

4차 산업혁명 선언 전후의 인공지능 연구 동향: 키워드 네트워크 분석

박현성, 최승일
 공주대학교 산업시스템공학과
 e-mail:sichoi@kongju.ac.kr

Trends in artificial intelligence research before and after the declaration of the Fourth Industrial Revolution: a keyword network analysis

Hyeon-Seong Park, Seungil Choi
 Dept. of Industrial & Systems Engineering, Kongju National University

요약

본 논문에서는 4차 산업혁명 선언 이전과 이후의 인공지능 연구 동향에 대해 학위 논문은 어떤 차이점이 있는지 알아보기 위해 국내 석/박사 논문 데이터를 수집하여 빈도 분석과 키워드 네트워크 분석을 해보았다. 그 결과 2016년 학위 논문 수 대비 2017년은 약 4배 증가하였고, 매년 약 1.4 ~ 2배의 성장세를 보임을 확인할 수 있었다. 선언 전에는 기술, 적용 연구가 주를 이루었으며 서로 고립되어 연구가 활발하지 못함을 확인하였고, 선언 후에는 융복합, 교육 연구가 주를 이루며 서로 연결되어 연구가 활발함을 확인할 수 있었다.

1. 서론

4차 산업혁명은 Klaus Schwab이 2016년 1월 스위스 다보스에서 개최한 세계경제포럼에서 “제4차 산업혁명의 이해”를 주제로 선언하면서 그 개념이 전 세계에 널리 확산되었다 [1].

이후 우리나라 정책 분야에서 빠르게 수용되었고, 중앙 부처에서 4차 산업혁명에 대한 대응책을 마련하고 2017년 대선에서 후보들이 관련 공약을 발표했다. 또한, 알파고와 이세돌의 대국을 통해 인공지능과 디지털 기술 발전이 각인되어 4차 산업혁명은 무엇인지 정확히 모르더라도 기술에 변화가 있음을 인지하게 되었다[2].

인구 고령화 및 저출산으로 인해 노동 인력 부족과 장비 필요가 증가하며, 완전 무인자동화, 자율주행 차량 및 선박 등의 기술 필요성이 부각되고 있다. 이에 4차 산업혁명 또는 인더스트리4.0의 핵심 요소 기술인 인공지능, 로봇, 빅데이터, 클라우드, 사물인터넷, 5/6G, 시뮬레이션, 적층 제조 등을 활용하여 혁신적인 산업 발전이 이루어져 이러한 기술을 통해 노동력 부족에 대응하는 방안이 필요하다[3].

인공지능 분야 연구 동향 분석은 많이 이루어지고 있는데 [4], 본 논문은 국내 석사 및 박사 학위 논문 데이터로 인공지능 연구 동향을 4차 산업혁명 선언 이전과 이후로 구분하여

빈도 분석과 키워드 네트워크 분석으로 확인하고자 한다.

2. 데이터 수집

데이터의 수집은 학술연구정보서비스에서 수행하였다. 저자 주제어에 인공지능을 포함한 논문으로, 한국어로 작성된 2010년 ~ 2023년 국내 석/박사 논문을 대상으로 하였다. 이에 따라 2010년 02월 ~ 2023년 02월 학위 논문을 확보하였고, 중복 논문을 제거하여 1,063건의 논문과 6,117개의 저자 주제어를 수집하였다.

[표 1]을 확인하면 선언 직후인 2016년 대비 4차 산업혁명 개념이 확산된 2017년은 약 4배의 성장세를 보이며 매년 1.4 ~ 2배의 성장세를 확인할 수 있다. 2023년은 후기 학위 논문이 포함되어 있지 않다.

[표 1] 4차 산업혁명 선언 전후 연도별 논문 수 비교

연도	2010	2011	2012	2013	2014	2015	2016	합계
	논문	1	6	3	8	4	2	6
연도	2017	2018	2019	2020	2021	2022	2023	합계
	논문	25	48	76	102	209	284	289

3. 데이터 전처리

키워드 분석에서는 서로 같은 단어라도 대/소문자, 단/복수

등의 차이로 서로 다른 키워드로 인식하기 때문에 데이터 전처리는 매우 중요한 과정이며 분석의 정확도 및 정밀도를 높이기 위해 정제하려는 키워드 분야의 지식을 일정 수준 이상 필요로 한다. 또한, 분석의 효율을 높이기 위한 과정이며 데이터 특성에 따라 전처리 과정을 생략해도 된다.

학술연구정보서비스에서는 DB를 상세내용 포함하여 엑셀 형식인 xls 확장자로 다운로드가 가능하며 해당 데이터를 엑셀 환경에서 오타 교정, 불필요한 띄어쓰기 제거, 특수문자 제거, 불용어 제거, 대/소문자 통일, 약어 통일, 국문, 한문, 영문을 국문으로 통일, 동의어, 유의어 통일, 복합어 분리 과정으로 키워드를 정제하였다.

4. 빈도 분석

정제 후 전체 키워드는 5,311개이고 4차 산업혁명 선언 이전인 2010년 ~ 2016년의 키워드는 134개이며 선언 이후인 2017년 ~ 2023년의 키워드는 5,177개이다. 선언 이전의 키워드는 주로 개념과 알고리즘에 관한 키워드가 출현하였으며 선언 이후의 키워드는 주로 알고리즘과 교육 및 융복합, 윤리, 리터러시에 관한 키워드가 출현하였다.

논문 데이터 검색에 사용한 주제어 인공지능을 제외하고 2번 이상 출현한 상위 10개의 키워드를 내림차순으로 나열하여 확인했을 때 선언 이전은 5개의 키워드를 제외하면 모두 빈도수가 한 번만 출현하였다.

[표 2] 2번 이상 출현한 키워드 상위 10개

4차 산업혁명 선언 이전		4차 산업혁명 선언 이후	
키워드	빈도	키워드	빈도
기계학습	3	심층학습	118
게임 인공지능	2	인공지능 교육	113
퍼지이론		기계학습	60
하이브리드 인공지능		인공지능 융합교육	56
감성조명		4차 산업혁명	42
-	-	초등학교	42
-	-	합성곱신경망	40
-	-	빅데이터	39
-	-	설명가능한 인공지능	39
-	-	인공지능 윤리	31

4차 산업혁명 선언 이후인 2017년 ~ 2023년의 키워드를 대상으로 1년 단위로 그룹을 나누고 주제어 인공지능을 제외하고 2번 이상 출현한 상위 3개의 키워드를 확인하여 [표 3]의 결과를 확인했다. 기술은 인공지능 개념 하위 분야인 기계학습(Machine learning)과 하위 분야인 심층학습(Deep learning)이 출현되고 그 외 인공지능을 활용한 교육과 인공지능 기술에 대한 교육이 출현하였다.

심층학습은 인공지능을 획기적으로 진화하게 만든 기술 중 하나이다. 인간의 뇌 구조에서 영감을 받아 여러 계층의 인공

신경망 즉, 많은 은닉층(Hidden layer)을 사용하는 기술이다. 기계학습은 인간이 개입하여 학습시켜야 하며 비정형 데이터는 처리하지 못하는 한계를 심층학습은 스스로 학습하며 비정형 데이터를 처리할 수 있다.

[표 3] 4차 산업혁명 선언 이후 2번 이상 출현한 키워드 상위 3개

2017년		2018년	
키워드	빈도	키워드	빈도
심층학습	3	4차 산업혁명	8
기계학습	2	기계학습	7
4차 산업혁명			
합성곱신경망			
빅데이터			
특징선택	-	지능정보사회	4
-	-	-	-
2019년		2020년	
키워드	빈도	키워드	빈도
심층학습	9	심층학습	10
인공지능교육	8	4차 산업혁명	8
기계학습	5	합성곱신경망	6
		빅데이터	
		인공지능스피커	
-	-	-	-
2021년		2022년	
키워드	빈도	키워드	빈도
심층학습	26	심층학습	38
인공지능교육	18	인공지능교육	27
기계학습	16	기계학습	14
-	-	-	-
2023년		-	
키워드	빈도	키워드	빈도
인공지능교육	62	-	-
인공지능융합교육	46	-	-
심층학습	30	-	-

5. 키워드 네트워크 분석

네트워크 분석의 장점은 주요 주제가 식별 가능하며 의미론적 관계와 시각적 표현 및 영향력 있는 키워드가 식별 가능하고, 단점은 데이터의 전처리 문제와 모호성 및 문맥 의존성과 제한된 범위 및 키워드 선택의 주관성이 있다.

네트워크는 노드와 노드간의 관계 또는 연결(Link)로 구성되어 있다. 분석의 중요한 개념 중 하나인 중심성 분석은 이러한 네트워크 안에서 중요한 역할을 하는 노드는 무엇인지를 파악하는 통계적 기술로 다양한 형태의 중심성 측정 지표를 사용한다. 본 논문에서는 연결 중심성 지표를 활용하였다.

[표 4] 대표적인 중심성 측정 지표 특징

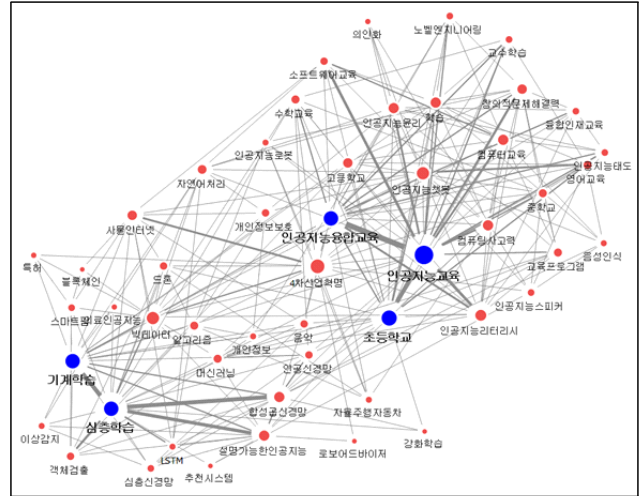
중심성 지표	특징
연결 중심성	값이 클수록 노드는 다른 노드와 많은 관계를 맺고 있어 중요한 역할을 한다
근접 중심성	값이 클수록 노드는 다른 노드에 빠르게 접근할 수 있어 정보 전파나 영향력 확산에 중요하다
매개 중심성	값이 클수록 노드는 정보 흐름이나 리소스 전달에 중요한 역할을 한다

네트워크는 2가지 유형이 있다. 첫째, 노드와 노드가 바로

연결되는 1-Mode Network. 둘째, 노드와 노드가 바로 연결되는게 아닌 중간에 서브노드가 존재하는 2-Mode Network가 있다. 본 논문에서 논문과 저자 주제어를 분석하기 위한 데이터는 2-Mode Network 구조로써 저자 주제어 노드가 바로 연결되지 않고 논문 제목의 서브노드를 통해 연결되며 링크는 저자 주제어가 몇 개의 논문과 연결되어 있는지의 수를 의미한다. 여러 네트워크 분석 툴 중 (주)사이람 NetMiner4를 채택하여 사용하였고 저자 주제어 검색에 사용한 키워드인 인공지능 키워드 제거와 8회 이상 출현한 키워드만 추출하였다. 또한 2-Mode Network를 변환하여 1-Mode Network에서 연결 중심성 지표를 확인하였다. 4차 산업혁명 선언 이후의 그룹에서 연결 중심성 값은 인공지능교육 0.58로 가장 높았고 초등학교 0.46 기계학습 0.44로 뒤를 이었다.

[표 5] 4차 산업혁명 선언 이후 연결 중심성 상위 5개

키워드	연결 중심성
인공지능교육	0.580
초등학교	0.460
기계학습	0.440
인공지능융합교육	0.420
심층학습	0.400
4차산업혁명	0.400



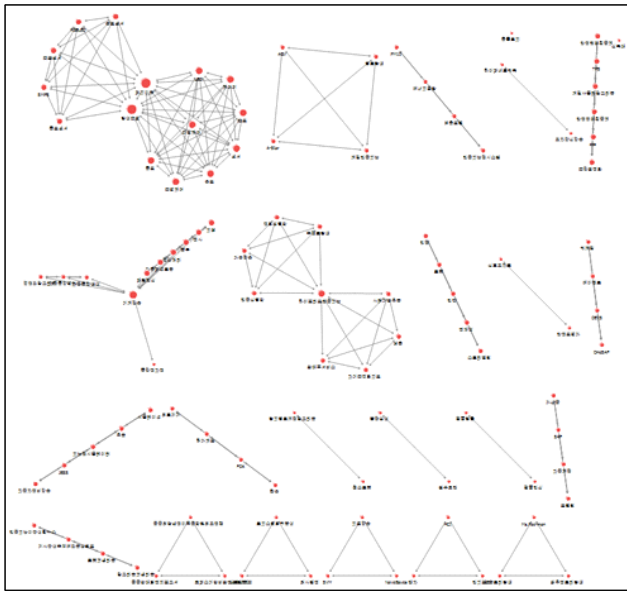
[그림 2] 4차 산업혁명 선언 이후 네트워크 분석

6. 결론

자본력과 기술을 바탕으로 인공지능을 선도하는 국가를 빠르게 추격하기 위해서는 연구 동향에 대한 파악과 비교가 필요하다. 본 연구에서는 4차 산업혁명 선언 이전과 이후의 국내 인공지능 연구 동향을 분석하였다. 결과적으로 선언 이후 국내 학위 논문 수가 상당히 증가하며 매년 성장세를 보이고 있다. 선언 전에는 기술 및 적용 연구가 주를 이루었으며 네트워크가 서로 고립된 연구 활동이었지만, 선언 이후에는 융복합 및 교육 그리고 심층학습, 기계학습이 주를 이루며 네트워크가 활발한 연구 활동을 확인했다. 이러한 결과는 국내 인공지능 연구의 성장과 다양화를 나타내며 4차 산업혁명 선언이 인공지능 연구 동향에 미친 영향을 보여준다. 향후 국내외 더 많은 데이터에 대한 심도 있는 연구가 이루어져, 정책 결정자 및 관련 연구자에게 유용한 정보가 되기를 기대한다.

참고문헌

- [1] Klaus Schwab, The Fourth Industrial Revolution, Crown, 2017.
- [2] 김석관, “산업혁명을 어떤 기준으로 판단할 것인가?”, 과학기술정책, 제1권 1호, pp. 113-141, 12월, 2018년.
- [3] 이수현, 김재운, “Industry 4.0 시대의 스마트 항만 효율성 평가”, 산업경제연구, 제35권 5호, pp. 1,075-1,096, 10월, 2022년.
- [4] 정명석, 박성현, 채병훈, 이주연, “논문데이터 분석을 통한 인공지능 분야 주요 연구 동향 분석” 디지털융복합연구, 제15권 5호, pp. 225-233, 5월, 2017년.



[그림 1] 4차 산업혁명 선언 이전 네트워크 분석