J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr

LDA 토픽모델링을 활용한 인공지능 관련 국가R&D 연구동향 분석☆

A Study on Analysis of national R&D research trends for Artificial Intelligence using LDA topic modeling

양 명 석[1] 이 성 희[2] 박 근 희[2] 최 광 남[2] 김 태 현[2*]
MyungSeok Yang SungHee Lee KeunHee Park KwangNam Choi TaeHyun Kim

요 약

특정 주제분야에 대한 연구동향 분석은 대부분 논문, 특허 등 문헌정보를 대상으로 한 키워드 추출을 통해 토픽모델링 기법을 적용하여 주요 연구주제와 연도별 추이 등을 살펴보는 방식을 활용하고 있다. 본 논문에서는 국가과학기술지식정보서비스(NTIS)에서 제공하는 인공지능 관련 국가연구개발사업 과제정보를 대상으로 LDA(Latent Dirichlet Allocation) 토픽모델링 기법을 활용하여 연구주제와 관련된 토픽들을 추출·분석하여 국가연구개발사업에 대한 연구주제와 투자방향에 대하여 분석하고자 한다. NTIS는 국가연구개발사업·과제정보를 비롯하여, 논문, 특허, 보고서 등 연구를 통해 생성된 주요 연구개발성과에 이르기까지 방대한 양의 국가R&D 정보를 제공하고 있다. 본 논문에서는 NTIS 통합검색에서 인공지능 키워드와 관련된 분류 검색을 수행하여 검색결과를 확인하고, 최근 3개년 과제정보를 다운로드 받아 기초데이터를 구축하였다. 파이썬에서 제공하는 LDA 토픽모델링 라이브러리를 활용하여 기초데이터 (연구목표, 연구내용, 기대효과, 키워드 등)를 대상으로 관련 토픽과 주제어를 추출하고 분석하여 연구투자방향에 대한 인사이트를 도출하였다.

☞ 주제어 : 토픽모델링, 인공지능, 국가연구개발사업, 연구주제변화, NTIS

ABSTRACT

Analysis of research trends in specific subject areas is performed by examining related topics and subject changes by using topic modeling techniques through keyword extraction for most of the literature information (paper, patents, etc.). Unlike existing research methods, this paper extracts topics related to the research topic using the LDA topic modeling technique for the project information of national R&D projects provided by the National Science and Technology Knowledge Information Service (NTIS) in the field of artificial intelligence. By analyzing these topics, this study aims to analyze research topics and investment directions for national R&D projects. NTIS provides a vast amount of national R&D information, from information on tasks carried out through national R&D projects to research results (thesis, patents, etc.) generated through research. In this paper, the search results were confirmed by performing artificial intelligence keywords and related classification searches in NTIS integrated search, and basic data was constructed by downloading the latest three-year project information. Using the LDA topic modeling library provided by Python, related topics and keywords were extracted and analyzed for basic data (research goals, research content, expected effects, keywords, etc.) to derive insights on the direction of research investment.

☞ keyword : topic modeling, Artificial Intelligence, National Research and Development Program, Research Trend, NTIS

1. 서 론

4차 산업혁명, COVID 19 등 팬데믹 상황에서 정보통신(IT)분야에 대한 연구와 투자가 지속적으로 증대하고 있다. 특히, 인공지능에 관한 연구는 기계학습, 딥러닝, 사물인터넷, 로봇, 빅데이터 등의 기술과 융합하여 다양한 주제로 발전하며 사회 전반에 걸쳐 많은 영향을 주고 있다.[1]

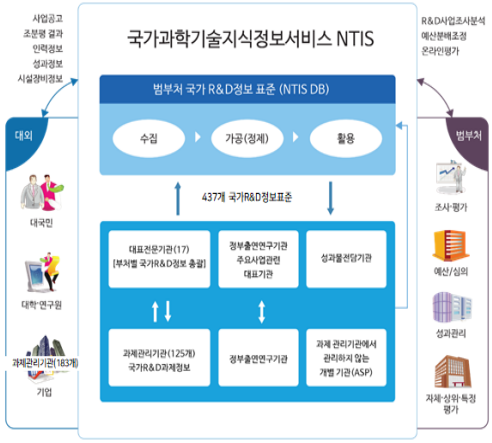
1 Dept. of Data-Centric Problem Solving Research, KISTI, Dae-jeon, 34141, Korea
2 Div. NTIS, KISTI, Dae-jeon, 34141, Korea
* Corresponding author (heemang@kisti.re.kr)
[Received 16 March 2021, Reviewed 17 March 2021(R2 9 August 2021), Accepted 21 August 2021]
☆ 본 논문은 2020년도 한국인터넷정보학회 추계학술발표대회 우수논문 추천에 따라 확장 및 수정된 논문임, 연구는 2021년도 한국과학기술정보연구원(KISTI) 기본사업 과제로 수행한 것입니다.

우리나라에서도 국가차원에서 전략적으로 인공지능 기술을 개발하고자 국가연구개발사업을 통한 많은 투자를 진행하고 있다.

특정 주제분야에 대한 연구동향 분석은 대부분 논문, 특허 등의 문헌정보를 대상으로 한 키워드 추출을 통해 토픽모델링 기법을 적용하여 관련주제 및 주제변화 등을 살펴보는 방향으로 수행되고 있다.

본 논문에서는 국가과학기술지식정보서비스(이하 NTIS, National Science & Technology Information Service)에서 제공하는 국가연구개발사업에 대한 과제정보를 대상으로 LDA 토픽모델링 기법을 적용하여 연구주제와 관련된 토픽들을 추출하고, 이러한 토픽을 활용하여 인공지능이라는 특정 분야와 관련한 국가연구개발사업에 대한 연구주제와 투자방향에 대하여 분석하고자 한다.

NTIS에서는 18개 부처·청으로부터 국가R&D사업에 대한 과제정보를 비롯하여 논문, 특허 등의 연구개발성과 정보를 수집하여 다양한 서비스를 제공하고 있다. 국가R&D정보뿐만 아니라 다양한 과학기술정보 서비스와 연계하여 약 1억 5천만 건의 과학기술관련 지식정보 콘텐츠를 구축하여 서비스하고 있다 [2].



(그림 1) NTIS 개념도
(Figure 1) NTIS concept

본 논문에서는 NTIS 통합검색에서 인공지능 키워드와 관련한 분류 검색을 수행하여 검색결과를 확인하고, 그중 최근 3개년 간 수행된 과제정보를 다운로드 받아 기초데이터를 구축하였다. 파이썬에서 제공하는 LDA 토픽모델링 라이브러리를 활용하여 기초데이터 (연구목표, 연구내용, 기대효과, 키워드 등)를 대상으로 관련 토픽과 주제어를 추출하고, 그 결과를 분석하여 연구투자방향에 대한 인사이트를 도출하고자 한다.

2. 관련 연구

2.1 국가연구개발정보

2011년 범부처 국가연구개발정보의 체계적인 수집·연계 및 공동활용 기반구축 등을 위해 중앙행정기관 대표 전문기관 정보관리시스템과 NTIS간 상호연계를 추진하는 「국가연구개발사업의 관리 등에 관한 규정」이 제정되었다. NTIS는 이 규정의 제25조 제3항에 따라 표1과 같이 국가연구개발정보표준을 제정하고 이에 따라 국가연구개발정보를 수집하여 서비스해왔다. 현재는「국가연구개발혁신법 [시행 2021. 2. 1.]」에 따라 「국가연구개발정보표준([과학기술정보통신부고시 제2021-6호, 2021. 1. 25., 폐지]」)은 폐지되고, 「국가연구개발혁신법」에 따른 연구개발정보의 수집·생산·관리 및 활용 등 연구개발정보의 처리에 관한 '국가연구개발정보처리기준'이 시행(2021.1.1.)된 이후, 이에 의거하여 국가연구개발정보를 구축하여 제공하고 있다.

NTIS에서는 공공데이터에 대한 개방요구정책에 맞춰 국가연구개발정보에 대한 개방대상과 항목, 이용범위 등을 명확히 하고 개방서비스를 대폭 확대하여 제공하고 있다. 누구나 쉽게 국가연구개발정보를 접근하고 활용할 수 있도록 2017년부터 검색결과를 직접 다운로드 받을 수 있도록 개선하였으며, 정보활용도를 높이기 위해 2019년 바로분석, 2020년 과제정보 시각화 등의 다양한 서비스를 지속적으로 개발하여 제공하고 있다.

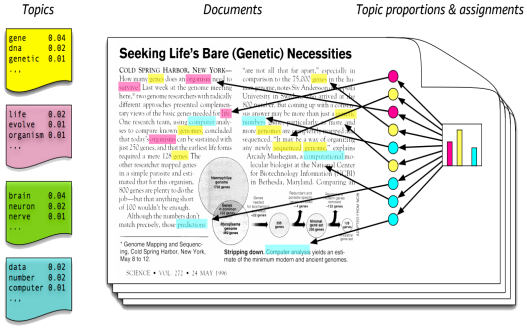
(표 1) NTIS 데이터 제공 항목(2020년 기준)
(TABLE 1) National R&D Information Standard

구분	과제	성과	인력	평가위원	시설장비	합계
항목수	164	168	39	11	55	437
개방 항목수	130	165	12	-	48	355
수집 관리	전문기관 연계 또는 연구자 입력(296개), 연구개발성과 연계(133개), NTIS자체관리(8개)					

2.2 토픽모델링

토픽모델링은 텍스트 마이닝 기법 중 하나로 비구조화된 문서 집합에서 잠재된 토픽들을 추출하는 확률적 모

델 알고리즘이다. 토픽모델링은 잠재 의미분석(LSA: Latent Semantic Analysis)을 시초로, 확률기반 잠재의미분석(pLSA:Probabilistic LSA)기법으로 발전되다, Blei, Ng and Jordan(2003)가 고안한 그림2의 LDA(Latent Dirichlet Allocation)기법을 토픽모델링에 많이 활용하고 있다. 최근에는 Teh et al.(2007)가 고안한 HDP(Hierarchical Dirichlet Process)기법이 토픽의 주제개수를 지정하지 않고 내부 알고리즘을 통해 결과를 추출해줘 관심이 집중되고 있지만 국내에서는 LDA 모델링을 주로 사용하고 있는 추세이다. [3]



(그림 2) LDA 토픽모델링의 이해 DaviD m. Blei 2012.04
(figure 2) Probabilistic topic models, DaviD m. Blei 2012.04

박준형 외 1 [4]의 연구에서는 연구동향 분석을 위한 토픽 모델링 기법 비교를 통해 LDA와 HDA를 비교하여 실험하였는데 LDA의 경우 해당분야의 거시적인 연구동향 파악에, HDA의 경우 세부적인 연구동향에 장점이 있는 것으로 파악되었다. 본 연구에서는 거시적인 관점에서 동향파악을 위해 LDA 토픽모델링을 활용하고자 한다.

박건철 외 1[5] 연구에서는 Scopus DB 및 Springer DB에서 스마트시티와 관련된 학술논문 11,527건의 제목과 초록, 발행연도 등의 정보를 수집하여 연구현황, 연구주제, 연구분야 추이 등을 LDA기반 토픽모델링 기법을 활용하여 분석하였다. 주제 간의 연관관계를 분석하여 향후 스마트 시티 관련 연구분야에 활용할 수 있도록 분석하였다.

남춘호 [6]의 연구에서는 전통적인 사료자료에 토픽모델링을 적용하여 농민일기 단행본에서 다루는 주제분야를 추출하여 핵심내용을 파악할 수 있도록 지원하는 연구를 수행하였다.

정진명 외 2 [7] 연구에서는 소셜 미디어 데이터를 대상으로 교육정책과 관련한 토픽모델링을 수행하여 여론에 대한 방향성을 주제분석하는데 활용하였다.

토픽모델링 기법은 방대한 양의 텍스트 데이터를 대상으로 주제분야에 대한 키워드를 분석하는데 많은 이점이 있다. 본 논문에서는 인공지능 분야에 대한 연구개발 투자방향과 주제 분야를 살펴보기 위하여 국가연구개발정보를 대상으로 LDA 기법을 적용하여 분석을 수행하고자 한다.

3. 연구방법

3.1 분석 대상 데이터

NTIS에서는 통합검색에서 이용자가 관심 있어 하는 국가연구개발정보에 대한 정보를 키워드, 분류 등을 활용해 검색하고, 그 결과를 다운로드 받아 활용할 수 있도록 하고 있다. 특히, 국가연구개발정보의 핵심인 과제정보에는 표2와 같이 사업명, 과제명, 연구목표, 연구내용, 기대효과, 키워드 등이 포함되어 있어 주제분야 분석에 유용하게 활용할 수 있다.

(표 2) 국가연구개발 과제정보 구성 항목
(TABLE 2) Items of National R&D Project Information

항목명	설명
부처명	연구개발사업의 기획, 평가 및 관리에 관한 제반사항을 주관하는 중앙행정기관의 명칭 예)과학기술정보통신부
사업명	사업명(세부사업코드의 사업명, 국가연구개발사업 조사·분석의 소분류 사업명)
과제고유번호	NTIS에서 발급하는 국가연구개발과제의 범부처 과제고유번호
과제명(국문)	신청 및 협약 과제를 기준으로 한 세부과제명의 정식명칭(국문과 영문과제명으로 구분)
과제명(영문)	
중략	
연구목표	개발하고자 하는 기술(공정 또는 제품 포함)의 수준·성능·품질 등 연구목표에 대한 요약
연구내용	연구내용에 대한 요약
기대효과	과제 수행 시 기대효과 요약
한글키워드	과제 내용을 대표하는 한글 키워드 5개 내외
영문키워드	과제 내용을 대표하는 영문 키워드 5개 내외

3.2 데이터 분석

데이터 분석 절차는 그림 3과 같다. 첫째, 데이터 수집

(그림 3) 데이터 분석 절차
(figure 3) Process of Data Analysis

단계에서는 NTIS에서 제공하고 있는 통합검색에서 키워드("인공지능", "Ariticial inteligence", "지능형" 등)와 분류검색에서 국가중점기술분류(인공지능, 빅데이터 분야)를 선택하여 검색하고, 그 결과를 엑셀파일로 다운로드 받아 최근 3개년(2018~2020년, 21년 1월기준)과제정보를 수집하였다. 과제고유번호를 활용하여 중복을 제거한 후 표3과 같이 분석대상 데이터를 구축하였다.

(표 3) 분석대상 데이터 건수(2018~2020)
(Table 3) Number of Analysis Data (2018-2020)

연도	2018	2019	2020	합계
과제수	5,267	6,550	8,942	20,759

둘째, 데이터 정제 단계에서는 분석대상 과제 데이터 20,759건에서 표 2와 같은 과제정보 항목 중 텍스트 분석과 토픽모델링을 수행하기 위해 필요한 항목(과제명, 연구목표, 연구내용, 기대효과, 키워드 등)에서 텍스트 데이터를 추출하고, 특수캐릭터, 결측치 정보 등을 제거하여 데이터를 구축하였다.

셋째, 토픽모델링 단계에서는 대상 데이터에 대한 워드클라우드 분석을 통해 국가연구개발과제에서 인공지능 분야 핵심 키워드가 무엇인지 살펴본 후, 정부부처별, 연도별 인공지능 분야 투자추이를 정량 분석하여 국가연구개발사업에 대한 투자 추이를 살펴보았다. 또한 인공지능 관련 주제분야를 파악하기 위해 토픽모델링을 수행하였다. 우선, NTIS에서 구축한 불용어 사전과 국가R&D 용어사전[8] 등을 활용하여 과제고유번호를 키값으로 하고, 텍스트 데이터를 대상으로 형태소 분석(mecab)을 수행[9]하여 관련 키워드 들을 추출하였다. 이를 바탕으로 파이썬(python)과 LDA 토픽모델링 라이브러리(gensim)를 활용[10]하여 토픽모델링을 수행하였다. LDA모델링의 특성상 분류하고자 하는 토픽(K)을 임의로 지정(10개)하여 1차적으로 수행하고, 토픽들의 분포 변화를 살펴보며 최적화된 주제분류를 파악하기 위해 모델링 최적화 작업을 수행하였다.

넷째, 모델링 최적화단계에서는 LDA 토픽모델링에 최적화된 주제 개수를 찾기 위해 주제 일관성(Topic Coherence)을 기준으로 주제를 찾아주는 모델링 최적화(mallet) 라이브러리를 활용하였으며, 주제일관성 계수가 높은 값을 선택하여 총 16개 토픽을 선정하였다.[11]

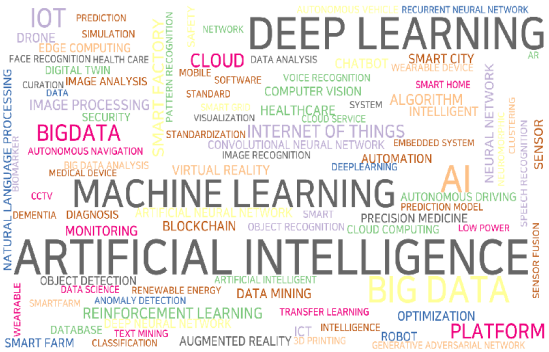
다섯째, 모델링 최적화를 통해 파악된 주제 개수(K=16)를 값으로 하여 토픽모델링을 다시 수행하여 그림 6과 같이 최종 결과를 얻은 후 토픽들을 중심으로 인공지능에 관한 연구주제 동향을 분석하였다.

4. 분석결과

4.1 핵심키워드

인공지능 분야 핵심키워드를 살펴보기 위해 과제정보 내의 한글키워드와 영문키워드 컬럼에 있는 단어를 중심으로 워드클라우드를 생성해 보았다. 한글키워드에서는 그림 4와 같이 "인공지능", "딥러닝", "빅데이터", "기계학습", "사물인터넷","머신러닝", "플랫폼", "클라우드", "자율주행", "지능형", "강화학습", "스마트팜", "음성인식", "가상현실" 등이 상위에 랭크됨을 알 수 있다. 또한 영문키워드의 경우, 그림 5와 같이 한글키워드와 비슷하게 "Artificial Intelligence", "Deep learning", "machine learning", "Bigdata", "IoT", "AI", "Cloud", "Platform",

(그림 4) 인공지능분야 워드클라우드(한글키워드)
(Figure 4) word cloud (korean Keyword)



(그림 5) 인공지능분야 워드클라우드(영문키워드)
(Figure 5) word cloud (english Keyword)

"Reinforcement learning", "Smart Farm" 등이 상위에 랭크됨을 알 수 있었다. 딥러닝, 머신러닝, 빅데이터, 플랫폼 등에 관한 기초 연구에서 스마트팜, 가상현실 등 응용분야에 까지 다양한 분야에 걸쳐 국가연구개발사업이 수행됨을 확인 할 수 있었다.

4.2 투자 동향

최근 3개년 간 인공지능 분야 국가연구개발사업 과제수와 연구비는 표 4와 같이 매년 지속적으로 꾸준히 증가됨을 알 수 있다. 특히 최근 20년 과제수와 연구비합계가 2018~2019년에 비해 크게 증가되었음을 확인할 수 있다.

그림 6과 같이 부처별 과제수 증감추이를 살펴보면 과학기술정통부, 교육부, 중소벤처기업부, 산업통상자원부 순으로 연도별로 과제가 지속적으로 증가함을 알 수 있다. 과제 수는 중소벤처기업부의 경우, 중소기업을 지원하는 성격이 강해 과제당 연구비의 규모는 작으나 많은 수의 과제를 공모하여 추진함에 따라 과제수가 많은 현상을 보인다. 또한 교육부의 경우에도 기초연구를 수행하는 대학이나 연구기관을 대상으로 과제를 공모하기 때문에 과제규모는 작지만 다양한 분야에서 많은 과제가 수행됨을 알 수 있다.

(표 4) 과제수와 연구비 분포((2018-2020)
(Table 4) Number of Prjoect & Total Cost

연도	과제수	연구비(합계, (단위:천원)
2018	5,267	2,220,183,479
2019	6,550	2,261,427,412
2020	8,942	3,189,833,683
합계	20,759	7,671,444,574

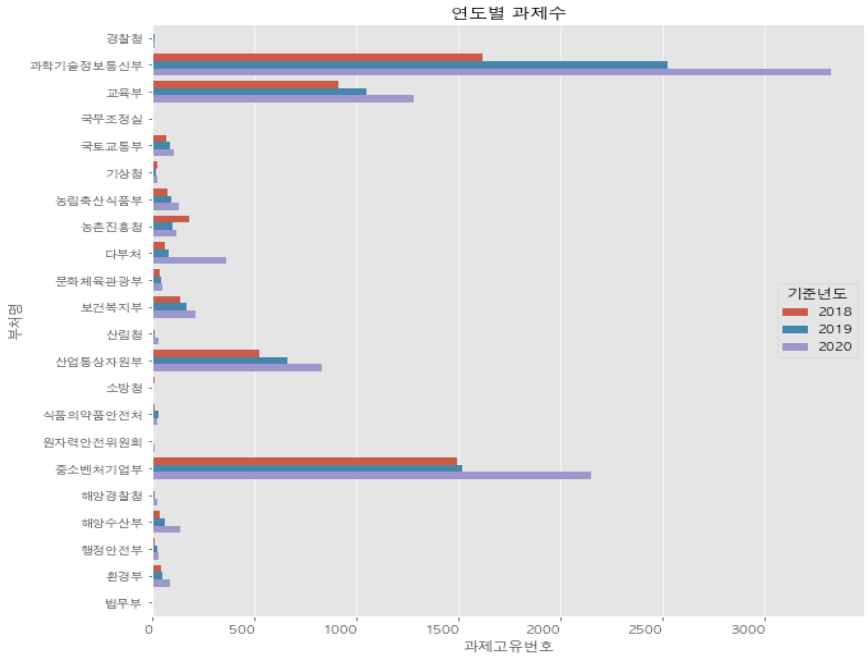
과제당 연구비의 분포는 부처별 연구비의 규모나 수준에 따라 달리 구성되므로 단순 과제수의 증가가 투자 추이를 반영한다고 보기는 어렵지만 과제수가 꾸준히 증가하는 것으로 보아 인공지능 분야에 대한 연구가 활발히 지속되고 있음을 알 수 있었다.

또한 그림 7과 같이 부처별 연구비 추이를 살펴보면 과학기술정보통신부, 산업통상자원부, 중소벤처기업부, 국토교통부 순으로 연구비가 많이 투입된 것으로 나타났고, 2018년도부터 2020년도에 이르기까지 각 부처에서 인공지능 분야 연구개발사업에 대한 투자를 지속적으로 확대해나가고 있음을 확인할 수 있다.

4.3 토픽모델링 결과 분석

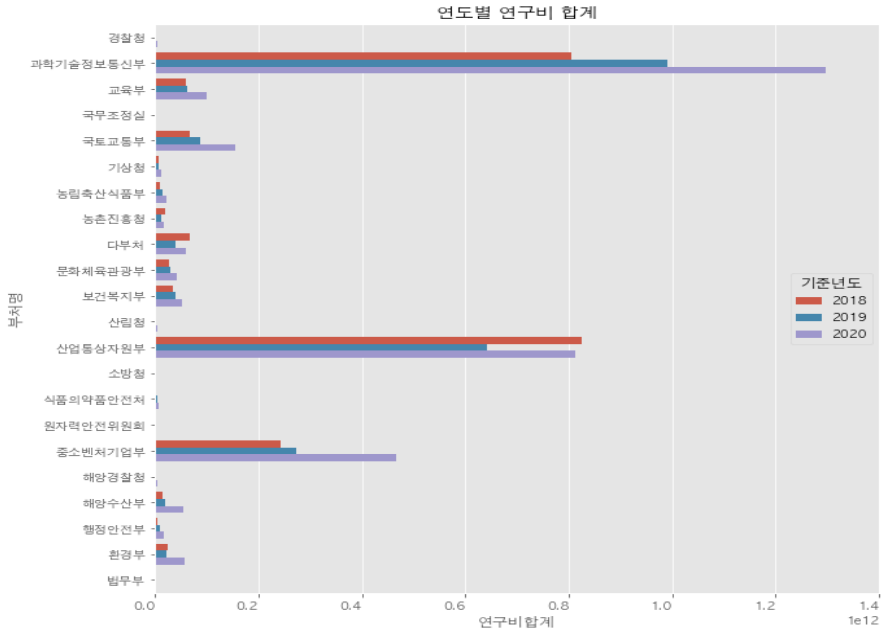
LDA 토픽모델링을 통해 파악한 주제영역은 표 5와 같이 총 16개 분야에 걸쳐 빅데이터, 영상인식, 음성인식, 인공지능기술 등 기초기술영역을 바탕으로 응용분야인 에너지, 교육, 소재, 의료, 해양선박, 서비스플랫폼, 시장제품, 재난안전, 스마트시티, 스마트 공장 등 다양한 산업분야로 확대되어 나가는 것을 확인할 수 있었다.

그림 8에서 보는 바와 같이 인공지능 분야에서는 독립된 연구주제 수행보다는 인공지능 분야의 핵심기술들을 바탕으로 다양한 주제영역에 걸쳐서 연구가 수행됨을 알 수 있었다.



(그림 6) 부처별 과제수 추이

(Figure 6) Annual number of projects according to government funding agencies



(그림 7) 부처별 연구비 추이(2018~2020)

(Figure 7) Annual total cost according to government funding agencies

(표 5) 토픽모델링 결과
(Table 5) result of Topic Modeling

주제분야	핵심키워드
에너지	에너지, 전력, 메모리, 설계, 연산, 구조, 회로, 하드웨어, 전기, 컴퓨팅
빅데이터	데이터, 분석, 수집, 빅데이터, 정보, 플랫폼, 모델, 서비스, 통합
영상인식	영상, 딥러닝, 인식, 이미지, 카메라, 알고리즘, 검출, 3차원, 자동
교육분야	교육, 소자, 융합, 사회, 핵심, 인력, 미래, 혁신, 창출
소재분야	소재, 물질, 규명, 분석, 세포, 구조, 특성, 유전체, 유전자, 약물
의료분야	진단, 환자, 의료, 치료, 임상, 분석, 질환, 평가, 측정
해양선박	위성, 선박, 금융, 해양, 챗봇, 문서, 수중, 문자, 상담, 서비스
정보 서비스	정보, 서비스, 개인, 분석, 사용자, 생활, 건강, 맞춤형, 수집, 공간
서비스 플랫폼	플랫폼, 서비스, 콘텐츠, 클라우드, 설계, 소프트웨어, 디지털, 모바일, 검증
시장제품	시장, 제품, 서비스, 기업, 고객, 상품, 추천, 해외, 디자인, 창의
음성인식	학습, 인식, 인간, 음성, 기계, 행동, 모델, 언어, 자동
재난안전	안전, 사고, 재난, 상황, 감지, 위험, 대응, 평가, 피해, 화재
스마트시티	정보, 운영, 환경, 스마트, 생산, 분석, 실증, 연계, 현장, 도시
로봇	로봇, 환경, 네트워크, 제어, 통신, 센서, 차량, 지능, 자율주행, 자율
기초기술	모델, 학습, 예측, 기법, 분석, 딥러닝, 데이터, 알고리즘, 기계
스마트 공장	설계, 센서, 제작, 모듈, 제품, 공정, 측정, 제어, 장치, 모니터링

(그림 8) 토픽 분포도
(Figure 8) Topic Distribution Plot

5. 결 론

본 연구에서는 앞서 살펴본 바와 같이, NTIS에서 제공하는 국가연구개발사업 과제정보를 대상으로 인공지능 분야의 최근 3개년 간 투자동향과 연구동향에 대하여 살펴보았다. 인공지능 분야의 경우에는 인공지능 관련 기초기술(딥러닝, 기계학습 등)에 관한 연구뿐만 아니라 인공지능 기술을 활용한 의료, 소재, 스마트시티, 로봇 등 다양한 영역에 걸쳐 연구가 폭넓게 진행되고 있음을 알 수 있었다. 또한 정부차원에서 인공지능 분야에 대한 기초기술연구를 비롯하여 산업 전반에 걸쳐 각 부처별 특성에 맞게 인공지능 기술을 활용한 다양한 분야에 투자를 지속적으로 확대해가고 있음을 알 수 있었다.

참고문헌(Reference)

[1] NIA, “Artificial Inteligence in Society”, 2019.

[2] MyungSeok Yang, WonKyun Joo, KiSeok Choi, YoungKuk Kim, YunJeong Kim, “Development of platform-based knowledge map service to get data insights of R&D institution on user-interested subjects”, Wireless Personal Communications, 98(40: 3265-3285, 2018.
https://doi.org/10.1007/s11277-017-5097-z

[3] Namgyu Kim, Donghoon Lee, Hochang Choi, William Xiu Shun Wong, “Investigations on Techniques and Applications of Text Analytics”, KICS, 42(2): 471-492, 2017.
https://doi.org/10.7840/kics.2017.42.2.471

[4] JunHyeong Park, Hyo-Jung Oh, “Comparison of Topic Modeling Methods for Analyzing Research Trends of Archives Management in Korea: focused on LDA and HDP”, kliss, 48(4), 235-258, 2017.
https://doi.org/10.16981/kliss.48.201712.235

[5] Keon Chul Park, Chi Hyung Lee, “A Study on the Research Trends for Smart City using Topic Modeling”, Journal of Internet Computing and Services(JICS), 20(3): 119-128, 2019.
https://doi.org/10.7472/jksii.2019.20.3.119

[6] ChunHo Nam, “Review of the applicability of topic modeling techniques in diary data research”, Journal of Cross-Cultural Studies, 22(1):89-135. 2016.

[7] Jin-myeong Chung Young-ho Park Woo-ju Kim, “Social Media Analysis Based on Keyword Related to Educational Policy Using Topic Modeling”, Journal of Internet Computing and Services(JICS), 19(4): 53-63, 2018.
https://doi.org/10.7472/jksii.2018.19.4.53

[8] TaeHyun Kim, MyungSeok Yang, KwangNam Choi , “A Study on the Construction of the Terminology Dictionary for National R&D Information Utilization”, Journal of Korea Contents Association, 19(10) :217-225, 2019.
https://doi.org/10.5392/JKCA.2019.19.10.217

[9] mecab, https://pypi.org/project/python-mecab-ko/

[10] gensim, https://pypi.org/project/gensim/

[11] Getting Started with Topic Modeling and MALLET, Shawn Graham, Scott Weingart, and Ian Milligan, https://programminghistorian.org/en/lessons/topic-modeling-and-mallet

◐ 저 자 소 개 ◑

양 명 석(MyungSeok Yang)
1999년 충남대학교 컴퓨터과학과(이학사)
2001년 충남대학교 대학원 컴퓨터과학과(이학석사)
2017년 충남대학교 대학원 컴퓨터공학과(공학박사)
2001년~2021년 한국과학기술정보연구원 NTIS센터 책임연구원
2021년~현재 한국과학기술정보연구원 데이터기반문제해결연구단장
관심분야 : 데이터베이스, 데이터마이닝, 네트워크분석
E-mail : msyang@kisti.re.kr

이 성 희(Sung-Hee Lee)
1995년 전남대학교 전산학과(이학사)
2006년 한국과학기술원 텔레콤경영(MBA)
2015년~현재 한국과학기술정보연구원 NTIS센터 선임연구원
관심분야 : 정보시스템, 데이터마이닝, 네트워크분석
E-mail : sunghee.lee@kisti.re.kr

박 근 희(Park Keun-Hee)
2007년 건양대학교 정보보호학과(이학사)
2016년 전북대학교 대학원 정보보호공학(공학석사)
2016년~현재 한국과학기술정보연구원 NTIS센터 선임연구원
관심분야 : 정보보안, 네트워크, 데이터베이스, 클라우드
E-mail : pkh7514@kisti.re.kr

최 광 남(Kwang-Nam Choi)
1992년 충남대학교 컴퓨터공학과(공학사)
1994년 충남대학교 대학원 컴퓨터공학과(공학석사)
2017년 배재대학교 대학원 컴퓨터공학과(공학박사)
1994년 한국과학기술정보연구원(KISTI, 당시 연구개발정보센터) 입사
2018년~2021년 KISTI NTIS센터 센터장
2021년~현재 한국과학기술정보연구원 국가과학기술데이터본부 본부장
관심분야 : 빅데이터, 과학계량학, 정보분석
E-mail : knchoi@kisti.re.kr

김 태 현(Tae-Hyun Kim)
1999년 충남대학교 컴퓨터과학과(이학사)
2001년 충남학교 대학원 컴퓨터과학과(이학석사)
2002년~2004년 한국전자통신연구원 연구원
2004년~ 한국과학기술정보연구원 선임연구원 /서비스혁신팀장
관심분야 : 정보검색, 정보분석, 전문용어사전구축, 소프트웨어공학
E-mail : heemang@kisti.re.kr