

인공지능발달 토픽 프레임 연구 -계열화(seriation)와 통합화(skeumorph)의 사회구성주의 중심으로- A Study on AI Evolution Trend based on Topic Frame Modeling

권상희*, 차현주**

성균관대학교 미디어커뮤니케이션학과*, 성균관대학교 미디어문화콘텐츠연구소**

Sang-Hee Kweon(skweon@skku.edu)*, Hyeon-Ju Cha(top1@skku.edu)**

요약

본 연구의 목적은 AI 기술 특허(전체)와 주요 신문에 나타난 AI 보도 프레임에 바탕으로 AI 발달과정 추세를 예측하고 이를 기술-설명하는 것이다. 이를 위해 지난 9년간 출원된 한국과 미국 기술특허 요약문과 국내 주요 신문의 AI(Artificial Intelligence) 뉴스 텍스트를 분석하였다. 본 연구는 빅데이터를 활용한 토픽모델링과 시계열회귀분석이 사용되었으며, 추가로 네트워크 의제 상관분석과 회귀분석 기법이 사용되었다. 본 연구결과는 다음과 같다. 첫째, Topic 모델링 분석결과, AI 기술특허 요약문에서는 인공지능, 알고리즘 5G(Hot AI 기술) 등의 순으로 확인되었으며, AI 뉴스보도에서는 산업 적용, 데이터 활용과 시장 적용 등의 순으로 확인되어 AI의 사회문화 보도 경향을 나타냈다. 둘째, 시계열회귀분석결과, 상승추세 토픽으로는 사회문화적으로 AI 일상적·문화적 이용과 산업적용 시작이 도출되었다. 하락추세토픽으로는 시스템, 하드웨어 기술 중심으로 나타났다. 셋째, 상관관계와 회귀관계를 활용한 QAP 분석 결과, AI 기술특허와 뉴스 보도 프레임 간의 상관관계는 높은 것으로 나타났다. 이를 통해 AI 발달에서 AI 기술특허와 뉴스 보도 프레임이 미디어 담론의 결정요인에 의해 사회적으로 구성되는 것을 알 수 있었다.

■ 중심어 : | 네트워크 분석 | AI 구성주의 | 토픽모델링 | 인공지능 | 계열화 | 통합화 | 시계열회귀분석 |

Abstract

The purpose of this study is to explain and predict trends the AI development process based on AI technology patents (total) and AI reporting frames in major newspapers. To that end, a summary of South Korean and U.S. technology patents filed over the past nine years and the AI (Artificial Intelligence) news text of major domestic newspapers were analyzed. In this study, Topic Modeling and Time Series Return Analysis using Big Data were used, and additional network agenda correlation and regression analysis techniques were used. First, the results of this study were confirmed in the order of artificial intelligence and algorithm 5G (hot AI technology) in the AI technical patent summary, and in the news report, AI industrial application and data analysis market application were confirmed in the order, indicating the trend of reporting on AI's social culture. Second, as a result of the time series regression analysis, the social and cultural use of AI and the start of industrial application were derived from the rising trend topics. The downward trend was centered on system and hardware technology. Third, QAP analysis using correlation and regression relationship showed a high correlation between AI technology patents and news reporting frames. Through this, AI technology patents and news reporting frames have tended to be socially constructed by the determinants of media discourse in AI development.

■ keyword : | Network Analysis | Topic Modeling | AI(Artificial Intelligence) | Seriation | Skeumorph | Time Series Regression Analysis |

I. 서론

인공지능이 사회 전반에 걸쳐 모든 분야로 확산되어 감에 따라 사회 변화의 핵심동력으로 인공지능의 중요성이 사회적인 이슈와 주요 의제로 등장하고 있다. 그러나 이러한 인공지능 기술 영역과 사회과학적인 이슈의 등장에도 불구하고, 관련 연구논문은 아직 상대적으로 미흡한 실정이다. 더욱이 지금까지 인공지능의 연구 동향은 인공지능 기술, 기술특허 등 '기술'을 고려한 연구가 주를 이루었다고 볼 수 있다. 이에 따라 사회과학적인 측면에서의 관련 연구의 필요성이 더욱 강조된다. 이에 본 연구는 국내 인공지능 연구 기술과 사회문화적인 프레임 동향을 살펴봄으로써 인공지능과 관련한 국내 연구가 어느 정도까지, 어떤 방향으로 진화하는지 살펴보고자 한다. 또한 인공지능의 기술 영역과 사회문화 이슈인 뉴스 영역의 프레임을 실증적으로 분석하고자 한다. 이는 지금까지 인공지능 관련 국내 연구들이 '기술' 위주의 연구라고 한다면, 본 연구는 사회과학적 접근에서 연구의 유용성을 확장한다는 점에서 차별화된다.

비정형 텍스트 데이터의 분석은 텍스트마이닝 기법을 통해 이루어질 수 있으며 이 기법 중 하나인 토픽모델링을 활용한 연구는 다양한 분야에서 활발히 진행되어 왔다. 토픽모델링이란 텍스트 마이닝 기법 중 비구조화 문서 집합에서 숨겨진 핵심적 요소들을 찾아내기 위한 자연어 처리 알고리즘인데, 특히 한 문건 속에 동시에 출현하고 있는 단어들과 의미 간의 관계를 분석하는 방법론의 하나이다. 이는, 이슈의 현저성을 찾고 또 이의 패턴 등을 파악함에 있어 유용성을 지닌다[1]. 그러므로 토픽모델링은 문헌에서 숨겨져 있는 주제들을 찾아내기 위해 개발된 통계 추론 모델로 선행연구들에서 학술 분야 예측, 테크놀로지 분야, 사회 분야 예측 등에 사용될 수 있다. 또한 시계열회귀분석은 토픽모델링 결과를 시간의 흐름에 따라 시각적으로 표현할 수 있는 대표적인 분석기법으로 토픽모델링과 함께 활용되고 있다[2].

이러한 텍스트 데이터의 분석을 위해 토픽모델링을 활용한 국내연구를 살펴보면 다음과 같다. 김창식 외[3]는 국내에 출판된 정보시스템분야 저널의 연구동향을

조사하기 위하여 2002년부터 2016년까지의 관련 논문 초록 1,245편을 대상으로 토픽모델링과 시계열회귀분석을 수행하였다. 분석 결과, 도출된 20개의 토픽 중에서 고객 충성도, 소통혁신 등 4개의 상승토픽과 시스템 구축 등 2개의 하락 토픽을 도출하였다. 백영민 외[4]은 위키리크스에 공개된 외교 전문 텍스트를 이용하여 LDA 분석을 진행하였다. 이들은 전체 문서 집합에서 7개의 토픽을 파악하였는데, '한국 내 정세', '북한 관련', '주한미군', '남북경협', '해외파병', '동북아 문제', '한미 통상' 등으로 토픽이 구성됨을 파악하였다. 윤소연, 윤동근[5]은 재난 및 안전관리 분야의 동향을 토픽모델링 기법을 활용하여 분석하였는데, 1990년부터 2016년까지 '재난' 및 '안전'이란 키워드가 들어간 뉴스 기사를 수집하고, 이를 기반으로 토픽모델링 분석을 진행하였다.

토픽 모델링을 활용한 국내 연구의 경우 초기에 문헌정보학 분야에서 토픽모델링을 활용하여 정보를 파악하는 연구가 진행되었다가 점차 언론학 분야에서도 적용되고 있다[6-8]. 진실아 외[6]는 트위터에서 추출한 데이터를 이용하여 토픽모델링 기법 중 LDA 분석을 진행하여 이를 통해 전체 데이터에서 열 개의 잠재 토픽을 파악하였다. 다만 해당 연구에서 토픽 모형은 연구검증의 보조 수단으로 활용되었음을 알 수 있다. 박자현과 송민[1]은 국내 문헌정보학 학술지에서 논문 초록 정보들을 수집한 후, 이를 기반으로 LDA분석을 실시하였고, 20개의 잠재 토픽을 추출하였는데, 이를 통해 문헌정보학 분야에서 주목받는 분야는 무엇이고, 또 하향세에 있는 분야는 어떠한 것들이 있는지를 시기적으로 구분하여 확인할 수 있었다. 박윤중[7]의 경우 중복 프레임의 검증을 위해 연구 기간 동안의 신문 기사를 수집하여, 기사 본문을 대상으로 토픽모델링 분석을 진행하였고, 토픽 간 의미연결망 분석을 통해 중복 프레임이 어떻게 드러나는지 규명하였다. 정다영[9]의 경우, 공정무역에 대한 기사를 대상으로 토픽모델링 분석을 진행하였고, 추출된 토픽과 시민단체의 문건에서 나타난 핵심 어휘 간의 경향성이 어떻게 드러나는지 살펴보았다. 문안나[10]의 경우, 기업 투자와 관련한 정보인 IR에 관하여 토픽분석을 진행하였는데, 토픽모델링 분석의 경우 IR 연구의 검증을 위한 보조적 수단에 머무

를 뿐 이를 주된 분석으로 활용하지는 않았다.

이상으로 언론 분야에서 토픽모델링을 이용하여 진행된 연구들이 아직 충분히 이뤄지지 않았고, 토픽모델링을 활용함에서도 보조적인 검증수단으로써 활용할 뿐, 이를 주된 검증 방법으로써 활용하는 연구는 미흡함을 알 수 있다. 원자력 이슈에 관하여 이태준 외[8]는 토픽모델링 기법을 이용하며 미디어의 개별 프레임을 파악하는 연구가 진행되었으나, 본 연구와 비교하여 연구 대상에 있어 차이가 있고, 또한 연구 시기에 있어서도 본 연구와 다른 시점을 가지기에 연구의 차별성을 갖는다.

한편, 2016년 알파고와 이세돌 9단의 바둑 대국으로 인공지능에 대한 국민적 관심이 높아지면서 2019년은 국내 많은 산학연구 기관에서 인공지능이란 화두로 새로운 시대를 예고하는 뉴스로 사회적인 프레임이 크게 고조되었다. 국내 인공지능 기술의 경우 2022년까지 기술 최강국과 비교해 85% 수준까지 높아질 예정이고, 국내 대학과 IT기업 및 연구소를 중심으로 인공지능 기술개발 및 연구가 진행 중이다. 따라서 이러한 현시점에서 인공지능 산업의 육성과 장기적 가치 창출을 위해서 기술 연구의 방향이나 투자 분야에 대한 불확실성을 줄일 필요가 있으며 보다 체계적이고 구체적인 로드맵이 더욱 필요한 시점이다.

이에 본 연구는 AI 기술과 문화적인 특성을 규명하고 AI 미디어의 특성과 관련 사업자들의 AI 기술 도입 현황을 분석하여 AI 미디어가 가져올 미디어 산업 생태계와 사회문화적 변화에 대해 AI 기술, AI 미디어 수용모형을 통해 그 패턴을 기술, 설명하고 예측하는 데 있다. 이를 위해 사용되는 이론적 논거는 토픽 모델(topic model)을 이용한 LDA 토픽모델링 기법을 활용하여 AI 기술의 동향 및 예측에 관한 연구이다. 또한 본 연구에서는 토픽모델링 기법과 더불어 의미연결망 분석을 통해 AI 이슈에 관한 토픽과 미디어 프레임이 어떻게 나타나고, 상호간의 연관성이 드러나는지 살펴보고자 한다. 본 연구를 위해 빅카인즈에서 수집되는 데이터를 활용하여 인공지능의 기술과 사회문화적인 진화 방향을 시계열 토픽모델링(Topic Modeling) 분석기법을 통해 규명하고, 이를 실증적으로 기술하고자 한다.

아를 위해 본 연구는 2000년부터 2019년까지의 AI

기술 특허(전체)와 주요 신문에 나타난 AI 보도 프레임을 바탕으로 세부 주제를 규명하고 이들의 동향과 구조적 특성을 파악하고자 한다. 구체적으로 수집된 AI 기술 특허(전체)와 주요 신문에 나타난 AI 보도 프레임을 분석 대상으로 하여 토픽 모델링을 통해 세부 주제를 추출하고 주제들을 대상으로 네트워크 분석을 수행하여 주제들의 구조적 특성을 분석하고자 한다. 또한 본 연구는 AI 기술특허와 뉴스의 키워드와 토픽 모형, 그리고 네트워크의 의미구조는 AI의 기술적 발달과 이를 이용하고 적용하는 사회 적용에 관한 상관관계, 의미네트워크의 회귀관계의 실증검증을 통하여 AI 발달의 장기적인 발달과 진화의 사회적인 공진화를 시기별로 검증하는 데 있다. 본 연구의 결과물은 AI 기술특허와 뉴스와의 상관관계를 밝히는 데 유용한 자료로서 기능할 수 있을 것이다. 이를 통해 인공지능의 기술과 사회문화적인 진화 방향을 좀 더 쉽게 이해할 수 있도록 도와 주고, 이에 관한 연구들이 활성화되는 계기를 마련할 수 있을 것이다.

II. 연구설계

1. AI 토픽모델 기술방식

AI 관련 기술 분야의 토픽모델과 사회영역의 토픽모델을 통한 기술 사회구성(social construction) 모형은 아래 그림으로 모형화(modeling) 할 수 있다. 계열화는 기술의 발달로 나타나는 테크놀로지 발달 경향을 제시해주는 용어이다. 그러므로 AI 기술 중심의 계열화(seriation)는 AI기술로 파생되고 새로운 문화를 통합화(skeumorphs)한 사회문화 영역의 상호작용을 통해 AI 기술의 진화를 기술할 수 있다. 이를 바탕으로 Hot AI 테크놀로지와 Cool AI 기술로 그 프레임 경향을 사회 구성주의 시각으로 분류하여 AI 미디어 기술을 예측하는 데 목적이 있다. 토픽모델을 도출하여 토픽 간의 연결성이 어떤 프레임을 형성하는지 살펴보고자 한다. 또한 토픽모델링 분석을 통해 연관성이 높은 핵심 어휘 간의 집합인 주제(topic) 간의 관계에서도 어떤 프레임이 강조되는지 파악하고 이를 의미연결망 분석 결과에서 나타나는 프레임 유형과 비교하여 AI의 미디어적인

예측을 기술하고자 한다.

인공지능의 창작, 인공지능으로 인한 저작권 이슈, AI로 인한 문화와 일상의 변화는 사회문화적인 영역에 해당한다. AI가 그린 미술품이 새로운 예술 장르로 등장하고 있다. 창의성 또한 심리적인 창의성(개인의 마음속에 독창적인 아이디어 발상)과 역사적인 창의성(새로운 아이디어를 내놓는 경우)으로 AI가 가져온 새로운 영역의 창의성이 나타난 것이다. 즉 방대한 양의 데이터를 수집 결합하여 새로운 영역을 만들어내는 것이다. AI는 제조업, 자동차, 금융, 의료, 방송, 영상 등에 적용되고 있다. 이러한 인공지능의 수행 능력에 따라 약한 인공지능(weak artificial intelligence)과 강한 인공지능(strong artificial intelligence)으로 나눌 수 있다. 학습을 통해 특정분야를 인터넷 강의처럼 풀 수 있는 수준의 인공지능, 즉 인공지능이 고차원적인 사고능력을 갖추고 어떤 문제를 실제로 해결할 수 있는 상태의 지식력과 자신의 인식능력을 갖추고 있는 것을 강한 인공지능(strong artificial intelligence)이라 한다. 특히 인공지능 기술은 IoT, Wearable device 등의 각종 기기 및 기술과 접목·사용될 수 있으며, 이를 통해 사람의 역량을 확대해 문제처리 능력, 미래예측 능력, 문화적 능력의 수준을 크게 향상할 수 있다. 강한 인공지능 분야는 인간보다 지능 수준이 높고 종합적인 판단을 할 수 있는 강한 인공지능에 관한 기술 개발이 진행되고 있다. 이는 개별적인 지적 활동을 통합적으로 처리하는 강한 인공지능과 인간처럼 계산하는 지능이 아니라 인간처럼 생각하는 지능에 대한 연구가 진행되고 있다. 강한 인공지능에 대한 연구를 위해서는 심리학, 신경과학, 뇌 과학 등의 연구와 함께 통섭적·학제적 연구가 필요하다. 미래부와 정보통신기술진흥센터(2016)는 인공지능 기술을 5개의 핵심기술(학습 및 추론, 상황이해, 언어이해, 시각이해, 인지인식 및 인지)로 구분(학습 및 추론) 데이터에 내재된 패턴, 규칙, 의미 등을 알고리즘을 기반으로 스스로 학습하게 하여, 새롭게 입력되는 데이터에 대한 결과를 예측할 수 있도록 하는 기술로 정리하고 있다.

2. 테크놀로지와 사회문화 토픽프레임 : 계열화(seriation)와 통합화(skeumorph)

계열화는 매체기술의 변화가 통시적으로 어떻게 변화해 왔는가를 규명하기 위해 고고인류학에서 차용된 것이다. 고고인류학에서 '문화유물'은 시간의 변화에 따라 변화하는 일련의 속성들로 규정되어진다. 하나의 계열화 지도(map)는 시간의 흐름에 따라 도입, 유지, 그리고 소멸되는 경향을 보여준다. 디지털기술의 경우, 디지털의 기술적 구성요소는 단일 과정에서 패치워크(patchwork) 경향으로 변화되어왔다 (가령, 이메일은 지속되고 고퍼(gopher)와 같은 것은 소멸하는 경우). 디지털과 관련된 계열화를 정리해 보면 [그림 1]과 같다.

이러한 선형적 변화모형에서 명백한 진보와 개선을 양산하는 디지털의 '새로운 버전'을 상상해 볼 수 있을 것이다. 그러나 이러한 계열화 관점에서 볼 때 디지털로 인한 맥락적인 진화(즉, 개별 전통 미디어 장르 적용, 통신, 그리고 산업체의 역할을 포함하는 기술에 대한 사회구조적 시각)를 반영하지 못한다. 하나의 속성이 새로운 버전의 기능성에 명확하게 기여하는 것은 없으나 여전히 새로운 버전 안에서도 재사용되고 있다면, 이러한 속성은 통합화(skeumorphs)로서 기능하는 것이다.

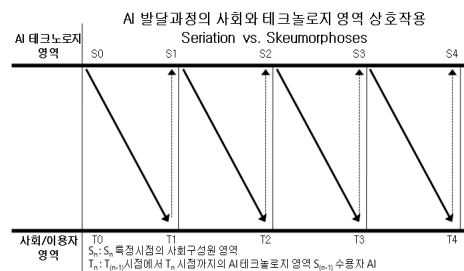


그림 1. AI 발달에 기술과 사회문화 영역의 공진화 모형

이에 AI 관련 신문기사는 계열화와 통합화 관점에 따라 서로 다른 주제, 키워드, 구성 요소, 톤, 초점 등을 표출해 왔다. 따라서 본 연구는 계열화는 기술적 발전을 설명하는 것으로 한정시키고 통합화는 사회·문화적 구조에 의해 설명되는 것으로 규정할 것이다. 즉 계열화는 AI 기술의 일련의 발전과정을 기술하는 것이고, 통합화(skeumorphs)는 AI 기술이 사회시스템과 사용자의 사용으로 만들어낸 결과로서 나타나는 문화적 양

상과 산업·콘텐츠 결과물을 의미한다. 따라서 테크놀로지 발달에 관한 중심언어인 계열화(seriation)와 테크놀로지와 사회문화 간의 연결 언어인 통합화(skeumorphs) 연구는 전통적으로 새로운 미디어(예컨대, 라디오, 텔레비전 나아가 인터넷, 모바일 등)가 등장할 때 어떠한 패러다임으로 진화해왔는가를 살펴보는 연구로 종종 등장해왔다[11].

사회구성주의는 버저와 루크만(Berger & Luchman, 1967)의 현실의 사회적 구성과 블루머(Blumer, 1969)의 상징적 상호작용, 슈츠(Schutz, 1972)의 현상학적 사회학 등에 기초하고 있는데, 이는 언론과 현실과의 관계를 설명하는 관점에 큰 영향을 주었다. 구성주의적 접근(constructionist approach)은 언론이 단순한 사실이나 사회상을 객관적으로 반영한다는 현실 반영론적 시각에서 현실 구성적 시각으로 전환되는 계기를 만들었다[12]. 즉 언론은 파편화된 사실정보를 특정한 틀이나 시각을 통해 선별하고 강조함으로써 사회적 현실을 구성한다는 것이다[13]. 뉴스 프레임은 이러한 언론의 사회적 현실 구성을 가장 잘 설명해주는 이론적 틀이라 할 수 있다. 즉 현실 인식, 해석, 선택, 강조, 그리고 배제를 통해 지속적으로 형성되는 패턴이 프레임을 만들어낸다는 것이다[14]. 전술한 바와 같이 언론이 사회적 현실을 구성하는 데 있어 선택과 배제, 그리고 현저성(salience)의 과정을 거친다. 여기서 선택은 두 가지 유기적 과정을 말한다 할 수 있다. 하나는 게이트키퍼(gatekeeping) 과정을 거쳐 특정 뉴스 아이템이 언론을 통해 보도되게 할 것인지 결정하는 것이고[15], 다른 하나는 뉴스 보도의 내용을 어떤 속성으로 구성할지에 관한 것이다[16]. 이러한 선택의 과정은 의제설정(agenda-setting)에서 말하는 대상의제(1차 의제)와 속성의제(2차 의제)로 나타난다고 할 수 있다. 언론에 의해 구성된 사회적 현실을 살피기 위해 여러 이론적 틀이 동원될 수 있으나 기본적으로 대상의제, 즉 보도의 주제가 무엇인지 파악할 필요가 있을 것이다.

이러한 테크놀로지 진화를 기술하는 기존 연구인 계열화와 통합화를 차용하여 본 연구는 특히 출연에 나타난 테크놀로지적인 발달과 AI 뉴스 기사를 사회문화적인 관점에서 바라보게 될 것이다. 즉, 전통 뉴스 매체에서는 AI 기술의 발전에 대해 시대적으로 어떻게 기록하

고 있는가? 라는 것과 특히 출연에서의 사회와 AI 기술 간의 관계에 대해서는 사회시스템이 사용해가는 과정에서 AI 테크놀로지를 어떠한 방식으로 통합해 가는가를 통해 전통 뉴스 미디어가 보도과정에서 틀짓기(frame)를 어떠한 경향으로 나타나고 있는가에 초점을 맞추었다.

3. AI 토픽모델 기술방식

본 연구는 인공지능 발달에 따른 개별 특어의 발달과 이를 보도하는 뉴스 텍스트에 나타난 경향을 기술 프레임과 네트워크 의제설정을 바탕으로 연구문제를 도출하였다. 본 연구를 통해 알아보고자 하는 연구문제는 다음과 같다.

연구문제 1. 인공지능(AI) 관련 뉴스 기사를 바탕으로 하는 발달 시기는 어떻게 구분되는가?

연구문제 2. 언론사 AI 보도와 특히 기술에 나타난 토픽(topic)의 특징은 무엇인가?

연구문제 3. 각 뉴스와 특히 시기별 프레임(frame) 경향은 어떻게 나타나는가?

연구문제 4. 뉴스와 특히 네트워크 간의 상관관계(network QAP)는 어떻게 나타나는가?

III. 연구방법

본 연구는 한국 DB 특히 문서 분석대상으로 AI(Artificial Intelligence) DB 특허문서를 택하였다. AI(Artificial Intelligence) DB 특허문서를 선택한 이유는 특허 출연에 나타난 테크놀로지적인 발달과 AI 뉴스 기사를 사회문화적인 관점에서 AI 관련하여 뉴스 프레임은 언론이 사회적 현실 구성을 가장 잘 설명해주는 이론적 틀이기 때문이다. 검색서비스인 키프리스(Kipris)에서 키워드 검색 후 온라인 다운로드 서비스를 통해 발명의 명칭, 출원 번호와 일자, 등록 번호와 일자, IPC(국제특허분류), 초록 등을 가져왔다. 특허 초록의 전처리를 통해 문서 내에 나타난 불필요한 특수문자, 관사, 조사 등의 불용어는 삭제하고 각 키워드들에 품사를 할당한 뒤에 명사들만 선택하였다. 도출된 키워드들의 레이블명을 정하기 위해 토픽프레임으로 표제

어선정(Lemmatization)을 하였다. 같은 방법으로 신문기사는 주요 신문을 데이터베이스화하고 있는 한국 언론진흥재단의 빅카인즈(Big Kinds) 시스템에서 'AI'를 키워드로 입력하여 추출하였다. 모두 807개의 기사가 분석되었고, 그중에서 키워드로 추출된 코퍼스는 총 658개였다. 텍스트 처리를 바탕으로 주요 단어의 빈도를 추출하였다. 이어 텍스트 분석을 바탕으로 20개의 AI 세부기술을 추출하여 미디어와 일반 AI기술 사이에 나타난 Topic모형을 분류하였다. 도출된 세부 AI 미디어 관련 토픽에 대해 핵심 AI기술과 뉴스에 나타난 핵심 토픽 주제를 파악하고, 2010년 이후 2018년 이전 AI의 진화 발전 단계에 나타난 Keywords를 NodeXL 기계언어로 분석하고 이를 Matrix 데이터로 연도별 비중 추이 분석을 하였다. 1차적으로 주요어의 빈도, 주요어의 네트워크 연관성, 주요단어의 중심성(매개중심성, 연결중심성, 아이겐벡터 중심성)을 통해 AI 발달의 경향을 파악하였다. 마지막으로 이 정제된 특허 Matrix와 뉴스 Matrix 데이터를 UCINET 분석에서 QAP 분석으로 상관회귀분석으로 특허기술과 뉴스 토픽간의 상관성을 분류하였다. 자료수집 세부 과정을 살펴보면 [그림 2]와 같다. 본 연구는 '빅카인즈'라는 뉴스 통합 데이터베이스를 이용하여 연구자료인 뉴스를 수집하였다. 구체적인 수집과정은 아래와 같다. 우선, 빅카인즈 공식 사이트에 들어간 후에 메인 페이지에서 '상세검색' 버튼을 찾아 클릭하면 원하는 내용을 상세하게 검색할 수 있는 화면이 나왔다. 검색유형은 '뉴스'를 선택하여 검색범위는 '형태소 분석/전체'로 선택하였다. 2011년-2018년간 인공지능과 관련된 보도의 상황을 정확하게 파악하기 위해, 상세검색어는 '인공지능'으로 입력하여 반드시 포함하는 단어는 매체보도에서 자주 쓰는 인공지능(Artificial Intelligence)을 의미한 'AI'로 입력하였다.

또한 인공지능(Artificial Intelligence)과 다른없이 같은 'AI'로 약자인 '조류인플루엔자(Avian Flu)'와 구별하기 위해 제외하는 단어는 '조류인플루엔자'로 입력하였다. 검색기간은 2010년에서 2018년까지 각각 한해 한해를 입력하였고 9년 동안 매년씩의 데이터를 수집하였다. 예를 들면, 2010년의 데이터를 수집했을 때 검색기간은 '2010-01-01'에서 '2011-12-31'로 입력하였다. 주의해야할 점은 본 연구의 2018년 데이터 수집은 2018-09-05까지 하였다.

다음으로 뉴스 데이터를 수집할 언론사 출처가 중앙지(경향신문, 국민일보, 내일신문, 문화일보, 서울신문, 세계일보, 한겨레, 한국일보를 포함), 경제지(매일경제, 머니투데이, 서울경제, 아시아경제, 파이낸셜뉴스, 한국경제, 해럴드경제를 포함), 방송사(MBC, OBS, SBS, YTN를 포함)와 전문지(디지털타임스, 전자신문를 포함)를 선택하였다. 검색 버튼을 누르면 검색결과를 나왔다. 해당 연도의 보도 건수가 화면에서 강조 표시로 보인다. 2010년에 인공지능과 관련된 보도 건수가 16건이다. 다운로드를 클릭하여 해당 연도의 자세한 뉴스 데이터를 엑셀 파일로 받았다.

IV. 연구결과

1. 연도별 특허와 뉴스에 나타난 키워드 네트워크

연구 결과, 기술적인 특허는 주로 HOT AI에 해당하는 알고리즘 텍스트마이닝(text mining), 네트워크, 5G 기술 등이 특허 기술로 도출되었다. 이런 기술들은 최근 AI 분야에서 활발하게 연구되는 핵심기술들이다. 사회문화적인 이용에 관한 AI 뉴스 보도는 알파고, Robotics, 의료, 미디어, 산업 등의 실제 사용토픽으로 구성된다.

본 연구에서는 AI 토픽모델링을 활용하여 뉴스와 특

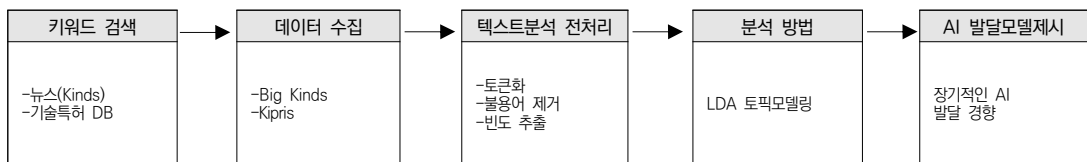


그림 2. 자료수집 세부 과정

허문서로부터 토픽을 도출하고 AI 주제 관련하여 특화와 뉴스 간 상관관계를 분석하였다. 여기에 사용된 토픽 모델링에 의해 추출된 토픽 간의 의미연결망 분석을 통해 토픽간의 관계성을 QAP 상관관계와 회계관계를 파악하였다. 이는 클러스터 구조 내의 데이터를 매트릭스 데이터화하여 UCINET 분석을 통하여 특화와 뉴스 프레임 간의 상관성과 연결성을 명확하게 보여주고 있다.

2. 기술특허에 나타난 연도별 주요 기술특허 토픽

특허에 나타난 AI 기술의 주요 단어는 Data/ Network/ Component/ 뉴스에 나타난 사회적 로봇, 또는 AI는 감정적, 정서적 교류를 통한 중요한 인공 행위자(artificial agent)로 진화하는 것을 볼 수 있다.

디지털 기술 위주의 계열화(seriation) 보도 방식에서 점차 특허 키워드, 연결망과 AI 관련되어 파생된 문화를 통합화(skeumorphs) 뉴스 형태로 분류하였다. 즉 특허 요약문은 주로 AI 관련 기술 중심(AI, 알고리즘, 5G)으로 프레임 될 수 있었으며, 뉴스 보도에서는 AI의 산업적 적용, 데이터분석 시장 적용 등 사회·문화 관련 보도 경향으로 발전한 것이다. 이를 바탕으로 그 프레임 경향을 사회구성주의 시각으로 분류하여 미디어 AI의 기술과 사회적인 연관성을 상관관계와 회귀관계로 LDA 토픽모델링 분석과 이를 프레임 분석 방법을 20개의 AI 세부 토픽 모형으로 추출하였다. 프레임을 도출해서 이를 연도별 비중 추이 분석을 통하여 AI 기술 발달의 계열화(seriation)와 통합화(skeumorphose)로 분류하였다. 텍스트 탐색, 컴퓨터 관리, 프로그래밍 구문, 네트워크 관리, 멀티미디어, 무선 네트워크 기술 등이 Hot 기술로 도출되었다. 이런 기술들은 최근 AI 분야에서 활발하게 연구되는 핵심 기술들이다.

본 연구는 AI 기술특허와 사회문화 간의 네트워크의 제설정 분석을 통해 상호의제 연결을 상관관계와 인과관계를 분석하고 이를 통해 AI 발달의 사회구성주의 미디어 발달에 대하여 실증적으로 검증을 시도한 결과는 다음과 같다.

소위 인공지능경망은 딥러닝이라는 토픽이 2010년부터 등장하는 것을 볼 수 있다. 이러한 AI는 더이상 인간 작동에 의해 작동하는 수동적인 존재가 아니라 스스로 판단과 결정을 통한 능동적인 행위자(agent)로 인간을 피동자(patient)로 진화한다. 디지털 기술위주의 계열화(seriation) 보도방식에 점차 특허 키워드, 연결망과의 AI 관련되어 파생된 문화를 통합화(skeumorphs) 뉴스 형태로 분류하였다. 즉 특허 요약문은 주로 AI 관련 기술 중심(AI, 알고리즘, 5G)으로 프레임 될 수 있었으며, 뉴스 보도에서는 AI의 산업적 적용, 데이터분석 시장 적용 등 사회·문화 관련 보도 경향으로 발전한 것이다. 이를 바탕으로 그 프레임 경향을 사회구성주의 시각으로 분류하여 미디어 AI의 기술과 사회적인 연관성을 상관관계와 회귀관계를 사용하여 LDA 토픽모델링 분석과 프레임 분석 방법을 통해 20개의 AI 세부 토픽 모형으로 추출하였다. 상기의 프레임을 도출하고 이를 연도별 비중 추이 분석을 통하여 AI 기술 발달의 seriation과 skeumorphose로 분류하였다.

본 연구는 AI 특허기술과 뉴스 보도간의 상호의제설정과 네트워크분석을 통해 상호의제 연결을 상관관계 분석과 회귀분석을 사용하여 상관관계와 인과관계를 분석하고 이를 통해 AI 발달의 사회구성주의 미디어 발달에 대하여 실증적으로 검증을 시도하였다.

분석은 한국어 언어분석 프로그램은 NodeXL과 연결망분석 프로그램인 UCINET을 활용하였다[표 1].

표 1. QAP메트릭스 데이터와 UCINET 이용 상관관계 및 회귀분석

	A	B	C	D	E	F	G	H	I
A		11	8	0	3	1	1	2	6
B	11		29	17	25	11	4	3	24
C	8	29		9	12	5	1	4	14
D	0	17	9		7	2	1	1	6
E	3	25	12	7		5	2	1	11
F	1	11	5	2	5		1	0	7
G	1	4	1	1	2	1		0	2
H	2	3	4	1	1	0	0		3
I	6	24	14	6	11	7	2	3	

	A	B	C	D	E	F	G	H	I
A		6	12	2	1	3	0	3	5
B	6		31	5	1	7	2	9	29
C	12	31		6	6	7	2	7	20
D	2	5	6		2	5	2	1	8
E	1	1	6	2		2	1	0	3
F	3	7	7	5	2		1	4	9
G	0	2	2	2	1	1		0	3
H	3	9	7	1	0	4	0		3
I	5	29	20	8	3	9	3	3	

AI 기술특허 Matrix data	AI 뉴스 Matrix data
---------------------	-------------------

표 2. 2010-2018 AI 기술특허와 뉴스의 QAP

특허 \ 뉴스	2010 news	2011 news	2012 news	2013 news	2014 news	2015 news	2016 news	2017 news	2018 news
2010 patent	0.337	0.255							
2011 patent	0.360	0.246	0.315						
2012 patent		0.305	0.300	0.336					
2013 patent			0.341	0.386	0.533				
2014 patent				0.318	0.465	0.485			
2015 patent					0.392	0.626	0.632		
2016 patent						0.710	0.639	0.369	
2017 patent							0.677	0.353	0.533
2018 patent								0.333	0.515

연구 결과, AI 기술의 발달 주요 키워드와 이를 바탕으로 한 토픽이 AI 발달에 따라 의미 연결망과 핵심어, 핵심어의 중심성이 시기별로 변화하는 것을 검증하였다. AI 뉴스의 네트워크의제의 유사성에 관한 검증에서는 AI 발달이 AI뉴스의 영향을 받는 것 보다는 AI 특허기술이 뉴스에 영향을 주는 회귀경향이 높게 나타났다. 반면 AI 뉴스와 AI 특허기술의 상관관계는 2014년 이후, 2016년-2018년에 AI 특허기술과 AI 뉴스와의 상관관계는 더욱 높아짐을 알 수 있었다.

수집된 AI 뉴스와 AI 특허 데이터를 대상으로, 년도별 토픽 프레임을 분석 후 정리하였다. 그래픽 결과는 [그림 3]과 같다.

[그림 3]은 AI 뉴스와 AI 특허 텍스트에 나타난 주요 토픽 프레임(topic frames)들과 그 프레임 하부에 포함된 주요 단어(key words)들과 토픽프레임의 네트워크를 제시하고 있다. 이차적으로 어떤 토픽을 반영하는지를 더 자세히 분석할 필요가 있다. 이를 위해 NodeXL의 군집화 기능이 있는 Clauset-Newman-Moore 알고리즘을 활용하여 관련 군집을 세밀하게 나누고 프레임을 비교 분석하였다. 사업자 유형에 따른 토픽 분석 그래픽은 대체로 토픽별로 비교가 가능하게 나타났다. 토픽 그래프의 G1이 가장 크기가 크고 중요한 토픽 그룹이다. G2, G3가 그다음으로 중요한 토픽이다. 이후 G4-G12에 이르기까지 중요도에 따르는 크기와 밀도를 의미연결망 형태로 보여주고 있다. 이를 토픽별로 중요한 하부 키워드어에 해당하는 중심성분석(연결중심, 매개중심, 에이젠벡터중심)을 통해 프레임의 특성을 파악하였다.

결론적으로, AI 특허기술과 AI의 사회적인 이용에 관한 뉴스는 상호적으로 의제상관관계를 시간(time) 변인에 따라 다르게 나타나고 높은 상관관계가 나타남을 확인하였다. 비인간/ 인간, 신체/정신, 인공/ 휴먼, 무기체/ 유기체, 복제/ 원본에 따른 기술과 인간 정신 사회적 상호작용 출현의 사례들이 등장하고 있다. 페퍼는 음성인식을 통한 의사소통 관계로 등장하고 나아가 얼굴표정, 제스처 정보를 융합하는 인식 정보처리 행위자로 등장하고 있다. 조라봇(Zora Bots), 리얼돌(Real Doll)이 통합인식 에이전트로 등장하고 있다.

본 논문에서 제시한 네트워크의제 방법론이나 자연어 처리를 통하여 매트릭스데이터 도출을 통한 기술과 사회영역의 상호 연관성을 검증하는 것은 사회의제나 기술혁신, 미디어 발달 등 다양한 분야에서의 동향분석이나 정책 도출 또는 새로운 방송/ 기술/ 모발일/VR/3D의 발달과 진화 수요 예측에 활용될 수 있을 것이다.

3. 특허와 뉴스 기수별 프레임(frame) 및 키워드 서술

뉴스와 특허에 나타난 주요 단어들을 분류하여 구성된 하위군집은 [표 3-표 8]과 같다. 연결 정도 중심성에 의하여 크게 13개 하위군집을 나누었다. [표 3-표 8]은 각 군집에 속한 주요 단어의 중심성 값을 나타낸 것이다.

[표 3]에 제시된 바와 같이, 특허 기(2010년~2013년) 분석 기간에 총 17개 프레임을 나타나고 전체 데이터에 나타난 주요 단어 중 '출현 빈도수'를 기준으로 단어를 선정한 결과는 다음과 같다.

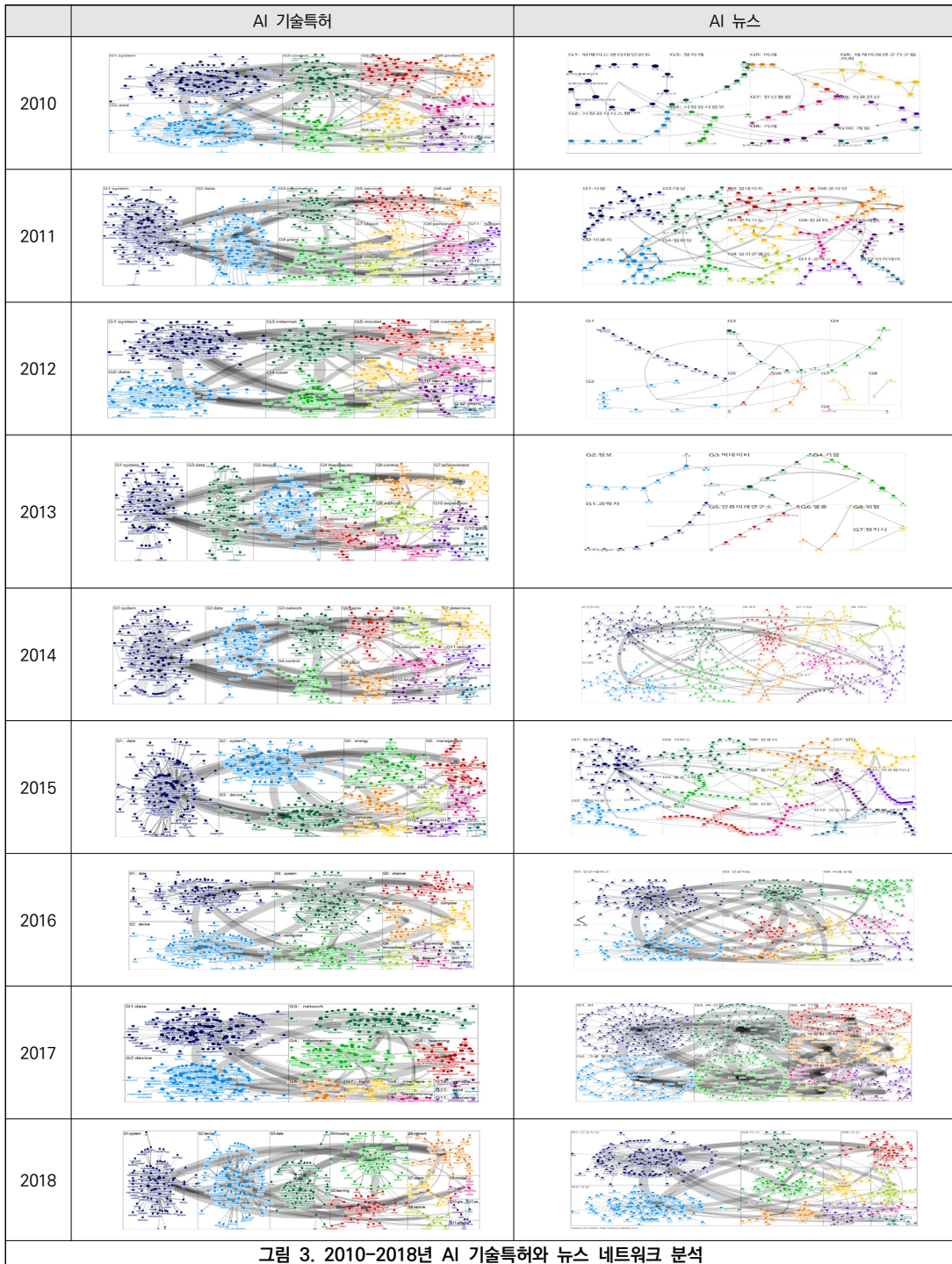


그림 3. 2010-2018년 AI 기술특허와 뉴스 네트워크 분석

표 3. AI 기술특허 1기

프레임 유형	군집	세부 토픽	상위 주요단어	중심성 값			
				연결 중심성	매개 중심성	아이겐벡터 중심성	
시스템 프레임	2010-G1	데이터 정보 방법	데이터	636	1917861.064	0.008	
			시스템	600	1790849.192	0.007	
			정보	424	1012242.390	0.006	
			방법	311	533454.505	0.005	
			장치	310	556059.585	0.005	
			다원성	259	503423.503	0.004	
			이미지	240	459390.280	0.004	
			컨트롤	218	422001.071	0.004	
			네트워크	216	320172.360	0.004	
	2011-G1	네트워크 컴퓨터 의사소통	네트워크	230	389322.095	0.004	
			컴퓨터	139	124737.683	0.003	
			의사소통	137	130315.634	0.003	
			가구	117	117700.925	0.002	
	2012-G2	시스템 장치 방법	시스템	713	2236481.312	0.006	
			장치	413	799205.918	0.005	
			방법	359	642087.951	0.004	
			네트워크	280	470005.999	0.004	
			의사소통	179	270753.794	0.003	
	2013-G1	시스템 네트워크 통제	구성단위	171	233058.931	0.002	
			시스템	580	1642170.710	0.006	
			네트워크	252	373488.458	0.004	
			통제	240	407925.061	0.003	
	데이터 프레임	2010-G2	이용자 기반 사용	제공	223	353636.887	0.004
				의사소통	182	268235.863	0.003
				이용자	371	905712.837	0.005
				기반	346	777455.536	0.005
		2011-G2	이미지 연관 신호	사용	205	421964.300	0.003
				사이	159	278768.723	0.002
이미지				241	382794.144	0.004	
연관				194	335406.007	0.003	
2012-G1		데이터 기반 다원성	신호	179	284940.957	0.003	
			입력	142	146800.966	0.003	
			데이터	731	2319447.973	0.007	
			기반	441	970212.857	0.004	
	다원성		352	759186.170	0.004		
	이미지		285	483669.684	0.004		
	신호		225	393937.905	0.003		
	상태		182	257680.936	0.003		
2013-G3	내용 시간 모형	환자	169	305979.535	0.002		
		현재	134	160498.540	0.002		
		에너지	118	170454.473	0.001		
		해결	15	7515.009	0.000		

컨트롤 프레임	2010-G3	방법 사용 발명	방법	163	209858.438	0.002
			사용	159	282749.949	0.002
			발명	137	198427.818	0.002
			관련	106	207277.964	0.002
			결정	106	117545.004	0.002
2013-G6	순서 적절하다 최고	순서	22	20432.394	0.000	
		적절하다	20	15433.950	0.000	
		최고	19	10540.833	0.000	
		자재	19	18304.008	0.000	
네트워크 프레임	2010-G4	몸 표면 운동	몸	69	90711.778	0.001
			표면	66	116158.383	0.001
			운동	54	77732.519	0.001
			전기	41	39959.768	0.001
			프레임	40	23773.396	0.001
2012-G3	원천 에너지 유체	원천	96	123342.365	0.001	
		에너지	95	102293.280	0.001	
		유체	78	121917.272	0.001	
2010-G5	카드 기록 기간	세포	77	127165.415	0.001	
		카드	38	41198.929	0.001	
		기록	24	20440.612	0.000	
		기간	24	24972.733	0.000	
2011-G4	센서 정도 몸	행동	22	15597.975	0.001	
		센서	115	145973.699	0.002	
		정도	90	162973.489	0.001	
2012-G12	레이저 보조의 마킹	몸	78	107372.307	0.001	
		레이저	22	32132.230	0.000	
		보조의	14	11270.883	0.000	
		마킹	9	8897.374	0.000	
2010-G6	최적화 터미널 동물	양자	9	13327.470	0.000	
		최적화	29	18907.034	0.000	
		터미널	24	34557.897	0.001	
		동물	19	14797.873	0.000	
		온라인	17	10796.129	0.000	
2010-G7	공기 세포의 존재	공기	20	28360.353	0.000	
		세포의	20	28793.143	0.000	
		존재	16	5138.767	0.000	
		해결	15	7515.009	0.000	

'시스템(713)'이 가장 많은 출현 빈도수로 1 순위를 차지했다. 다음으로 '데이터(636)'가 많이 언급되고 이어서 '시스템(600)', '정보(424)', '방법(311)', '네트워크(280)', '이미지(241)' 등으로 나타났다. 등장한 단어를 살펴보면 주로 AI 기술과 관련한 시스템(네트워크, 컴

퓨터, 의사소통, 기구, 발명, 모듈), 데이터(이용자, 이미지, 신호, 모형, 탐지), 네트워크(전기, 프레임, 측정, 원천, 에너지) 등 프레임으로 구성한 내용이다. 그리고 구체적인 많이 나타난 기술의 토픽은 '의사소통(137)', '신호(179)', '에너지(118)', '전기(41)', '자원(44)', '보안(44)' 등이 있는데 주로 기계 기술과 자연이나 자원의 기술 분야에 다룬다. 이를 통해 2010년에서 2013년까지 AI 특히 분야는 대부분 시스템 컴퓨터 기술 발달에 관한 것임을 알 수 있다. [표 4]에 제시된 바와 같이, 특히 2기(2013년~2016년) 분석 기간에는 총 13개 프레임을 나타냈다.

표 4. AI 기술특허 2기

프레임 유형	군집	세부 토픽	상위 주요단어	중심성 값		
				연결 중심성	매개 중심성	아이겐 벡터 중심성
시스템	2014-G1	정보 이미지 연관	정보	551	1181016.622	0.005
			이미지	293	573440.227	0.003
			연관	292	464894.820	0.003
			내용	267	393124.955	0.003
			신호	220	373798.921	0.003
			시간	186	293551.421	0.002
			물건	161	160412.696	0.002
	2015-G2	시스템 장치 방법	시스템	845	2310251.991	0.005
			장치	764	1727592.436	0.004
			방법	523	912549.364	0.003
			네트워크	384	598894.906	0.003
			모듈	316	379489.505	0.002
			구성된	305	407054.179	0.002
			표면	189	240467.764	0.001
	2016-G3	표면 에너지 시각적	에너지	150	108529.527	0.001
			시각적	119	108129.611	0.001
			구	118	129400.495	0.001
			구성원	107	96380.555	0.001
			기계 장치	106	119125.253	0.001
			인간	102	114138.999	0.001
			방법	379	646526.121	0.004
데이터	2014-G2	방법 다원성 네트워크	다원성	332	591227.857	0.003
			네트워크	315	521994.649	0.004
			사용	254	404078.106	0.003
			세트	242	350230.493	0.003
			모듈	217	281357.547	0.002
			방법	204	239873.778	0.002
			이미지	376	618377.152	0.003
	2015-G1	이미지 연관	이미지	376	618377.152	0.003

컨트롤	2016-G1	데이터 정보 다원성	연관	353	520672.797	0.003
			신호	302	462637.421	0.002
			시간	273	405710.217	0.002
			물건	236	249241.595	0.002
			결정	214	181451.523	0.002
			데이터	1193	3408475.595	0.005
			정보	794	1542178.689	0.004
	2014-G4	이용자 빛 인간	다원성	621	1124092.819	0.003
			연관된	449	677504.863	0.002
			세트	447	657123.123	0.003
			이미지	443	702018.143	0.003
			이용자	539	1226288.036	0.005
			빛	75	111403.361	0.001
네트워크	2014-G3	환자 세포 몸	인간	66	51911.750	0.001
			교통	55	42205.917	0.001
			현지	50	37471.589	0.001
			환자	112	139614.886	0.002
	2014-G6	뒤로 항구 머리	세포	89	130598.273	0.001
			몸	89	110967.002	0.001
			탐지	84	143135.367	0.001
			인공	76	72135.337	0.001
			뒤로	22	14732.157	0.000
			항구	19	2748.333	0.000
생명 요인	2015-G6	활동적 결합 압축	머리	17	37649.147	0.000
			밸브	17	8712.628	0.000
			나타내는	15	1907.150	0.000
			활동적	59	78357.986	0.001
	2015-G12	직선의 퇴보적인 발전기	결합	28	14806.226	0.000
			압축	27	22050.096	0.000
			늘기	21	2510.222	0.000
			기술	21	1586.908	0.000
			게임	20	9459.488	0.000
			직선의	12	1697.602	0.000
기억(력)/인식	2016-G11	영향 개발중, 생활양식	퇴보적인	12	14510.612	0.000
			발전기	11	10350.272	0.000
			비행기	10	8937.264	0.000
			영향	18	17286.527	0.000
	2015-G12	개발 중 생활양식	개발 중	14	2200.694	0.000
			생활양식	12	6280.309	0.000

'데이터(1193회)'는 가장 많은 출현 빈도수로 1순위를 차지했다. 특히 2기에는 시스템, 데이터, 컨트롤, 네트워크 프레임은 1기와 같이 나타났는데 비교적 감소했고 새로운 토픽이 생겼다. 생명 요소, 기억/인식, 가상, 파워, 경영, 환경 프레임은 비중이 더 높게 차

지했다. 각 군집에 들어간 주요 단어들이 모여 구성된 토픽 내용에 의하여 군집이 어떠한 프레임 유형에 속하는지를 판단할 수 있다. 생명 요소 프레임에서 '머리(17)', '팔(13)' '활동적(59)' ; 기억(력) / 인식 프레임에는 '발진기(11)', '배열(9)', '거울(6)' 등이 나타났으며, 가상 프레임에서는 '역사(26)', '생산(22)', '회복(13)' 등 단어가 등장했다. 이를 통해 AI 특허 2기에서 AI 특허와 관련 프레임은 과학 기술에 집중하는 것에 벗어나 생명 요인이나 기억력, 인식, 환경 등 전통기술과 달리 새로운 기술에 집중하는 경향을 보였다.

[표 5]에 제시된 바와 같이, 특허 3기(2016년~2018년) 분석 기간에 총 8개 프레임을 나타낸다. 여기서 특허 1기에서 계속 있었던 시스템, 데이터, 네트워크, 프레임은 계속 유지되는 것으로 보이며 세부 토픽이 변화가 있는 것으로 나타났다. 시스템 프레임의 세부 토픽은 '이미지(950)', '시간(677)' '차량(645)'; 데이터 프레임의 세부 토픽은 '다원성(862)', '모형(718)', 네트워크 프레임의 세부 토픽은 '몸(272)', '빛(256)' '타이밍(34)'으로 변했다.

표 5. AI 기술특허 3기

프레임 유형	군집	세부 토픽	상위 주요단어	중심성 값		
				연결 중심성	매개 중심성	아이겐벡터 중심성
시스템	2018-G1	이미지 시간 차량	이미지	950	1752110.814	0.002
			시간	677	1067031.143	0.002
			차량	645	918574.189	0.002
			센서	638	867551.666	0.002
			힘	456	509433.875	0.001
			가치	410	456953.942	0.001
			산출량	401	373812.980	0.001
			연결된	395	512448.510	0.001
데이터	2017-G1	다원성 이미지 시간	다원성	862	1617539.629	0.003
			이미지	612	905532.947	0.002
			시간	423	581413.889	0.002
			물건	421	516668.884	0.002
			모형	410	501556.878	0.002
			목표	352	378630.566	0.002
			가치	320	423229.710	0.001
	2018-G3	데이터 시스템 모형	데이터	2026	6353093.119	0.003
			시스템	1839	5475793.859	0.003
			모형	718	1210657.094	0.002
			번호	337	429507.637	0.001
			개인	221	200924.956	0.001

네트워크	2017-G3	부분 몸 빛	치료	169	161951.604	0.001
			부분	323	365457.081	0.001
			몸	272	397725.704	0.001
			빛	256	336586.212	0.001
			표면	253	277426.248	0.001
시간	2017-G6	옵션 의전 회의	에너지	211	253854.032	0.001
			옵션	53	25551.570	0.000
			의전	52	83535.831	0.000
다바이스	2017-G2	시스템 장치 정보	회의	31	7665.765	0.000
			시스템	1337	3804741.513	0.004
			장치	1285	3183863.008	0.004
			정보	1074	2277719.587	0.003
			방법	839	1429228.145	0.003
	2018-G2	이용자 네트워크 모듈	네트워크	621	1112129.674	0.002
			모듈	579	795418.561	0.002
			이용자	1399	3301608.632	0.003
			네트워크	830	1368771.357	0.002
			모듈	813	1206750.655	0.002
			통제	744	1094819.514	0.002

이는 특허 3기(2016년~2018년)에는 전통 기술 프레임이 유지하면서 새로운 기술 분야도 발전하고 있는 것을 알 수 있다. 또한 특허 3기에는 '자동차 프레임'과 '학습 프레임'은 처음 등장하고 그 중 자동차의 세부 토픽은 '종양학(16)', '생산성(8)', '차량(123)' 등이 나타났으며, 학습 프레임에는 '이용자(1032)', '방식(55)', '소비(44)', '수면(39)', '빠른(30)' 등이 나타났다[표 5]. 이를 통해 특허 3기에서 전통 기술에 관련 특허는 감소하고 새로운 기술이 발전해 나가는 방향이며, 자동차와 인공지능 학습 기술도 매우 발전하고 있는 추세를 알 수 있다.

[그림 4]에 제시된 AI 기술특허 고빈도 키워드 연도별 그래프에서 보는 바와 같이, 2010년-2018년의 키워드 빈도수를 살펴보면, 계속 나오는 단어들이 있다.

특히, 데이터, 시스템, 디바이스, 정보, 방법, 이미지, 네트워크, 의사소통, 센서, 자동차, 서비스, 환경, 컴퓨터, 비디오, 모바일, 게임 등 단어가 계속 상승세가 보이고 미래에도 계속 상승할 것이라고 예상된다. 전통 기술 분야(시스템, 데이터, 네트워크 등)가 계속 발전하고 있을 뿐만 아니라 환경, 비디오, 모바일, 게임 등 새로운 토픽은 매우 빠른 상승세를 보여서 주목할 필요가 있다. 또한 [그림 5]에 제시된 AI 뉴스 고빈도 키워드 연

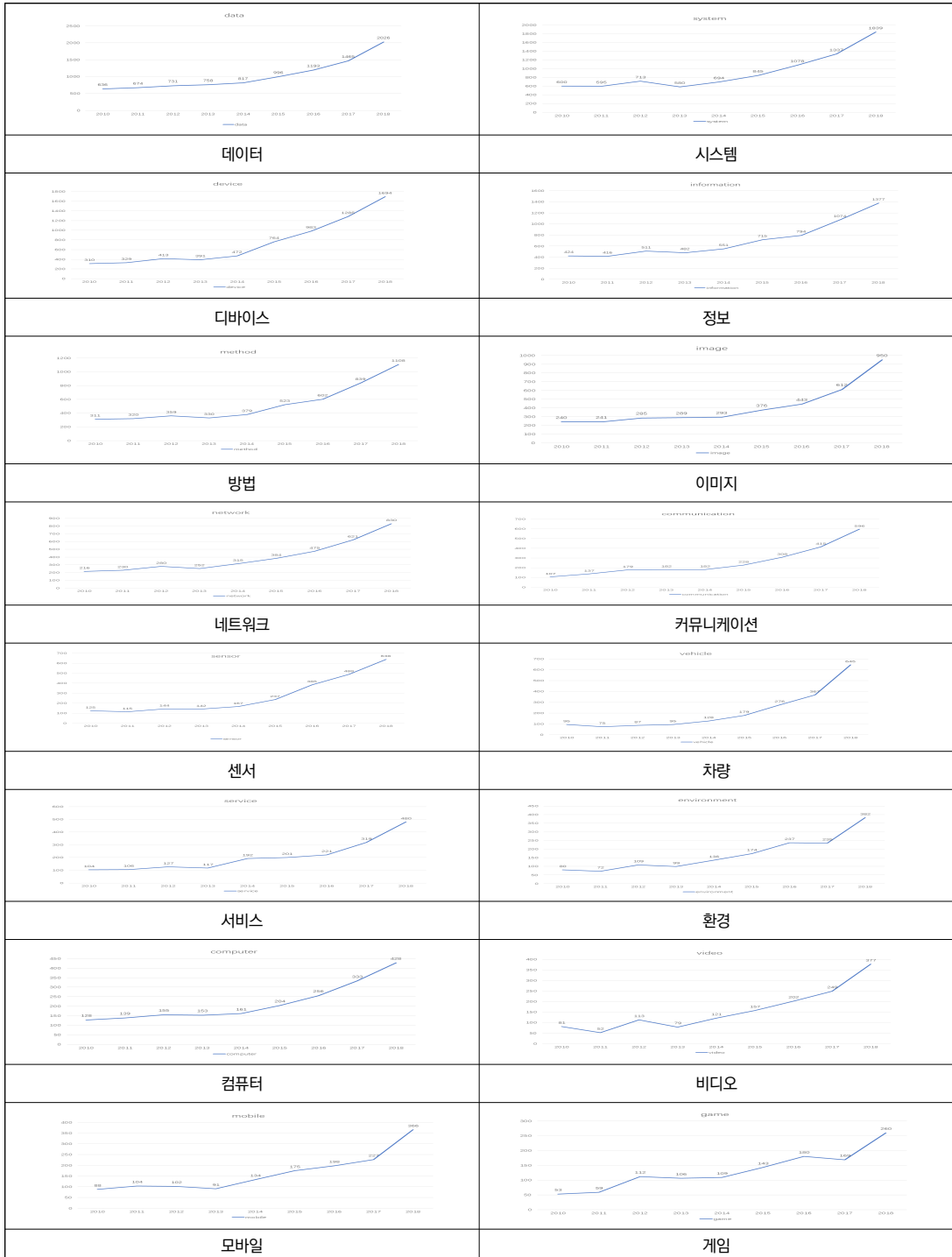


그림 4. 시기술특허 고빈도 키워드 연도별 그래프

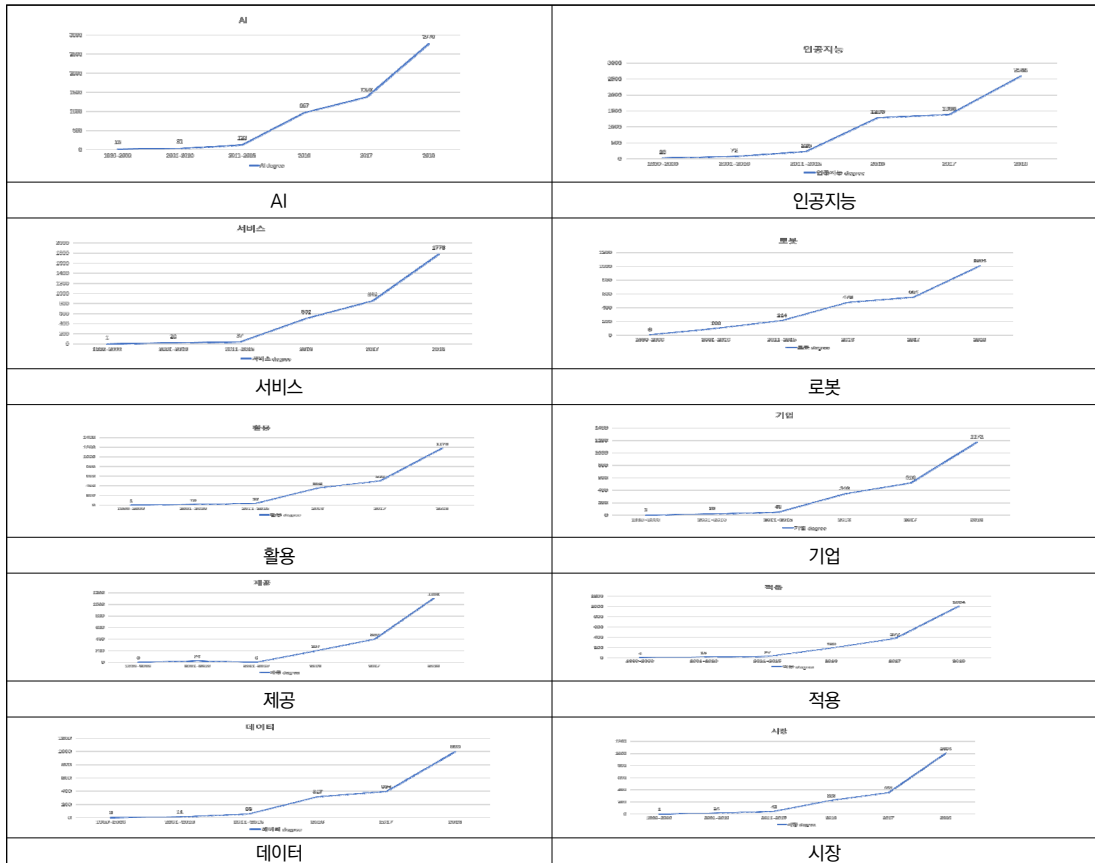


그림 5. AI뉴스 고빈도 키워드 연도별 그래프

도별 그래프에서 보는 바와 같이, 2000-2018년의 뉴스 보도를 살펴보면, 2010년 이전의 현저성이 약함으로 유의미한 데이터를 갖고 있지 않았으나 2010년부터 뉴스 보도에서 'AI, 인공지능' 두 개의 키워드는 압도적인 빈도수와 증가속도를 보이나 '서비스, 로봇, 활용, 적용, 기업, 제공, 데이터, 시장' 등 키워드 또한 고빈도 키워드로 나타났다. 이를 살펴보면 AI와 관련한 뉴스 보도는 AI의 상업적 발전 및 경제적 영역에 중점을 두고 있다는 것을 파악할 수 있다.

뉴스 1기의 키워드 군집을 살펴보면, [표 6]에 제시된 바와 같다. 뉴스 1기의 주된 키워드 군집들이 구성된 프레임은 “AI 관련 산업의 발달” 프레임, “AI 기술개발” 프레임, “AI 발달 지역” 등 3개의 프레임으로 추출할 수 있다.

우선, 2010-G2에서는 '시장 감시 시스템(2)', '시장 감

시위원회(2)', 2010-G4는 '시장 검사업무(2)', '新시장 감시 시스템(2)', 2010-G8은 '청문회(2)', '검증절차(2)' 등 상위 단어들로 AI와 관련된 시장감사 산업의 발전을 설명하고 있다[표 6]. 온라인(4), '존(2)' 등으로 AI와 관련한 온라인 산업에 대한 보도인 것을 알 수 있다. 이를 바탕으로 1기의 첫 번째 프레임을 “AI 관련 산업의 발달” 프레임으로 명명할 수 있다.

표 6. AI 뉴스 1기

프레임 유형	군집	세부 토픽	상위 주요단어	중심성 값		
				연결 중심성	매개 중심성	아이겐벡터 중심성
AI 관련 산업의 발달' 프레임	2010-G2	서울시육, 여의도, 시장 감시시스템	고도	2	158.500	0.000
			서울시육	2	130.500	0.000
			여의도	2	106.500	0.000
			시장감시위원회	2	91.500	0.000
			시장감시시스템	2	114.500	0.000

AI 기술 개발 프레임	2010-G4	시장 감시 업무, 신시장 감시 시스템	시장감시업무	2	252.000	0.000	
			불공정거래	2	156.000	0.000	
			적출기법	2	120.000	0.000	
			新시장감시시스템	2	42.000	0.000	
	2010-G8	청문회 검증 절차	청문회	2	60.000	0.000	
			검증절차	2	55.000	0.000	
			행적	2	28.000	0.000	
	2010-G10	신종 거래	신종	2	358.500	0.000	
			거래	2	342.500	0.000	
			불공정거래	2	42.000	0.000	
			한국거래소	1	0.000	0.000	
	2011-G6	인공 지능 온라인	인공지능	11	17200.992	0.064	
			온라인	4	770.625	0.006	
			무료	2	763.775	0.016	
			존	2	3312.200	0.004	
	2011-G5	협동 모드 업데이 트	협동모드	6	5163.467	0.038	
			업데이트	6	1457.500	0.022	
			대규모	3	462.650	0.012	
			영어	2	381.767	0.016	
풋볼			2	347.050	0.010		
아이체트			2	76.233	0.005		
채팅			2	172.600	0.004		
2011-G7			기능 청소 음성 인식	기능	6	1786.383	0.034
				청소	4	683.275	0.012
				원격조정	4	236.167	0.006
	모니터링	3		145.342	0.014		
	음성인식	2		137.483	0.018		
	기술	2		305.383	0.016		
2011-G8	실리콘 밸리 강의	생활가전	2	231.000	0.010		
		실리콘밸리	6	2276.617	0.000		
		강의	5	7620.633	0.020		
		스탠퍼드대	4	4782.733	0.006		
		수강생	2	342.500	0.005		
2011-G8	실리콘 밸리 강의	화제	2	396.150	0.001		

[표 7]에서 보는 바와 같이, 뉴스 2기에서는 뉴스 1기에 비해 “AI 관련 산업의 발달” 프레임, “AI 기술개발” 프레임, “AI 발달 지역” 등 3개의 프레임 기초위에 “AI 로봇” 프레임, “인간과의 대조” 프레임이 추가되었다. 키워드의 유형과 범위 또한 넓어졌다. 2014년 군집2는 '영화(18)'가 가장 높은 중심성 값을 보여주며 '미국(8)', '온라인(7)' 등과 연결된 것으로 나타났다. 2015년 군집 2에서도 '3D'이라는 단어가 상위 단어에 차지하고 있다. 두 군집 모두 영화 산업과 관련한 토픽인 것을 확인

할 수 있다. 2015년 군집5는 '미국(3)', '애플(2)'이 상위 단어를 차지함으로 스마트폰에 관한 토픽인 것을 알 수 있다. 2016년 군집4는 '서울(102)'이 최상위 단어로 나타났다으며 '미래부(49)', '포럼(48)', '러닝(36)' 등 단어들로 헬스 토픽을 구축했다. 군집10은 '신경(18)', '검사(13)', '질환(13)' 등 단어들로 건강과 관련된 토픽을 형성했다. 2016년 군집4와 군집10은 AI와 관련한 건강 산업을 다루는 것임으로 이들은 “AI 관련 산업의 발달” 프레임에 속한다고 할 수 있다. 2016년 군집5는 '북한(25)', '무기(22)'위주로 AI와 군사를 연결시켰다. 이외에도 '부동산(21)', 'CPU(15)' 등 단어들로 다양한 산업의 발달을 보여주었다. 키워드의 변화를 살펴보면 뉴스 2기는 뉴스 1기에 비해 AI 산업 발달이 한층 더 발전한 것을 알 수 있다.

이와 동시에 AI 응용이 예술, 의료 등 산업을 넘어서 군사에도 응용되며 폭이 넓어진 것을 알 수 있다. 또한, 국가별의 AI 산업 중점의 차이가 보인다. 2014년 군집 4에는 '로봇(8)', '자율주행(4)', '바이두(2)' 등 단어들로 AI 로봇의 응용 토픽을 형성했다. 한편, 2016년 군집7은 '개념(48)'과 '대책(16)', '예방(16)', '재난(15)' 등 단어로 로봇이 재난의 예방 및 응대 방향에서의 응용을 제시했다. 이 시기에 '로봇' '자율주행' 등 키워드가 처음으로 뉴스보도에서 나타났다. 2016년 군집1은 '인공지능(1286)'이 가장 높은 중심 값을 보여주며 '서비스(502)', '구글(487)' '투자(251)', '시장(226)' 등 단어를 통해서 AI가 원래 인간들의 활동 또는 작업 영역에서 어떠한 역할을 하고 있는지를 제시했다. 이로 인해 군집1을 “인간과의 대조” 프레임에 분류했다. 알파고가 이세돌과의 대결에서 승리를 거두면서, 인공지능에 대한 사람들의 관심이 폭발적으로 증가했다고 한다. 2016년 군집2는 '인간(697)'과 '알파고(636)'를 둘러싼 인간과 AI의 역할을 비교했다. 이 군집 안에 '바둑(365)', '9단(362)', '사람(341)', '컴퓨터(262)' 등 단어는 그 당시의 빅이슈인 인간과 알파고의 바둑 대전을 설명한 것으로 이들이 “인간과의 대조” 프레임에 속한다. 2014년 군집5에서는 '영국(14)', '물리학자(4)', '매체(4)' 등 단어를 통하여 영국의 AI 발전 현황 토픽을 구성했다. 2015년 군집7도 영국 및 스웨덴의 AI 발전을 제시했다. 2015년 군집10은 '일본(7)', '구글(3)' 위주로

AI의 발전 지역 토픽을 구성했다. 이외에도 '미국기업(2)', '일본기업(2)', '선도(2)', 'AI 분야(2)' 등 단어로 선도국의 AI 발달을 제시했다. 한편, 2016년 군집8은 '전공(33)', '공학(26)', '연구진(25)', '도쿄대(15)' 등 상위 단어를 통하여 대학교 안에서의 AI 발달연구를 제시했다. 2기의 “AI 기술개발” 프레임은 2016년 군집9에서의 '프로그래밍(23)' 등 토픽으로 구성되었으며 1기보다 언급이 적은 것으로 나타났다[표 7].

표 7. AI 뉴스 2기

프레임 유형	군집	세부 토픽	상위 주요단어	중심성 값		
				연결 중심성	매개 중심성	아이겐 벡터 중심성
2014-G2	영화	미국	영화	18	70367.107	0.002
			미국	8	23734.238	0.000
			온라인	7	61962.005	0.001
2015-G2	세계	3D	세계	7	15687.679	0.010
			3d	6	7431.067	0.000
			각국	2	1366.300	0.002
2015-G5	미국	애플	미국	3	1802.216	0.001
			애플	2	918.837	0.011
			우승	2	586.550	0.000
2016-G4	서울	포럼	서울	102	436528.182	0.001
			미래부	49	121726.782	0.001
			포럼	48	64224.196	0.001
			개최	44	81890.453	0.000
2016-G10	신경	검사	신경	18	60059.740	0.000
			검사	13	15440.396	0.000
			질환	13	34047.300	0.000
			질환	13	34047.300	0.000
2016-G5	북한	무기	북한	25	60993.238	0.000
			무기	22	72970.445	0.000
			부동산	21	16579.364	0.000
AI 로봇 프레임	2014-G4	로봇 개발	gpu	19	29967.896	0.000
			로봇	8	32921.789	0.008
			개발	4	15859.473	0.003
			자율주행	4	4838.130	0.000
			자율주행	4	4838.130	0.000
2016-G7	개념 과학자	과학자	바이두	2	18895.708	0.010
			개념	48	81307.585	0.001
			과학자	18	16472.624	0.000
			방지	17	46617.880	0.000
인간과의 대조 프레임	2016-G1	인공지능 서비스	대책	16	10931.217	0.000
			대형	16	16248.041	0.000

AI 발달 지역 프레임	2016-G2	인간	인공지능	1286	8092266.060	0.008		
			서비스	502	1772508.930	0.004		
			구글	487	1831950.167	0.004		
			활용	352	946132.870	0.003		
			기업	340	1036411.928	0.003		
			데이터	317	867006.096	0.003		
			정보	306	848259.469	0.003		
			투자	251	671158.134	0.002		
			시장	226	525700.584	0.002		
			인간	697	2860960.067	0.005		
			알파고	636	2441649.481	0.004		
			바둑	365	1130851.350	0.003		
			9단	362	1192737.374	0.002		
			사람	341	849205.052	0.004		
AI 발달 지역 프레임	2014-G5	영국	세계	333	1190003.160	0.003		
			미래	333	924464.431	0.003		
			미국	315	1532916.859	0.002		
			컴퓨터	262	659305.190	0.003		
			영국	13	26889.526	0.001		
			사진	4	2291.592	0.003		
			스티븐	4	5532.698	0.001		
			현지시각	4	2355.816	0.001		
			물리학자	4	2383.402	0.000		
			매체	4	17615.088	0.000		
			BBC	3	1178.395	0.000		
			2015-G7	현지시각	현지시각	5	3601.446	0.000
					영국	4	3688.327	0.000
			AI 발달 지역 프레임	2015-G10	일본	스웨덴	2	319.167
일본	7	2639.260				0.002		
구글	3	3740.137				0.001		
미국기업	2	2283.792				0.001		
ai분야	2	1928.292				0.000		
선도	2	1572.792				0.000		
일본기업	2	881.958				0.000		
2016-G8	전공	전공				33	57756.463	0.000
		공학				26	42038.806	0.000
		연구진				25	36029.814	0.000
		사피엔스				18	47071.745	0.000
		도쿄대				15	9520.883	0.000
		기술원				11	5262.059	0.000
		가천대				10	20031.361	0.000
		AI 기술개발 프레임	2016-G9	프로그래밍 SF	프로그래밍	23	9889.954	0.000
sf	22				69626.009	0.000		
심사	18				31315.911	0.000		
유전자	18				50282.887	0.000		
문학상	12				9855.113	0.000		

[표 8]에 제시한 바와 같이, 뉴스 3기의 “AI 관련 산업의 발달” 프레임은 2기와 비교하면 발전의 중점을 알 수 있다. 우선, 2017년 군집3에서 'AI 인공지능(1386)'은 가장 높은 중심성 값을 보여주면서 'AI(1376)' 및 '로봇(554)', '컴퓨터(143)'과 함께 인공지능이 컴퓨터 및 자동화 영역의 발전을 나타냈다. 또한, 2017년 군집5의 '3D(46)'는 2기 보다 강한 중심성을 보이며 인공지능이 3D 입체영화 영역의 발전과 함께 점점 활발해진다는 것을 알 수 있으며 이는 2017년 군집6의 '디자인(73)', '사운드(22)' 및 2017년 군집9의 'GPU(32)' '그래픽(17)'과 함께 인공지능이 시각과 청각의 다양한 영역에서 역할을 하고 있는 것을 설명했다.

표 8. AI 뉴스 3기

프레임 유형	군집	세부 토픽	상위 주요단어	중심성 값		
				연결중심성	매개 중심성	아이겐 벡터 중심성
AI 관련 산업의 발달 프레임	2017-G3	인공지능 AI 로봇	인공지능	1386	10732056.559	0.006
			AI	1376	11382828.172	0.006
			로봇	554	2812318.591	0.003
			교수	226	1735075.578	0.001
			예측	184	587822.928	0.001
			컴퓨터	149	311609.435	0.001
			인류	108	244941.924	0.001
	2017-G5	장비 3D	장비	46	79046.826	0.000
			3D	41	123961.696	0.000
			장기	26	106207.249	0.000
			미래부	22	100878.606	0.000
			3차원	19	18684.202	0.000
			미래창조과학부	19	60975.732	0.000
			통신망	16	3704.556	0.000
	2017-G6	디자인 전국 사운드	디자인	73	188107.321	0.001
			전국	46	99732.534	0.000
			조정	26	61605.871	0.000
			사운드	22	52635.220	0.000
			피해	21	69270.588	0.000
	2017-G9	GPU 그래픽	접수	21	34087.345	0.000
			GPU	32	30491.925	0.000
			부여	22	54598.442	0.000
			그래픽	19	46911.785	0.000
			신청	17	21130.614	0.000
	2018-G4	AI 대상	AI	2770	43349572.623	0.003
			대상	383	2040591.642	0.001

AI 기술개발 프레임	2018-G10	의원 관광 축제	호텔	203	1173241.804	0.000
			병원	195	1043436.100	0.000
			연구원	140	1223589.151	0.000
			편의점	120	500519.023	0.000
			부동산	108	508252.063	0.000
	2017-G4	정도 대비	의원	73	593049.226	0.000
			관광	50	76400.376	0.000
			축제	30	147134.054	0.000
			광주	20	144946.448	0.000
			정도	144	447653.251	0.001
2018-G2	서비스 활용 제공 적용	대비	134	351444.183	0.001	
		최대	110	277899.606	0.001	
		달러	102	467458.666	0.001	
		매출	79	149436.519	0.001	
		서비스	1776	20573838.890	0.003	
		활용	1178	10634639.063	0.002	
		제공	1102	9923245.422	0.002	
인간과의 대조 프레임	2018-G8	공장, LG 디스플레이	적용	1004	9274023.564	0.002
			데이터	996	8179459.279	0.002
			정보	871	7561285.445	0.002
			공장	152	574797.768	0.000
			LG 디스플레이	42	78859.076	0.000
AI 발달 지역 프레임	2018-G6	자율, 공항	신형	41	104747.516	0.000
			자율	87	310942.919	0.000
			공항	63	142670.109	0.000
			입주	46	209136.595	0.000
			항공	39	81030.714	0.000

2018년 군집4 및 군집10의 '호텔(203)', '병원(195)', '의원(73)', '부동산(108)', '관광(50)', '축제(30)' 등 프레임을 통해서 인공지능은 인간 삶의 의식주 곳곳에서 스며들고 있다는 사실을 알 수 있다. 3기의 AI 기술개발 프레임은 2017년의 군집4 및 2018년의 군집2로 구성되었다. 2017년의 군집4의 '최대(110)', '달러(102)' 등 프레임을 보면 이 시기의 인공지능 기술에 대한 보도는 기술개발 자급에 초점을 맞추고 있다는 점을 알 수 있다. 이와 동시에 2018년 군집2의 '서비스(1776)' 및 '플랫폼(859)'은 인공지능 기술이 플랫폼 서비스 등 상업 용도에서 크게 발전되었다는 것을 알려주었다. 2018년 군집8에서는 '공장(152)'이 가장 높은 값을 차지했으나 이는 'LG디스플레이(42)' 및 '롯데제과(35)' 등 국내 회사의 프레임이 연결되어 AI와 인간이 제조업 등 실업 영역에서의 협력 혹은 대립적인 입장을

나타났다. 마지막의 2018년 군집6은 '공항(63)' 및 '남북(31)' 두 개의 프레임으로 이시기 뉴스보도의 “AI 발달 지역” 프레임이 주로 남북 간의 비교 관계를 나타낸 것을 알 수 있다[표 8].

4. 뉴스와 기술특허 프레임의 의 상관관계(correlation):

뉴스-기술특허 QAP 상관관계

뉴스와 기술특허의 상관관계는 사회문화적으로 1기(2010년~2013년)는 2기(2014년~2016년)와 3기(2016년~2018년)에 비교하여 낮은 상관관계를 보이는 것으로 나타났다[표 2]. 아래 [그림 6]에 제시된 바와 같이, 2010년~2013년의 뉴스와 기술특허의 QAP 상관관계를 분석해보면 2010년 뉴스 대 2011년 기술특허의 상관관계는 0.360이며 2011년 뉴스-2012년 기술특허의 상관관계는 0.305로 소폭 줄어든 것을 알 수 있으나 2012년 뉴스-2013년 기술특허의 상관관계를 살펴보면 뉴스가 기술특허에 대한 영향력이 다시 0.341까지 소폭 증가했다. 이 시기를 전체적으로 보면, 전년도도의 뉴스 보도가 다음 연도의 기술특허에 대한 영향이 0.3~0.4의 범위에서 유지하고 있으며 부동이 작은 것으로 알 수 있다.

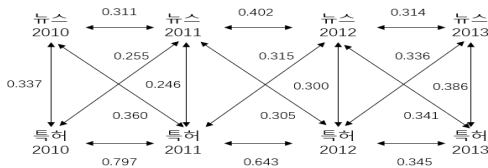


그림 6. 2010-2013년 뉴스-특허 QAP 상관관계

한편, 전년도도의 특허가 다음 년도의 뉴스에 대한 영향력이 0.255에서 0.336까지 증가했다. 비교해보면 기술특허가 뉴스에 대한 영향력보다 빠른 속도로 증가하고 있다는 것을 파악할 수 있다. [그림 7]에 제시된 바와 같이, 2013년~2016년을 보면, 전 시기보다 특허가 뉴스에 대한 영향 값이 0.5이상까지 증가했으나, 뉴스가 특허에 대한 상관관계는 0.3~0.4 범위에 유지하다가 0.710까지 급속 증가한 것을 볼 수가 있다. 동시에 같은 년도의 뉴스와 특허 간의 상관관계도 전 시기보다 꾸준히 증가한 것으로 나타났다.

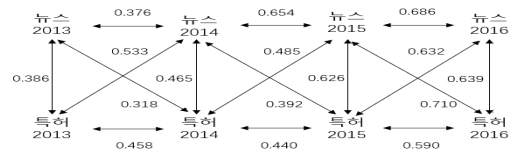


그림 7. 2013년-2016년 뉴스-특허 QAP 상관관계

[그림 8]에 제시된 바와 같이, 이 시기의 그래프를 보면 2017년도의 뉴스와 특허 간의 상관관계 및 2016 특허-2017뉴스의 상관관계 값이 모두 전보다 급속히 감소한 것을 보이며 이는 다음 년도인 2018년에는 다시 0.5이상의 수준으로 회복한 것을 알 수가 있다. 이와 반대로 뉴스가 이듬해 특허에 대한 상관관계 값은 선 상승 후 하락의 변동 추세로 보인다.

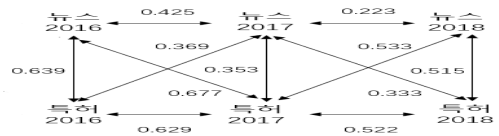


그림 8. 2016-2018년 뉴스-특허 QAP 상관관계

V. 결론 및 함의

1. 논문 요약

본 연구는 인공지능(AI) 진화에 따른 기술적 발달 경로와 이와 관련한 뉴스 보도의 경향을 1) 토픽모형(topic model)과 2) 시기별 프레임(frame) 분석을 통하여 인공지능 발달 경향을 분석하였으며, 두 영역 간의 프레임의 특성을 통하여 인공지능 관련 특허가 초기의 기술적인 차원에서 산업적인 차원으로 진화하고 있는 것을 실증적으로 검증하였다. 더불어 3) 프레임과 특허 간의 네트워크 상관관계 분석(QAP)을 통해 기수별 상관성의 특징이 1기(2010년~2013년)에는 기술 중심의 토픽, 2기(2013년~2016년)에서는 AI의 사회 문화적인 높은 상관관계가 있음을 확인하였다. 3기(2016년~2018년)에서는 기술보다는 AI 이용에 대한 프레임이 우세한 영역으로 나타났다.

2. 연구 함의

본 연구를 통해 기술과 사회문화 간의 구성주의적인 상관관계가 AI 발달에 중요한 역할을 하는 것으로 나타났다. 전체적으로 2010년에서 2018년에 이르는 AI 뉴스토픽과 AI 특허 토픽뉴스는 증가하는 뉴스 프레임으로 나타났으며 12개의 토픽이 상호 연관성이 있는 그룹으로 매칭을 시켜 정리하였다. AI 특허는 대체적으로 기술 중심의 토픽으로 구성되고 있으며, AI 뉴스보도는 AI를 통한 사회문화적인 연결 토픽으로 구성된다고 평가된다. AI 특허와 뉴스 프레임은 시스템, 데이터를 중심으로 하는 인공지능 기술에 대한 개발을 중심으로 하고 있으며, AI를 이용한 사회문화적인 발생 뉴스는 미래와 이를 이용하는 과학기술계의 정보 활용방안에 대한 프레임으로 정리할 수 있다. AI 특허와 AI 뉴스에 나타난 토픽의 특성은 AI 특허는 주로 시스템 데이터 등인 반면, AI 뉴스는 시장과 엔터테인먼트와 연결되는 첫째는 산업적 토픽프레임으로, 인공지능을 새로운 산업사회로 가는 기회이자 새로운 사회구성을 위한 도구로 보는 프레임이다. 지능에 기반한 4차 산업사회는 새로운 경제적 부가가치를 창출하고 미래 사회를 위한 성장 동력으로 작동하리라는 시각이다. 이러한 뉴스 프레임은 4차 산업사회의 방향성과 새로운 플랫폼 창조를 위해 인공지능과 4차 산업 관련 분야에 정부, 교육, 연구의 상호 연관성을 강조한다. 두 번째는 AI는 개인과 국가의 생존기술의 프레임으로, 인공지능 시대를 살아갈 로봇, 디지털 전환과 인공지능 상호작용은 우리가 살아갈 미래로 프레임하고 있다. 인공지능 시스템으로 개인화된 콘텐츠 추천과 산업 시스템의 인공지능 채택은 인간을 대체할 생산시스템의 새로운 플랫폼의 등장으로 인간과 AI의 상호작용으로 인해 나타나는 커뮤니케이션, 문화콘텐츠의 특성을 이해하는 프레임이다. AI 시대 사회적 또는 개인적인 실용적인 지식과 교육 방향으로 나타나는 산업구조 변화와 교육과 학습, 교육체계의 변화 등에 관한 프레임이다. 세 번째는 인간과 인간, 그리고 인간과 기계 사이에 나타나는 새로운 방식의 커뮤니케이션 프레임이다. 생산과 문화에서 인간-인간, 인간-AI, 그리고 AI-AI 간의 생산해내는 문화, 산업적인 현상을 이해하는 프레임이다. 순수한 인간에 의한 생산, 소비보다는 다양한 형태의 소통과 생산, 소비현상을 이해하는 프레임이 등장하는 것이다.

토픽 1은 AI, 인공지능 관련 기술단어와 사회문화적인 단어의 지속적인 증가세이다. 4차 산업혁명과 더불어 IT를 기반으로 한 혁신은 지속적으로 확대될 것이라고 판단되는 시점에 국내 AI, 인공지능관련 연구동향 분석은 시의적절하다고 판단된다. 본 연구의 결과에서는 인공지능 관련 분야에서 지속적으로 관심을 두어야 할 토픽으로 로봇과 서비스가 도출되었다. 이 토픽들은 지속적으로 중요도가 높아지고 있다. 다음으로는 기업과 활용이고 그 다음으로는 적용과 제공 마지막으로 시장에서의 데이터 활용에 관한 것이다. 이는 AI 기술이 급속도로 발전하고 ICT 기술들과 상승작용을 하면서 미디어의 범주와 영향력을 더욱 확장하여 새로운 개념의 미디어로 진화시키고 있음을 알 수 있다. 또한 미디어의 생산과 소비의 주체가 '사람'과 '기계'로 확장되고, '사람과 사람', '기계와 기계', '사람과 기계'간의 소통을 매개하는 역할을 하는 미디어-(AI와 미디어의 결합) AI 기술이 인간의 지적 활동, 즉 시각, 언어, 감각, 이해, 학습, 추론 등의 능력을 구현·재현함으로써 미디어의 생산과 소비 과정에 인간을 돕고, 인간을 대체하고, 인간을 능가하며 참여 AI 기술과 빅데이터의 결합은 산업과 사회를 크게 바꾸는 원동력이 될 것으로 판단된다.

둘째, 기존의 많은 연구동향 분석에 있어서 활용되던 언어 네트워크분석기법의 한계점을 극복한 토픽모델링, 시계열회귀분석과 같은 연구기법을 적용하였고, 이는 연구동향을 더 깊이 있게 이해할 수 있게 한다. 특히, 시계열회귀분석을 활용한 토픽 분석은 연구동향분석에 통계적 유의성 검정을 추가할 수 있다는 점에서 더욱 의미가 있다.

이상으로 4차 산업혁명을 주도하는 기술들(AI, 빅데이터, IoT, 자율주행차, 3D프린팅 등)에서 공통 핵심요소는 '데이터'-양털 생산이 기계 엔진을 만나 잉여생산을 낳고, 인클로저 운동을 촉진하며, 1차 산업혁명의 불씨가 되었던, 데이터는 4차 산업혁명의 토대이며 근간을 형성할 것이다. 데이터는 그 자체로는 원자재의 역할을 하며 AI와 만나 정보가 되고, 지식이 되고, 지혜로 변환되어 가치를 창출할 것이며 파급력이 높은 '미디어'와 'AI' 기술이 결합된 'AI 미디어'는 우리의 생활, 가치관, 사회체계 그리고 국가의 역할 및 기능까지 변화시킬 것으로 예상된다.

참 고 문 헌

[1] 박자현, 송민, “토픽모델링을 활용한 국내 문헌정보학 연구동향 분석,” 정보관리학회지, 제30권, 제1호, pp.7-32, 2013.

[2] 윤혜진, 김창식, 광기영, “지역관광 연구 트렌드 분석: 텍스트마이닝 활용,” 한국관광레저학회학술발표대회, 제12권, pp.275-282, 2017.

[3] 김창식, 최수정, 광기영, “토픽모델링과 시계열회귀분석을 활용한 정보시스템분야 연구동향 분석,” 한국디지털콘텐츠학회논문지, 제18권, 제6호, pp.1143-1150, 2017.

[4] 백영민, 최문호, 장지연, “한미 정권교체에 따른 주한 미 대사관 외교문서의 주제와 감정표현 변화,” 서울대학교 언론정보연구소, 제51권, 제1호, pp.133-179, 2014.

[5] 윤소연, 윤동근, “토픽모델링을 이용한 재난 및 안전관리 동향분석,” 대한공간정보학회, 제25권, 제3호, pp.75-85, 2017.

[6] 진설아, 허고은, 정유경, 송민, “트위터데이터를 이용한 네트워크 기반 토픽 변화 추적 연구,” 정보관리학회, 제30권, 제1호, pp.285-302, 2013.

[7] 박윤중, *프레임의 진화 : 토픽모델링과 연결망분석을 활용한 중북 프레임의 진화과정연구*, 연세대학교 대학원, 석사학위논문, 2015.

[8] 이태준, 이승배, 오창동, “원자력 이슈에 대한 정부와 언론의 커뮤니케이션 전략 비교 연구: 토픽모델링,” 한국지역언론학회, 제17권, 제3호, pp.172-229, 2017.

[9] 정다영, “빅데이터 시대의 개인정보 자기결정권,” 경북대학교IT와법연구소, 제14권, pp.151-209, 2017.

[10] 문안나, “한국 사회의 IR(Investor Relations) 개념 : 1994~2014년 신문기사 내용분석 및 의미연결망 분석,” 홍보학연구, 제20권, 제1호, pp.50-78, 2016.

[11] 권상희, “디지털 미디어 발달 모형 검증,” 한국디지털콘텐츠학회논문지, 제18권, 제2호, pp.423-428, 2005.

[12] 양승목, “주류 언론학의 제도화와 패러다임 변화,” 언론과정보, 제5권, 제1호, pp.67-103, 1999.

[13] McQuail, D, *Mass communication theory* (5th ed.), 양승찬, 이강형 (역) (2008), <매스커뮤니케이션 이론>, 파주: 나남, 2005.

[14] T. Gitlin, *The whole world is watching*, LA: University of California Press, 1980.

[15] P. Shoemaker, *Gatekeeping*. Thousand Oaks, CA: Sage, 1991.

[16] M. E. McCombs, *Setting the Agenda: The mass media and public agenda*, Malden, MA: Blackwell, 2004.

저 자 소개

권 상 희(Sang-Hee Kweon)

정회원



- 1985년 2월 : 서울대학교 언론정보학과
 - 1997년 8월 : 캘리포니아주립대학교 R-TV-Film (석사)
 - 2002년 8월 : 남일리노이대학교 Mass Com & Media Arts 학과 (박사)
 - 2003년 3월 ~ 현재 : 성균관대학교 미디어커뮤니케이션학과 교수
- <관심분야> : 소셜미디어, 사이버커뮤니케이션, AI Big Data

차 현 주(Hyeon-Ju Cha)

정회원



- 1988년 2월 : 중앙대학교 교육학과
 - 2010년 8월 : 성균관대학교 예대학 석사
 - 2020년 2월 : 성균관대학교 사회과학대학 철학박사
 - 2007년 3월 ~ 현재 : 글로벌인재개발교육연구소장
- <관심분야> : 심리측정평가, AI Big Data, 인성 및 생애교육, 리더쉽과 진로교육, 미디어문화콘텐츠