Journal of the Korea Industrial Information Systems Research Vol. 23 No. 3, Jun. 2018 :87-95
http://dx.doi.org./10.9723/jksiis.2018.23.3.087 ISSN:1229-3741

Latent Dirichlet Allocation (LDA) 모델 기반의 인공지능(A.I.) 기술 관련 연구 활동 및 동향 분석

(Systemic Analysis of Research Activities and Trends Related to Artificial Intelligence(A.I.) Technology Based on Latent Dirichlet Allocation (LDA) Model)

정 명 석[1], 이 주 연[2]*
(Chung Myoung Sug and Lee Joo Yeoun)

요 약 최근 인공지능(Artificial Intelligence; A.I.)의 기술 발전과 함께 이에 대한 관심이 증가하고 있으며 관련 시장도 비약적으로 확대되고 있다. 아직은 초기단계이지만 2000년 이후 현재까지 계속 확장되고 있는 인공지능 기술 분야의 연구방향과 투자 분야에 대한 불확실성을 줄이는 것이 중요한 시점이다. 이러한 기술 변화와 시대적 요구에 따라서 본 연구는 빅데이터(Big Data) 분석방법 중 텍스트 마이닝(Text Mining)과 토픽모델링(Topic Modeling)을 활용하여 기술동향을 살펴보고, 핵심기술과 성장 가능성이 있는 연구의 향후 방향성을 제시하였다. 본 연구의 결과로부터 인공지능의 기술동향에 대한 이해를 바탕으로 향후 연구 방향에 대한 새로운 시사점을 도출할 수 있으리라 기대한다.

핵심주제어 : 인공지능, 빅데이터, LDA, 텍스트분석, 기술동향

Abstract Recently, with the technological development of artificial intelligence, related market is expanding rapidly. In the artificial intelligence technology field, which is still in the early stage but still expanding, it is important to reduce uncertainty about research direction and investment field. Therefore, this study examined technology trends using text mining and topic modeling among big data analysis methods and suggested trends of core technology and future growth potential. We hope that the results of this study will provide researchers with an understanding of artificial intelligence technology trends and new implications for future research directions.

Key Words : Artificial Intelligence, Big Data, LDA, Text Analysis, Technology Trend

* This study was supported by Ministry of Trade, Industry and Energy(MOTIE) and Korea Institute of Energy Technology Evaluation and Planning(KETEP). (No.20162010104570).
* This study is the result of research carried out by the Ministry of Commerce, Industry and Energy and the Korea Industrial Technology Development Agency (N0001083)
* 본 논문은 한국산업정보학회가 참가하는 2018년 경영정보관련 "춘계통합학술대회"에서 발표한 정명석, 이주연의 논문(Laten Dirichlet Allocation (LDA) 모델 기반으로 한 인공 지능 (A.I.) 기술의 연구 활동 및 동향에 대한 체계적인 분석)을 보완한 것임을 밝혀둔다.
* Corresponding Author : jooyeoun325@ajou.ac.kr
Manuscript received April 24, 2018 / revised June 11, 2018 / accepted June 13, 2018
1) 아주대학교 산업공학과, 제1저자
2) 아주대학교 산업공학과, 교신저자

1. 서 론

인공지능이란 인간이 가지고 있는 고유 능력이라 할 수 있었던 학습, 추론, 지각, 탐색 등의 능력들을 컴퓨터의 기술로 풀어낸 것이라 할 수 있다. 인공지능이란 용어는 1956년 미국 다트머스 대학의 존 매카시 교수가 다트머스의 한 학회에서 처음 사용하면서 등장하였다. [1] 세계적으로 인공지능에 대한 연구는 대학을 중심으로 한 학계뿐만 아니라 및 민간기업의 연구기관에서 활발히 진행되고 있으나, 우리나라는 아직까지 초기 단계이다. 그렇기 때문에 인공지능 산업에서 우리나라가 지향해야 할 기술연구 방향과 투자분야에 대한 불확실성을 감소시켜줄 수 있는 방안이 절대적으로 필요한 시점이다.

본 연구는 시스템적 체계에 의해서 인공지능 기술과 관련된 논문데이터의 핵심 키워드들을 분류하여, 최근 기술적 동향을 파악할 수 있는 방법을 고찰한다. 또한, 2000~2017년에 진행된 인공지능의 연구 분야 및 관련 동향을 분석하여 향후 연구나 투자의 방향성에 대한 시사점을 제시하고자 한다.

2. 관련 연구

우리나라는 인공지능 분야의 국가정책을 제시하고자, 2016년 3월 관련 산업을 육성하기 위한 정부기관 내 인공지능 총괄 팀을 신설하였다. 주요 민간 기업들과의 협력을 통한 인공지능 개발 컨트롤타워 설립 등 다각적으로 노력하고 있다. 그러나 인공지능의 선도국가라고 할 수 있는 미국, 유럽, 중국 등에 비해 투자규모가 1/10에 그치는 등의 한계를 가지고 있는 실정이다.[2],[3]

따라서 우리나라가 제한적인 조건에서 인공지능 분야를 선도하고 있는 국가 및 기업의 기술력을 추격하여 경쟁력을 갖기 위해서는 인공지능 연구 주제와 내용을 가지고 우리나라의 기술수준을 정확하게 분석하여, 관련 기술 개발 과정을 최적화하기 위한 한국형 기술 로드맵 마련이 필수적일 것으로 판단된다.

이를 체계적으로 접근하는 방법론 중의 하나로는 다양한 분야에서 관련된 주제를 정리하고 검색 및 이해를 도와주는 토픽모델링이 있다. 토픽모델링(Topic Modeling)은 텍스트 마이닝(Text Mining) 기법 중의 하나로 텍스트 정보에 숨겨진 주제들을 추론하기 위해 고안된 통계적 추론 기법이다. 일종의 통계 알고리즘으로 구조화 되지 않은 방대한 빅데이터인 큰 문서 집합에서 핵심이 되는 주제를 탐색하는 것이다.[4] 비정형 텍스트의 주제를 찾아서, 주제 또는 토픽별로 단어의 분포, 문서별 토픽의 분포를 추정하고, 주제 간의 연관도 분석과 시간에 따른 변화를 분석하여 준다. 이의 전개는 (1) 빅데이터(Big Data)에서 특정 주제를 찾아내고, (2) 검색된 주제에 따른 특정 주석을 표기하고, (3) 문헌의 주석을 중심으로 문서를 예측, 정리해 주는 단계를 거친다. 토픽모델링 기법 중 주제들을 확률모형을 기반으로 문헌의 특정 주제를 찾아내는 LDA (Latent Dirichlet Allocation)가 많이 이용되고 있다.[5]

Griffiths는 논문의 초록을 대상으로 토픽모델링을 적용하였다. 미국국립과학회보(Proceedings of the National Academy of Sciences of the United States of America)에서 시계열 분석을 통해 최근의 관심이 증대되는 연구주제와 관심이 저조한 연구주제를 분석하여 연구 동향을 파악하고자 하였다.[6]

Client는 키워드 기반 검색의 단점을 보완하고자, 개념 또는 주제 검색을 기반으로 하는 LDA를 통한 토픽모델링을 사용하였다. 전자 발견(Electronic Discovery) 문제에 대해 빠르게 해결할 수 있는 방안으로 문서 군집화(Clustering), 분류, 순위 지정 및 요약 자료를 LDA 로 만든 실제 응용 프로그램을 사용하여 검증하였다.[7]

따라서, 본 연구에서는 문서의 주제 분류에 의한 토픽모델을 사용하여 인공지능 기술의 주제와 내용을 분류하고, 시간 축 분석을 통하여 향후 연구 동향을 예측하고자 한다.

3. 연구 방법론

3.1 분석 절차

본 연구는 인공지능 관련 빅데이터(Big Data)

를 활용하여 다음과 같은 단계적 절차에 따라 분석하였다. 첫째로 데이터 전 처리 과정을 통한 분석의 용이성 확보, 두 번째로 빅데이터 분석 방법 중 하나인 토픽모델링 기법을 적용하고, 마지막으로 키워드들의 출연빈도와 이에 대한 시계열 분석을 수행한다.

Fig. 1은 이와 같이 구성된 분석 절차를 도식화한 것이다. 데이터 수집은 데이터베이스에서 논문데이터를 수집하는 과정을, 데이터 전 처리는 노이즈 제거 및 데이터 표준화 작업을 의미한다. 다음으로는 토픽모델링 방법의 절차를 적용하여 인공지능 분야의 연구 동향을 분석한다.

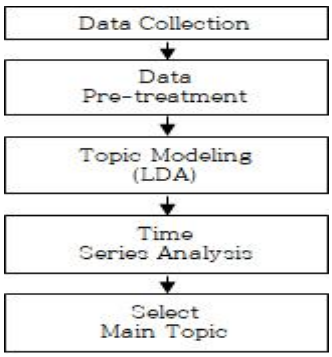

Fig. 1 Data Analysis Process

3.2 분석 대상

인공지능 관련 연구동향 및 주요 연구 분야를 파악하기 위하여 본 연구에서는 Web of Science를 통하여, 2000년부터 2017년까지 SCI(E) 저널에 게재된 논문데이터를 수집하였다. 1차적으로 I.S.I(Institute for Science Information) 사에서 제공하는 인용색인 데이터베이스에서 'Artificial Intelligence' 의 키워드로 주제 분야를 검색하여 11,085건의 논문데이터를 수집하였다.

1차 수집된 데이터 중에서 실제 해당 논문 키워드를 포함하는 데이터를 추출하고, 그 다음 R과 SPSS 프로그램을 이용하여 데이터의 전 처리 작업을 거쳤다. 중복제거 및 인공지능 관련 키워드에 맞는 7,978건의 논문데이터를 최종 분석 대상으로 선정하였다. 논문 키워드에 맞는 데이터를 대상으로 논문데이터의 제목과 대표 키워드를 활용한 토픽모델링을 실시하였다.

3.3 분석 방법

논문데이터에서 주제가 되는 키워드를 추출하기 위해서 LDA (Latent Dirichlet Allocation) 기반의 토픽모델링을 사용하였다. LDA는 일관성을 가지고 의미 있는 주제를 군집화 할 수 있는 장점으로 비정형 텍스트 분석에 사용되고 있다. LDA 모델을 Fig. 2에서 구체적으로 살펴보면, 문서 정보에 대해 주제들의 분포 Θ와 각 주제별 단어가 생성될 분포 β가 주어질 때, 특정 문서가 만들어질 확률을 선택한다. 선택된 주제에 대한 파라미터들을 통해 앞단에서부터 단어를 채울 때마다 하나의 주제를 선택하고 다시 그 주제로부터 단어를 선택하는 확률적 방식으로 주제를 추론하는 모델이다.[8]

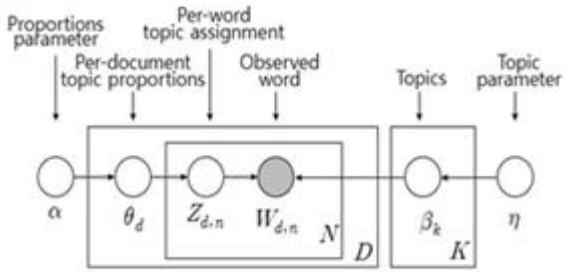

Fig. 2 Latent Direct Allocation Model (Blei, 2012) [14]

K: 전체 토픽 수
N: d번째 문서의 단어 수
W: 단어 수
네모 박스: 해당 횟수만큼 반복하는 것을 의미
동그라미: 변수를 표현함

Fig. 2에서 파라미터 a, β, Θ, Z는 숨겨진 파라미터이며 Dirichlet분포에 의해 결정된다. a와 β는 단어를 추출하기 위한 모수이다. 전체 문서에 동일한 값을 갖는다. Θ_d는 d번째 문서가 가진 토픽 비율이다. $Z_{d,n}$은 문서 d의 n번째 단어가 어떤 토픽에 해당하는지를 할당 한다. $W_{d,n}$은 문서 d의 n번째 단어이다. 여기에서 관찰 가능한 변수는 단어 $W_{d,n}$이며, 이 정보만을 가지고 하이퍼파라미터(사용자 지정) a, β를 제외한 모든 잠재변수를 추정하게 된다.[8]

인공지능의 세부 기술을 추출하기 위해 전처리 과정을 거친 논문데이터를 대상으로 LDA 알고

리즘을 통한 토픽모델링을 수행하였다. 이를 위해 R 패키지 중의 하나인 'Tophicmodels'를 사용하였다. LDA 모델의 파라미터 추정은 Gibbs Sampling 방식을 활용하였다. 또한 토픽모델링 결과를 시각화 하여 보여줄 수 있는 LDAvis package를 활용하였다.

LDA를 적용하여 논문데이터를 가장 적합하게 분류할 수 있는 토픽 주제가 몇 개인지 추정하였다. 논문데이터에 나타난 문헌정보로부터 토픽 주제 7 개를 추출하고, 추출된 주제 각각에 대한 적절한 주제명을 선정하였다. 마지막으로 선형회귀분석을 실시하여 시간의 흐름에 따른 인공지능 산업의 트렌드를 분석하고, 회귀계수의 유의성과 부호를 기준으로 최근 부상하는 토픽과 그렇지 않은 토픽을 분류하였다.

4. 결 과

4.1 인공지능기술 분야 연구 동향

본 연구에서는 인공지능기술 관련 향후 효과적인 연구의 방향성과 기술 개발을 위한 연구 투자 결정을 위한 근거자료를 제시하고자, Web of Science 데이터베이스의 논문데이터를 활용하여 주요 연구 분야 및 최신 동향 등을 분석하였다.

3장에서 기술한 대로 'Artificial Intelligence' 등의 키워드로 주제 분야 검색을 수행하여 1차 수집한 후 실제 논문키워드에 해당 키워드를 포함한 데이터를 최종 분석 대상으로 선정하였다.

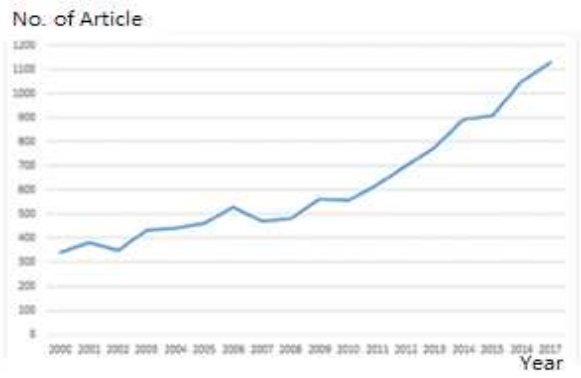

Fig. 3 Artificial Intelligence Paper Output (Web of Science, 2000-2017)

연도별 논문 생산량은 2000년 이후 관련 논문의 생산량은 지속적인 증가 추세를 보였다. Fig. 3에서 보듯이, 2000년 340건에서 2017년 1,128건으로 3.3배 가량 증가한 것을 알 수 있다.

인공지능 분야의 논문이 게재된 학술지는 Table 1과 같이 요약된다. 구체적으로 살펴보면, 'Computer Science Artificial Intelligence'가 3,471건으로 가장 많은 것으로 나타났다. 그 다음으로는 'Engineering Electrical Electronic (1,575건)', 'Computer Science Interdisciplinary Application (1,039건)', 'Computer Science Information System (816건)' 등의 순으로 나타났다.

Table 1 Number of Articles Published by Artificial Intelligence Journals(Year 2000-2017)

No.	Name of Journal	No. of Articles
1	Computer Science Artificial Intelligence	3,471
2	Engineering Electrical Electronic	1,575
3	Computer Science Interdisciplinary Application	1,039
4	Computer Science Information System	816
5	Computer Science Theory Methods	800
6	Operation Research Management Science	721
7	Engineering Multidisciplinary	549
8	Computer Science Software Engineering	532
9	Automation Control System	520
10	Energy Fuels	409
…	…	-
119	Behavioral Science, Biophysics	27
120	Polymer Science, Spectroscopy	26
121	Genetics Heredity	25

4.2 토픽 모델링

데이터 전 처리 과정을 거친 7,978건의 논문데이터를 대상으로 Gibbs Sampling 방식의 LDA 알고리즘을 통한 토픽모델링을 수행하였다. 이 때 토픽의 개수와 반복 횟수는 일반적으로 해석이 용이하고, 더 의미 있는 결과를 위해 연구자가 직접 판단하는 것이 바람직하다. [9] 본 논문에서는 샘플링을 1000 회, 3000 회, 5000 회 및 10,000 회 반복하여, 최종적으로 샘플링 회수로서 10,000 회를 결정 하였다.

LDA의 민감도 분석을 통해 본 연구의 논문데이터에서 총 7 개의 주제가 추출되었다. 주제 수는 5, 7, 10 및 20 개를 K 값에 적용하여 분석하였다. 결과적으로, 가장 적합한 분류가 가능한 K 값은 7 로 확인되었다. α값은 1 / K 로 산정하였고, β값은 200/W로 산정하였다. 여기서는 각 주제와 키워드의 관련성을 정의하는 가중치 매개변수 λ ($0 \leq \lambda \leq 1$)를 정의한다. [9] 최적의 파라미터 튜닝의 적합성 계산 결과는 λ를 1.0으로 정의함으로써 Fig. 4와 같이 얻어졌다.

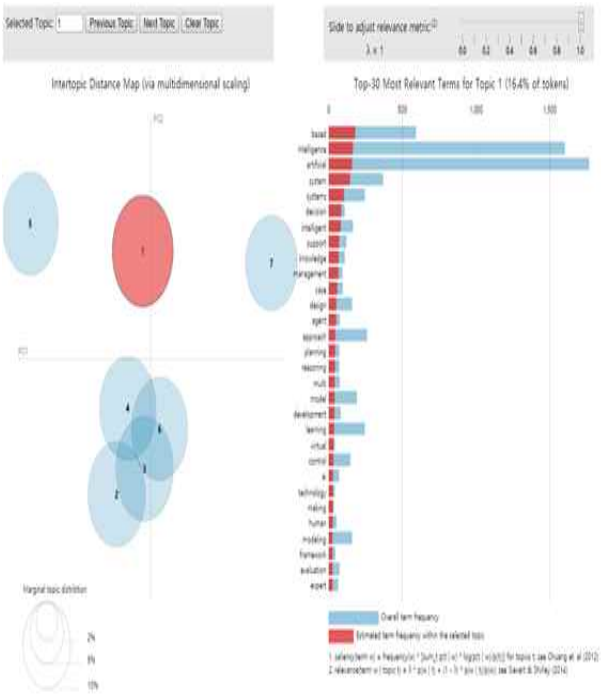

Fig. 4 As a Result of LDAvis, visualize the Relation with Keyword according to Topics

이상과 같은 토픽에 대한 분석 후, 선행연구에서 서술한 인공지능 분야의 기술 범위에 따라 주제를 분류하였다. 7개의 토픽을 통해서 1차로 주제를 유추하였으며, Table 2에 인공지능 기술 로드맵에 따른 기술의 발전 방향에 맞추어 주제를 선정하였다. Forrester의 인공지능 TechRadar 보고서를 바탕으로 포브스에서는 유망한 인공지능 기술을 정의한 바 있으며,[10, 11] 본 연구에서도 다음과 같이 토픽별 주제 선정에 활용하였다.

1) 토픽 1 딥 러닝 플랫폼 (Deep Learning Platforms): 다중 추상화 계층을 가진 인공 신경망으로 구성된 기계 학습의 한 형태이다. 주로 분류 응용 프로그램이나 패턴 인식, 행동인지, 컴퓨터 비전, 영상 레이블링 사물 인식 등의 분야에서 사용된다.

Table 2 Topic Modeling Result of Article Data

Topic no.	Topic Word	Subject
Topic 1	artificial, Intelligence, neural, analysis, data, classification, recognition, system, networks, diagnosis, image, techniques, machine, network, detection, application, automated, cancer	Deep Learning Platforms
Topic 2	artificial, intelligence, power, systems, system, control, forecasting, time, techniques, detection, analysis, fault, application, real, diagnosis, networks, intelligent, method, neural	Robotic process automation
Topic 3	artificial, intelligence, special, applications, issue, research, conference, introduction, engineering, science, review, future, international, human, editorial, cognitive, techniques, computing, advances, computing, advances, preface	Natural language generation
Topic 4	artificial, intelligence, machine, learning, neural, support, prediction, vector, analysis, data, diagnosis, network, classification, system, techniques, networks, application, approach	Machine Learning Platforms

Topic 5	fuzzy, learning, analysis, data, reasoning, system, case, recognition, logic, knowledge, classification, model, information, systems, selection, algorithm, image, method, approach	Speech recognition
Topic 6	artificial, intelligence, power, system, systems, neural, control, techniques, networks, forecasting, application, hybrid, detection, time analysis, network, intelligent, design, energy	AI-Optimization Hardware
Topic 7	optimization, algorithm, problem, genetic, search, approach, learning, programming, design, swarm, multi, particle, scheduling, algorithms, hybrid, dynamic, networks, solving, system	Biometrics

2) 토픽 2 로봇 프로세스 자동화(Robot Process Automation): 효율적인 비즈니스 프로세스를 지원하기 위해 스크립트 및 기타 방법을 사용한다. 인간이 직접 작업이나 프로세스를 실행하기에는 비효율적인 곳에 인간 활동을 대신하여 사용된다. 단순동작을 반복하는 공장의 로봇 팔에서 나아가 기획된 목표를 달성하기 위해 스스로 계획과 실행을 하는 로봇들이 연구되고 있다.

3) 토픽 3 자연 언어 생성(Natural Language Generation): 컴퓨터 데이터에서 텍스트를 생성하는 것으로, 고객 서비스, 보고서 생성 및 비즈니스 인텔리전스 보고서 등에 사용된다. 언어처리에 활발히 이용되는데, 구글의 경우 현재 음성으로 모바일 검색의 20%를 구현하고 있다. 향후 실시간 통역의 가능성을 열어두고 있다. 대화를 통해 인간과 상호 작용이 가능한 시스템에 대한 연구를 진행 중이다.

4) 토픽 4 머신 러닝 플랫폼(Machine Learning Platforms): 인간이나 생물이 지식과 사물에 대한 개념을 학습하듯이 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술 분야이다. 대부분은 예측 또는 분류를 포함하며, 엔터프라이즈 응용 프로그램에서 사용된다. 최근 머신 러닝은 대량의 학습 데이터를 처리하는 방향으로 진행되고 있다. 애플의 시리를 이용한 문의, 구글을 통한 검색과 테슬라 자동차를 운전하는 것 모두가 기계학습의 데이터가 될 수 있다.

5) 토픽 5 대화 인식(Speech Recognition): 컴퓨터가 인간의 말을 응용 프로그램 형식으로 변환한다. 모바일 응용 프로그램과 대화형 음성 응답 시스템에 사용된다.

6) 토픽 6 Topic 6 AI 최적화 하드웨어(AI Optimization H/W): AI 지향 연산 작업을 효율적으로 실행하기 위해 설계된 그래픽 처리 장치 (GPU) 및 어플라이언스를 포함한 기술이다. 인공지능의 워크로드 과제를 가속화 할 수 있는 인프라 솔루션의 개발이 활발하다. 엔비디아 GPU가 탑재된 IBM 딥러닝 전용 서버 등이 그 예이다.

7) 토픽 7 생체인식(Biometrics): 인간과 기계 간의 자연스러운 상호 작용을 가능하게 하며. 이미지와 터치 인식, 언어 및 신체 언어를 포함한다. 생체인식 솔루션은 인간의 얼굴, 홍채, DNA, 정맥, 지문 등과 같이 인체의 특정 부위에 독특하고 측정 가능한 특성을 사용한다. 수집된 정보를 AI 시스템이 이해할 수 있는 코드로 변환하여 준다. 생체인식 솔루션은 일반적으로 기업 및 정부 기관의 보안 및 액세스 제어에 사용된다. [11], [12].

상기 정의된 7개의 토픽별 주제의 세부 내용에서 알 수 있듯이 "Artificial Intelligence"에는 다양한 기술과 방법론들이 포함되어 있다.

4.3 연도별 토픽 회귀분석

인공지능 기술(토픽)들의 연도별 추세를 알아보기 위해, 시계열 선형회귀분석을 실시하였다. 이 때 독립변수로는 연도를, 종속변수로는 인공지능 기술의 연도별 비중 평균값을 사용하였다. 각 토픽이 부상하는 분야인지 아닌지에 대한 판단은 회귀계수의 부호와 유의성을 기준으로 이루어졌다. 회귀계수가 유의수준 5 %에서 유의한 양의 값을 가지면 Hot 토픽으로, 음의 값을 가지면 발전이 상대적으로 저조한 Cold 토픽으로 구분하였다.

Fig. 5는 부상하는 추세의 Hot 토픽과 그렇지 않은 Cold 토픽의 회귀선을 보여준다. 토픽 1 딥러닝 플랫폼, 토픽 3 자연 언어 생성은 시간 경과

에 따라 연구 활동의 감소 경향을 보인 반면, 토픽 2 로봇 프로세스 자동화와 토픽 6 AI 최적화 하드웨어는 지속적으로 증가하는 추세로 나타났다.

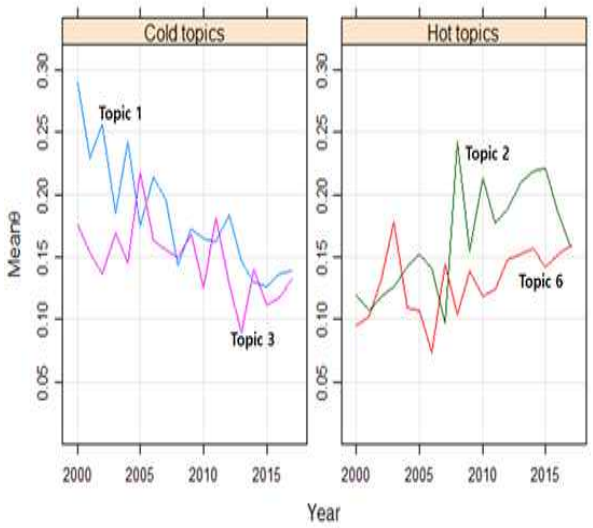

Fig. 5 Hot & Cold Topic Analysis

4.4 토픽 시계열 분석과 기술 트렌드 비교

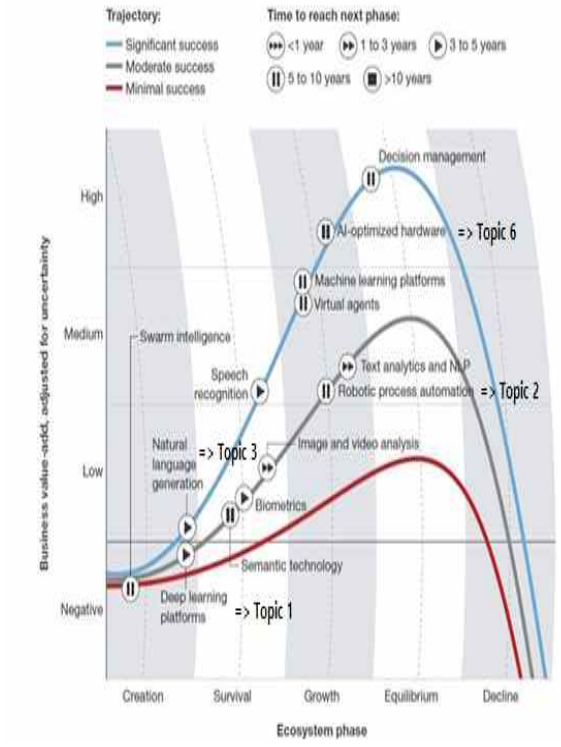

Fig. 6 TechRadar: Artificial Intelligence Technologies, Q1 '17 [10]

연도별 시계열 회귀분석의 결과로 부상하는 분야 또는 저조한 분야로 도출된 토픽들을 Fig. 6에서 Forrester의 인공지능 기술 트렌드 분석 그래프와 비교하여 보았다. Fig. 6의 인공지능 TechRadar 에 따르면, 토픽 2, 6은 최근 급부상하는 기술로 인정받고 있으며, 토픽 1, 3은 저조한 발전 단계로 분류되고 있다. 본 연구의 토픽별 시계열 분석 결과가 Forrester가 2017년 발표한 기술 트렌드와 수렴함을 확인 할 수 있었다.

5. 결 론

인공지능 기술의 개발과 연구 활동이 급격히 확대되어 왔지만, 지금까지는 인공지능 분야의 연구 동향에 대해 키워드 분석과 토픽모델링에 대한 다양한 연구가 부족하였다. 이에 본 연구는 인공지능 기술 분야에서 어떠한 주제의 연구가 어떠한 추세로 진행되어 왔는지에 대한 흐름을 파악하고자 하였다. 다시 말하면, 기존의 인공지능 관련 논문에서 다루어졌던 주제들을 파악하는 것에서 나아가, 상대적으로 연구 개발이 활발한 세부 분야와 저조한 세부 분야를 함께 분석하였다. 2000년부터 18년 동안 Web of Science 에서 발간된 SCI(E) 저널에 게재한 논문 중 인공지능 분야 관련 기술을 포함한 논문을 대상으로 저자 키워드 분석을 통하여 시간의 흐름에 따른 주요 연구 분야를 도출한 결과는 다음과 같다.

최근 인공지능 관련 유망 기술 중 인간 활동을 대신하여 인간이 직접 작업이나 프로세스를 실행하기에는 비효율적인 곳에 사용 되는 로봇 프로세스 자동화와 AI 지향 연산 작업을 효율적으로 실행하기 위해 설계된 AI 최적화 하드웨어의 실용적인 연구가 활발히 진행 되고 있음을 알게 되었다.

또한 인공 신경망으로 구성된 기계 학습의 한 형태인 딥 러닝 플랫폼 (Deep Learning Platforms)과 컴퓨터 데이터에서 텍스트를 생성하는 자연 언어 생성(Natural Language Generation)은 아직은 상대적으로 저조한 발전분야 중의 하나로 판명 되었다.

인공지능 관련 기술 동향 및 주요 연구 주제를 살펴본 본 연구의 결과를 바탕으로 인공지능의

기술 동향과 연구 활동 추이 등에 대한 분석이 가능하다. 관련 기술 개발 및 연구 주제에 대한 시계열 분석에서 도출된 결과로부터 향후 인공지능 분야 연구의 투자와 발전 방향을 시사 받을 수 있다. 따라서 본 연구의 분석 결과는 인공지능 분야의 연구자들에게 앞으로 연구방향을 설정하는 데 도움을 줄 수 있을 것으로 기대 된다. 또한 산업 현장에서는 기술 트렌드에 맞는 투자 방향을 결정하는 데에도 기여 할 것이다.

본 연구는 논문서지 데이터에 대한 정량적인 분석만 이루어졌다는 한계점이 있다. 본 연구에서 분석된 논문데이터에 대한 정량적인 분석을 보완 할 수 있는 Delphi 분석 등 정성적인 분석과 연계된 전문가들의 의견 수렴이 부족하다는 것이다. 향후에는 인공지능 전문가 그룹과의 공동 연구를 통해 메타데이터인 특허 및 논문 분석을 통해 인공지능 기술의 유망 연구 분야를 도출하는 방법론을 개발하고, 그 타당성을 제시하는 발전된 연구가 지속되어야 할 것이다.

References

[1] Kim, S. R., “The Current Situation and the Suggestions of AI and Law Research in Legal Reasoning,” *Research of IT and Law, 5,* pp. 319-346, 2011.

[2] Yun, Y. D., Yang, Y. W. and Lim, H. S., “A SNS Data-driven Comparative Analysis on Changes of Attitudes toward Artificial Intelligence,” *Journal of Digital Convergence, Vol. 14, No. 12,* pp. 173-182, 2016.

[3] Shin, K. S., Choi, H. R. and Lee, H. C., “ Topic Model Analysis of Research Trend on Renewal Energy,” *Journal of the Korea Academia-Industrial Cooperation Society, Vol. 16, No. 9,* 2017.

[4] Steyvers, M., and Griffiths, T., “Probabilistic Topic Models,” *Handbook of latent semantic analysis, Vol. 427, No. 7,* pp. 424-440, 2007.

[5] Griffiths, T. L., and Steyvers, M., “Finding Scientific Topics,” *Proceedings of the National Academy of Sciences, Vol. 101, No. 1,* pp. 5228-5235, 2004.

[6] Pazhayidam George, Clint., “Latent Dirichlet Allocation: Hyperparameter Selection and Applications to Electronic Discovery,” Ph.D. diss., University of Florida. 2015.

[7] Hyundai Research Institute. “VIP Report,” *Vol. 646,* pp. 1-14, 2011.

[8] Blei, D. M., “Probabilistic Topic Models,” *Communications of the ACM, Vol. 55, No. 4,* pp. 77-84, 2012.

[9] Carson Sievert, Kenneth E, Shirley., “LDAvis: A Method for Visualizing and Interpreting Topics: Proceedings of the Workshop an Interactive Language Learning,” *Association for Computational Linguistics,* pp. 63-70, 2014.

[10] TechRadar, “Artificial Intelligence Technologies, Q1 ’17,” FORRISTER RESEARCH, Report. 2017.

[11] Gill Press, “Top 10 Hot Artificial Intelligence (AI) Technologies,” Forbes. 2017.

[12] “One Hundred Year Study on Artificial Intelligence (AI100),” Standford University, 2016, https://ai100.standford.edu.

[13] Mimno, D. M. and McCallum, A., “Topic Conditioned on Arbitrary features with Dirichlet- Multinomial Regression,” *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI2008),* pp. 411-418, 2012.

[14] Hornik, K., and Grün, B., “Topic Models: An R Package for Fitting Topic Models,” *Journal of Statistical Software, Vol. 40, No. 13,* pp. 1-30, 2011.

[15] Chung, M. S. Park, S. H., Chae, B. H., and Lee, J. Y., "Analysis of Major Research Trends in Artificial Intelligence through Analysis of Thesis Data," *The Society of Digital Policy and Management,* pp. 225-233, 2017.

[16] Jo, Y. H., Kim, D. W., Moon, I. C. and Oh, Alice H., "Learning Influence Propagation of Personal Blogs with Content and Network Analyses: 2010 IEEE Second International Conference on Social Computing," IEEE. 2010.

[17] Kim, H S., Park, C.Y., Kim, I.C., "Implementation of Temporal Reasoning Services Using a Domain-Independent AI planner," Analysis, *Korea Society of Industrial Information Systems, Vol. 14, No. 4,* pp. 37-48, 2009.

[18] Park, J.S., Kim N.R., Han, E.J., "Analysis of Trends in Science and Technology Using Keyword Network," Analysis, *Korea Society of Industrial Information Systems, Vol. 23, No. 2,* pp. 63-73, 2018.

[19] Joo, S.H., Ha, S.H., Park, S.H., "Technology Keyword Network and Cognitive Map Analysis : to Prospect Promising Technology of UAV(Unmanned Aerial Vehicle) Airframe Industry," *Korea Society of Industrial Information Systems, Vol. 21, No. 5,* pp. 55-72, 2016.

[20] Lau J.H., Grieser K., Newman D, Baldwin T., "Automatic Labelling of Topic Models." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies," Vol. 1,* pp. 1536 - 1545, 2011.

정 명 석 (Chung Myoung Sug)

- 정회원
- 아주대학교 산업공학과 박사 과정 수료
- 서강대학교 경영전문대학원 경영 학석사
- 현)Oracle 상무
- 전)SAP 본부장
- 관심분야 : 인공지능(AI), 융합기술, 데이터 분석

이 주 연 (Lee Joo Yeoun)

- 종신회원
- 인하대학교 경영대학 경영학박사
- 현)아주대학교 공과대학 산업 공학과 교수
- 현)한국빅데이터서비스학회 회장
- 관심분야 : 융합기술연구, 비즈니스인텔리전스