

언론은 인공지능(AI)을 어떻게 다루는가?: 뉴스 빅데이터를 통한 한국과 미국의 보도 경향 분석*

박종화** · 김민성*** · 김정환****

<목 차>

I. 서론	IV. 연구결과
II. 이론적 배경	4.1 주제어 분석
2.1 인공지능 개요	4.2 토픽모델링
2.2 언론 보도 분석	4.3 언어 네트워크 분석
III. 연구 방법	V. 논의 및 결론
3.1 데이터 수집	참고문헌
3.2 데이터 전처리	<Abstract>
3.3 데이터 분석	

I. 서론

최근 기술의 진화는 미디어 소비 전반에 영향을 미치고 있다. 인공지능(AI)은 미디어 제작과 생산, 유통, 소비 등 미디어 가치사슬 전 과정에 다양한 형태로 스며들었다. 뉴스, 방송, 영화, 음악, 쇼핑 등 주요 미디어 분야에서 알고리즘이 적용된 서비스가 확대되고 이용자들에게 맞춤형 선택지를 제공함에 따라 이용자 편의가 확대되고 만족도도 증가하고 있다. 하지만 이와

함께 블랙박스와 같은 알고리즘의 처리 및 운영 방식에 대한 의문이 제기되면서 편향성에 대한 우려가 함께 야기되고 있다. 여기서 알고리즘은 “컴퓨터 또는 디지털 대상이 특정 업무를 수행하는 방법에 대한 설명으로, 명확하게 정의된 한정된 개수의 규제나 명령 집합”(Goffey, 2008)으로, 데이터를 처리하는 규칙을 포괄적으로 의미한다.

알파고, 이루다 사태 등 사회적 영향이 큰 굵직한 이슈의 경우 언론사는 엄청난 양의 기사

* 이 논문은 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임.
(NRF-2019S1A3A2099973)

** 공주대학교 상업정보교육과, jhpark@kongju.ac.kr(주저자)

*** 고려대학교 미디어산업연구센터, okminsung@naver.com

**** 부경대학교 미디어커뮤니케이션학부, mediakim@pknu.ac.kr(교신저자)

를 쏟아내며, 이들이 만들어낸 담론은 곧 잠재적 이용자들의 인식에도 강한 인상을 남긴다. 즉 미디어에서 어떤 이슈를 선택하여 다양한 담론이나 이미지를 구성하며 다루는 과정을 통해 이용자들의 이슈에 대한 해석이나 인식, 태도 및 판단 등이 달라질 수 있는 것이다(양혜승, 2008; 이준웅, 2001; Entman, 1993; Lippmann, 1922; Shah et al, 1996). 인공지능의 경우 4차 산업혁명 시대의 코어 기술이자 미래 산업 발전을 위한 기반 기술로, 전 산업에 걸친 혁신과 지능형 융합 견인을 위해 필수적으로 갖추어야 할 유망한 기술로 주목받고 있다. 하지만 우리나라의 경우 사실상 인공지능 후발주자로 주요 경쟁국 대비 상용서비스가 활성화되어 있지는 않은 단계이다(과학기술정보통신부, 2018; 한국과학기술기획평가원, 2021). 즉 이용자가 일상 생활에서 인공지능을 경험할 수 있는 서비스는 AI스피커나 챗봇 정도에 그치고 있기에 이용자들은 주로 미디어를 통해 인공지능 관련 소식을 접하고 이에 대한 인식을 형성하게 된다. 하지만 사회 전반의 인식은 기술에 대한 긍정적 가치와 기대보다 우려와 걱정이 앞서고 있는 것이 현실이다. 실제로 최근에 이루어진 4차산업혁명위원회(2021)의 대국민 조사에 따르면 인공지능 제품 및 서비스를 잘 활용하고 있다는 응답은 22.9%에 불과했으며, 인공지능 기술을 신뢰한다는 응답은 40.5%에 머물렀다. 또한 프라이버시(51.8%)나 AI 기술 안전성에 대한 우려(45.3%), AI 자체가 불러올 수 있는 일자리 문제와 같은 영역에 대한 우려(49.0%)도 해결되지 않은 것으로 나타났다. 본 연구는 이러한 경향이 언론의 보도에서도 그 원인을 찾을 수 있다는 점에서 출발한다.

인공지능에 대한 두려움은 규제 논의로 확산되고 있는 추세이다. 보다 구체적으로는 알고리즘에 대한 규제 논의를 대표적인 사례로 들 수 있는데, 유럽연합(EU)의 ‘개인정보보호규정(GDPR: General Data Protection Regulation)’에서 본격적으로 다뤄지기 시작하였다. GDPR에 알고리즘 투명성(transparency)과 관련한 내용이 포함됐는데, 프로파일링 등에 대해 설명을 요구할 권리를 적극 포함시킨 바 있다. 알고리즘이 데이터를 학습하면서 편향성을 야기할 가능성이 존재함에 따라 오바마 정부는 ‘빅데이터: 알고리즘 시스템, 기회와 시민권(Big data: A report on algorithm systems, opportunity, and civil rights)’이라는 보고서를 통해 알고리즘에 따른 편향성의 위험을 지적하기도 하였다. 최근 공정거래위원회는 공정거래법 개정을 통해 알고리즘 담합, 데이터 독점 등 새로운 유형의 불공정거래 행위를 제재할 수 있는 수단을 만들겠다고 밝히는 등 국내에서도 알고리즘 규제와 관련된 논의가 확대되기 시작하였다.

하지만 공학과 사회과학에서 인공지능을 바라보는 관점이 다르고, 사업자와 정부의 입장 역시 일부 상반된 양상을 보이는 등 인공지능 기술을 중심으로 한 생태계 내 주요 이해관계자들의 인식과 입장에는 차이가 존재한다. 인공지능을 실제로 연구하는 기술 현장에서는 인공지능을 정의함에 있어 기술적 기능을 강조하지만 정책 입안자의 경우 인공지능을 인간의 사고 및 행동과 비교하는 측면에서의 정의를 주로 사용하며, 이러한 격차로 윤리나 규제 측면에서의 노력이 현재의 기술 상황에 비해 우려를 지나치게 강조할 수 있다는 연구 결과도 확인할 수 있다(Krafft et al., 2020). 이에 본 연구

는 이해관계자들의 입장 차이에 대한 연구의 시작으로, 인공지능 기술의 소비 주체인 소비자가 가장 쉽게 접할 수 있는 정보원인 언론 보도에서 나타나고 있는 경향과 핵심 의제들을 살펴보고자 한다. 특히 인공지능 분야 선발주자이 관련 논의가 가장 많이 진행되어 있다고 평가되는 미국과 한국의 보도행태를 비교함으로써 바람직한 논의의 방향과 목적에 대한 제언을 도출해 보고자 한다.

II. 이론적 배경

2.1 인공지능 개요

인공지능은 인간의 학습능력과 추론능력, 지각능력, 자연어 처리능력 등을 컴퓨터 프로그램으로 구현한 기술을 의미한다. 이를 구성하는 요소로는 머신러닝, 알고리즘, 컴퓨팅 파워, 빅데이터 등을 들 수 있다(정준화, 2019). 알파고와 이세돌의 대국 이후 인공지능에 대한 사회적, 정책적 관심이 급증하였으며 향후 인간의 미래를 바꿀 수 있는 가장 대표적인 기술로 떠올랐다. 미국의 주요 IT 기업들을 중심으로 관련 기술이 고도화되고 있으며, 동시에 인공지능이 접목된 다양한 서비스들이 출시되고 있다. 즉, 인공지능 기술은 의료, 교육, 커머스, 유통, 자율주행 자동차 등 다양한 분야에 접목되어 각 영역에서 주요 활용 기술로 자리 잡아 가고 있다.

인공지능의 대표적인 특성으로는 학습성, 다양성, 범용성 등을 들 수 있다(김광호 등, 2020). 이를 바탕으로 신경망 학습을 통해 원하는 결

과를 보다 효율적으로 달성할 수 있으며, 목적에 따라 다양한 형태와 모델을 구현할 수 있다. 특히, 현재 인공지능 기술은 기계학습을 통해서 스스로 규칙을 만들어내는 머신러닝과 딥러닝으로 대표되는데 이는 인공지능 분야의 성장을 선도하고 있다.

이러한 기술을 기반으로 인공지능 관련 시장 규모는 기하급수적으로 증가할 전망이다. 미국 시장분석 기관 트라티카(Tractica, 2019)에 따르면, 전 세계 인공지능 시장규모가 2022년에는 약 51억 달러 규모에 달할 것이며, 2025년에는 약 126억 달러로 급격한 성장을 이룰 것이라 예측된다. 그뿐만 아니라 인공지능의 도입은 기업뿐 아니라 국가의 생존에 영향을 미칠 수 있을 것으로 예상된다. 인공지능 기술 도입은 국가별 부가가치(GVA)에 상당한 영향을 줄 것으로 예상되는 가운데(Accenture Research and Frontier Economics, 2017), 국내의 경우 인공지능 기술 도입으로 GVA가 2035년까지 약 4.9% 증가할 것으로 전망된다.

정부 차원에서도 인공지능 활용 확대를 통해 스마트 공장, 건설, 로봇, 의료, 스마트팜, 스마트 양식 등의 산업 고도화를 추진하고 있다. 이를 통해 전 산업에 인공지능과 주요 기술을 기반으로 한 융합의 확산을 심어주고자 노력 중이다. 특히 미디어 주요 분야에도 인공지능이 접목된 사례들이 등장하고 있다. 저널리즘 영역에서 기사 발굴 및 작성, 개인 맞춤형 편집 및 추천을 비롯해 영상 분야에서는 기획, 제작, 편집, 추천 전 과정에 인공지능이 활용된 사례들이 소개되고 있다.

2.2 언론 보도 분석

언론 보도는 다양한 공적 의제들을 매개하고 조정하는 공론장으로서 역할을 수행하기 때문에(Habermas, 1991), 언론 보도를 분석하는 것은 어떠한 사회 현상에 대한 공적 의제를 도출하기에 적합하다고 할 수 있다. 또한, 언론 보도 분석을 통해 사회적 논의구조 혹은 사회적 담론을 도출할 수 있다는 점에서 언론 보도 분석을 진행하는 것은 의미가 있다고 할 것이다. 따라서 본 연구에서는 한·미 양국의 인공지능에 대한 주요 의제를 도출하고, 차이점을 비교하기 위해 언론 보도 분석을 진행하고자 한다.

특히 한국과 미국을 비교분석하는 이유와 관련해, 미국은 인공지능 분야에서 가장 앞선 기술력을 갖고 있으며(곽현, 이성원, 2019; ITIF, 2021), 이를 바탕으로 충분한 뉴스 기사가 보도되었다는 점을 들 수 있다. 또한, 미국의 주요 사회적·정치적 의제는 미국 매체를 통하여 정보의 형태로 만들어져 한국 매체에 전달되어 정치, 사회, 경제 등 국내의 다양한 영역에 영향을 미친다(박지용·이현우·노기섭, 2021). 따라서 인공지능 기술을 선도하고 있는 미국 매체의 담론 분석을 바탕으로 인공지능에 대한 주요 의제를 도출한다면 인공지능 기술에 대한 가장 발전된 의제를 파악할 수 있고, 이를 한국 매체 담론 분석과 비교한다면 현재 국내의 인공지능 기술의 현주소와 국내 인공지능 기술과 언론 보도가 나아가야 할 방향을 제시할 수 있을 것이다.

선행연구를 살펴보면 최근 인공지능에 관한 언론 기사를 분석하는 연구들이 일부 진행되고 있다. 최현중(2020)은 2015년 1월부터 2020년

4월까지 인공지능 관련 국내 신문 보도 경향을 빈도분석 및 토픽모델링을 통해 탐색하였다, 이를 통해 인공지능과 관련된 기사의 수와 분야가 매년 증가하고 있음을 확인하였다. 아울러, 인공지능과 맥락에서 노설현(2020)은 2016년과 2019년 사이의 ‘인공지능’을 키워드로 한 3,889건의 국내 기사를 바탕으로 토픽모델링 분석방법을 토대로 인공지능에 대한 토픽 10가지를 도출하였다. 또한, 박성욱(2019)은 Textom을 활용하여 2015년 11월부터 2018년까지 3년간 구글과 네이버 신문기사의 ‘Data Technology’ 키워드를 바탕으로 관련 핵심 키워드를 도출하였다. 황서이와 김문기(2019)는 1985년부터 2018년까지 출판된 총 1,691편의 학술논문을 토픽모델링과 의미연결망 분석을 통해 학계에서 논의된 인공지능에 대한 주요 토픽을 도출하였다. 이러한 기존의 선행 연구결과들은 인공지능 및 4차 산업혁명과 관련된 키워드 및 주제를 도출하고자 노력하였다는 점에서 의의가 있지만, 단순 빈도분석과 토픽모델링만을 분석방법으로 삼았다는 점에서 한계가 있다. 키워드 빈도분석과 토픽모델링은 키워드 사이의 맥락적 정보를 고려하지 못해 깊이 있는 해석이 어렵다는 단점이 있기 때문이다(Lenci, 2008). 따라서 텍스트 내 연결 패턴을 분석하여 키워드 사이의 의미 관계를 파악하는 언어 네트워크 분석을 보완적으로 사용한다면 맥락적 정보를 고려하여 더욱 풍부한 설명을 할 수 있다(이수상, 2014). 따라서 본 연구에서는 기존 토픽모델링 기반 연구의 한계를 극복하고, 인공지능에 대한 언론 보도행태를 보다 풍부하게 설명하기 위해 주제어 빈도분석 및 토픽모델링 뿐만 아니라 언어 네트워크 분석을 보완적으로

수행하고자 한다.

한편, 기존의 미국 언론 보도는 주로 질적 분석방법을 바탕으로 주로 연구되어왔다. 남재일과 박재영(2020)은 탐사보도 측면에서 한국 기사 65건과 미국 기사 46건을 내용분석 하여, 한국 탐사보도가 나아가야 할 대안적 방안을 제시하였다. 또한, 김민성과 윤영민(2015)은 한국, 미국, 영국 기사의 내용분석을 바탕으로 보도 프레임의 국가별 차이를 검증하였다. 이러한 내용분석 방법은 국가별 보도 프레임을 파악하는데 용이하지만, 특정 주제에 대해 전반적인 의제를 도출하고 제공된 텍스트 내 정보의 의미를 파악하는데는 한계가 있다. 또한, 빅데이터 분석방법을 적용한 국가별 비교 연구는 사드를 둘러싼 한국과 중국의 갈등 프레임 연구(정원준, 2018), 한류에 대한 한국과 일본의 담론 차이 연구(정수영·황경호, 2015)와 같이 주로 중국과 일본의 언론과 국내 언론을 비교한 연구가 주를 이루었지만, 한국과 미국의 언론 보도행태를 비교한 빅데이터 기반 연구는 찾아보기 어렵다. 하지만 인공지능과 관련 기술은 명백히 미국이 선도하고 있으며, 이러한 인공지능 분야 기술을 선도하고 있는 미국의 언론 보도와 국내 언론 보도를 비교한다면 국내 인공지능의 현주소와 미래를 파악할 수 있으리라 판단하였다. 따라서 본 연구에서는 한국과 미국의 주요 일간지의 뉴스 기사 제목을 바탕으로 인공지능에 대한 한·미 양국의 주요 의제를 도출하고, 나아가 양국의 인공지능 보도행태에 대한 차이점까지 살펴보하고자 한다.

기사 제목만으로 주요 의제가 도출이 가능한 것은 기사 제목이 갖는 ‘예독가능성(previewability)’ 때문이다(임영호, 2005). 즉,

독자는 기사 제목을 통해 뉴스 내용을 미리 파악하고 원하는 정보나 관심 있는 기사를 선별할 수 있다. 아울러, 기사 제목은 해당 기사 내용의 핵심 논점을 제시하고 기사의 중요성을 드러내 독자들의 관심을 끌어들이는 방식으로 작성되기 때문에 뉴스의 제목에서 반복적으로 등장하는 단어를 도출하고, 이들 사이의 네트워크를 살펴본다면 해당 이슈나 주제에 대한 주요 의제를 도출할 수 있다.

따라서 본 연구는 한국과 미국 언론 매체의 뉴스 제목을 바탕으로 주제어 빈도분석, 토픽모델링, 언어 네트워크 분석을 수행하여, 인공지능에 대한 한국과 미국 언론 매체의 주요 의제를 도출하고 이를 비교 분석하고자 한다. 보다 구체적으로, 한국과 미국의 주요 일간지 기사 제목에서 자주 쓰인 핵심 단어들은 무엇인지 도출하고 핵심 단어들 사이의 연결구조를 시각화하여, 한·미 양국의 인공지능에 대한 보도가 어떻게 다루어졌는지 탐색하고자 한다. 이를 바탕으로 본 연구에서는 아래와 같은 두 가지의 연구문제를 설정하고 이에 대한 해답을 도출하고자 한다.

연구문제 1. 인공지능(AI) 관련 보도에서 나타나는 주요 의제들은 무엇인가?

연구문제 2. 인공지능(AI) 관련 보도는 국가간(한국과 미국)에 어떠한 차이를 보이는가?

Ⅲ. 연구 방법

본 연구는 빅데이터 방법을 활용하여 한국과 미국의 언론 보도를 분석하고 국가별 언론에서

다루는 인공지능에 대한 인식의 차이를 살펴보고자 하였다. 구체적으로, 본 연구는 한국과 미국의 주요 언론들의 뉴스 기사의 제목을 대상으로 텍스트마이닝 방법 중 토픽모델링과 언어 네트워크 분석을 활용하여 인공지능과 관련 보도에서 나타나는 주요 의제를 도출하고 나아가 인공지능 관련 보도에서 한국과 미국 언론의 인식 차이를 살펴보고자 하였다.

텍스트마이닝 분석은 데이터 수집, 데이터 전처리, 데이터 분석을 포함하는 일련의 과정으로 (황명화, 2014), 본 연구에서는 텍스트마이닝 분석 절차를 1) 데이터 수집, 2) 데이터 전처리, 3) 데이터 분석 단계로 세분화하여 설명하고자 한다.

3.1 데이터 수집

한국과 미국의 주요 언론의 인공지능에 관한 기사 수집에 앞서, 매체의 영향력과 정치적 중립성을 고려하여 한국과 미국의 주요 언론사를 선정하고자 하였다. 매체의 영향력과 정치적 성향을 고려하여 국내 종합일간지 구독률 순위 상위 5개 <조선일보>, <중앙일보>, <동아일보>, <한겨레신문>, <경향신문> 언론사 및 대표적인 경제지인 <매일경제>를 포함하여 국내 주요 보수지, 진보지, 경제지의 주요 언론사를 모두 포괄하고자 노력하였다(구정우·조성권, 2017; 민진영, 이애리, 2021; 함승경 등, 2021). 아울러, 미국의 주요 신문사 역시 매체의 영향력을 고려하여 최근 3년간 미국의 판매 부수 상위 6개 일간지 USA투데이<USA Today>, 월스트리트저널<Wall Street Journal>, 뉴욕타임스

<New York Times>, LA타임스<Los Angeles Times>, 뉴욕포스트<New York Post>, 시카고 트리뷴<Chicago Tribune>을 포함하였다(Statista, 2021). 아울러, 정치성향 측면에서는 중도 성향의 언론사로 평가받는 USA투데이, 월스트리트저널, 진보 성향 언론사로 뉴욕타임스, LA타임스, 보수 성향 언론사로 뉴욕포스트, 시카고트리뷴을 선정하여 데이터를 수집하였다.¹⁾ 데이터의 수집 기간은 인공지능에 관한 관심의 시발점이 되었던 ‘알파고’가 등장한 2016년 3월부터 2021년 2월까지 5년간 한국과 미국의 인공지능에 대한 담론을 살펴보았다.

한국의 온라인 뉴스 기사 수집을 위해 한국 언론진흥재단의 ‘빅카인즈’를 활용하였다. 기사 제목에 “인공지능” 혹은 “AI”를 포함하고 있는 주요 일간지 보도 자료를 수집하였다. 포토, 속보, 단신, 조류독감(AI), 중복된 기사 등 본 연구의 목적에 맞는 정보를 포함하고 있지 않은 기사를 제외하고, 총 11,116건의 기사를 수집하였다. 언론사별 기사 수는 <조선일보> 2,211건, <중앙일보> 1,835건, <동아일보> 1,305건, <한겨레신문> 458건, <경향신문> 803건, <매일경제> 4,504건으로 나타났다.

미국의 온라인 뉴스는 렉시넥시스 학술(LexisNexis Academic) 데이터베이스를 활용하여 수집하였다. LexisNexis Academic은 세계 각국의 신문·잡지·통신 기사, 뉴스레터, TV 방송 내용 등 디지털 자료를 다양한 언어(영어, 프랑스어, 스페인어, 독일어 등)로 제공하고 있다. 검색 키워드는 “artificial intelligence” 혹은 “A.I.”를 기사 제목에 포함하고 있는 보도

1) 미국 언론의 정치 편향성에 대한 정보를 제공하는 웹사이트 올사이즈(AllSides)를 참고하여 미국 언론사의 정치 편향성을 선정하였다. URL: <https://www.allsides.com/media-bias/media-bias-ratings>

<표 1> 연도별 한·미 기사수

연도	한국 기사수	미국 기사수
2016	769	621
2017	1,399	1,271
2018	2,533	1,615
2019	2,575	1,538
2020	3,360	984
2021	480	150

자료를 수집하였다. LexisNexis Academic의 경우 포도, 속보 등의 기사는 포함하지 않기 때문에 단신 및 중복된 기사를 제외한, 총 6,179건의 기사를 수집하였다. 언론사별 기사 수는 <USA Today> 361건, <Wall Street Journal> 430건, <New York Times> 4,093건, <New York Post> 167건, <Chicago Tribune> 577건으로 나타났다. 미국의 주요 일간지의 인공지능 관련 뉴스 보도(총 6,179건)는 한국(총 11,116건) 대비 약 45% 수준의 양적 차이가 존재하였다. 수집된 한국과 미국의 연도별 기사수는 <표 1>과 같다.

3.2 데이터 전처리

본 연구를 위해 수집한 데이터는 텍스트 기반의 비정형 데이터이기 때문에 분석이 쉬운 형태로 전처리(preprocessing) 하는 과정이 필수적이다. 특히 전처리 방식에 따라 연구결과에 큰 영향을 미칠 수 있다는 점에서 텍스트마이닝 분석방법에서 데이터 전처리는 매우 중요하다. 따라서 텍스트마이닝을 활용한 연구는 수집한 텍스트에 관한 전처리 작업을 상세하게 소개하는 과정이 양적 분석과 비교해 더욱 중요하다(이소정, 2020), 본 연구에서는 텍스트 데이터의 전처리 과정을 자세히 설명하고자 노력하였다. 본 연구에서 데이터 전처리를 위해 R

프로그래밍의 ‘dplyr’, ‘KoNLP’, ‘stringr’, ‘textclean’, ‘tidyr’, ‘tidytext’, ‘tm’, ‘widyr’ 등의 패키지를 활용하였고, 텍스트 데이터 전처리 과정은 다음과 같다.

먼저, ‘KoNLP’와 ‘tm’ 패키지를 활용하여 한글과 영어의 품사를 분리하였고, ‘tidytext’ 패키지를 활용하여 각 기사의 제목을 품사별로 토큰화하였다. 다음으로 ‘tm’ 과 ‘stringr’ 패키지를 활용하여 각 토큰에 포함된 구두점과 괄호 등 기호를 제거하고, 축약어(예: ‘인공지능’, ‘AI’를 전부 ‘AI’로 처리), 유사어(예: ‘Cars’, ‘Car’를 전부 ‘Car’로 처리)를 하나의 단어로 통일하였다. 또한, 띄어쓰기를 기준으로 세세하게 토큰화 된 단어(예: ‘4차’와 ‘산업혁명’)로 구분된 단어를 ‘4차 산업혁명’으로 처리) 역시 유의어(thesaurus)로 지정하는 과정을 거쳤다. 아울러, 분석과정에서 의미를 부여할 필요가 없는 불용어(stopword)(예: 오늘, 내년, we, her 등)를 사전에 추가하는 과정을 반복하여 수집한 뉴스 기사 제목을 분석 목적에 맞게 필터링하였다.

또한, 토픽모델링과 언어 네트워크 분석을 수행하기 위해 기사 제목을 의미의 최소 단위인 형태소로 나누는 전처리 과정을 진행하였다. 특히 한글의 경우 어휘 형태소(명사, 형용사, 동사 등)와 문법 형태소(조사, 어미 등)로 구분된다. 본 연구에서는 ‘KoNLP’와 ‘tm’ 패키지를

통해 형태소의 의미와 기능에 따라 분리하고, 그중에서 명사, 형용사, 동사만을 분석을 위해 추출하였다. 이러한 전처리를 거친 말뭉치(corpus)를 데이터 분석 단위(unit of analysis)로 선별하였다.

3.3 데이터 분석

본 연구의 목적은 텍스트마이닝을 활용하여 인공지능에 대한 한국과 미국의 주요 의제를 도출하고, 양국 간 보도행태의 차이를 탐색하는데 있다. 먼저, 본 연구에서는 주요 의제를 도출하기 위해 주제어 빈도분석을 수행하였다. 주제어 빈도분석을 통해 텍스트 내에서 자주 언급되는 주제어를 도출하고, 이들이 언급되는 빈도를 중요도로 판단하여 분석하는 기법이다(정근하, 2011). 주제어 빈도는 단순히 특정 단어가 텍스트 내에서 얼마나 자주 출현하는지를 보여주는 단어 빈도(Term Frequency: TF)로 정의할 수 있다. 간단히 말해서 TF 값이 클수록 주제어 내에서 중요도가 높다고 할 수 있다.

주제어 빈도분석은 대상 텍스트 내 주요 주제어를 도출할 수 있다는 점에서 의미가 있지만, 주제어들이 서로 어떤 연관이 있는지 파악하기 힘들다. 이에 따라 학계에서는 주제어 빈도분석을 발전시킨 연관어 분석(keyword based association analysis)을 주로 사용한다. 연관어 분석은 텍스트 내에서 함께 언급된 주제어의 동시-출현(co-occurrence) 빈도 혹은 상관관계 값을 바탕으로 연결 관계를 분석하면 언론 보도상의 주요 관심 토픽을 도출할 수 있고, 나아가 주요 토픽 간의 관련성을 파악할 수 있다.

본 연구에서는 연관어 분석방법 중 토픽모델

링과 언어 네트워크 분석을 활용하였다. 토픽모델링은 해당 텍스트를 대표하는 특정 주제 그룹을 추출하여 주는 분석방법이다. 여기에서 토픽이란 같이 등장할 확률이 높은 단어들의 집합으로 정의할 수 있는데, 이는 전체 텍스트를 관통하는 잠재적인 주제를 추출할 수 있다는 점에서 사회 문제 및 현상에 관한 주요 의제를 도출하는 데 유용하다. 본 연구에서는 여러 가지 토픽모델링 방법 중 잠재디리클레할당(Latent Dirichlet Allocation: LDA) 모형을 사용하였다. LDA 모형은 다른 토픽모델링 모형에 비해 결과 해석이 쉽고(Blei, 2012), 과적합 문제를 해결하여 방대한 데이터 내에서 여러 가지 토픽을 추출할 수 있기 때문이다(Blei et al., 2003). 본 연구에서는 R 프로그래밍의 'topicmodels' 패키지를 사용하여 토픽모델링을 실시하였다. 한국과 미국의 최적 토픽 수는 Griffiths와 Steyvers(2004)의 기준에 의해 결정하였다. 인공지능에 관한 한국 뉴스 기사 토픽 수는 총 7개, 미국 뉴스 기사 토픽 수는 총 6개로 도출되었다.

한편, 본 연구에서 언어 네트워크 분석은 동시-출현 빈도를 활용하여 주제어 사이의 관계를 네트워크 형태로 표현하였다. 언어 네트워크 분석을 통해 주제어들이 어떤 맥락에서 함께 사용되었는지 파악할 수 있다. 즉, 언어 네트워크 분석은 단어 간의 의미 관계를 도출할 수 있다는 점에서 토픽모델링과 상호보완적으로 활용할 수 있다. 본 연구에서는 R 프로그래밍의 'tidygraph'와 'tidytext' 패키지를 활용하여 연결 중심성(degree centrality)을 중심으로 네트워크 분석을 수행하였다.

요약하면, 본 연구의 연구 방법은 다음과 같

다. 빅인즈와 LexisNexis Academic과 같은 디지털 뉴스 데이터베이스를 활용하여 총 17,295건(한국: 11,116; 미국 6,179)의 뉴스 기사를 수집하였다. 이후 형태소(명사, 형용사, 동사) 추출, 불용단어 및 어구 삭제, 비슷한 단어 통합과 같은 전처리 과정을 거쳐 텍스트마이닝 기반의 주제어 빈도분석, 언어 네트워크 분석, LDA 토픽모델링을 수행하였다. 이를 통해 언론 매체에서 한국과 미국의 인공지능에 관한 주요 의제가 어떻게 담론화되고 있는지를 파악하고, 양국 간의 언론 보도의 주요 의제 차이를 살펴봄으로써 한·미 언론 보도행태에 대한 함의를 도출하고자 하였다.

한·미 주요 일간지 뉴스 기사 제목에서 인공지능 관련 주요 주제어 빈도수 상위 15개를 정리하면 <표 2>와 같다. 당연하게도 국내 기사에서 인공지능 혹은 미국 기사에서 artificial intelligence는 가장 빈도수가 높은 키워드였지만, 모든 기사 제목에 인공지능 혹은 artificial intelligence가 포함되어있어 키워드 빈도분석을 수행할 때 두 단어를 제거하였다.

먼저, 한국에서 가장 많이 언급된 주제어는 '바둑'이었고, 다음은 '삼성', '기술', '시대', '로봇' 순으로 높은 빈도수를 보였다. 국내 언론 보도는 인공지능에 관한 관심을 최초로 유발한 '알파고'와 관련된 '바둑'에 가장 높은 관심을 보였다. 또한, 국내에서 인공지능 기술이 가장 앞서서 '삼성' 및 '기술', '시대', '로봇'과 같은 인공지능 관련 주제어를 가장 많이 언급한 것으로 나타났다. 국내 언론 보도는 주로 인공지능 도입 및 관련 서비스 개발에 대한 주제를 더욱 중요하게 다룬 것으로 나타났다.

한편, 미국 언론 보도는 'China'가 가장 많이

IV. 연구결과

4.1 주제어 분석

2016년 3월 1일부터 2021년 2월 28일까지

<표 2> 한·미 뉴스 기사의 단어 빈도

한국			미국		
순위	주제어	빈도수	순위	주제어	빈도수
1	바둑	620	1	China	380
2	삼성	545	2	Google	337
3	기술	541	3	Facebook	241
4	시대	512	4	Robot	226
5	로봇	494	5	Trump	203
6	서비스	460	6	Technology	172
7	세계	437	7	Business	150
8	기업	403	8	Amazon	136
9	스피커	395	9	Future	128
10	기반	388	10	World	122
11	개발	345	11	Silicon Valley	103
12	활용	330	12	Microsoft	102
13	인간	329	13	Apple	101
14	미래	314	14	Data	95
15	예측	311	15	Tech Giants	93

언급되었고, 그다음은 ‘Google’, ‘Facebook’, ‘Robot’, ‘Trump’ 순으로 많이 언급되었다. 미국 언론 보도는 중국과의 인공지능 패권 전쟁과 관련된 이해당사자로 ‘China’와 ‘Trump’를, 인공지능 기술의 대표적인 IT 기업인 ‘Google’과 ‘Facebook’을 주제로 많이 언급한 것으로 나타났다. 미국 언론은 인공지능과 관련된 정치적 이슈와 대형 IT 기업 중심의 인공지능을 좀더 주요하게 보도했다고 볼 수 있다.

4.2 토픽모델링

인공지능에 대한 한국과 미국의 주요 언론사의 토픽을 LDA 토픽모델링으로 도출하였다. 토픽별 키워드는 해당 토픽에서 가장 대표적인 주제를 베타 값을 기반으로 도출하였다. 토픽별 베타 값은 하나의 토픽 내에서 해당 단어가 동시에 출현할 확률을 의미하는데, 베타 값이 클수록 토픽 내의 주요 키워드임을 의미하며 토픽 내 다른 단어들과 관련성이 높다는 것을 뜻한다. 한국과 미국의 토픽별 주요 키워드 10개와 베타 값은 <표 3>과 <표 4>를 통해 제시하였다. 또한, 각 토픽은 매개 단어(초록색으로 표시)를 중심으로 연결되어 있는데, 이는 토픽이 매개 단어를 중심으로 서로 연결되어 있음을 의미한다.

우선, 한국 언론 매체의 주요 토픽은 총 7가지로 <그림 1>과 같다. 도출된 토픽에 대한 명칭은 토픽별 키워드들을 기반으로 명명하였다. 토픽1은 ‘4차 산업혁명 시대 전략’, 토픽2는 ‘디지털 플랫폼 구축’, 토픽3은 ‘미래 인재 양성’, 토픽4는 ‘활용 시스템 개발’, 토픽5는 ‘챗봇 서비스 도입’, 토픽6은 ‘음성인식 스피커 출시’,

토픽7은 ‘알파고 대국’으로 명칭을 부여하였다. 국내 언론은 인공지능을 새로운 시대의 기술로서 개발해야 할 역량으로 이해하고 있다. 또한, 국내 언론의 인공지능에 대한 보도는 주로 ‘챗봇 서비스’, ‘음성인식 스피커’ 등 미시적인 수준에서 서비스 출시를 주요 토픽으로 보도하는 것으로 나타났다.

국내 언론 보도의 각 토픽을 연결하는 매개 단어를 중심으로 결과를 살펴보면 토픽1, 토픽2, 토픽3은 ‘빅데이터’라는 매개 단어로 서로 연관되어 있다. 즉, ‘4차 산업혁명 시대 전략’, ‘디지털 플랫폼 구축’, ‘미래 인재 양성’에서 ‘빅데이터’ 기술이 인공지능과 관련된 가장 중요한 기술로 보도되었음을 추측할 수 있다. 토픽1과 토픽4는 ‘예측’과 ‘활용’을 매개로 연결되어 있으며, 토픽3과 토픽4를 매개하는 단어는 ‘국내’로 나타나 국내 중심의 인공지능 산업과 인재 양성을 주로 다룬 것으로 나타났다. 또한, 토픽2와 토픽3을 매개하는 단어는 ‘정부’로서 한국의 인공지능 관련 사업 및 인재 양성은 정부 주도로 이루어지고 있음을 나타낸다. 한편, 토픽5과 토픽6은 ‘출시’를 매개로 서로 연결되어 있으며, 토픽5와 토픽7은 ‘도입’을 중심으로 서로 매개되어 있다. 이는 서비스 출시와 도입과 같은 미시적인 수준에서 국내 인공지능 보도가 비중 있게 강조됐음을 의미한다. 특히 챗봇과 음성인식 스피커의 경우 기초적인 수준의 인공지능을 구현한 서비스로써, 국내의 인공지능 관련 서비스가 아직 초기 단계에 머물러 있음을 암시한다.

미국 언론 매체의 주요 토픽은 총 6가지로 이를 시각화하면 <그림 2>와 같다. 한국 토픽모델링 결과와 마찬가지로 각 토픽에 대한 명칭

은 도출된 키워드들을 바탕으로 연구자들의 상의를 통해 명명되었다. 토픽1은 ‘미래 기술의 명암’, 토픽2는 ‘기술 패권 전쟁’, 토픽3은 ‘모빌리티의 미래’, 토픽4는 ‘AI와 일상’, 토픽5는 ‘소셜 미디어와 허위정보’, 토픽6은 ‘로봇의 등장과 직업의 미래’로 명칭을 부여하였다. 미국 언론은 인공지능 기술의 긍정적 효과뿐 아니라 부작용에 대한 논의를 비중 있게 다루고 있음을 알 수 있다. 아울러, 인공지능 기술을 둘러싼 중국과의 기술 패권 전쟁을 역시 중요한 의제로서 다루어진 것으로 분석되었다.

미국 언론 보도의 각 토픽을 연결하는 매개 단어를 중심으로 결과를 살펴보면 토픽1, 토픽3, 토픽6을 연결하는 매개적인 키워드는 ‘Future’로 인공지능이 미래에 다양한 영역에 어떠한 영향을 미칠지에 대한 보도가 활발하게

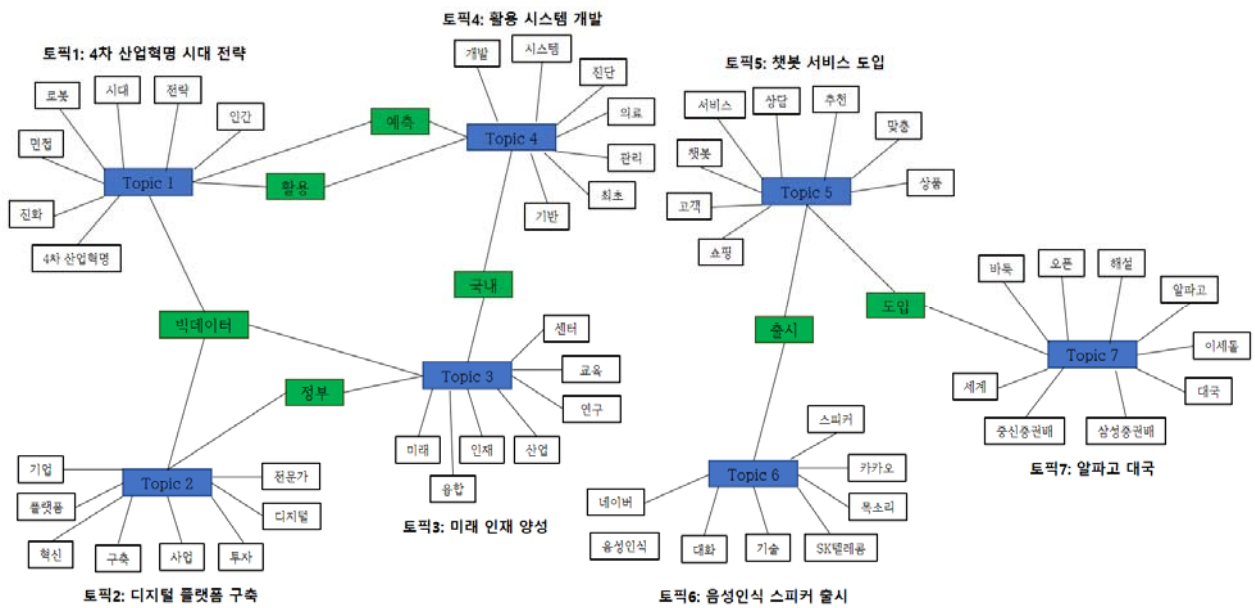
전개되고 있다는 것을 추론할 수 있다. 토픽1과 토픽4는 ‘Data’와 ‘Device’로 연결되어 있는데, 이는 일상이 된 인공지능 기술의 명암은 인공지능 기기와 데이터에 의존한다는 것을 의미한다. 한편, 토픽1과 토픽5는 ‘Privacy’와 ‘Tech Giants’, 토픽2와 토픽5는 ‘Security’로 서로 연관되어 있음을 알 수 있는데, 이는 미국 언론은 아마존, 페이스북과 같은 IT 대기업들에 의한 프라이버시 염려와 보안 문제를 비중 있게 강조하며 이를 쟁점화하고 있음을 알 수 있다. 미국 언론에서 주로 다루고 있는 인공지능 관련 토픽들은 인공지능 기술의 명암, 중국과의 인공지능 기술 패권 전쟁, 인공지능이 일자리에 미치는 영향 등으로 국내 언론 대비 보다 거시적인 수준에서 인공지능 기술이 논의되고 있음을 알 수 있다.

<표 3> 한국 언론 보도 토픽모델링 결과

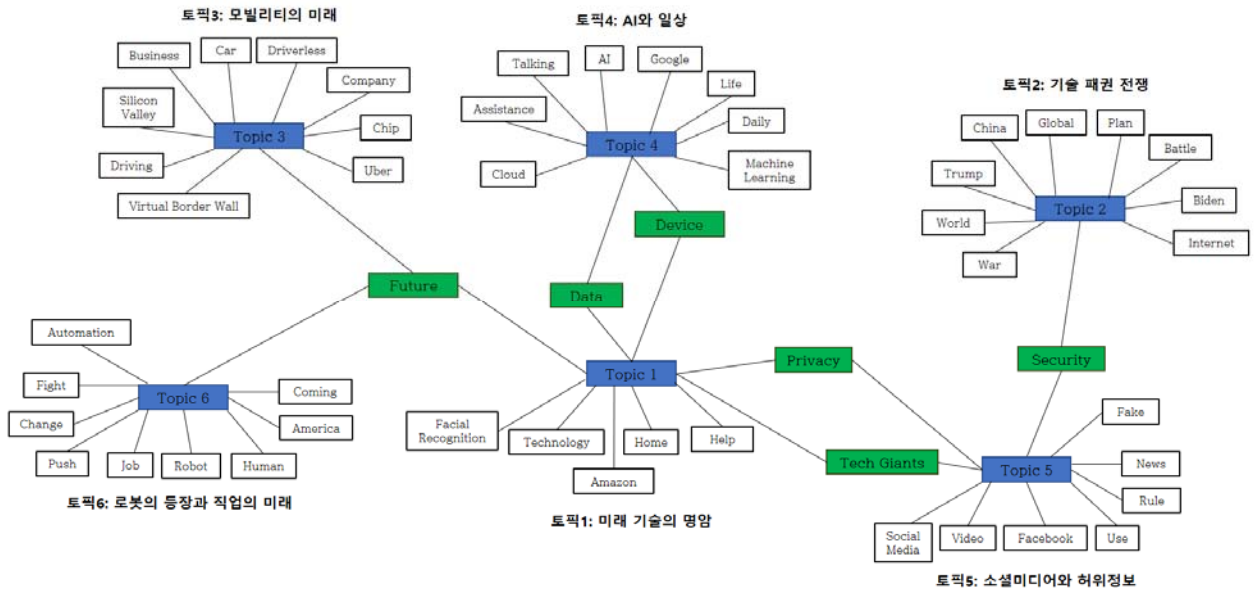
토픽1		토픽2		토픽3		토픽4		토픽5		토픽6		토픽7	
키워드	Beta	키워드	Beta	키워드	Beta	키워드	Beta	키워드	Beta	키워드	Beta	키워드	Beta
시대	0.0523	기업	0.0412	미래	0.0418	기반	0.0586	서비스	0.0329	스피커	0.0478	세계	0.0555
로봇	0.0443	플랫폼	0.0357	인재	0.0247	개반	0.0570	챗봇	0.0236	네이버	0.0388	바둑	0.0464
인간	0.0258	빅 데이터	0.0347	연구	0.0238	국내	0.0495	고객	0.0214	기술	0.0251	중신 증권배	0.0459
예측	0.0150	혁신	0.0260	교육	0.0208	시스템	0.0206	도입	0.0194	카카오	0.0230	삼성 증권배	0.0221
전략	0.0114	사업	0.0227	산업	0.0193	최초	0.0137	추천	0.0174	음성 인식	0.0150	알파고	0.0216
4차 산업혁명	0.0084	투자	0.0190	융합	0.0170	예측	0.0132	맞춤	0.0155	출시	0.0145	오픈	0.0198
면접	0.0082	디지털	0.0180	양성	0.0151	활용	0.0115	상품	0.0132	대화	0.0135	이세돌	0.0163
빅 데이터	0.0075	구축	0.0124	빅 데이터	0.0141	진단	0.0100	쇼핑	0.0123	목소리	0.0127	해설	0.0145
활용	0.0067	전문가	0.0113	센터	0.0124	관리	0.0095	출시	0.0120	SK 텔레콤	0.0120	도입	0.0094
진화	0.0060	정부	0.0112	국내	0.0110	의료	0.0088	상담	0.0110	구글	0.0118	대국	0.0088

<표 4> 미국 언론 보도 토픽모델링 결과

토픽1		토픽2		토픽3		토픽4		토픽5		토픽6	
키워드	Beta	키워드	Beta	키워드	Beta	키워드	Beta	키워드	Beta	키워드	Beta
Amazon	0.0196	China	0.0544	Business	0.0205	A. I.	0.0385	Facebook	0.0206	Robot	0.0259
Technology	0.0181	Trump	0.0289	Silicon Valley	0.0114	Google	0.0250	Social Media	0.0119	Job	0.0105
Data	0.0115	World	0.0144	Driving	0.0137	Life	0.0206	Use	0.0116	Human	0.0100
Home	0.0109	War	0.0131	Car	0.0123	Daily	0.0089	Fake	0.0111	Future	0.0090
Help	0.0093	Global	0.0115	Future	0.0088	Device	0.0070	News	0.0105	America	0.0080
Future	0.0086	Internet	0.0098	Company	0.0086	Talking	0.0070	Privacy	0.0083	Coming	0.0073
Privacy	0.0086	Battle	0.0068	Driverless	0.0085	Assistance	0.0059	Tech Giants	0.0077	Fight	0.0070
Facial Recognition	0.0083	Security	0.0068	Border Wall	0.0085	Cloud	0.0047	Security	0.0065	Automation	0.0069
Tech Giants	0.0073	Plan	0.0065	Uber	0.0082	Machine Learning	0.0047	Rule	0.0064	Change	0.0069
Device	0.0062	Biden	0.0059	Chip	0.0078	Data	0.0045	Video	0.0058	Push	0.0063



<그림 1> 한국 언론 보도 토픽모델링 결과



<그림 2> 미국 언론 보도 토픽모델링 결과

4.3 언어 네트워크 분석

LDA 토픽모델링은 전체 텍스트를 구성하는 키워드들과의 관련성을 베타 값에 근거해서 분석하는 방법임에도 불구하고(Blei et al. 2003), 각 키워드 사이의 맥락적 의미를 파악하기 힘들다는 점에서 한계가 있다. 언어 네트워크 분석은 텍스트에 포함된 키워드들 사이의 맥락적 정보와 의미를 나타냄으로써, LDA 토픽모델링이 갖는 한계점을 극복할 수 있게 해주는 분석 기법이다(남인용·박한우, 2007). 언어 네트워크 분석 내에서 키워드들의 영향력을 파악하기 위해 중심성(centrality) 분석을 주로 이용한다. 중심성이 높은 키워드는 다른 키워드 대비 상대적으로 동시 출현 빈도가 높아 텍스트의 의미를 구성하는데 핵심적인 역할을 한다고 볼 수 있다. 본 연구에서 중심성이 높은 키워드는 한국 언론 보도 네트워크 내에 빨간색 노드로

표시하였으며, 미국 언론 보도 네트워크에서는 파란색 노드로 표시하였다.

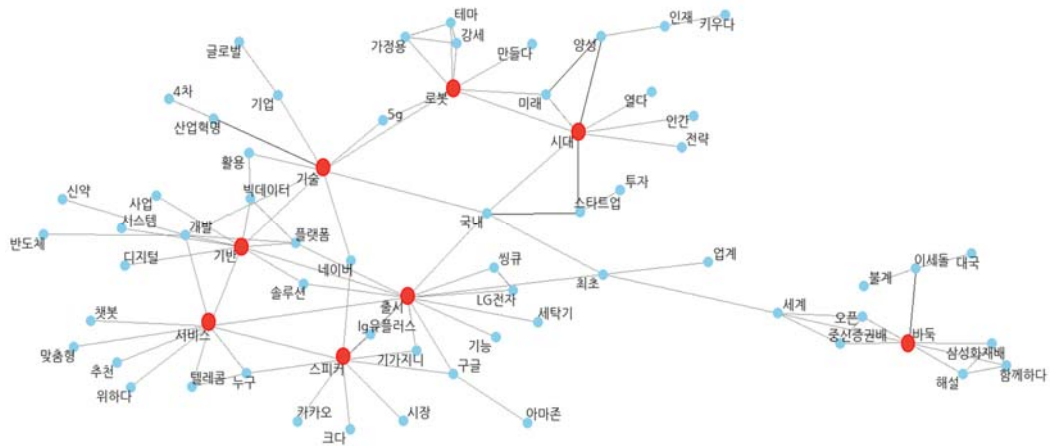
국내 언론 보도의 언어 네트워크 분석결과는 <그림 3>과 같다. 국내 언론 보도의 언어 네트워크 분석에서 중심성이 큰 키워드는 ‘출시’, ‘서비스’, ‘기반’, ‘로봇’, ‘시대’, ‘바둑’으로 나타났다. ‘기술’, ‘로봇’, ‘시대’, ‘기반’을 중심으로 이루어진 키워드 네트워크는 국내 언론은 인공지능 기술을 미래 먹거리로 상정하고 이를 위한 인재 양성, 미래 전략, 4차 산업혁명 기술, 인공지능 기반 플랫폼과 시스템 개발을 주로 다룬 것으로 나타났다. 또한, ‘출시’, ‘서비스’, ‘스피커’를 중심으로 이루어진 키워드 네트워크는 ‘누구’, ‘기가지니’, ‘쌍큐’, ‘세탁기’, ‘업계’, ‘최초’ 등 키워드들과 연관되어 있으며, 이는 국내 언론의 인공지능에 대한 보도가 ‘SK텔레콤’, ‘LG유플러스’, ‘카카오’, ‘네이버’ 등 대기업의 인공지능 제품 및 서비스 광고를 크게

다른 것으로 나타났다. 한편, ‘바둑’, ‘스피커’, ‘챗봇 서비스’로 나타나는 키워드 네트워크에서 국내 언론이 주로 다루고 있는 인공지능 관련 서비스 영역은 바둑, 스마트스피커, 챗봇 서비스를 이추론할 수 있다.

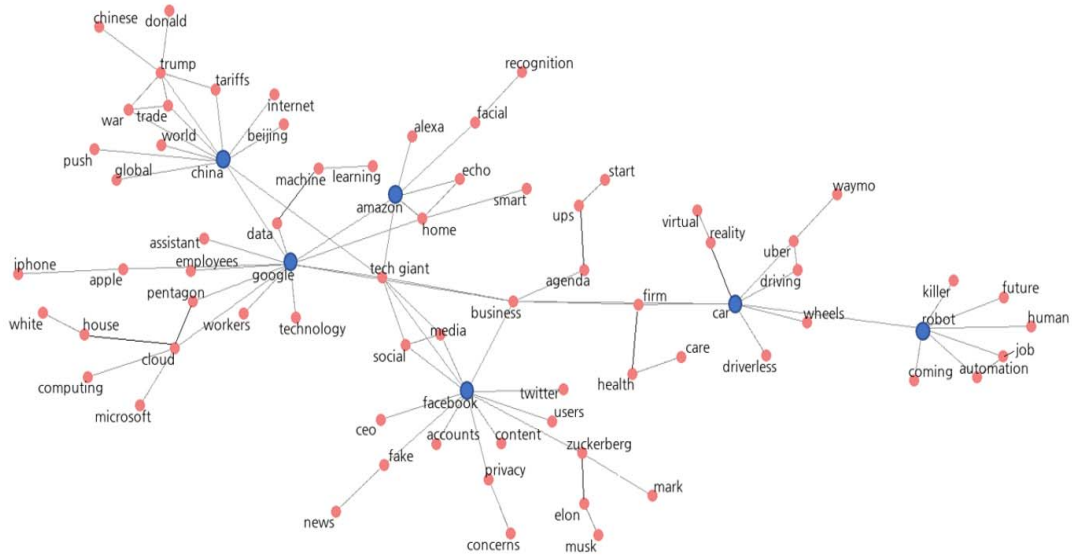
한편, 미국 언론 보도의 언어 네트워크 분석 결과는 <그림 4>와 같다. 미국 언론 보도의 언어 네트워크 분석에서 중심성이 큰 키워드는 ‘Google’, ‘Amazon’, ‘Facebook’, ‘China’, ‘Car’, ‘Robot’으로 나타났다. ‘Google’, ‘Amazon’, ‘Facebook’을 중심으로 이루어진 키워드 네트워크는 미국의 인공지능 산업은 ‘Tech Giants’로 대표되는 IT 대기업에 의해 주도되고 있음을 암시한다. 또한, ‘China’를 중심으로 보이는 키워드 네트워크는 앞서 미국 언론의 토픽모델링 결과(토픽2)를 보충 설명한다. 미국 언론 매체는 미국이 중국과의 인공지능 기술 패권 전쟁에 이기기 위해 중국에 무역 관세를 부과하는 것에 대하여 비중 있게 다루고

있음을 알 수 있다. 또한, ‘Car’를 중심으로 이루어진 키워드 네트워크는 미국 언론에서 주로 다루고 있는 인공지능 관련 서비스 영역은 자율주행 자동차임을 나타냈다.

로봇의 경우 양국 모두에서 중심성이 높은 키워드로 도출되었지만, 관련된 키워드 네트워크를 살펴보면 차이점이 나타났다. 국내 언어 네트워크 분석결과를 살펴보면, 로봇은 ‘가정용’, ‘만들다’, ‘강제’, ‘기술’ 등 키워드와 연결되어 있다. 이는 국내 언론 매체는 로봇에 대한 보도를 가정용 로봇 기술의 상용화에 초점을 맞춰 쟁점화했다고 추론된다. 반면에 미국 언어 네트워크 분석결과는 ‘Robot’이 ‘Job’, ‘Killer’, ‘Automation’, ‘Coming’, ‘Human’, ‘Future’와 연결되어 있음을 나타냈다. 미국 언론 매체는 로봇에 대한 보도를 자동화 및 미래 인간 직업의 암살자로서 역할에 초점을 맞춰 쟁점화하였음을 알 수 있다.



<그림 3> 한국 언론 보도 언어 네트워크 분석



<그림 4> 미국 언론 보도 언어 네트워크 분석

V. 논의 및 결론

본 연구는 ‘한·미 주요 언론사의 인공지능에 대한 보도에서 나타나는 주요 의제는 무엇이 있으며, 양국의 주요 의제 간의 차이는 존재하는가’라는 연구문제를 갖고 시작되었다. 이러한 연구문제를 해결하기 위해 양국 주요 언론사의 뉴스 기사 제목을 기반으로 주제어 빈도분석, 토픽모델링, 언어 네트워크 분석 등 다양한 텍스트마이닝 방법을 활용하여, 한·미 언론 보도 내용 및 차이를 분석하였다. 본 연구의 결과를 요약하면 다음과 같다.

첫째, 한·미 주요 언론사의 뉴스 기사 제목에서 나타난 공통적인 의제는 ‘인공지능에 대한 미래’이다. ‘인공지능에 대한 미래’라는 공통적인 의제에도 불구하고, 한·미 양국 주요 언론 보도에 따른 미래에 대한 의제는 양국 간 차이가 있었다. 한국 언론 매체는 인공지능에 대

한 미래를 대비하여 인재 교육 및 산업 육성에 초점을 맞추고 보도하는 반면에, 미국 언론 매체는 미래 인공지능 기술의 명암, 인공지능으로 인한 인간 직업의 미래, 모빌리티의 미래 등 미래 인공지능 기술에 대한 긍정/부정적 영향을 다양한 시각에서 보도하고 있음을 보여주었다. 이는 미국 대비 한국은 아직 인공지능 산업이 초기 단계에 머물러 있음을 나타낸다. 특히 아직 한국의 경우 인공지능 관련 인재가 부족하여 인공지능 관련 미래 인재 양성에 초점을 맞추고 있으며, 챗봇 서비스, 음성인식 스피커, AI 기반 냉장고 등 기초적인 수준에서 구현된 인공지능 제품을 홍보하는데 보다 초점을 맞추고 있다. 반면 미국의 경우 인공지능 기술의 발전에 대한 인간의 삶에 대한 의제를 주로 다루고 있다. 이는 미국의 인공지능 기술이 한국의 인공지능 기술보다 높은 수준에서 구현되어 있으며, 이에 대한 사회적인 부작용이 나타나고 있

음을 암시한다.

둘째, 한국의 인공지능은 정부 주도로 성장을 꾀하는 반면에, 미국의 인공지능은 소위 ‘Tech Giants’라 불리는 Google, Amazon, Facebook 등 IT 기업의 주도로 성장하고 있음을 확인할 수 있었다. 특히 국내 언론은 인공지능 교육 및 연구, 인공지능 관련 디지털 플랫폼 구축에 정부의 역할을 기대하고 있다. 한편, 미국의 경우 모든 토픽에서 ‘Tech Giants’가 포함되어 있다. 예컨대, 소셜미디어와 허위정보는 Facebook 위주로 발전하고 있고, 자율주행 자동차 등 모빌리티의 미래는 ‘Uber’를 필두로 발전하고 있으며, 일상 영역에서는 Google이 주도하여 인공지능 기술을 발전시키고 있음을 나타낸다. 이는 한·미 양국의 인공지능 발전 주체에 대한 인식의 차이가 존재함을 방증한다고 할 수 있다.

셋째, 한국 언론의 인공지능에 대한 담론은 미시적인 수준에서 진행되고 있지만, 미국 언론의 인공지능에 대한 담론은 거시적인 수준에서 이루어지고 있다. 미국 언론의 경우 자율주행 자동차, 안면인식, 인공지능 스피커 등 구체적인 인공지능 기술에 대한 보도부터 인공지능 기술로 인한 직업의 미래, 허위정보로 인한 문제, 인공지능 기술을 놓고 다투는 패권 전쟁 등 정치적인 이슈까지 포괄적으로 다루고 있다. 한편, 한국의 경우 챗봇, 스마트스피커 등 초기 단계의 인공지능 기술과 인공지능 관련 스타트업 투자 등에만 국한되어 미국 대비 미시적인 수준에서 인공지능에 대한 의제가 논의되었다고 할 수 있다. 또한, 같은 로봇이라는 단어를 가지고도 한국은 ‘가정용 로봇’, ‘로봇 기술’ 등 미시적인 수준에서의 로봇을 다루고 있다면, 미국

은 자동화를 기반으로 인간의 일자리를 뺏을 수 있는 로봇의 미래까지 다루고 있음을 알 수 있다. 이는 미국 언론이 한국 언론보다 한 단계 높은 차원에서 인공지능 기술을 다루고 있음을 나타낸다.

넷째, 인공지능 기술의 관심 적용 분야에서도 차이가 존재하는 것을 확인할 수 있었다. 한국 언론은 ‘알파고’로 대표되는 바둑 영역과 앞서 언급한 챗봇 및 스마트스피커를 인공지능 서비스의 주요 의제로 보도하였다. 한편, 미국 언론은 자율주행차 영역과 허위정보 탐지 등 소셜 미디어 영역에서 인공지능 서비스를 주요 의제로 보도하였다. 이는 한국은 아직도 여전히 알파고의 충격에서 벗어나지 못하고 있으며, 알파고 이후 챗봇 및 스마트스피커 등 기초적인 수준에서의 인공지능 기술 적용에 초점을 맞추고 있음을 방증한다. 반면에 미국은 허위정보와 자율주행 자동차, 가상현실, 얼굴 인식 등 생활에 밀접한 인공지능 기술 적용에 관심 두고 있음을 확인하였다. 이는 인공지능 활용 서비스에 대한 한·미 양국의 관심사의 차이 역시 존재한다고 할 수 있다.

다섯째, 한국 주요 일간지는 미국 주요 일간지 대비 인공지능 관련 제품 및 서비스에 관한 광고성 기사를 많이 보도하고 있음을 확인할 수 있었다. 언어 네트워크 분석결과를 보면 구체적인 제품명에 대한 보도와 ‘업계 최초’, ‘출시’, ‘인공지능 기반’ 등 단어를 통해 인공지능 기반 제품에 관한 광고성 기사가 미국 대비 많은 것을 확인할 수 있었다. 반면에 미국 주요 일간지는 광고성 기사보다는 프라이버시 및 보안, 기술 패권 전쟁, 로봇 자동화 등장에 따른 일자리 문제 등 사회적 이슈에 좀 더 집중하여

보도하는 경향을 보였다.

이를 바탕으로 본 연구의 함의는 다음과 같다. 첫째, 본 연구는 인공지능에 대한 담론을 뉴스 기사 분석을 통해 확인했다는 점이다. 언론 보도는 다양한 공적 의제들이 생성되고 조정되는 공론장으로서 사회적인 담론을 포함한다(Habermas, 1991). 즉, 특정 주제에 대한 언론 보도를 분석하는 것은 관련 주요 의제들과 사회적 담론들을 파악하는 데 기여한다. 특히 본 연구는 한국과 미국 언론 보도를 빅데이터 분석을 통해 인공지능에 대한 의제들이 어떻게 구성되어있고, 나아가 한·미 양국 간 인공지능에 대한 주요 의제들의 차이를 도출했다는 점에서 의의가 있다. 둘째, 본 연구는 한국과 미국 양국에서 보도한 뉴스 기사를 실증적으로 분석하여 기존 연구의 한계를 보완하고 관련 연구 확장성을 넓혔다는 점이다. 기존 한·미 뉴스 비교는 내용분석 기반의 질적인 방법에 기초해 분석되었다(남제일·박재영, 2020). 또한, 기존의 나라별 언론 보도 빅데이터 비교 분석은 한국과 중국(정원준, 2018), 한국과 일본(정수영·황경호, 2015)에서 주로 진행되었다. 하지만 담론 비교 분석에서 한·미 언론을 기반으로 텍스트마이닝 방법을 적용한 연구는 거의 없다. 정보 전달 채널로써 미디어의 역할을 고려했을 때 특정 주제에 대한 다양한 국가의 담론을 살펴보는 것은 의미가 있다. 따라서 본 연구는 핵심 정보 전달 채널인 매체의 담론을 분석하는 연구의 양적 및 질적 확장에 이바지함으로써 학문적 기여가 있다고 할 것이다. 셋째, 본 연구는 온라인저널리즘의 윤리적 이슈를 다시 한번 상기시켰다는 점에서 의의가 있다. 특히 인공지능에 대한 한국 언론의 보도행태는

광고매출의 극대화를 위해 뉴스를 광고의 도구로 삼는 ‘극단적 상업 저널리즘’과 맥을 같이한다(임종섭, 2015). 광고와 뉴스의 혼재는 언론의 신뢰도 하락을 떨어뜨리는 치명적인 요소가 되어 저널리즘의 존립 자체를 위협할 수 있다는 점에서 한국의 주요 일간지의 광고성 기사 보도행태는 지양할 필요가 있다. 따라서 이를 방지하기 위해 현직 및 예비 언론인을 대상으로 온라인저널리즘 관련 교육을 지속해서 이루어질 필요가 있다. 넷째, 본 연구는 국내 인공지능 기술에 대한 다양한 사회과학적 논의에 대해 제시했다는 점에서 의미가 있다. 현재 국내 인공지능 기술은 아직 기초적인 단계(챗봇, 스마트스피커 등)에 머물러 다양한 논의가 이루어지지 않고 있지만, 향후 직업의 미래, 허위 정보, 자율주행 자동차, 가상현실 등 미래 사회를 위해 인공지능에 기술에 대해 국내 실정에 맞는 사회과학적인 논의가 충분히 이뤄져야 할 필요가 있다.

이러한 함의에도 불구하고, 본 연구는 몇 가지 한계점이 있어 후속 연구를 통해 보완될 필요가 있다. 첫째, 본 연구의 분석 단위는 뉴스 기사의 제목만 분석 대상으로 삼았기 때문에, 기사 전체 내용을 고려하지 못했다. 따라서 후속 연구를 통해 기사 전체 내용을 고려한 분석이 진행된다면 본 연구가 보다 풍부하게 설명될 수 있을 것이다. 둘째, 본 연구에서는 소셜 미디어와 뉴스 댓글과 같은 개인 기반 담론을 함께 분석하지 못했다. 특히, 소셜 미디어와 뉴스 댓글은 개인의 생각을 온전히 담고 있어 담론으로서 분석 가치가 크다. 따라서 후속 연구를 통해 페이스북, 인스타그램, 트위터, 뉴스 댓글, 온라인 커뮤니티 게시글 등 다양한 개인들

의 생각과 공적 의제를 다룬 뉴스 보도의 차이를 본다면 더 많은 함의를 도출할 수 있을 것이다. 셋째, 본 연구는 기사 내의 감정과 시간 흐름에 따른 의제 및 논조의 변화를 고려하지 못했다. 인공지능 기술의 경우 아직 기초적인 단계이기 때문에 국내 기사 논조(tone)는 중립적일 가능성이 크다. 하지만 향후 인공지능 기술이 발달하고 이에 대한 긍정적, 부정적 영향에 관한 기사가 본격적으로 나오는 기술의 성숙기에 접어들게 된다면, 후속 연구를 통해 감정 분석 및 시간 흐름에 따른 의제와 논조의 변화 등 다양한 추가 분석을 수행할 수 있을 것이라 기대한다. 이를 바탕으로 한·미 양국 기사의 감정을 파악하고, 시간 흐름에 따른 인공지능에 대한 언론 보도의 논조를 파악할 수 있는 연구는 인공지능 기술에 대한 보다 풍부한 설명을 할 수 있을 것으로 생각한다.

참고문헌

- 과학기술정보통신부. “I-Korea 4.0 실현을 위한 인공지능(AI) R&D 전략,” 2018.
- 곽현, 이성원, “특허분석을 통한 인공지능 기술 분야 경쟁력 분석: 특허 시장성과 기술력 질적 분석을 중심으로,” 정보시스템연구, 제28권, 제3호, 2019, pp. 141-158.
- 구정우, 조성권, “언론을 통해 본 한국의 ODA: 토픽 모델링을 활용한 신문기사 분석, 1993~2016,” 국제지역연구, 제26권, 제3호, 2017, pp. 173-210.
- 김광호 등, AI 시대의 미디어, 북스타, 2020.
- 김민성, 윤영민, “잊혀질 권리에 대한 언론 보도 분석: 한국과 미국, 영국의 보도 프레임 을 중심으로,” 언론과학연구, 제15권, 제3호, 2015, pp. 33-66.
- 노설현, “토픽모델링을 활용한 인공지능 관련 이슈 분석,” 디지털융복합연구, 제18권, 제5호, 2020, pp. 75-87.
- 남인용, 박한우, “대권 예비후보자 관련 신문기사의 네트워크 분석과 홍보전략,” 한국 정당학회보, 제6권, 제1호, 2007, pp. 79-107.
- 남재일, 박재영, “한·미 탐사보도 성격 비교 연구,” 언론과학연구, 제20권, 제3호, 2020, pp. 5-45.
- 민진영, 이애리, “‘좋아요’와 ‘싫어요’ 같은 간접적 사회적 정보의 방향과 강도는 온라인 뉴스 콘텐츠 댓글의 속의의 결과 어떤 관련이 있는가? 토픽 모델링을 이용한 토픽 다양성 분석,” 정보시스템연구, 제30권, 제4호, 2021, pp. 303-327.
- 박성욱, “빅데이터 기법을 활용한 Data Technology의 키워드 분석,” 기술혁신학회지, 제22권, 제2호, 2019, pp. 265-281.
- 박지용, 이현무, 노기섭, “미국 대선 후보로 알아본 미국 언론이 한국 언론에 미치는 영향 분석: 텍스트마이닝을 중심으로,” 한국지능시스템학회논문지, 제31권, 제6호, 2021, pp. 510-518.
- 양혜승, “인터넷 뉴스 댓글의 견해와 품질이 독자들의 이슈에 대한 태도에 미치는 영향,” 한국언론학보, 제52권, 제2호, 2008, pp. 254-281.
- 이소정, “잠재디리클레할당 분석을 이용한 ‘노인일자리’ 관련 신문기사 토픽분석,” 디

- 지털융복합연구, 제18권, 제10호, 2020, pp. 537-546.
- 이수상, “언어 네트워크 분석 방법을 활용한 학술논문의 내용분석,” 정보관리학회지, 제31권 제4호, 2014, pp. 49-68.
- 이준웅, “갈등적 이슈에 대한 뉴스 프레임 구성 방식이 의견형성에 미치는 영향: 내러티브 해석모형의 경험적 검증을 중심으로,” 한국언론학보, 제46권, 제1호, 2001, pp. 441-484.
- 임중섭, “온라인 신문의 초기화면 광고 게재에 나타난 극단적 상업 저널리즘 현상 분석: 2011년과 2014년을 중심으로,” 지역과 커뮤니케이션, 제19권 제2호, 2015, pp. 61-96.
- 임영호, 신문원론, 한나래, 2005.
- 정근하, “텍스트마이닝과 네트워크 분석을 활용한 미래예측 방법 연구,” 2011 (한국과학기술기획평가원 연구보고서), Available: https://www.kistep.re.kr/boardDownload.es?bid=0002&list_no=24667&seq=10172.
- 정수영, 황경호, “한·일 주요 일간지의 한류 관련 뉴스 프레임과 국가 이미지: 기사 헤드라인에 대한 의미연결망 분석을 중심으로,” 한국언론학보, 제59권, 제3호, 2015, pp. 300-331.
- 정원준, “사드 (THAAD) 이슈를 둘러싼 한국과 중국 간 갈등 쟁점의 변화 추이 연구: 빅데이터를 이용한 시계열 토픽 모델링과 언어 네트워크 분석 기법으로,” 한국광고홍보학보, 제20권, 제3호, 2018, pp. 143-196.
- 정준화, “인공지능(AI) 정책의 주요 쟁점과 과제,” 한국지역정보화학회 동계학술발표대회논문집, 2019, pp. 39-50.
- 최현중, “토픽 모델링을 이용한 인공 지능 관련 신문 기사의 보도 경향 분석,” 한국디지털콘텐츠학회논문지, 제21권, 제7호, 2020, pp. 1293-1300.
- 한국과학기술기획평가원, “2020년도 예비타당성조사 보고서: 사람 중심 AI 강국 실현을 위한 차세대 인공지능 핵심원천기술 개발사업”, 2021.
- 함승경, 김혜정, 김영옥, “코로나 19 언론보도 경향에 대한 빅데이터 분석: 이슈 주기 및 언론사 정치적 지향에 따른 주제 분석과 언어 네트워크 분석 적용,” 한국언론학보, 제65권, 제1호, 2021, pp. 148-189.
- 황명화, “어른 모니터링을 위한 비정형 빅데이터 시공간분석 방법론 연구,” 2014. (국토연구원 연구보고서), Available: <https://www.krihs.re.kr/publica/reportView.do?num=000003600230>.
- 황서이, 김문기, “국내 인공지능분야 연구동향 분석 -토픽모델링과 의미연결망분석을 중심으로-,” 한국지털콘텐츠학회논문지, 제20권, 제9호, 2019, pp. 1847-1855.
- 4차산업혁명위원회, “2021년도 인공지능 이용인식조사,” 2021.
- Accenture Research and Frontier Economics, “Artificial Intelligence Has Potential to Increase Corporate Profitability in 16 Industries by an Average of 38 Percent by 2035,” 2017, Available: <https://newsroom.accenture.com/news/a>

- ccenture-report-artificial-intelligence-has-potential-to-increase-corporate-profitability-in-16-industries-by-an-average-of-38-percent-by-2035.htm.
- Blei, D. M., Ng, A. Y., and Jordan, M. I., "Latent Dirichlet Allocation," *The Journal of Machine Learning Research*, Vol. 3, 2003, pp. 993-1022.
- Blei, D. M., "Probabilistic Topic Models," *Communications of the ACM*, Vol. 55, No. 4, 2012, pp. 77-84.
- Entman, R. M., "Framing: Toward Clarification of a Fractured Paradigm," *Journal of Communication*, Vol. 43, 1993, pp. 51-58.
- Griffiths, T. L., and Steyvers, M., "Finding Scientific Topics," *Proceedings of the National Academy of Sciences*, Vol. 101, 2004, pp. 5228-5235.
- Habermas, J., *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*, MIT press, Massachusetts, 1991.
- Krafft, P. M., Young, M., Katell, M., Huang, K., and Bugingo, G., "Defining AI in Policy versus Practice," *In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, pp. 72-78.
- Lenci, A., "Distributional Semantics in Linguistic and Cognitive Research," *Italian Journal of Linguistics*, Vol. 20, No. 1, 2008, pp. 1-31.
- Lippmann, W., *Public Opinion*, Macmillan, New York, 1922.
- Shah, D. V., Domke, D., and Wackman, D. B., "To Thine Own Self be True: Values, Framing, and Voter Decision-making Strategies". *Communication Research*, Vol. 23, 1996, pp. 509-560.
- Tractica, *Artificial Intelligence Market Forecasts Report*, 2019, Available: <https://omdia.tech.informa.com/OM000840/Artificial-Intelligence-Market-Forecasts>.

박 종 화 (Park, Jong Hwa)



울산과학기술원에서 테크노경영학사와 경영공학 석사와 박사학위를 취득하였다. 현재 공주대학교 상업정보교육과 조교수로 재직하고 있으며, 주요 관심분야는 프라이버시, 전자상거래 플랫폼, 비즈니스 애널리틱스 등이다.

김 민 성 (Kim, Min Sung)



고려대학교에서 법학사, 미디어학석사와 박사학위를, 인디애나대학교 로스쿨에서 LL.M.을 취득하였다. 현재 고려대학교 미디어산업연구센터 객원연구원으로 활동하고 있다. 주요 관심분야는 미디어 산업, 법제, 개인정보와 프라이버시 등이다.

김 정 환 (Kim, Jung Hwan)



고려대학교에서 미디어학사, 석사, 박사학위를 취득하였다. 현재 부경대학교 미디어 커뮤니케이션학부 조교수로 재직 중이며, 주요 관심분야는 플랫폼 생태계와 미디어경영 등이다.

<Abstract>

How Does the Media Deal with Artificial Intelligence?: Analyzing Articles in Korea and the US through Big Data Analysis

Park, Jong Hwa · Kim, Min Sung · Kim, Jung Hwan

Purpose

The purpose of this study is to examine news articles and analyze trends and key agendas related to artificial intelligence(AI). In particular, this study tried to compare the reporting behaviors of Korea and the United States, which is considered to be a leader in the field of AI.

Design/methodology/approach

This study analyzed news articles using a big data method. Specifically, main agendas of the two countries were derived and compared through the keyword frequency analysis, topic modeling, and language network analysis.

Findings

As a result of the keyword analysis, the introduction of AI and related services were reported importantly in Korea. In the US, the war of hegemony led by giant IT companies were widely covered in the media. The main topics in Korean media were ‘Strategy in the 4th Industrial Revolution Era’, ‘Building a Digital Platform’, ‘Cultivating Future human resources’, ‘Building AI applications’, ‘Introduction of Chatbot Services’, ‘Launching AI Speaker’, and ‘Alphago Match’. The main topics of US media coverage were ‘The Bright and Dark Sides of Future Technology’, ‘The War of Technology Hegemony’, ‘The Future of Mobility’, ‘AI and Daily Life’, ‘Social Media and Fake News’, and ‘The Emergence of Robots and the Future of Jobs’. The keywords with high centrality in Korea were ‘release’, ‘service’, ‘base’, ‘robot’, ‘era’, and ‘Baduk or Go’. In the US, they were ‘Google’, ‘Amazon’, ‘Facebook’, ‘China’, ‘Car’, and ‘Robot’.

Keyword: Artificial Intelligence, Big Data, Topic Modeling, Language Network Analysis

* 이 논문은 2021년 11월 19일 접수, 2021년 12월 10일 1차 심사, 2022년 2월 24일 2차 심사, 2022년 3월 16일 게재 확정되었습니다.