

토픽모델링을 활용한 인공지능 관련 이슈 분석

노설현

안양대학교 ICT융합공학부 통계데이터과학전공 조교수

Analysis of Issues Related to Artificial Intelligence Based on Topic Modeling

Seol-Hyun Noh

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

요 약 본 연구는 국내의 인공지능과 관련된 기사들을 LDA 알고리즘에 기반한 토픽모델링 기법으로 분석하여 인공지능 관련 주요 이슈들을 도출하고 세부적으로 분석함으로써 인공지능 기술이 전(全) 산업 분야와 융합을 통해 창출할 수 있는 새로운 가치를 통찰하고, 인공지능 기술을 지식 경영에 적용할 수 있는 분야를 도출하는데 유용한 정보를 생산하고자 하였다. 본 연구에서는 '인공지능'을 검색어로 하여 추출된 11개의 중앙지와 8개의 경제지, 주요 방송사의 2016년부터 2019년까지 3,889건의 기사를 대상으로 오픈 소프트웨어인 R을 활용한 토픽모델링 기법을 사용하여 토픽 별 키워드들을 추출하였다. 각 토픽의 키워드 간 연관성을 나타내는 PMI(Pointwise Mutual Information) 측도를 높이고록 relevance 파라미터 λ 를 최적화하여 토픽 별 키워드를 추출하였으며, 키워드들로부터 타당한 근거를 바탕으로 토픽 명을 추론하였다. 추출된 토픽들은 인공지능 기술의 응용 분야와 사회, 경제, 산업, 문화 전반에서 일어나고 있는 변화 및 정부의 지원 정책과 비전을 폭 넓게 나타냈다.

주제어 : 토픽모델링, LDA 알고리즘, 인공지능, 이슈 분석, 지식 경영, 빅데이터

Abstract The present study determined new value that can be created through the convergence between artificial intelligence technology (AIT) and all industries by deriving and thoroughly analyzing major issues related to artificial intelligence (AI). This study analyzes domestic articles related to AI using topic modeling method based on LDA algorithm. Keywords were extracted from 3,889 articles of eleven metropolitan newspapers, eight business newspapers and major broadcasting companies; articles were selected by searching for the keyword "artificial intelligence". Keywords were extracted by optimizing the relevance parameter λ to improve the measure of pointwise mutual information (PMI), which shows the association among the keywords of each topic, and topic names were inferred from keywords based on valid evidence. The extracted topics widely showed changes occurring throughout society, economy, industries, culture, and the support policy and vision of the government.

Key Words : topic modeling, LDA algorithm, AI, issues analysis, knowledge management, big data

*Corresponding Author : Seol-Hyun Noh(shnoh@anyang.ac.kr)

Received March 25, 2020

Accepted May 20, 2020

Revised April 9, 2020

Published May 28, 2020

1. 서론

인공지능(Artificial Intelligence, AI)은 학습, 이해, 추론 및 상호작용과 같은 인간과 같은 인지 과정을 수행하는 기계를 설명하는 데 사용하는 용어로서, 우리나라는 2016년 12월 공개한 「제4차 산업혁명에 대응한 지능정보사회 중장기 종합대책」에서 인공지능을 “인간의 인지 능력(언어·음성·시각·감성 등)과 학습, 추론 등 지능을 구현하는 기술로 인공지능 SW/HW, 기초기술(뇌과학·산업 수학 등)을 포괄”한다고 정의하였다. 1990년대 중반부터 인터넷의 폭발적 확산으로 생성된 방대한 양의 데이터와 GPU를 비롯한 컴퓨팅 자원의 발달로 인해 제프리 힌튼(Geoffrey Hinton) 교수가 제안한 딥러닝(Deep Learning) 기술이 실용화되기 시작했다. 2016년 구글의 알파고 1.0(16만 기보 지도학습기반 심층강화학습과 확률적 샘플링 기반 의사결정)이 이세돌 9단을 이기고, 2017년 알파고 2.0(비지도학습 소량데이터 기반 자가학습)이 커제 및 최고 수준의 바둑기사들에게 승리하면서 인공지능 기술이 일반인들에게도 확실히 인식되고 재조명되는 계기가 되었다. 영국의 컨설팅 회사인 프라이스워터하우스쿠퍼스(PwC)의 연구에 따르면 사물인터넷(IoT)에서 생성된 방대한 데이터는 인공지능 기술과 결합되어 차세대 디지털 혁명을 일으킬 것으로 예상되며, 인공지능을 사용해 일상적인 작업의 자동화를 실시함으로써 생산성이 향상되어 2030년까지 글로벌 GDP가 14%(15.7조 달러 상당)까지 증가할 것으로 추정된다[1]. 그러나 생산성과 혁신의 격차에서 인공지능 혜택의 불균형 분배에 대한 우려도 제기되고 있어 인공지능이 경제, 사회적으로 미칠 영향에 대한 논의가 활발하게 전개되고 있다.

이에 본 연구에서는 국내의 인공지능과 관련된 뉴스 기사들을 토픽모델링 기법으로 분석하여 인공지능 관련 주요 이슈를 도출하고 세부적으로 분석함으로써, 인공지능 기술의 응용 분야와 사회, 경제, 산업, 문화 전반에서 인공지능 기술이 일으키고 있는 혁신과 국민의 관심 분야를 도출하였다. 뉴스 미디어는 사회 구성원들의 높은 수준의 동의가 나타난 뉴스로서의 가치가 있는 의제들을 보도하므로, 뉴스 기사를 대상으로 토픽모델링 기법을 통해 높은 빈도로 추출된 주제(의제)들은 사회 구성원들이 공통으로 갖고 있는 관심과 의견 수렴 및 합의에 의해 도출된 주제(의제)들로 볼 수 있다. 따라서 본 연구는 인공지능 기술이 전(全) 산업 분야와 융합을 통해 창출할 수 있는 새로운 가치를 통찰하는 데 유용한 정보를 생산한다는 점에서 중요한 연구라고 할 수 있다.

2. 선행연구 및 연구자료

2.1 선행연구

토픽모델링 기법은 특히 빅데이터에 대한 동향을 파악하는데 유용하게 활용되고 있는 기법이다. 박자현, 송민(2013)은 토픽모델링을 통해 국내 문헌정보학 연구 동향을 분석하였고[2], Flora(2011)는 토픽모델링을 이용하여 웹 블로그의 콘텐츠 동향을 분석하였다[3]. Lau et al.(2012)은 트위터 데이터를 토픽모델링을 통해 분석하여 온라인 트렌드를 분석하였으며[4], Jeong, Song(2014)은 컴퓨터 사이언스와 의학 분야에서 논문, 특허, 웹 뉴스 등 다양한 매체를 통해 토픽모델링을 이용하여 연구 동향을 분석하였다[5].

박주섭, 홍순구, 김종원(2017)은 미국 특허 문서 중 AI(Artificial Intelligence) 초록을 대상으로 토픽모델링 기법을 활용하여 20개의 AI 세부기술을 추출하고, 도출된 세부기술에 대해 핵심기술을 파악하였으며, 연도별 비중 추이 분석을 통하여 Hot기술과 Cold기술을 분류하였다[6]. 정명석, 박성현, 채병훈, 이주연(2017)은 한국인 저자가 게재한 SCIE(E) 저널의 논문 중 인공지능 관련 논문을 수집하여 빈도분석과 키워드 네트워크 분석을 수행하여 시간의 흐름에 따라 인공지능 관련 연구는 이론적 연구에서 실용, 응용적 연구가 많아지는 것을 확인하였다[7]. 정명석, 정소희, 이주연(2018)은 국내외 특허데이터 중 ‘인공지능’ 키워드 검색 결과 도출된 데이터를 대상으로 키워드 네트워크 분석 및 IPC 분류 기준 공백기술 분석을 수행하여 국내 인공지능 관련 기술 중 데이터 인식 기술과 디지털 정보 전송기술이 상대적으로 부족함을 확인하였다[8]. 정명석, 이주연(2018)은 토픽모델링 기법을 활용하여 논문데이터의 핵심 키워드들을 분류하였고 2000년부터 2017년까지 진행된 인공지능의 연구 분야를 분석하였다[9]. 황서이, 김문기(2019)는 토픽모델링 기법을 사용하여 1985년부터 2018년까지 총 1,691편의 학술논문에서 서명, 주제어, 초록을 대상으로 인공지능 분야의 연구동향을 파악하였다[10]. Table 1은 선행연구와 본 연구와의 차이점을 정리한 것이다.

현재까지의 연구에서는 기술적 매체인 특허와 학문적 매체인 논문을 대상으로 토픽모델링 기법을 활용하여 인공지능에 대한 기술동향 및 연구동향을 분석하였다. 본 연구에서는 사회적 매체인 뉴스 기사를 대상으로 이슈 분석을 수행하여 인공지능 기술이 사회, 경제, 산업, 문화 전반에 어떠한 영향을 끼치고 있는지, 사회 구성원들이

Table 1. Comparison of previous studies with this study

Prior research	Research data	Analysis algorithm	Result of analysis	Remark
J. S. Park, S. G. Hong, J. W. Kim (2017)	U.S. patents collected using AI as a keyword	Topic modeling method using LDA algorithm	Identify detailed technology and core technology related AI, Classify Hot/Cold Technology	Technology Trend Analysis (Optimization of parameters of the LDA algorithm is not considered)
M. S. Chung, S. H. Jeong, J. Y. Lee (2018)	Domestic and foreign patents collected using AI as a keyword	Keyword network analysis	Derive keywords, Analyzing blank technology based on IPC classification	Technology Trend Analysis
M. S. Chung, J. Y. Lee (2018)	Papers collected using AI as a keyword	Topic modeling method using LDA algorithm	Analysis of AI-related technology trends and major research topics	Analysis of technology trends and research trends (Optimization of parameters of the LDA algorithm is not considered)
S. Hwang, M. Kim (2019)	Academic papers in the field of AI in Korea	Topic modeling method using LDA algorithm, Semantic network analysis	Extracting AI-related research topics, Analysis of relationship between research topics	Research Trend Analysis (Optimization of parameters of the LDA algorithm is not considered)
This study	AI-related articles collected from major media in Korea	Topic modeling method using LDA algorithm	Identification of issues in all areas of society related to AI	Social issue analysis (Optimization of parameters of the LDA algorithm)

인공지능 기술에 대하여 어떤 분야의 관심을 갖고 어떤 의견을 가지고 있는지에 대한 분석을 통해 인공지능 기술의 응용 분야와 사회, 경제, 산업, 문화 전반에서 인공지능 기술이 일으키고 있는 혁신과 전 산업 분야와 융합을 통해 창출할 수 있는 새로운 가치를 통찰하는 데 유용한 정보를 생산하고자 하였다. 본 연구에서는 ‘인공지능’을 검색어로 하여 추출된 11개의 중앙지와 8개의 경제지, 주요 방송사의 2016년부터 2019년까지 3,889건의 뉴스를 대상으로 오픈 소프트웨어인 R을 활용한 토픽모델링 기법을 사용하여 토픽 별 키워드를 추출하였다. 각 토픽의 키워드 간 연관성을 나타내는 PMI(Pointwise Mutual Information) 측도를 높이도록 relevance 파라미터 λ 를 최적화하여 토픽 별 키워드를 추출하였으며, 키워드들로부터 타당한 근거를 바탕으로 토픽 명

을 추론하였다. 본 연구에서는 54개 매체의 뉴스 기사를 분석 대상으로 하여 토픽모델링 기법을 통해 높은 빈도로 추출된 주제들은 사회 구성원들이 공통으로 갖고 있는 관심과 의견 수렴 및 합의에 의해 도출된 주제들로 볼 수 있으므로 새로운 시장이 창출될 분야를 도출하고, 인공지능 기술이 산업에 적용되어 일으키고 있는 혁신을 살펴봄으로써 인공지능 기술을 지식 경영에 적용할 수 있는 분야를 모색하는데 유용한 정보를 생산하고자 하였다.

2.2 연구자료

본 연구에서는 2016년 1월 1일부터 2019년 12월 31일까지를 분석기간으로 설정하였다. 분석 대상은 한국언론진흥재단의 기사데이터베이스인 빅카인즈(www.kinds.or.kr)에서 ‘인공지능’을 검색어로 하여 조선일보, 중앙일보, 동아일보를 비롯한 11개의 중앙지와 매일경제, 서울경제를 비롯한 8개의 경제지, 각종 지역종합지와 주요 방송사에서 추출된 3,889건의 기사이다.

3. 이론 및 연구방법론

본 연구에서는 문서들의 주제를 도출하기 위해 토픽모델링 방법을 사용하였다. 토픽 모델링은 확문적 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)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[11]. LDA 알고리즘의 개요는 다음과 같다.

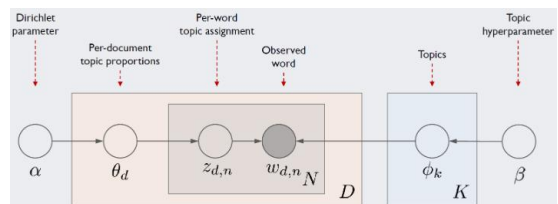


Fig. 1. Architecture of LDA algorithm

(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 라이브리리인 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 [12].

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

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 [13].

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

$p(w_i)$ 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 타당한 토픽을 추출하기가

용이해지며 토픽 간 구분이 명확해진다. Fig. 2는 2.2의 연구자료에 대하여 LDA 알고리즘을 사용한 토픽 모델링을 통해 도출한 한 토픽에 대해 출현 빈도가 높은 상위 30개의 키워드들에 대하여 λ 에 따른 PMI의 변화를 관찰한 결과이다.

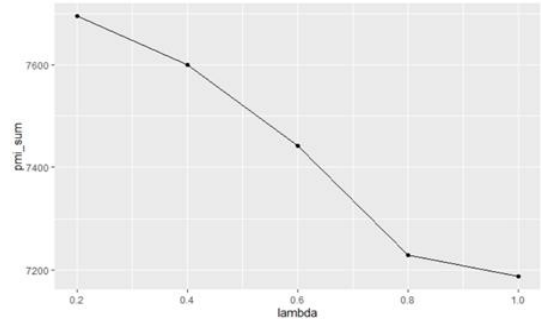


Fig. 2. Trend of PMI between keywords for a topic by λ

본 연구에서는 LDA 알고리즘을 활용한 토픽 모델링 기법을 활용하여 2.2의 연구자료에 대해 Fig. 3과 같은 연구 절차를 통해 국내의 인공지능 관련 주요 이슈를 세부적으로 분석하였으며, λ 를 0.2로 설정하여 키워드 간의 연관성이 커지게 함으로써 키워드들로부터 명확하게 구분되는 타당한 토픽 명을 추출하였다.

빅카인즈(www.kinds.or.kr)에서 수집된 3,889건의 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 척도로서 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 백

터틀 $\mathbf{w}_d = (w_1, w_2, \dots, w_{N_d})$ 라 하면,

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

(3)

이다[11]. K 의 값이 5, 7, 10, 15, 20, 25, 30인 경우에 대해 실험하여 perplexity를 측정하였으며, Fig. 4의 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. Fig. 5는 효율적인 perplexity를 나타내도록 토픽 수를 15로 설정하여 추출된 토픽들이 중복되는 부분이 작게 전체 말뭉치(corpus)에서 고르게 분포함을 보여주는 Intertopic Distance Map이다.

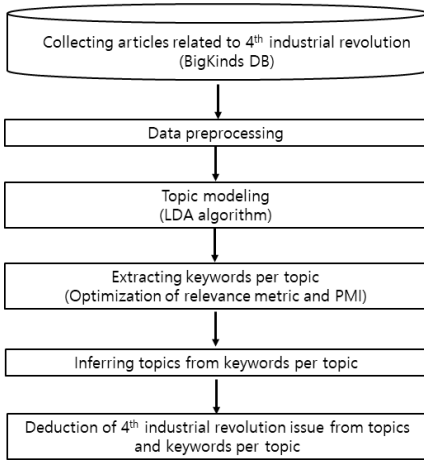


Fig. 3. Research process

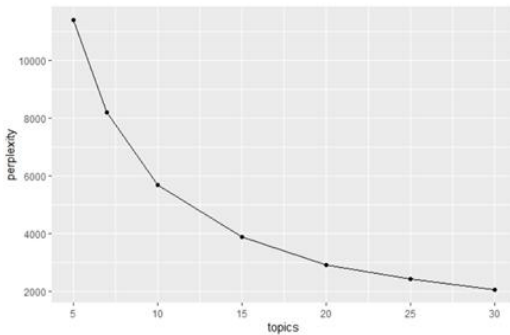


Fig. 4. Perplexity of number of topics

4. 연구결과

4.1 연구 문제

본 연구에서는 사회적 매체인 뉴스 기사를 대상으로

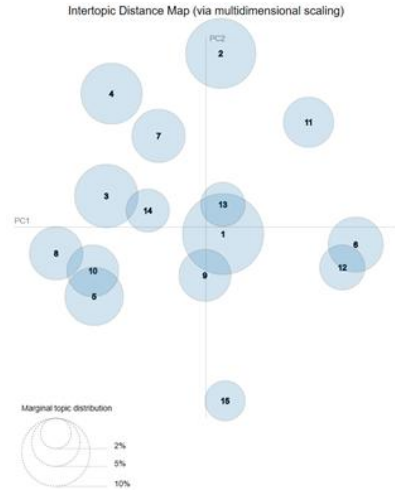


Fig. 5. Intertopic Distance Map ($K=15$)

토픽모델링 기법으로 이슈 분석을 수행하여 인공지능 기술이 사회, 경제, 산업, 문화 전반에 어떠한 영향을 끼치고 있는지, 사회 구성원들이 인공지능 기술에 대하여 어떤 분야에 관심을 갖고 어떤 의견을 가지고 있는지를 살펴보고자 하므로 연구문제 1을 설정하였다. 본 연구에서는 LDA 알고리즘에 기반한 토픽모델링 기법을 사용하여 토픽들이 명확하게 구분되고 타당한 토픽이 추출되도록 LDA 알고리즘의 파라미터들을 최적화하기 위해 연구문제 2를 설정하였다. 또한 토픽모델링 기법을 통해 높은 빈도로 추출된 주제들은 사회 구성원들이 공통으로 갖고 있는 관심과 의견 수렴 및 합의에 의해 도출된 주제들로 볼 수 있으므로, 이로부터 인공지능 기술로 인해 산업에서 나타나고 있는 혁신을 살펴봄으로써 지식 경영에 적용할 수 있는 인공지능 기술을 모색하기 위해 연구문제 3을 설정하였다.

- 연구문제 1: 인공지능을 키워드로 검색한 뉴스 기사들에 나타난 주제별 키워드들은 무엇이며, 키워드들로부터 도출된 주제는 무엇인가?
- 연구문제 2: 뉴스 기사를 대상으로 각 토픽별 키워드를 추출하고, 키워드들로부터 토픽명을 추론하는 분석 과정에서 사용한 독창적인 연구 기법은 무엇인가?
- 연구문제 3: 연구문제 1의 분석 과정과 결과에서 인공지능 기술을 지식 경영에 적용할 수 있는 분야를 모색하는데 필요한 정보는 어떤 것이 있는가?

4.2 연구 분석

Table 4는 ‘인공지능’을 검색어로 하여 수집된 3,889건의 기사들을 대상으로 토픽모델링을 통해 도출한 주제별 키워드들을 나타내며 Table 3은 주제별 키워드들로부터 유추한 토픽명을 나타낸 것이다. Table 3과 4는 ‘연구문제 1’에 대한 분석 결과이다. 각 토픽의 임의의 두 키워드 쌍에 대한 단어 간 연관성을 나타내는 척도 PMI가 커지도록 $\lambda = 0.2$ 로 설정하여 키워드들을 도출함으로써 각 토픽이 특징적으로 분명하게 추출될 수 있도록 하였다. Table 3의 Frequency는 전체 문서에서 각 토픽을 구성하는 단어들이 등장한 빈도를 나타낸다.

Table 3. topic name inferred by topic modeling analysis

No.	Topics	Freq. (%)
1	The Role of AI and Government Support Policies in the 4th Industrial Revolution	14.1
2	Coexistence of AI and Humanity	10.5
3	Trend of AI Leading Company	8.6
4	Introduction and Development of AI	8
5	Disaster Prediction and Response Using AI	7.3
6	Impact of AI on the Economy	6.5
7	AI in Autonomous Vehicles and Home Appliances	6.1
8	AI in Healthcare and Finance	5.8
9	AI Technology Support and National Vision	4.4
10	Concerns about AI Arms Competition and Hegemony Competition	4.2
	total	75.5

Table 4. Keywords inferred by topic modeling analysis

No.	Keywords
1	4th industry, industrial revolution, artificial intelligence, fusion, creation, upbringing, forum, regulation, field, innovation, ICT, SW, change, new industry, competitiveness, R&D, regulatory reform, Davos, government, governance, science technology, industry, paradigm, our country, policy, ministry, basic science, strategy
2	author, novel, work of art, love, story, world, civilization, movie, publication, literature, people, Homo sapiens, Harari, self, earth, today, desire, planet, church, death, mind, drama, religion, premiere, history, extinction, philosophy
3	Amazon, service, enterprise, Google, company, game, startup, market, platform, Apple, customer, cloud, dollar, representative, store, mobile, CEO, employee, Tencent, delivery, business, on-line, sales, product, contents, Uber, Kakao

4	Artificial Intelligence, robot, human, AI, AlphaGo, machine, computer, person, go, Lee Se-Dol 9th State, intelligence, learning, algorithm, humanity, replacement, prediction, ability, job, deep learning, development, artificial, match, Google, decision, appearance, automation, professor, field, go player
5	5G, fire, disaster, CCTV, intense hear, Smart City, KT, weather, earthquake, forecast, intelligent, mobile communication, warning, emergency dispatch, control, Meteorological agency, facility, safety, communication, fire fighting, LG U+, safety industry, base station, El Nino, evacuation, smart, natural disasters, real time
6	economy, job, nation, government, congress, president, society, policy, enterprise, income, labor, support, enlargement, republic of Korea, growth, politics, major company, country, future, amendment, youth, people, innovation, budget, employ, invest, think, worker
7	vehicle, autonomous driving, operator, autonomous vehicles, driving, CES, car, Smart Home, home appliances, LG electronics, TV, air conditioner, Genesis, IFA, Hyundai Motor, race, navigation, display, Galaxy, resolution, Bixby, mobility, battery, load, SUV, Tesla, electric car, road
8	data, Blockchain, finance, big data, medical treatment, bank, uses, service, information, Fintech, patient, privacy, insurance, regulation, hospital, Bitcoin, deal, loan, currency, Healthcare, robo advisor, individual, task, operation, customer, offer, telemedicine, cure, product, ICO
9	nation, embracing, peace, police, finance, fair, economy, innovation, president, budget, government, Korean peninsula, enlargement, congress, effort, fair economy, country of inclusion, 2 years, reform, vitality, growth, support, respect, unfair, result
10	China, Huawei, Hong Kong, president Xi Jinping, drone, Dubai, air Force, flight, Navy, Killer robot, fighter plane, pilot, Trump, DJI, submarine, Macron, weapon, trade war, Beijing, weapon system, terrorism, Boeing, Tiananmen, battle, customms, communist party, protesters, tank

Topic 11 ~ topic 15는 주제 별로 추출된 키워드들이 명확한 토픽을 나타낸다고 판단하기 어려워 분석 대상에서 제외하였다. Topic 1 ~ topic 10의 토픽명을 주제별 키워드들을 바탕으로 다음과 같이 유추하였으며, 각 토픽들은 인공지능 관련 국내의 주요 이슈들을 나타낸다.

Topic 1의 토픽명을 ‘4차 산업혁명 시대에 인공지능의 역할과 정부의 지원 정책’으로 추론한 근거는 다음과 같다. 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 책자를 통해 4차 산업혁명의 정의, 4차 산업혁명을 이끄는 기술, 4차 산업혁명의 영향력, 4차 산업혁명의 기술적·사회적 이슈에 대한 설문조사 등을 포괄적으로 다루면서 시작되었다[14]. 4차 산업혁명은 디지털 기술이 각 산업과 사회에 침투해 융합하는 디지털 전환(digital transformation)의 심화과정으로 이해될 수 있다. 4차 산업혁명의 추동기술로 언급되는 빅데이터, 인공지능, 사물인터넷, 나노기술, 바이오 기술 등의 많은 기술 중 3차 산업혁명과의 구분자 역할을 하는 기술은

단연 인공지능이라 할 수 있다. 우리나라는 「제4차 산업혁명에 대응한 지능정보사회 중장기 종합대책」에서 인공지능을 “인간의 인지능력(언어음성시각감성 등)과 학습, 추론 등 지능을 구현하는 기술로 인공지능 SW/HW, 기초과학(뇌과학산업수학 등)을 포괄한다”고 정의하였다. 지능정보사회로 나아갈수록, 빅데이터 플랫폼과 인공지능 알고리즘의 의존성이 증대되므로 4차 산업혁명의 경쟁력을 갖추고 신산업을 창출하기 위해서는 사물인터넷, 인공지능, 빅데이터 기술이 융합될 수 있는 환경을 조성하는 것이 중요하다. 2016년 8월 정부는 제2차 과학기술전략회의에서 인공지능을 비롯한 9대 국가전략프로젝트 추진과제를 확정하였다. 그리고 지능정보사회 구현과 4차 산업혁명을 주도할 인공지능 핵심 기술을 개발하고 국가 인공지능 기술 역량을 제고해 글로벌 시장을 선점하기 위한 10개년 계획을 수립하여 2026년까지 인공지능 산업의 글로벌 경쟁력 확보를 목표로 삼고 있다 과학기술정보통신부는 SW컴퓨팅산업원천기술개발사업(2013년~2023년)을 중심으로 인공지능 핵심 SW 기술과 HW 기술을 개발하고 있다. 세부 사업으로 언어지능 프로젝트인 엑소브레인(Exobrain), 시각지능 프로젝트인 딥뷰(DeepView), 음성지능 개발 프로젝트를 진행하고 있다. 또한 인공지능 알고리즘 구동과 성능을 좌우하는 핵심 자원인 데이터 관련 규제 개선의 필요성을 정부에서도 인지하고 있다. 「데이터 산업 활성화 전략」에서 우리나라가 세계적으로도 엄격한 개인정보 규제로 인해 데이터 활용이 크게 위축되어 있고, 2016년 비식별화조치 가이드라인을 수립하였으나 활성화되지 못하고 있음을 지적하고 이를 개선하기 위한 방안을 제시하고 있다.

Topic 2의 토픽 명을 ‘인공지능과 인류의 공존’으로 추론한 근거는 다음과 같다. 호모 사피엔스가 현생인류로 등장하게 된 결정적 이유는 사피엔스 즉, 지혜에 있다. 지혜는 생각하는 힘, 관계를 추상화시키고 논리화시키는 능력이다. 보이지 않는 공간 너머의 것을 상상하고 죽음 이후를 생각하는 관념적 사고방식은 종교와 예술을 탄생시켰고 이는 문명과 제도로 발전하였다. 인공지능이 인간보다 뛰어난 영역이 있지만, 인간이 하는 모든 역할을 대신할 수는 없다는 것이 다수의 견해이다. 인공지능은 속도와 확장성, 압기력, 계산능력 등의 정량적 역량이 뛰어나지만 인간 고유의 리더십, 창의력, 사회적 기술, 공감 등의 역량이 부족하므로 인간과 기계가 서로 협력할 때 가장 큰 성과 향상을 이룬다는 연구 결과가 있다[15]. 신경생물학자이자 인지과학자인 안토니오 다마지오(Antonio Damasio)는 인간이 외부 환경과 상호작용하며 내리는

의사결정 능력은 전 신체를 통해 쌓인 다양한 경험적 감각의 자원들과 감정에 기반하고 있으며 이는 인간 고유의 특징임을 확인하였다[16]. 인공지능과 인간의 공존을 위한 안전장치를 마련한다면, 어렵고 힘든 단순반복적인 업무는 인공지능 로봇이 담당하고 인간은 인공지능이 할 수 없는 인간 고유의 장점을 살리는 업무에 집중함으로써 인류의 삶과 가치를 높일 수 있을 것으로 기대된다. 지금의 시대는 산업사회가 발견하지 못한 인간의 고유성이 무엇인지, 기계와의 차이를 토대로 인간이 기계와 공존하는 것이 어떤 것인지 성찰하는 기회라고 할 수 있다.

Topic 3의 토픽 명을 ‘인공지능 선도기업의 동향’으로 추론한 근거는 다음과 같다. 2016년 알파고가 전 세계의 주목을 받은 이후 2017년 구글의 CEO인 선다 피차이(Sundar Pichai)는 회사 전략이 ‘Mobile first’에서 ‘AI first’로 전환되었다고 발표하였으며 마이크로소프트 역시 2017년 연차 보고서를 통해 ‘Mobile-Cloud first’에서 ‘AI first’로 회사 전략의 중심축이 이동했음을 보여주었다[17]. 인공지능 기술과 시장 환경의 급격한 변화로 기업들은 자체 연구개발만으로 기술과 서비스 혁신의 속도를 따라잡을 수 없게 되어, 기술과 인재 획득을 위한 M&A, 스타트업에 대한 투자와 오픈소스를 통한 개방형 R&D를 적극적으로 추진해 왔다. 글로벌 ICT 기업인 구글, 아마존, 페이스북, 마이크로소프트, IBM, 엔비디아, 바이두, 알리바바, 텐센트, 삼성전자, 네이버는 2012년부터 2017년까지 약 50개의 인공지능 기업 M&A를 진행했으며 피인수 기업의 기술과 인재를 자사 연구개발 인프라에 흡수한 후 추가적인 연구개발을 통해 인공지능 제품 및 서비스 상용화를 앞당겼다[17]. 투자한 기업들은 이미지 분석, 음성 인식, 자연어 처리, 데이터 분석 및 추론 기술을 구현하거나 관련 서비스들을 상용화한 기업들이다. 글로벌 ICT 기업들은 외부의 연구 역량을 자사 연구개발과 제품 상용화에 활용하기 위하여 인공지능 API(Application Programming Interface)와 알고리즘 소스코드 등 기업의 핵심기술과 개발 도구를 전 세계 개발자들에게 개방하고 있다[17].

우버는 ‘우버풀’ 서비스를 통해 빅데이터와 인공지능 기술을 활용하여 사회에 흩어져 존재하는 잉여 자원과 자산을 체계화하고 소비자의 수요를 파악하여 연결해주는 공유비즈니스를 구축하였으며, 카카오는 통합 인공지능 플랫폼인 ‘Kakao i’를 통해 대화, 번역, 금융, 교통 등 다양한 분야에서 서비스를 제공하고 있다.

Topic 4의 토픽 명을 ‘인공지능의 등장과 발전 과정’으로 추론한 근거는 다음과 같다. 1950년 영국의 수학자

앨런 튜링(Alan Turing)이 논문을 통해 기계가 생각할 수 있는지 테스트하는 방법, 지능적 기계의 개발 가능성, 학습하는 기계에 대해 기술한 이후 존 폰 노이만(John von Neumann) 교수에게 영향을 주어 현대 컴퓨터 구조의 표준을 구축한 시점을 인공지능 역사의 시작으로 본다. 1958년 코넬대학교의 심리학자 프랭크 로센블라트(Frank Rosenblatt)가 뇌신경을 모사한 인공 신경 뉴런인 퍼셉트론(Perceptron)을 개발한 후 인공지능 연구는 부흥기에 접어들었다가 기술의 한계로 암흑기를 거쳐 2000년대 들어와 새로운 부흥기를 맞이했다. 1990년대 중반부터 인터넷이 폭발적으로 확산되면서 방대한 양의 데이터가 생성되었고, GPU를 비롯한 컴퓨팅 자원의 발달로 인해 제프리 힌튼(Geoffrey Hinton) 교수가 제안한 딥러닝(Deep Learning) 기술이 실용화되기 시작했다[17]. 2016년 구글의 알파고 1.0(16만 기보 지도학습 기반 심층강화학습과 확률적 샘플링 기반 의사결정)이 이세돌 9단을 이기고, 2017년 알파고 2.0(비지도학습 소량데이터 기반 자가학습)이 커제 및 최고 수준의 바둑기사들에게 승리하면서 인공지능 기술이 일반인들에게도 확실히 인식되고 재조명되는 계기가 되었다. 미국의 정보 기술 자문회사인 가트너(Gartner)는 현재 스마트폰 상의 가상비서 어플리케이션과 스피커형 기기로 상용화되고 있는 음성인식 기술의 경우 '생산성 안정단계(Plateau of Productivity)'에 접어들었다고 판단하였고, 딥러닝과 기계학습, 자연어 처리 등 주요기술은 '기대 거품의 정점(Peak of Inflated Expectations)' 단계로서 성공 사례와 실패 사례들이 공존하며 일부 기업들이 사업에 착수하는 단계로 보고 있다[17]. 산학연의 인공지능 전문가를 대상으로 측정한 5대 인공지능 기술인 이미지영상 인식/분석, 신호 인식/분석, 텍스트언어 인식/분석, 데이터 보유/처리, 컴퓨팅의 중요도와 성숙도가 모두 리커트 7점 척도에서 4.0을 넘어 중상위 수준에 도달하고 있어 많은 기업들이 현존하는 인공지능 기술을 바탕으로 다양한 제품 및 서비스를 상용화하는 움직임은 타당하다고 볼 수 있다[17]. 인공지능이 특정 분야에서 활용되는 약인공지능(weak AI)을 넘어 인간의 모든 지적 활동을 대체하는 강인공지능(strong AI)으로 발전할지는 아직 미지수이지만, 딥러닝 분야 최고 권위자인 얀 르쿤(Yann LeCun) 교수는 2018년 삼성 AI 포럼에서 지도학습(Supervised Learning)을 통해 주어진 데이터를 바탕으로 주변 환경과 조건까지 예측해 인지하는 AI를 개발할 수 있으므로, 세상에 대한 배경지식을 쌓고 예측하며, 인간과 소통할 수 있게 하는 것이 미래 AI 연구 방향이 되

어야 할 것이라고 말했다.

Topic 5의 토픽 명을 '인공지능을 활용한 재난-재해 예측 및 대응'으로 추론한 근거는 다음과 같다. 기후변화와 도시화는 인류에게 대규모 홍수와 가뭄, 폭염, 한파 등 자연재해를 전 세계에 일어나게 하고 있다. 뿐만 아니라 화재, 붕괴, 폭발 및 환경오염 사고 등 뜻하지 않은 재해는 에너지, 통신, 교통 등 국가기반체계 마비와 전염병 같은 사회적 재난으로 이어지고 있는 추세이다. 재난-재해를 예측하고 대응하는 분야에 인공지능(AI), 사물인터넷(IoT), 빅데이터 등 최신 ICT 기술이 접목되고 있다. 유럽에서는 홍수 피해를 막기 위해 사물인터넷 기술을 활용한 '홍수 조기 경보시스템'을 운영하고 있으며 일본은 지진과 쓰나미를 예측하기 위하여 바다 위 부유물에 GPS를 장착하여 감시 시스템을 운영하고 있다. 미국의 911(미국 긴급구조 번호)에서는 인공지능 왓슨(Watson)을 활용하여 응급상황과 출동 요청에 대한 우선순위를 자동으로 매기고 대응하며, 응급 의료 서비스에 적용해 최적의 대처 방안을 제공 받는다. 구글은 2019년 인공지능 기반 홍수 예측 시스템을 개발하였다. 홍수 예측 AI 알고리즘은 홍수가 일어날 시간, 장소 및 홍수 규모를 예측하며 정확도가 90% 수준에 이른다. 2018년 과학기술정보통신부는 행정안전부와 협력해 인공지능, 빅데이터, 사물인터넷, 클라우드, 가상현실(VR), 증강현실(AR)의 혁신성장동력 분야별 기술을 재난안전 분야에 활용하기 위한 '혁신성장동력 재난안전 활용 시행계획'을 확정하였다. 정부는 혁신성장동력 6대 분야 기술을 활용한 전(全)주기 스마트 재난안전관리체계를 구축하는데 5년간 6,153억원을 투입할 계획이다. 재난 전조를 감지하고 예측하는 체계를 만들기 위해 인공지능, 사물인터넷, 클라우드, 빅데이터, 모바일 기술이 도입되며, 재난대비 교육 및 훈련체계를 위해 가상현실과 증강현실이 도입된다. AI를 활용한 효율적인 의사결정 지원 체계를 구축하여 재난이 발생했을 때 신속하게 대응할 수 있도록 하고 재난 현장의 인명탐지, 정보탐색 등 재난 복구 시에는 지능형 로봇을 활용할 예정이다.

Topic 6의 토픽 명을 '인공지능이 경제에 미치는 영향'으로 추론한 근거는 다음과 같다. 인공지능 기술의 발전과 채택이 확대되면서 인공지능이 경제사회적으로 미칠 영향에 대한 논의가 활발히 진행되고 있다. 영국의 컨설팅 회사인 프라이스워터하우스쿠퍼스(PwC)는 인공지능 기술의 채택이 확산되고 기술 개발이 가속화됨으로써 2030년까지 글로벌 GDP가 14%(약 15.7조 달러)까지 증가할 것으로 추정하였으며, 미국의 컨설팅 회사인 맥킨

지(McKinsey & Company)는 인공지능이 2030년까지 전세계 GDP를 약 13조 달러 증가시키고 매년 1.2%씩 증가시킬 것으로 추정하여 향후 인공지능 기술이 생산성 향상과 경제성장의 원동력이 될 것으로 전망하였다[1]. 동시에 인공지능은 고용시장 양극화, 불평등 증가, 구조적 실업 및 바람직하지 않은 새로운 산업 구조의 출현과 같은 위험을 초래할 수 있다는 우려도 제기되고 있다[18]. 전 세계의 현재 인공지능 채택 수준은 다양하므로 선진국과 후발국 간의 격차가 확대될 수 있고, 이에 따라 생산성과 혁신의 격차가 발생하여 인공지능 혜택의 불균형 분배 문제가 발생할 수 있다는 것이다. 그러나 유럽 위원회의 보고서 「Economic impacts of artificial intelligence(AI)」에서는 인공지능이 경제사회에 미칠 부정적 영향은 어쩔 수 없는 것이 아니며, 신중한 정부의 정책 결정을 통해 부정적인 영향을 억제하면서 인공지능 개발을 강화하는 것이 바람직하다고 밝히고 있다[18].

Topic 7의 토픽 명을 ‘자율주행차와 가전의 인공지능’으로 추론한 근거는 다음과 같다. 2015년 미국 라스베이거스에서 개최된 세계 최대 가전전시회인 CES(Consumer Electronics Show)에서는 자율주행차가 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. 인공지능 기술은 자율주행차의 카메라 및 레이더와 라이더(LIDAR) 센서에서 수집된 정보와 차량 간 통신에서 수집된 정보를 활용하여 물체와 환경의 인식, 차량 제어, 정밀지도 생성, 사용자 음성 인식, 사용자 모니터링 등의 기능을 수행하며, 현재 개발된 딥러닝(Deep Learning) 인공지능 기술을 통한 차량 인식률은 89%로 나타나고 있다[19]. 2018년 1월 미국 라스베이거스에서 개최된 CES(Consumer Electronics Show)는 인공지능 기술 향연의 장이었다. 가전 기업 및 ICT 기업들은 인공지능 기반의 지능형 개인비서를 냉장고, 세탁기, 청소기 등에 탑재하였고, 완성차 업체와 통신사들은 3단계 수준의 자율주행을 구현한 자동차를 선보였다[17]. 휴머노이드, 강아지, 인형 등 갖가지 형상의 인공지능 기반 로봇들은 산업 현장을 넘어 일상생활에서도 인간과 기계의 공존이 얼마 남지 않았음을 보여주었다[17]. 현재 대표적인 인공지능 제품과 서비스는, ‘구글 렌즈(Google Lens)’, 지능형 개인비서 기기인 ‘구글 홈(Google Home)’, 인공지능 비서 서비스인 ‘구글 어시스턴트(Google Assistant)’, 아마존의 지능형 개인비서인 ‘알렉사(Alexa)’와 알렉사가 탑재된 ‘아마존 에코(Echo)’,

마이크로소프트의 지능형 개인비서 ‘코타나(Cortana)’, 의료·법률·금융 분야에서 인지 비즈니스(cognitive business) 서비스를 제공하고 있는 IBM의 ‘왓슨(Watson)’, 자율주행차 오픈소스 플랫폼인 바이두의 ‘아폴로(Apollo)’, 음성 비서 서비스인 삼성전자의 ‘빅스비’가 있다.

Topic 8의 토픽 명을 ‘의료와 금융의 인공지능’으로 추론한 근거는 다음과 같다. 선진국의 많은 인공지능 전문가들과 한국 AI 스타트업들은 의료계에서 인공지능의 필요성을 절감하고 여러 분야에서 혁신적인 노력을 기울이고 있다. 의료영상을 인공지능을 통해 분석하고 판단해서 빠른 시간 내에 정확한 진단을 하고 있으며, 의료 현장에서 벌어지는 모든 음성 정보를 음성 인식 인공지능을 통해 문자로 전사하여 전자의무기록을 작성하고 환자의 영상, 음성, 전자의무기록 데이터를 기반으로 한 질병 예측 및 예방, 치료 등에 대한 종합적인 솔루션을 의료진과 환자에게 제공하고 있다. 구글의 발표에 의하면 자사 인공지능의 정확도가 유방암의 경우 99%, 폐암의 경우 95%이며 전이암의 경우에도 높은 수준을 기록한 것으로 보고되었다. 의료영상 데이터를 분석하는 인공지능의 개발을 위해서는 정제된 대량의 데이터가 필요하므로 의료기관들 간 데이터 교환이 가능하도록 규제를 완화할 필요가 있다.

금융(finance)과 기술(technology)의 결합을 의미하는 핀테크(fintech)는 모바일 결제와 송금, P2P(peer-to-peer) 대출, 로보어드바이저(robotadvisor) 등 디지털 기술에 기반한 혁신적 금융 서비스를 의미한다[20]. 핀테크 분야에서의 인공지능은 기존 금융업 분야에서는 볼 수 없었던 새로운 형태의 비즈니스 모델을 등장시키고 있다. 인공지능을 활용한 개인 맞춤형 서비스, 투자자문 및 자산관리, 신용평가, 부정거래 탐지, 챗봇(chatbot) 서비스가 등장하였다. 로봇(robot)과 어드바이저(advisor)의 합성어인 로보어드바이저(robotadvisor)는 개인의 투자 성향 및 관련 데이터를 인공지능 알고리즘을 통해 분석한 후 최적화된 포트폴리오를 구성하고, 시장 상황에 따라 이를 재구성해 운용해주는 온라인 자산관리 서비스를 의미한다[21]. 블록체인(blockchain)은 관리 대상인 블록(block)이라고 하는 소규모 데이터들이 P2P 방식을 기반으로 생성된 체인(chain) 형태의 연결고리 기반의 분산 데이터 저장환경에 저장되어 누구도 임의로 수정할 수 없고 누구나 변경의 결과를 열람할 수 있는 분산 컴퓨팅 기술 기반의 데이터 위변조 방지 기술이다[22]. 블록체인은 기술적으로 공개키 알고리즘과 해시(Hash) 암호화 기술, 분산처리구조를 이용해 중앙 집중식 원장 구조를 저비용의 분산 원

장 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.

Topic 9의 토픽 명을 '인공지능 기술 지원과 국가비전'으로 추론한 근거는 다음과 같다. 세계 경제는 지속가능한 성장을 위한 패러다임의 대전환이 이뤄지고 있다. 기업의 자본 투자에 의존한 '산업화 모형'에서 벗어나 협업과 상호네트워크를 기반으로 아이디어와 혁신기술의 가치가 창출되는 '지식 경제'로 이행 중이다. 이에 정부는 소득주도 성장, 혁신성장, 공정경제를 3대 축으로 하는 경제 정책을 추진하여 '혁신적 포용국가' 실현을 국정 비전으로 제시하고 있으며, 초연결 지능화, 스마트공장, 스마트팜, 핀테크, 에너지 신산업, 스마트시티, 드론, 미래 자동차의 8대 선도 프로젝트를 추진하고 있다. 정부는 2019년 12월 17일 문재인 대통령 주재로 열린 제 53회 국무회의에서 과학기술정보통신부를 비롯한 전 부처가 참여하여 마련한 '인공지능(AI) 국가전략'을 발표하였다. 주요 내용은 첫째, 세계 최고의 메모리 반도체 경쟁력을 활용하여 AI 반도체 경쟁력 세계 1위를 목표로 AI 반도체 핵심기술 확보와 신개념 반도체(PIM) 개발에 전략 투자를 강화하고, 둘째, 전 국민이 AI 기초 역량을 습득할 수 있는 교육체계를 구축하고, 셋째, 전자정부를 넘어서는 AI 기반 차세대 지능형 정부로 탈바꿈하여 수준 높은 공공서비스를 제공하고, 넷째, AI의 혜택이 기술과 자본을 가진 계층에 집중되지 않고 모든 국민이 고루 누릴 수 있도록 일자리 안전망 구축, AI 윤리 정립 등을 통해 사람 중심의 AI 시대를 구현하겠다는 것이다. 문재인 대통령은 2020년 1월 16일 한국전자통신연구원(ETRI)에서 열린 과학기술정보통신부, 방송통신위원회 대상 신년 업무계획 보고의 모두 발언을 통해 과학기술과 정보통신의 힘으로 미래 먹거리를 확보하고 혁신적 포용국가 시대를 앞당겨야 하며, 과학기술 강국, 인공지능 일등국가가 그 기준이라고 밝혔다.

Topic 10의 토픽 명을 '인공지능 무기경쟁과 패권경쟁에 대한 우려'로 추론한 근거는 다음과 같다. 네덜란드에 기반을 두고 있는 국제 평화운동 단체인 팩스(Pax)는 보고서 'State of AI'에서 중국, 러시아, 미국 등 군사 강국 간의 인공지능 무기 경쟁에 대한 우려를 제기했다. 보고서에 따르면 반응시간을 단축하고 타격을 보다 정확하게 하기 위한 자율 살상무기 개발은 무력 사용에 관한 의

사 결정을 단축시키고 대량 살상 무기의 위험을 증가시킨다[23]. 킬러 로봇은 사람의 간섭 없이 공격이 가능한 완전 자율 살상 무기를 일컫는다. 미국, 중국, 러시아 등 10여 개국은 킬러 로봇을 개발하고 있다. 러시아는 2016년 11월, 국경에 주변 6km 내의 사람과 물체를 저장할 수 있는 킬러 로봇을 개발해 배치하여 경찰 무인 드론을 타깃으로 하고 있다. 이스라엘도 12kg의 소형 킬러 로봇을 운영하고 있으며 미국은 적을 공격할 수 있는 무인 드론, 무인 전투함 등을 개발한 상태이다. 영국의 타라니스 드론(Taranis drone), 미국 해군의 자율운항 무인 함정 '시 헌터(Sea Hunter)', 보잉의 무인잠수정 '에코 보이저(Echo Voyager)', 러시아의 무인 탱크 'MK-25', 한화의 'SGR-A1 센트리 건(SGR-A1 Sentry gun)' 등이 대표적인 킬러 로봇으로 손꼽힌다. 최근 미국-중국 간의 무역전쟁을 첨단 기술을 중심으로 전개되는 패권 경쟁으로 보는 시각도 있다. SW 기술을 둘러싸고 전개되는 미국-중국의 기술 갈등은 과거 OS에서 인터넷, 최근 인공지능 경쟁으로 범위가 확산되고 있다는 견해이다[24]. 인공지능은 일반목적기술(General Purpose Technology)의 특성이 있어 전 산업영역에서 응용되기 때문에 글로벌 인공지능 경쟁은 기술보호주의에서 기술안보주의의 추세를 보이고 있다는 견해이다.

4.3 연구 결과

연구문제 1에 대한 결론은 다음과 같다. Table 3과 Table 4에서 보듯이 '인공지능'을 검색어로 수집된 3,889건의 기사들에서 추출된 키워드들과 토픽들은 인공지능 기술의 응용 분야와 사회, 경제, 산업, 문화 전반에서 일어나고 있는 변화 및 정부의 지원 정책과 비전을 폭 넓게 나타냈다. 이 중 topic 3: '인공지능 선도기업의 동향' 토픽, topic 5: '인공지능을 활용한 재난-재해 예측 및 대응' 토픽, topic 7: '자율주행차와 가전의 인공지능' 토픽, topic 8: '의료와 금융의 인공지능' 토픽이 전체 토픽 빈도 1,575,134 건 중 436,348 건으로 27.7%를 차지하여, 인공지능 기술과 그 응용분야 및 산업 동향에 대한 관심이 가장 크게 나타났다. 다음으로는 topic 2: '인공지능과 인류의 공존' 토픽, topic 4: '인공지능의 등장과 발전 과정' 토픽, topic 6: '인공지능이 경제에 미치는 영향' 토픽이 전체 토픽 빈도 1,575,134 건 중 393,832 건으로 25.0%를 차지하여, 인공지능의 등장으로 인한 큰 틀에서의 사회, 경제 변화와 인공지능 기술로 인한 인류의 삶의 변화에 대한 관심이 두 번째로 큰 것으로 나타났다. Topic 1: '4차 산업혁명 시대에 인공지능의 역할과

정부의 지원 정책' 토픽과 topic 9: '인공지능 기술 지원과 국가비전' 토픽이 전체 토픽 빈도 1,575,134 건 중 290,299 건으로 18.4%를 차지하여, 정부의 인공지능 기술 지원 정책과 이를 통해 구현하고자 하는 미래 사회상에 대한 관심이 세 번째로 큰 것으로 나타났다. 그 외에도 topic 10: '인공지능 무기경쟁과 패권경쟁에 대한 우려' 토픽이 전체 토픽 빈도 1,575,134 건 중 66,116 건으로 4.2%를 차지하여, 인공지능 기술이 세계 질서에 끼치는 부정적인 영향에 대한 우려도 상당히 존재하는 것으로 나타났다. 도출된 키워드들과 토픽들로부터 인공지능 기술은 산업과 과학기술분야에서 만의 변화가 아니라 사회, 경제, 문화, 세계질서 전반에 변화를 일으키고 있음을 알 수 있다.

Topic 3, topic5, topic 7, topic 8에 나타난 것처럼 인공지능 기술은 다양한 산업에 적용되어 기술을 혁신시킴으로써 제조 산업에서 새로운 제품을 창출하고 생산성을 향상시키고 있으며, 의료, 금융, 서비스 산업에서 소비자에게 질 높은 경험을 제공하고 있을 뿐 아니라 재난·재해 예방과 이에 대한 신속한 대응 및 복구에도 활용되고 있다. Topic 2, topic 4, topic 6에 나타난 것처럼 인공지능 기술의 발전으로 인한 긍정적인 측면을 살펴보면 인간의 불필요한 노동력을 감소시키고 그 노동력을 창의성이 요구되는 생산적인 일에 사용하여 인류의 삶과 가치를 높일 수 있을 것으로 기대되며, 생산성 향상으로 인해 글로벌 GDP가 증가할 것으로 전망된다. 한편 인공지능이 정착되는 과정에서 발생하는 부정적 영향을 법적 제도를 마련하고 AI 윤리 정립을 통해 예방하고 억제해야 하는 과제가 남아 있다.

연구문제 2에 대한 결론은 다음과 같다. 본 연구에서는 LDA 알고리즘에 등장하는 파라미터 α , β , K , λ 의 값을 최적화하여 토픽 간 구분이 명확해지고 타당한 토픽이 추출될 수 있도록 하였다. LDA 알고리즘을 적용할 때 초기값으로 설정하는 Dirichlet 파라미터 α 와 β 의 값이 0.02, 0.04, 0.06, 0.08, 0.1인 경우에 대해 토픽모델링을 수행하여 도출되는 키워드들에 큰 변화가 없음을 확인하였고 토픽 별 키워드 구분이 명확해지도록 α 와 β 의 값을 설정하였다. LDA 알고리즘의 입력 값인 토픽 수 K 의 값을 5, 7, 10, 15, 20, 25, 30으로 변화시키면서 perplexity를 측정하여 전체 말뭉치(corpus)의 단어가 토픽에 등장할 확률이 커지도록 하면서 R 프로그램의 실행속도가 저하되지 않도록 최적의 K 값을 설정하여 토픽모델링 분석을 효율적으로 수행하였다. 각 토픽의 키워드

들의 순위를 조정하는 relevance metric의 파라미터 λ 를 작게 설정하면 각 토픽의 임의의 두 키워드 쌍에 대한 단어 간 연관성을 나타내는 척도인 PMI가 커진다는 사실을 발견하여 토픽 간 구분이 명확해지고 타당한 토픽이 추출될 수 있도록 λ 의 값을 설정하였다.

연구문제 3에 대한 결론은 다음과 같다. 경영과학은 최적화를 목표로 하는 학문이라고 할 수 있다. 기존의 지식을 활용한 창의경영에 더하여 4차 산업혁명 시대에는 빅데이터를 활용한 과학 경영과 인공지능 기술을 지식 경영에 적용함으로써 의사결정을 유연화하고 경영 최적화를 이룰 수 있는 방법이 다양해지고 있다. 기업의 정보 처리는 데이터의 처리에서 정보의 처리로 그리고 지식의 처리 단계로 발전했고 이제 지혜 또는 지능에 대한 처리 단계의 초입에 와 있다[25]. 기업이 인공지능 도입을 고려할 때는 현재 갖고 있는 문제 중에서 해결해야 할 가장 중요한 문제를 선택한 후 이를 해결할 여러 인공지능 기법들과 비인공지능 기법들을 대안에 올려놓고 실험과 평가를 통해 그 문제를 해결할 수 있는 기술을 선택하고 구현해야 할 것이다. 가장 주안점을 둘 부분은 topic 3, topic 5, topic 7, topic 8의 토픽 추론 근거에 나타난 것처럼 새로운 시장을 창출하기 위해 인공지능을 활용하는 것이다. 기업들이 인공지능을 응용해 온 분야를 세계 인공지능학회의 '혁신적 인공지능 응용상'을 수상한 사례를 통해 살펴보면, 인공지능을 조직 실무에 응용해서 성과를 낸 업무 중 가장 많은 경우가 자원 할당, 기획을 포함한 일정계획(스케줄링) 업무이고, 두 번째로는 사기 적발(fraud detection), 감시(monitoring), 감사(audit), 사전 경고(advanced warning) 등의 업무였다. 기업/조직 내 학습, 협력, 교육 등의 업무가 뒤를 이었으며 그 밖에 고장 진단과 검출, 데이터 필터링과 선택/순서 매기기, 제품 추천, 데이터 인식 후 변환/번역, 예측과 추정, 상황적 대응, 자료 추출, 큐레이션, 데이터 정제와 분석, 성능 최적화, 자연어 응대 등이 있었다[26]. 이 사례들은 인공지능 기술이 발전할수록 새로운 제품과 서비스를 위해 인공지능이 사용되는 경우가 많아질 것을 보여준다. 즉 새로운 시장의 창출, 블루오션의 창출을 위해 인공지능을 활용하는 것이다. 본 연구의 분석 과정과 결과는 새로운 시장이 창출될 분야와 인공지능 기술을 지식 경영에 적용할 수 있는 분야를 모색하는데 필요한 정보를 제공할 수 있을 것으로 기대된다.

5. 결론

본 연구는 ‘인공지능’을 키워드로 검색하여 국내 언론 사들에서 수집된 기사들로부터 토픽모델링 기법을 통해 핵심 주제들을 찾아내어 인공지능 기술의 영향으로 사회에서 일어나고 있는 변화와 국민들의 관심 분야를 도출했다는 점에서 의의를 지닌다. 뉴스 미디어는 사회 구성원들의 높은 수준의 동의가 나타난 뉴스로서의 가치가 있는 의제들을 보도하므로, 뉴스 기사를 대상으로 토픽모델링 기법을 통해 높은 빈도로 추출된 주제들은 사회 구성원들이 공통으로 갖고 있는 관심과 의견 수렴 및 합의에 의해 도출된 주제들로 볼 수 있다.

연구 방법으로는 LDA 알고리즘에 기반한 토픽모델링 기법을 사용하였으며 Dirichlet 파라미터인 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 토픽모델링 분석이 효율적으로 진행될 수 있는 파라미터 α, β, K 의 값을 결정하였다. 본 연구에서는 LDA 알고리즘의 학습 결과를 시각적으로 표현하는 라이브리리인 LDAvis의 가중치(weight parameter) λ 가 작을수록 각 토픽의 임의의 두 키워드 쌍에 대한 단어 간 연관성을 나타내는 척도 PMI가 커져 각 토픽이 특징적으로 분명하게 추출된다는 사실을 발견하였으며, 이로부터 파라미터 λ 를 최적화하여 키워드들과 토픽을 추출하였다.

본 연구에서는 토픽모델링 기법으로부터 도출된 주제별 키워드들로부터 타당한 근거를 바탕으로 토픽 명을 유추하여, 인공지능 기술의 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 인공지능의 역할과 정부의 지원 정책’, ‘인공지능과 인류의 공존’, ‘인공지능 선도기업의 동향’, ‘인공지능의 등장과 발전 과정’, ‘인공지능을 활용한 재난·재해 예측 및 대응’, ‘인공지능이 경제에 미치는 영향’, ‘자율주행차와 가전의 인공지능’, ‘의료와 금융의 인공지능’, ‘인공지능 기술 지원과 국가비전’, ‘인공지능 무기경쟁과 패권 경쟁에 대한 우려’의 10가지로서 인공지능 기술의 응용 분야와 사회, 경제, 산업, 문화 전반에서 일어나고 있는 변화 및 정부의 지원 정책과 비전을 폭 넓게 나타냈다. 인공지능 기술과 그 응용분야 및 산업 동향에 대한 관심이 가장 크게 나타났으며 인공지능의 등장으로 인한 큰

들에서의 사회, 경제 변화와 인공지능 기술로 인한 인류의 삶의 변화에 대한 관심이 두 번째로 큰 것으로 나타났다. 정부의 인공지능 기술 지원 정책과 이를 통해 구현하고자 하는 미래 사회상에 대한 관심이 세 번째로 큰 것으로 나타났다.

4차 산업혁명 시대에는 빅데이터를 활용한 과학 경영과 인공지능 기술을 지식 경영에 적용하여 의사결정을 유연화하고 경영 최적화를 이룰 수 있는 방법이 다양해지고 있다. 본 연구의 토픽 추론 근거에서 설명한 인공지능 기술로 인해 산업에서 나타나고 있는 혁신과 그 과정 및 기업들이 인공지능을 응용해 온 분야를 살펴봄으로써 인공지능 기술을 지식 경영에 적용할 수 있는 분야를 모색하는데 필요한 정보를 생산하였다.

향후에는 산업에서 인공지능 기술이 일으키고 있는 혁신을 나타내는 topic 3: ‘인공지능 선도기업의 동향’ 토픽, topic 5: ‘인공지능을 활용한 재난·재해 예측 및 대응’ 토픽, topic 7: ‘자율주행차와 가전의 인공지능’ 토픽, topic 8: ‘의료와 금융의 인공지능’ 토픽들에 대해 ‘경영’을 검색어에 추가하여 세부 주제 토픽모델링을 수행함으로써, 기업이 인공지능 기술을 활용하여 창출할 수 있는 새로운 시장과 지식 경영에 인공지능 기술을 적용할 수 있는 분야를 구체적으로 도출하고, 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.

REFERENCES

- [1] KSBI. (2019). *Global SME Policy Trend*, 2(9).
- [2] J. Park & M. Song. (2013). A study on the research trends in library & information science in Korea using topic modeling. *Journal of the Korean society for information management*, 30(1), 7-32. DOI : 10.3743/KOSIM.2013.30.1.007
- [3] S. T. Flora. (2017). Blogger-link-topic model for blog mining. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 28-39.
- [4] J. H. Lau, N. Collier & T. Baldwin. (2012). On-line trend analysis with topic models: #twitter trends detection topic model online. *Proceedings of COLING 2012*, 1519-1534.
- [5] D. H. Jeong & M. Song. (2014). Time gap analysis by the topic model-based temporal technique. *Journal of Informetrics*, 8(3), 776-790.

- DOI : 10.1016/j.joi.2014.07.005
- [6] J. S. Park, S. G. Hong & J. W. Kim. (2017). A Study on Science Technology Trend and Prediction Using Topic Modeling, *Journal of the Korea Industrial Information Systems Research*, 22(4), 19–28.
DOI : 10.9723/jksis.2017.22.4.019
- [7] M. S. Chung, S. H. Park, B. H. Chae & J. Y. Lee. (2017). Analysis of major research trends in artificial intelligence through analysis of thesis data, *Journal of Digital Convergence*, 15(5), 225–233.
DOI : 10.14400/JDC.2017.15.5.225
- [8] M. S. Chung, S. H. Jeong & J. Y. Lee. (2018). Analysis of major research trends in artificial intelligence based on domestic/international patent data, *Journal of Digital Convergence*, 16(6), 187–195.
DOI : 10.14400/JDC.2018.16.6.187
- [9] M. S. Chung & J. Y. Lee. (2018). Systemic Analysis of Research Activities and Trends Related to Artificial Intelligence(A.I.) Technology Based on Latent Dirichlet Allocation(LDA) Model, *Journal of the Korea Industrial Information Systems Research*, 23(3), 87–95.
DOI : 10.9723/JKSIS.2018.23.3.087
- [10] S. Hwang & M. Kim. (2019). 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*, 20(9), 1847–1855.
- [11] D. M. Blei, A. Y. Ng & M. I. Jordan. (2003). Latent Dirichlet allocation, *Journal of Machine Learning Research*, 3, 993–1022.
DOI : 10.14400/JDC.2016.14.12.1
- [12] C. Sievert & E. Shirley. (2014). LDAvis: A method for visualizing and interpreting topics, *Proceedings of the Workshop on Interactive Learning, Visualization, and Interfaces*, 63–70.
- [13] D. Newman, S. Karimi & L. Cavedon. (2009). External evaluation of topic models, *Proceedings of th 14th Australasian Document Computing Symposium*, 11–18.
- [14] S. Song. (2017). Historical development of industrial revolutions and the place of so called the fourth industrial revolution, *Journal of Science and Technology Studies*, 17(2), 5–40.
- [15] P. R. Daugherty & H. J. Wilson (2018). Collaboration Intelligence, Human and AI Join Forces. *Harvard Business Review Korea Magazine*, July, 2018.
- [16] A. Damasio. (1995). *Descartes's Error: Emotion, Reason, and the Human Brain*. Penguin Books.
- [17] H. Yang et al. (2018). A Prospective Analysis of Artificial Intelligence(AI) Technology and Innovation Policies: Focused on Improving Korea's National AI R&D Policy. *Policy Research 2018–13*.
- [18] EPRS. (2019). Economic impacts of artificial intelligence(AI). *European Parliamentary Research Service PE637.967*.
- [19] COMPA. (2018). Autonomous Vehicles. *S&T Market Report*, 65.
- [20] J. H. Lee. (2017). Impact of fintech spread on financial sector and audit implications, AIRI Research Report 2017–014.
- [21] G. Y. Lee. (2016). Domestic and Overseas RoboAdvisor Trend and Status Analysis. *e-Finance and Financial Security 2016–04*.
- [22] ETRI Standard Research Division. (2017). Blockchain, *ETRI Standardization Trend 2017–2*.
- [23] PAX. (2019). *State of AI: Artificial intelligence, the military and increasingly autonomous weapons*.
- [24] J. Y. Kim. (2019). Historical Patterns in the US–China AI Hegemonic Rivalry and the Latest Trend, *SPRI SW Industry Trend*.
- [25] G. Lim. (2016). *The digital management innovation of company in AI era, the direction to go forward*, Seoul Economy Daily, <https://www.sedaily.com/NewsView>
- [26] G. Lee. (2017). *Artificial intelligence, a way to optimize management*, Dong-A Business Review, https://dbr.donga.com/article/view/1101/article_no/8179

노 설 현(Seol-Hyun Noh)

[장학원]



- 1996년 8월 : 고려대학교 수학과(이학사)
- 1998년 8월 : 고려대학교 수학과(이학석사)
- 2003년 2월 : 고려대학교 수학과(이학박사)
- 2019년 3월 ~ 현재 : 안양대학교 통계

데이터과학전공 조교수

· 관심분야 : 빅데이터, 인공지능 알고리즘 및 SW, 확률론적 모델링

· E-Mail : shnoh@anyang.ac.kr