중앙집중형 대규모 발전소와 태양광과 풍력 등 분산형 소규모 신재생에너지 발전원들을 하나의 융합 시스템으로 사용할 수 있게 될 전망이다. 소규모 친환경 분산형 전원이 확대되면서, 점진적으로 화석연료 발전을 대체해나가는 과정에서 저탄소 에너지 시스템으로의 전환이 이루어지고 있다. 빅데이터를 통해 건물에서 사용하는 에너지를 효율적으로 관리하고 재생에너지로 생산된 전력을 스마트 기술을 통해 적절히 분배하고 조절하는 등 에너지 산업과 디지털의 융복합은 이미 이루어지고 있으며, 클라우드 기반의 소프트웨어를 이용하는 소규모 신재생에너지 발전 설비와 에너지저장장치는 이미 상용화되고 있다.

Topic 8의 토픽 명을 '4차산업 선도기업에 대한 투자'로 추론한 근거는 다음과 같다. 4차 산업혁명이 가시화되면서 증권업계는 새로운 투자 아이디어와 종목을 찾는 데 골몰하고 있다. 매출과 자산을 바탕으로 기업을 평가하고 대출하던 기업 금융을 벗어나 기업의 기술력과 미래 성장성을 복합적으로 평가하고 투자하는 방향으로 변화하고 있다. 전 세계 시가총액 상위를 차지하고 있는 애플, 구글, 마이크로소프트, 아마존, 페이스북이 모두 4차 산업혁명과 관련된 기업이다. 4차 산업은 이미 선진국에서는 자리 잡은 지 오래다. 알파고를 탄생시키고 자율주행자동차에 많은 투자를 하고 있는 알파벳(Alphabet, 구글의 지주회사), 클라우드 등 다양한 분야에서 두각을 보이고 있는 아마존(Amazon), 일본의 화낙(FANUC), 스위스의 ABB 등의 기업들은 4차산업 관련 꾸준한 수익을 내고 있고, 여러 나라에서 향후 인류의 먹거리 산업으로 4차산업을 적극 육성하고 있다. 우리나라는 아직 4차 산업혁명의 태동기나 ICT 역량이 뛰어나 머지않아 4차 산업혁명 관련 수익을 내는 기업들이 등장할 것으로 전망된다. 최근 글로벌 주식시장은 역사적인 변곡점을 맞이하고 있다. 최근 수년간 4차산업 관련 기업들이 매년 20%~30%씩 성장했고 시장이 고성장 중에 있다. 2018년 8월 애플이 증가 기준 시가총액 1조 달러를 넘어섰으며 2019년 5월에는 마이크로소프트 역시 같은 반열에 올랐다. 아마존 역시 2018년 9월 장중 시가총액이 1조 달러에 달한 바 있다. 글로벌 주식시장 흐름이 4차 산업으로 옮겨가는 모양새다. 현재 유가증권시장에서 거래되고 있는 4차 산업혁명 상장지수펀드(ETF)는 미래에셋자산운용의 'TIGER 글로벌 4차산업혁신기술', KB자산운용의 'KBSTAR 글로벌 4차산업 IT', 삼성자산운용의 'KODEX 글로벌 4차산업 로보틱스', 한국투자신탁운용의 'KINDEX 미국 4차산업 인터넷', 한화자산운용의 'ARIRANG 미국 나스닥 기술주'의 5개가 운용 중이다.

Topic 9의 토픽 명을 '4차 산업혁명에 대응한 교육의 변화'로 추론한 근거는 다음과 같다. 4차 산업혁명 시대 산업구조 변화와 직무 변화에 따른 일자리 변동이라는 역동에 대해 국내외 국가들은 교육 분야에서 인재양성과 대학교육의 혁신을 통해 대응하고 있다. 교육부는 '지능정보화사회에 대응한 중장기 교육정책의 방향과 전략'에서 '한국형 온라인 공개강좌(K-MOOC) 운영 사업'과 '4차 산업혁명 혁신선도대학 사업'을 추진하고 있으며 과학기술정보통신부에서는 'SW 중심대학 지원사업'을, 고용노동부에서는 '4차 산업혁명 선도인력 양성사업'을 추진하고 있다. 개별 대학들은 정부의 재정지원 사업에 맞추어 융합전공을 신설하는 등의 교과목 재설정, 교수 방법의 다양화, 교육 환경의 IT인프라를 추진하고 있다[16]. 고등교육에서도 기술변화에 끊임없이 적응할 수 있는 인재를 육성하기 위해 자기주도 학습역량과 창조적 문제해결역량, 소통기반 협력역량을 갖추도록 교수학습방식을 전환할 필요가 있다. 따라서 프로젝트학습과 수행평가를 중요시하는 교수학습방식을 개발하면서 대학입시에서도 대학의 자율성을 확대하는 방향으로의 변화가 필요하다[17]. 또한 5년 단임제의 한계를 넘어서는 '교육개혁위원회'에서 미래형 대입제도 및 국가교육과정을 차세대의 미래와 국가의 장기발전을 최우선하여 제시하고, 교육부와 교육청의 구조를 개방적이고 수평적으로 개혁하여 민간과 타 부처 및 공공조직들과 소통하고 협력할 수 있도록 교육개혁이 전방위로 추진될 필요가 있다[17].

Topic 10의 토픽 명을 '블록체인 기술과 금융'으로 추론한 근거는 다음과 같다. 블록체인(blockchain)은 관리 대상인 블록(block)이라고 하는 소규모 데이터들이 P2P 방식을 기반으로 생성된 체인(chain) 형태의 연결고리 기반의 분산 데이터 저장 환경에 저장되어 누구도 임의로 수정할 수 없고 누구나 변경의 결과를 열람할 수 있는 분산 컴퓨팅 기술 기반의 데이터 위변조 방지 기술이다[18]. 블록체인은 기술적으로 공개키 알고리즘과 해시(Hash) 암호화 기술, 분산처리구조를 이용해 중앙 집중식 원장 구조를 저비용의 분산 원장 구조(Distributed Ledger)로 대체할 수 있다. 블록체인이 활용될 수 있는 분야는 금융거래, 재산권 설정(공증, 증명, 보험, 등기서비스, 인증, 특허 등), 헬스케어 등의 사적인 부문과 전자투표 등 거버넌스와 관련된 공공분야 전반이다. 특히 블록체인 기반의 암호화폐인 비트코인을 시작으로 현재는 많은 종류의 암호화폐가 개발, 보급되어 있다. 국내외 금융기관들은 각종 컨소시엄을 구성하여 금융산업 내 블록체인 기술에 관한 표준과 프로토콜(protocol)에 관한 기본틀/framework)을 마련 중이다. 2015년 9월에 결성된 블록체인 컨소시엄 R3 CEV는 씨티그룹, 뱅크오브아메리카, JP모건체이스, 골드만삭스, 모건스탠리, UBS 등 전 세계 43개 글로벌 금융회사들이 블록체인 기술의 표준화를 위해 구성한 협력체(consortium)이다. 여기에는 LG CNS, KEB하나은행, KB국민은행, 신한은행 등 국내 기업도 참여하고 있다.

[표 1]에서 보듯이 '4차 산업혁명'을 검색어로 수집된 4,389건의 기사들에서 추출된 키워드들과 토픽들은 4차 산업혁명의 핵심기술 분야와 경제, 산업, 교육, 문화 전반에서 일어나고 있

는 변화 및 정부의 대응 정책들을 폭 넓게 나타낸다. 이 중 ‘4차 산업혁명 시대에 인문학의 역할’ 토픽이 전체 토픽 빈도 1,772,137건 중 187,613건으로 10.6%를 차지하였고, ‘빅데이터와 개인정보 규제 완화’ 토픽이 1,772,137건 중 183,029건으로 10.3%, ‘4차 산업혁명 시대의 혁신을 위한 정부와 기업의 활동’ 토픽이 1,772,137건 중 179,578건으로 10.1%를 차지하여 4차 산업혁명으로 인해 야기될 큰 틀에서의 사회변화 토픽들에 대한 관심이 가장 크게 나타났다. 다음으로는 ‘4차 산업혁명의 핵심기술 분야’ 토픽이 1,772,137건 중 138,683건으로 7.8%를 차지하였고, ‘스마트시티 조성 및 산업진흥’ 토픽이 1,772,137건 중 133,735건으로 7.5%, ‘4차 산업혁명 시대의 일자리 문제’ 토픽이 1,772,137건 중 117,708건으로 6.6%, ‘4차 산업혁명 시대 에너지 패러다임의 전환’ 토픽이 1,772,137건 중 114,748건으로 6.5%를 차지하여 4차 산업혁명으로 인해 야기될 산업의 변화에 대한 관심이 두 번째로 큰 것으로 나타났다. 다음으로는 ‘4차산업 선도기업에 대한 투자’ 토픽이 전체 토픽 빈도 1,772,137건 중 102,018건으로 5.8%를 차지하였고 ‘4차 산업혁명에 대응한 교육의 변화’ 토픽이 1,772,137건 중 67,289건으로 3.8%, ‘블록체인 기술과 금융’ 토픽이 1,772,137건 중 58,174건으로 3.3%를 차지하였다. 도출된 키워드들과 토픽들로부터 4차 산업혁명은 산업과 과학기술분야에서만만의 변화가 아니라 문화 전반에 변화를 일으키고 있음을 알 수 있다. 토픽 추출 결과에 비추어 볼 때, 정부가 「제4차 산업혁명에 대응한 지능정보사회 중장기 종합대책」에서 첫째, 글로벌 경쟁에서 뒤처지지 않고 지능정보기술을 이끌어 나갈 수 있는 자체 기술력 및 데이터-네트워크 인프라 확보와 둘째, 공공서비스 및 민간산업 전반에 지능정보기술 도입을 조기 확산하여 생산성 향상 및 국가 경쟁력 확보, 그리고 셋째, 변화하는 사회상을 반영한 교육-고용-복지 제도를 통해 소외계층 없이 국민 모두가 혜택을 누리는 안전한 지능정보사회 구현을 정책목표로 하여 세부과제를 추진해 나갈 것이라고 밝힌 부분은 매우 바람직한 정책방향이라고 판단된다.

III. 결 론

본 연구는 ‘4차 산업혁명’을 키워드로 검색하여 국내 언론사에서 수집된 기사들로부터 핵심 주제들을 찾아내고자 하는 토픽모델링 기법을 통해 4차 산업혁명의 영향으로 사회에서 일어나고 있는 변화와 국민들의 관심 분야를 도출했다는 점에서 의의를 지닌다.

연구 방법으로는 LDA 알고리즘에 기반한 토픽 모델링 기법을 사용하였으며 Dirichlet parameter인 α 와 β 의 값을 변화시키면서 추출되는 키워드들을 관찰하였고, LDA 알고리즘의 입력 값인 토픽 수 K 의 최적 값을 결정하기 위해 K 의 값을 변화시키면서 perplexity를 측정하여 토픽 모델링 분석이 효율적으로 진행될 수 있는 파라미터 α , β , K 의 값을 결정하였다. 각 토픽

의 임의의 두 키워드 쌍에 대한 단어 간 연관성을 나타내는 척도 PMI(Pointwise Mutual Information)가 커지도록 λ 의 값을 설정하여 각 토픽이 특징적으로 분명하게 추출될 수 있도록 하였다.

본 연구에서는 토픽 모델링 기법으로부터 도출된 주제 별 키워드들로부터 타당한 근거를 바탕으로 토픽 명을 유추하여, 4차 산업혁명의 영향으로 사회에서 일어나고 있는 변화와 국민들의 관심 분야를 객관적으로 도출하였다. 연구 결과 핵심 주제로 도출된 토픽들은 전체 문서에서 차지하는 토픽 빈도순으로 나열하면, ‘4차 산업혁명 시대에 인문학의 역할’, ‘빅데이터와 개인정보 규제 완화’, 4차 산업혁명 시대의 혁신을 위한 정부와 기업의 활동’, ‘4차 산업혁명의 핵심기술 분야’, ‘스마트시티 조성 및 산업진흥’, ‘4차 산업혁명 시대의 일자리 문제’, ‘4차 산업혁명 시대 에너지 패러다임의 전환’, ‘4차산업 선도기업에 대한 투자’, ‘4차 산업혁명에 대응한 교육의 변화’, ‘블록체인 기술과 금융’의 10가지로서 4차 산업혁명의 핵심기술 분야와 경제, 산업, 교육, 문화 전반에서 일어나고 있는 변화 및 정부의 대응 정책들을 폭 넓게 나타내었다. 토픽 추출 결과에 비추어 볼 때, 정부가 「제4차 산업혁명에 대응한 지능정보사회 중장기 종합대책」에서 글로벌 수준의 지능정보기술 기반 확보와 전 산업의 지능정보화 촉진, 사회정책 개선을 통한 선제적 대응을 정책목표로 하여 세부과제를 추진해 나갈 것이라고 밝힌 부분은 매우 바람직한 정책방향이라고 판단된다.

향후에는 relevance 파라미터 λ 가 작아질수록 각 토픽의 임의의 두 키워드 쌍에 대한 단어 간 연관성을 나타내는 척도 PMI가 커진다는 사실을 수학적으로 증명한다면 관련 연구들이 양적, 질적으로 더욱 풍부해질 것이다.

참고문헌

- [1] S. Song, "Historical development of industrial revolutions and the place of so called the fourth industrial 'revolution,'" *Journal of Science and Technology Studies*, Vol. 17, No. 2, pp. 5-40, 2017.
- [2] K. Schwab, *The fourth industrial revolution*, Portfolio Penguin, 2016.
- [3] J. Park and M. Song, "A study on the research trends in library & information science in Korea using topic modeling," *Journal of the Korean society for information management*, Vol. 30, No. 1, pp. 7-32, 2013.
- [4] S. T. Flora, "Blogger-link-topic model for blog mining," *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 28-39, 2017.
- [5] J. H. Lau, N. Collier and T. Baldwin, "On-line trend analysis with topic models: #twitter trends detection topic model online," *Proceedings of COLING 2012*, pp. 1519-1534, 2012.
- [6] D. H. Jeong and M. Song, "Time gap analysis by the topic

model-based temporal technique,” *Journal of Informetrics*, Vol. 8, No. 3, pp. 776-790, 2014.

- [7] S. Hwang and M. Kim, “An analysis of artificial intelligence (A.I.)_related studies' trends in Korean focused on topic modeling and semantic network analysis,” *Journal of Digital Contents Society*, Vol. 20, No. 9, pp. 1847-1855, 2019.
- [8] J. Hong and H. Moon, “Study on social issue with semantic network analysis of news on 4th industrial revolution,” *Integrated Academic Conference of Korean Academic Society of Business Administration*, pp. 180-201, 2017.
- [9] D. M. Blei, A. Ng and M. I. Jordan, “Latent Dirichlet allocation,” *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, 2003.
- [10] C. Sievert and K. E. Shirley, "LDavis: A method for visualizing and interpreting topics", *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pp. 63-70, 2014.
- [11] D. Newman, S. Karimi, and L. Cavedon, "External evaluation of topic models," *Proceedings of the 14th Australasian Document Computing Symposium*, pp. 11-18, 2009.
- [12] D. Lee, “4th industrial revolution and innovation policy,” *SCIENCE & TECHNOLOGY POLICY*, Vol. 27, No. 5, pp. 46-51, 2017.
- [13] TTA, *Core fusion cases of the fourth industrial revolution: Smart city concept and standardization status*, Standardization issue 2018-1, 2018.
- [14] KISDI, *Fourth industrial revolution and jobs*, KISDI Premium Report 2017-9, 2017.
- [15] H. Oh, “The fourth industrial revolution and job crisis in the Korean economy,” *The Korean Economic Forum*, Vol. 11, No. 2, pp. 93-115, 2018.
- [16] E. Kang, “The fourth industrial revolution and the future of university education,” *Journal of Educational Innovation Research*, Vol. 29, No. 1, pp. 279-297, 2019.
- [17] J. Lee, *Educational reform through project learning(II)*, KDI Research Report 2017-04, 2017.
- [18] ETRI Standard Research Division, *Blockchain*, ETRI Standardization Trend 2017-2, 2017.



노설현(Seol-Hyun Noh)

1998년 : 고려대학교 대학원
(이학석사)

2003년 : 고려대학교 대학원
(이학박사)

2003년~2005년: 삼성전자

2019년~현 재: 안양대학교 통계데이터과학전공 조교수

※관심분야 : 빅데이터(Big Data),
인공지능 알고리즘 및 SW(AI Algorithms and SW),
확률론적 모델링(Probabilistic Modeling) 등