

인공지능 기술 랜드스케이프 : 기술 구조와 기업별 경쟁우위[†]

A Technology Landscape of Artificial Intelligence:
Technological Structure and Firms' Competitive Advantages

이왕재(Wangjae Lee)*, 이학연(Hakyeon Lee)**

목 차

I. 서 론	IV. 인공지능 기술구조 분석
II. 선행 연구	V. 인공지능 기술역량 분석
III. 연구 방법	VI. 결 론

국 문 요 약

본 연구는 특허 데이터를 활용하여 인공지능 기술의 구조를 파악하고 주요 글로벌 IT 기업들의 인공지능 기술역량을 분석한다. 2007년부터 2017년까지 미국 특허청에 등록된 2,589개의 인공지능 특허를 바탕으로 LDA 토픽모델링을 수행하여 인공지능 분야의 20개의 기술 토픽을 도출하였다. 인공지능 기술 분야 중 언어이해, 음성처리보다는 시각이해, 데이터분석, 동작제어, 그리고 기계학습 분야의 연구 개발이 최근 활발한 것으로 나타났다. 또한 기업별 인공지능 기술 역량을 분석하여 인공지능 기술 분야 별로 우수 역량을 보유한 기업을 도출하고, 기업별로 강점을 가지고 있는 세부 기술 분야를 도출하였다. 본 연구 결과는 인공지능 기업들의 기술기획 및 전략 수립에 유용하게 활용될 수 있을 것으로 기대된다.

핵심어 : 인공지능, 특허 분석, 기술동향, 토픽모델링, LDA(Latent Dirichlet Allocation)

※ 논문접수일: 2019.2.18, 1차수정일: 2019.4.9, 게재확정일: 2019.4.30

* 서울과학기술대학교 IT정책전문대학원 박사과정/삼성전자 삼성리서치 연구원, proalex@naver.com, 02-6147-8527

** 서울과학기술대학교 글로벌융합산업공학과 교수, hylee@seoultech.ac.kr, 02-970-6469, 교신저자

† 이 연구는 서울과학기술대학교 교내연구비 지원으로 수행되었음(2018-0868).

ABSTRACT

This study analyzes the technological structure of artificial intelligence (AI) and technological capabilities of AI companies based on patent information. 2589 AI patents registered in USPTO from 2007 to 2017 were collected and analyzed by the Latent Dirichlet Allocation (LDA) to derive 20 AI technology topics. Analysis of technology development trends by AI technology reveals that visual understanding, data analysis, motion control, and machine learning are growing, while language understanding and speech technology are sluggish. In addition, we also investigated leading companies in each sub-field of AI as well as core competencies of global IT companies. The findings of this study are expected to be fruitfully used for formulation and implementation of technology strategy of AI companies.

Key Words : Artificial Intelligence (AI), Patent Analysis, Technology Trend, Topic Modeling, Latent Dirichlet Allocation (LDA)

I. 서론

인공지능(artificial intelligence, AI)은 지능을 가진 존재가 하는 일을 컴퓨터 또는 로봇이 할 수 있도록 만드는 기술이다(Britannica, 2019). 인공지능은 4차 산업혁명을 주도하는 혁신 기술로 전 세계적으로 주목받고 있다. 인공지능 기술은 고용, 생산, 소비 등 사회 전반에 큰 파급 효과를 일으키고 있으며 향후 인류에게 보다 많은 혜택을 가져다 줄 것으로 기대되고 있다(IITP, 2016; Russell et al., 2015). 선도적인 인공지능 기술 확보가 기업 경쟁력을 결정짓는 핵심 요인으로 인식됨에 따라, 많은 기업이 인공지능 관련 연구개발과 특허 확보에 사활을 걸고 있다(Cincera, 1997; Drucker, 1995; Fujii and Managi, 2018; Maresch et al., 2016; 김봉선·김연수, 2014). Google은 ‘인공지능 우선(AI first)’ 이라며 인공지능 기술에 집중하고 있고, Microsoft는 ‘인공지능이 기술의 궁극적 혁신(ultimate breakthrough)’이라고 강조하고 있다. 이들은 인공지능 기술에 대규모 연구개발 투자를 진행하고 있으며, Amazon, Apple, Facebook도 마찬가지이다(Brynjolfsson et al., 2017).

이처럼 인공지능은 기업의 새로운 성장 동력이자 핵심 기술로 자리매김하였다. 이에 따라 인공지능 기술 개발을 위한 연구는 매우 활발히 이루어지고 있으나, 인공지능 기술 자체를 분석 대상으로 삼는 연구는 아직 부족한 실정이다. 특정 기술 분야가 성장기에 진입하면 대상 기술에 대한 구조와 특성 등 기술 자체에 대한 연구가 수행되어 관련 연구개발 전략 수립 및 기술 기획에 활용 될 수 있어야 한다. 일부 연구들이 특허 및 논문 정보 등을 활용하여 인공지능 기술 구조 및 기술 동향에 대한 분석을 수행한 바 있으나(정명석·이주연, 2018; Fujii and Managi, 2018; Kim et al., 2018; 박재용, 2018), 대부분 거시적 차원의 기술 동향 분석에 그치고 있다. 기업의 기술 기획 및 기술 전략 수립을 위해서는 인공지능 기업들의 기술 역량 및 기술 개발 추세 등에 대한 보다 미시적인 관점에서의 연구가 필요하다. 이에 본 연구에서는 특허 정보를 바탕으로 토픽 모델링 기법을 적용하여 인공지능의 세부 기술을 도출한 후, 상세 기술영역별로 주요 글로벌 IT 기업들의 기술 역량 및 강약점을 분석한다.

이후 본 논문의 구성은 다음과 같다. 2장에서는 선행연구를 살펴보고, 3장에서는 본 연구의 연구방법을 설명한다. 4장에서는 토픽모델링 기법을 바탕으로 인공지능 기술 구조를 제시하고, 5장에서는 글로벌 IT 기업들의 인공지능 기술역량을 분석한다. 마지막으로 6장에서는 본 연구의 의의 및 향후 연구방향을 포함한 결론을 제시하고자 한다.

II. 선행 연구

인공지능 기술이 차세대 핵심 성장 동력으로 자리매김함에 따라, 다양한 인공지능 분야별로 활발한 연구개발이 이루어지고 있다. 이에 인공지능 기술 자체의 특성 및 기술 동향을 분석하려는 시도가 일부 이루어져 왔다. 임지연(2016)은 세계 주요 국가와 글로벌 ICT 기업의 인공지능 연구개발 동향을 기술과 시장 분류에 따라 분석하였다. 하지만 이 연구는 정성적이고 주관적인 관점으로 인공지능 기술의 동향을 분석하였다는 점에서 한계가 있다. 특히 데이터를 활용하여 실증적인 관점에서 인공지능 기술을 분석하려는 시도들도 일부 이루어져 왔다. 박재용(2018)은 학습, 인식, 추론으로 인공지능의 세부 분야를 정의하고, 특히 데이터를 활용하여 각 분야별 인공지능 기술 동향을 분석하였다. Kim et al.(2018)는 통계적 분석 방법론을 활용하여 인공지능 분야의 지속 가능한 기술 분야를 도출하였다. Fujii and Managi(2018)는 인공지능 연구개발 방향이 지식 기반에서 수학적 모형 등으로 변화함을 파악하였다. 하지만 이 연구들은 사전에 정의된 인공지능 세부 분야별로 분석을 진행하였기 때문에, 빠르게 영역을 확장하며 신기술이 끊임없이 등장하는 인공지능 분야의 복잡한 기술구조를 반영하지 못한다는 한계가 있다.

인공지능 기술의 영역을 정의하고 세부 기술을 도출하여 계층 구조를 규명하기 위해서는 기술 문서에 근거한 실증적 접근이 필요하다. 특허와 논문 같은 기술 문서를 바탕으로 세부 기술을 탐색하고 분류하는 방법은 이산 할당(discrete assignment)과 비율 할당(fractional assignment)으로 구분할 수 있다(Lee and Kang, 2018). 이산할당은 하나의 문서가 하나의 주제만을 포함한다고 가정하고, 각 문서들을 하나의 대표 분야로만 분류하는 방법이다. 기술 구조 분석 목적으로 가장 많이 활용되어 온 동시인용(co-citation) 분석은 문서 간의 동시 인용 관계를 바탕으로 유사한 문서들을 클러스터링하여 세부 기술 분야를 도출하는 방식으로, 이산 할당 방식을 채택하고 있다. 그러나 일반적으로 기술 문서는 여러 가지 주제를 포함하기 마련이므로, 하나의 문서를 특정 단일 주제로만 분류하게 되면 정보의 손실이 발생하게 된다. 따라서 한 문서 내에 포함된 복수의 주제를 그 비율에 따라 추출하여 분류하는 비율 할당 방식이 보다 이산 할당 방식에 비해 선호되고 있으며(강전학·이학연, 2018), 이러한 비율 할당 방식을 활용하는 토픽 모델링 기법이 널리 활용되고 있다.

토픽 모델링은 대량의 비정형 문서에 잠재된 주제를 탐색하는 통계적 방법론이며(Blei, 2012), 특허, 논문, 기사 등 다양한 유형의 텍스트 데이터 분석에 활용되고 있다. 토픽 모델링 기법 중 가장 많이 활용되고 있는 LDA(latent Dirichlet allocation) 기법은 분석 대상 문서에 대한 사전 정보와 정성적 판단이 불필요하며, 알고리즘을 통해 주제를 자동적으로 추출할 수

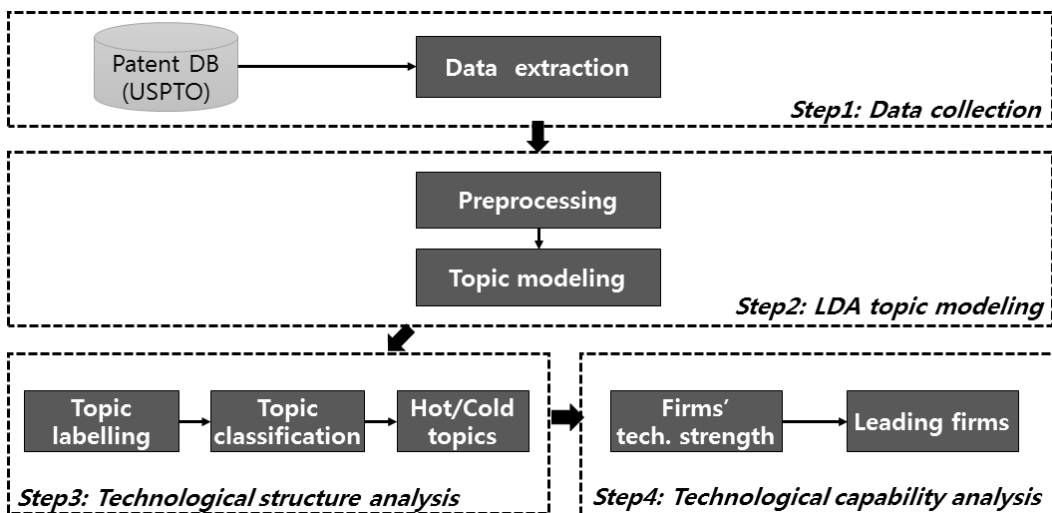
있다는 장점이 있다. 이로 인해 LDA 토픽 모델링은 클라우드컴퓨팅(강전하·이학연, 2018), 핀테크(김태경 외, 2016), 산업공학(정보권·이학연, 2016), 기술경영(Lee and Kang, 2018), 제 품서비스시스템(Lee et al., 2018) 등 다양한 기술과 학문 영역의 세부 주제를 탐색하는데 활용 되어 왔다.

인공지능 분야에도 토픽 모델링을 적용하려는 연구가 일부 이루어졌다. 정명석·이주연(2018) 은 인공지능 관련 논문을 대상으로 LDA를 적용하여 일곱 개의 인공지능 분야를 도출하고 연도 별 연구 동향을 분석하였다. 박주섭 외(2017)는 인공지능 기술 특허 초록을 바탕으로 LDA를 활용하여 20개의 인공지능 기술 토픽을 추출한 후 토픽별 트렌드를 분석하였다. 그러나 위 연구들은 논문에 근거한 학술적 측면의 분석이거나, 거시적 관점에서 전반적 기술동향만을 파악 했을 뿐, 기업 수준에서의 기술개발 역량 분석은 이루어지지 않았다. 이에 본 연구는 기업별 특허 정보를 바탕으로 인공지능 기술 관련 기업의 역량과 강약점을 분석한다.

III. 연구 방법

1. 연구 프레임워크

본 연구의 연구 프레임워크(framework)는 (그림 1)과 같이 데이터 수집, 토픽모델링 적용,



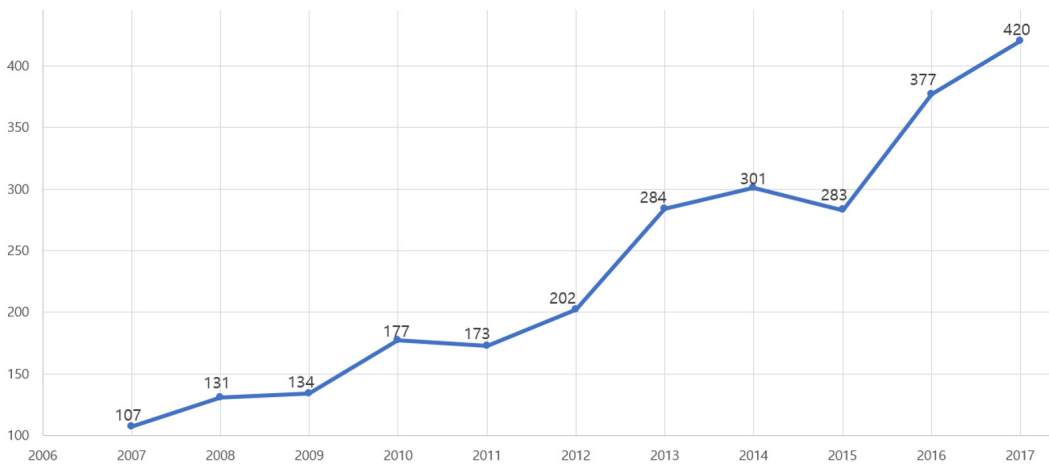
(그림 1) 연구 프레임워크

기술구조 분석, 기업 기술역량 분석으로 구성된다. 첫 번째 데이터 수집은 미국 특허청 데이터 베이스에서 인공지능 기술 특허를 추출한다. 다음으로 토픽 모델링 적용은 모델링에 필요한 데이터 전처리(preprocessing) 후 정제된 데이터를 이용하여 LDA기반의 토픽 모델링을 수행한다. 기술구조 분석에서는 인공지능 기술에 대한 특허 데이터와 토픽을 바탕으로 기술 토픽을 정의하고 기술 동향을 분석한다. 마지막으로 기업 기술역량 분석에서는 기업별 토픽비중 분석과 주요 기업의 기술우위를 분석한다.

2. 데이터 수집

미국 특허청(USPTO)에 등록된 2007년부터 2017년까지의 특허 데이터에서 ‘artificial intelligence’ 등 8개의 키워드가 제목과 초록에 포함되는 특허 데이터를 추출하였다. 검색에 사용한 키워드는 ‘artificial intelligence, machine learning, cognitive computing, deep learning, predictive APIs, natural language processing, image recognition, speech recognition’이며 트랙티카(Tractica)의 인공지능 기술·시장 분류에 나오는 기준을 활용했다(임지연, 2016). 이 특허 데이터에서 보유수량 기준 30대 기업의 특허 2,589건을 이용해 분석을 위한 데이터베이스를 구성하였다. 데이터베이스 특허정보는 제목, 초록, 발행 연도, 양수인 등 분석에 필요한 상세 필드를 포함한다.

(그림 2)는 인공지능 특허의 연도별 증가 추이를 나타낸다. 주요 기업의 인공지능 특허 등록 규모는 2007년 107개에서 2017년 420개로 꾸준히 증가했다. 특히 2016년에서 2017년, 2년간



(그림 2) 글로벌 IT기업의 연도별 등록 특허 수

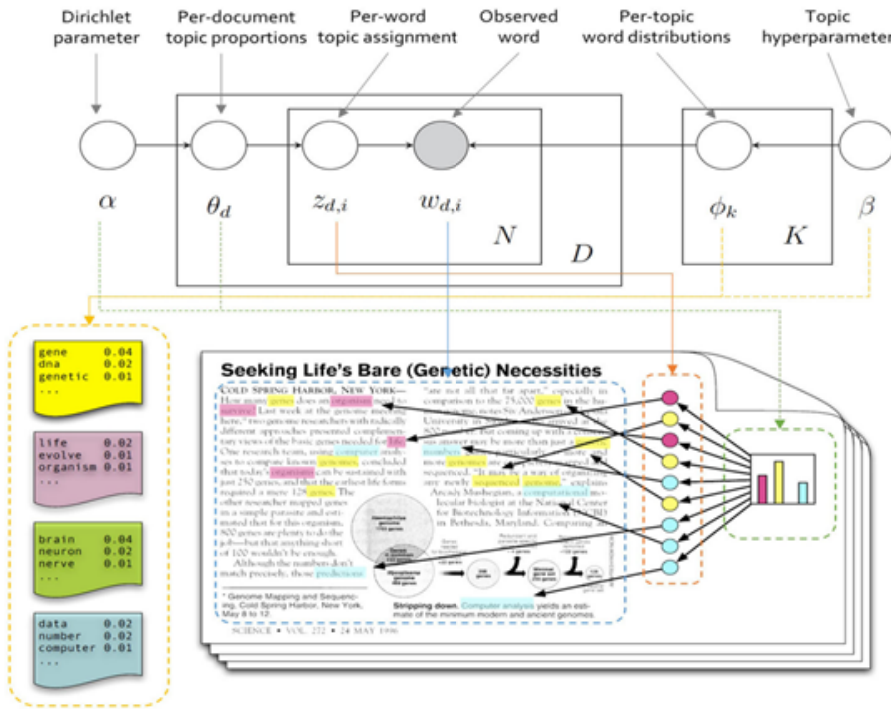
연 평균 20%가 넘는 특허등록 증가를 보여 기업들의 인공지능 기술 개발이 최근 활발히 이루어짐을 확인할 수 있다.

3. 토픽 모델링(topic modeling)

토픽 모델링을 수행하기 위해 제목과 초록 정보를 대상으로 데이터 전처리를 수행하였다. 먼저 모든 구두점과 숫자를 제거하였고 영문 대문자는 소문자로 변환하였다. 두 번째는 토큰화(tokenization) 과정으로 문장을 단어들로 분리하였다. 세 번째는 앞에서 분리한 단어들 중 분석에 불필요한 불용어(stop words)를 제거하였다. 제거된 단어들은 모든 a, an, the와 같은 관사나 in, by, with와 같은 전치사를 포함한다. 네 번째는 표제어 추출(lemmatization) 과정을 진행하였다. 단어는 문법에 따라 여러 형태가 될 수 있는데 이 변형된 형태 단어들 중 하나의 기본 단어인 표제어로 바꾸어야 한다. 예를 들어 단어 interact는 동사로 interact, interacts, interacted 형태로, 명사로 interaction, interactions, 형용사로 interactive 등과 같이 다양하게 사용될 수 있다. 표제어 추출에서는 앞의 예에서 다양한 동사는 interact, 다양한 명사는 interaction 같이 표제어로 선택되게 된다. 마지막으로 각 문서는 백오브워드(bag of words) 표현으로 변환했다. 코퍼스(corpus)가 N개의 단어를 가졌다면 각 문서는 N차원의 벡터 형태로 나타나게 된다.

앞에서 처리한 데이터를 활용하여 LDA(latent Dirichlet allocation) 토픽 모델링을 진행하였다. 토픽 모델링은 규모가 크고 구조화되지 않은 문서 집합에서 내재된 주제를 찾아내는 통계적 알고리즘이다. 이 알고리즘은 문서 생성 메커니즘을 설명하는 수학적 원리에 기반하고 있다. 문서를 사전 구분할 필요가 없으며 전문가의 도움 없이 문서 분석이 가능하다(Blei, 2012). LDA는 가장 널리 채택되는 토픽모델링 방법이라 할 수 있다(Lee and Kang, 2018). LDA에서는 문서를 구성하는 각 단어들에 주제 집합에 대한 확률분포를 가지고 있으며 이 분포에 따라 각 문서를 구성하게 된다고 가정한다. (그림 3)은 LDA 토픽 모델링의 문서 생성 과정을 예시로 표현한 것이다. $W_{d,i}$ 는 d번째 문서의 i번째 단어이고 $Z_{d,i}$ 는 $W_{d,i}$ 의 토픽 할당이다. θ_d 는 d번째 문서의 토픽 비율이고 ϕ_k 는 k번째 토픽의 단어 분포이다. 문서가 생성되는 프로세스는 다음과 같다: (1) 토픽 당 단어 분포(ϕ_k)가 전체 코퍼스(corpus)에서 정해진다. (2) 문서 당 토픽 비율(θ_d)이 결정된다. (3) 각 단어는 토픽 할당($Z_{d,i}$)에 의해 선택된다. α 와 β 는 θ_d 와 ϕ_k 의 디리클레 하이퍼파라미터(Dirichlet hyper-parameter)이다. N은 문서의 단어 수, D는 코퍼스의 문서 수를 나타내며 K는 코퍼스의 토픽 수를 의미한다(Blei et al., 2003).

수집된 2,589건의 특허 데이터에 대해 전처리를 수행한 후 LDA 추론을 시행하였다. LDA



(그림 3) 문서 생성 프로세스(Lee and Kang, 2018; Blei, 2012)

토픽 모델링 수행에는 R 패키지 'topicmodels'(Hornik and Grün, 2011)를 수정 활용했다. 깃스 샘플링(Gibbs sampling) 방법을 이용하였으며 반복 횟수는 1000회를 진행했다. α 는 코 퍼스에 의해 정해진 값을 β 는 기본값인 0.1을 적용했다.

토픽의 개수를 정하고 토픽명을 정의하기 위해 5인의 인공지능 전문가(대학교수 2인, 기업 연구소 소속 실무 전문가 3인)가 참여하는 워크숍을 진행하였다. 토픽 개수 K를 결정함에 있어 Perplexity와 같은 정량적 지표를 활용할 수 있다. 그러나 Perplexity는 모델링 관점에서 학습이 얼마나 잘 되었는가를 평가하는 지표일 뿐, 해석의 용이성을 담보하지는 않는다. 이 때문에 토픽의 해석이 중요한 연구들에서는 도메인 전문가들이 가장 적절한 토픽의 개수를 정성적으로 판단하는 것이 권장된다(Andrzejewski et al., 2007). 토픽 개수를 10, 15, 20, 25, 30로 다르게 적용하여 추출한 토픽별 단어 분포를 검토한 결과, 토픽 개수가 20개일 때 설명력이 가장 우수하다고 판단하였다. 추출된 20개의 토픽에 대해 전문가들의 협의를 통해 토픽명을 정의하였다. 각 토픽에서 높은 비중을 차지하고 있는 키워드를 바탕으로 토픽명을 정의하되, 키워드만으로 토픽명을 도출하기 어려운 경우 관련된 특허 제목과 초록을 참고하였다.

IV. 인공지능 기술구조 분석

1. 인공지능 기술토픽

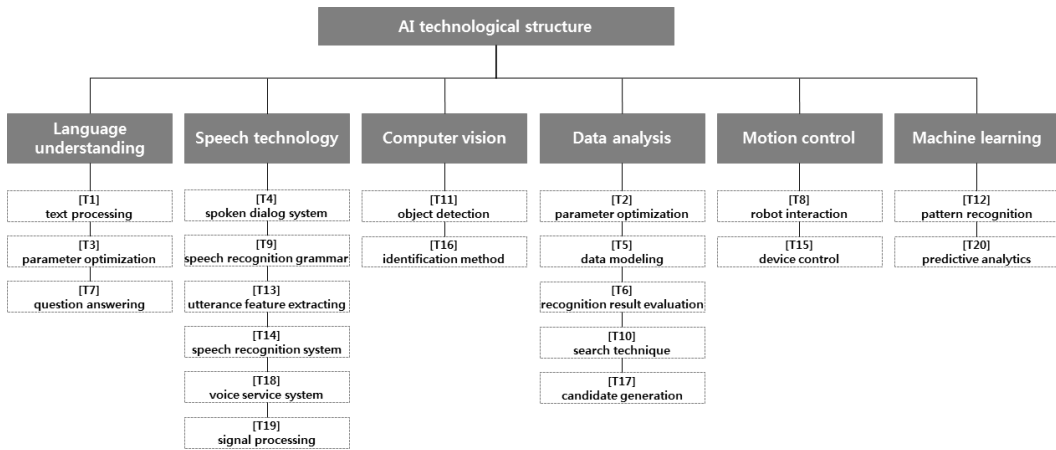
도출된 인공지능 기술의 20개 토픽과 토픽명, 토픽별 핵심단어, 비중은 <표 1>과 같다. 20개의

<표 1> 인공지능 기술 토픽

	Topic	Relevant words	Proportion (%)
[T1]	text processing	text, record, portion, receive, system, correction, convert, select, edit, send	4.80
[T2]	parameter optimization	score, function, weight, probability, parameter, threshold, adjust, calculate, distribution, measure	5.05
[T3]	machine translation	word, language, sequence, dictionary, pronunciation, phrase, vocabulary, translation, sentence, lexicon	4.97
[T4]	spoken dialog system	system, speak, dialog, action, transcription, storage, automatic, select, human, conversational	5.11
[T5]	data modeling	datum, model, update, generate, statistical, adaptation, build, predictive, supervise, unsupervise	4.93
[T6]	recognition result evaluation	information, recognition, result, determine, perform, error, condition, evaluate, criterion, reliability	4.97
[T7]	question answering	language, process, natural, question, feedback, parse, answer, understand, mechanism, linguistic	4.97
[T8]	robot interaction	component, method, system, interaction, analyze, agent, communication, automatic, robot, processor	4.67
[T9]	speech recognition grammar	grammar, module, speech, segment, recognition, engine, generate, application, accuracy, interpretation	4.87
[T10]	search technique	search, query, content, document, context, term, semantic, index, time, keyword	4.91
[T11]	object detection	image, object, capture, detect, recognition, area, vehicle, display, sensor, camera	5.13
[T12]	pattern recognition	machine, learn, rule, analysis, pattern, attribute, detection, analyze, behavior, adaptive	4.91
[T13]	utterance feature extracting	feature, utterance, speak, vector, extract, base, characteristic, generate, method, expression	4.73
[T14]	speech recognition system	speech, recognition, apparatus, check, recognizer, packet, transmit, server, interval, interactive	5.87
[T15]	device control	device, compute, control, command, mobile, location, point, determine, electronic, receive	5.04
[T16]	identification method	identify, method, determine, compare, match, video, category, relate, accuracy, algorithm	4.95
[T17]	candidate generation	input, unit, output, select, candidate, list, configure, recognition, generate, calculate	5.05
[T18]	voice service system	voice, application, call, system, response, interface, service, provide, communication, automate	5.12
[T19]	signal processing	audio, signal, noise, sound, source, processor, microphone, frequency, reduce, quality	4.91
[T20]	predictive analytics	train, set, learn, machine, classification, sample, algorithm, decision, prediction, performance	5.05

기술토픽은 인공지능의 다양한 분야를 포괄하고 있다. 특히 [T14] speech recognition system, [T11] object detection, [T18] voice service system, [T4] spoken dialog system 등이 높은 비중을 차지하고 있는 것으로 나타났으며, 이들은 최근 각광을 받고 있는 인공지능 스피커나 자율주행 기술과 밀접한 관계를 갖고 있는 것으로 보인다.

도출된 20개의 기술 토픽들은 각각 인공지능의 세부 기술을 나타낸다고 할 수 있으나, 보다 효과적으로 인공지능 기술구조를 파악하고 기업별 기술역량을 비교하기 위해서는 추출된 기술 토픽을 몇 개의 분야(sector)로 분류할 필요가 있다. 이에 본 연구에서는 인공지능 연구개발 실무에서 활용되는 기술 분류 체계를 바탕으로 전문가 토론을 통해 (그림 4)와 같이 6개의 인공지능 기술 분야를 정의하고, 20개의 기술 토픽을 분류하였다. 음성처리(speech technology) 분야가 토픽이 6개로 가장 많고, 데이터분석(data analysis) 분야 토픽은 다음으로 많은 5개였다. 언어이해(language understanding), 시각이해(computer vision), 동작제어(motion control), 기계학습(machine learning) 분야가 3개 이하로 상대적으로 적은 수의 토픽을 포함하고 있다.



(그림 4) 인공지능 기술 구조

언어이해 분야는 사람의 표현과 그 의미를 연결하는 것이다(Mejia, 2019). 사람이 문장을 보거나 들으면 지식과 지능을 활용해 이해한다. 이 때 문법만 아니라 단어와 구문 의미를 해석 하며, 특히 문장에 담긴 주제를 파악한다. 컴퓨터를 통한 언어이해를 위해서는 문법, 의미론, 추론 등을 상호 연결해 연구해야한다(Winograd, 1972) 언어이해 연구는 1950년대부터 언어학 과과 협조하여 발전되어 왔다. 하지만, 언어이해 기술은 새로운 단어 같은 언어정보의 지속 증가로 쉽지 않은 연구 분야라 할 수 있다(Bates, 1995). 언어이해 분야에서 도출된 토픽은

다양한 전자적 텍스트의 처리 기술인 [T1] text processing, 컴퓨터를 통한 번역을 다루는 [T3] machine translation, 인간이 제기한 질문에 대한 자동응답을 다루는 [T7] question answering 이다.

음성처리 분야는 인간 음성을 다루는 기술이다. 음성은 사람 사이에 기본적인 소통의 수단이며 정보를 교환하는 가장 자연스럽게 효율적인 방법이다(Gaikwad et al., 2010). 음성인식 기술은 컴퓨터 프로그램으로 구현된 알고리즘에 의해 음성 신호를 문자열로 바꾸는 것이다(Anusuya and Katti, 2009). 음성처리 분야에서는 음성대화 시스템 기술인 [T4] spoken dialog system, 음성인식 문법과 시스템관련 기술인 [T9] speech recognition grammar, [T14] speech recognition system, 언어를 음성으로 나타내는 발화관련 기술인 [T13] utterance feature extracting, 음성 서비스 시스템 관련 기술인 [T18] voice service system, 음성신호 처리 기술인 [T19] signal processing 토픽이 도출되었다.

시각이해 분야는 컴퓨터가 디지털 이미지나 비디오를 이해하고 분석하는 기술이다. 사람은 주변의 꽃과 같은 형체를 3차원 구조물로 쉽게 인식한다. 시각이해 분야에서는 이미지를 이용해 수학적으로 물체의 3차원 모양이나 외관을 분석하는 연구를 수행한다(Szeliski, 2010). 시각이해 분야에서 도출된 토픽은 디지털 이미지나 비디오에서 객체를 인식하는 기술인 [T11] object detection, 얼굴, 지문, 차량 같은 개별 객체를 식별하는 기술인 [T16] identification method 토픽이 나타났다.

데이터분석 분야는 필요한 정보를 얻기 위해 데이터를 검사, 정리, 변화, 모델링하는 것과 관련된 기술이다. 데이터분석을 위해서는 데이터 처리가 수반되는데 데이터처리는 의미를 갖는 정보를 산출하기 위해 데이터를 모으고 다루는 것이다(French, 1996). 데이터분석은 데이터 이용하는 다양한 기술을 포괄한다. 데이터분석으로 분류된 토픽은 파라미터 최적화관련 [T2] parameter optimization, 논리적 데이터 모델 구조를 만드는 [T5] data modeling, 인식 결과에 대한 평가와 관련된 [T6] recognition result evaluation, 데이터를 검색해 필요한 결과를 신속하게 찾아내는 [T10] searching technique, 데이터 선택과 관련된 [T17] candidate generation 토픽이 드러났다.

동작제어 분야는 기계 장치를 자동으로 제어하는 기술 분야이다. 자율주행 자동차, 로봇, IoT(Internet of Things)와 관련된 분야라 할 수 있다. 동작제어 분야에는 인간과 기계 상호작용관련 기술인 [T8] robot interaction, 장치를 제어하는 기술인 [T15] device control 토픽이 나타났다.

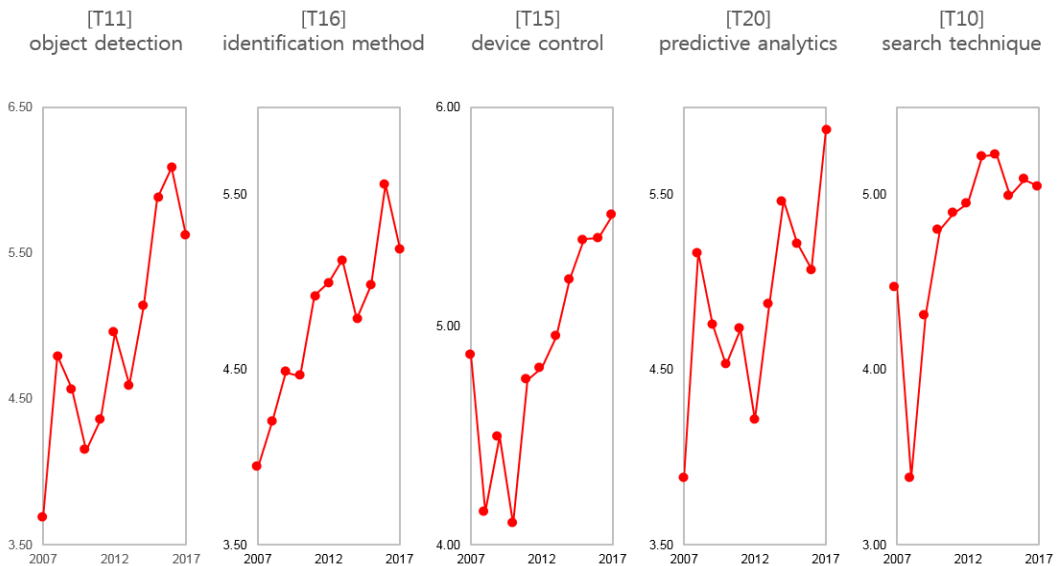
기계학습 분야는 컴퓨터로 수행하는 작업의 성능향상을 위한 알고리즘 연구라 할 수 있다. 이 알고리즘은 예측이나 의사결정을 위해 학습데이터(training data)를 통해 수학적 모델을

만든다. 경험으로부터 학습할 수 있게 컴퓨터를 프로그래밍하면 결국 상세한 프로그래밍 노력을 줄일 수 있다(Samuel, 1959). Samuel(1959)의 연구에서 기계학습이란 용어가 널리 사용되기 시작했다. 기계학습 분야에서는 데이터에서 패턴을 자동 인식하는 기술인 [T12] pattern recognition 토픽이 확인되었는데, 이것은 기계학습과 밀접하게 관련되어 있다 할 수 있다 (Bishop, 2006; Witten et al., 2016). 또한, 과거 데이터를 분석해 미래를 예측하는 기술인 [T20] predictive analytics 도 기계학습 분야 토픽으로 선정되었다.

2. 유망 토픽과 쇠퇴 토픽

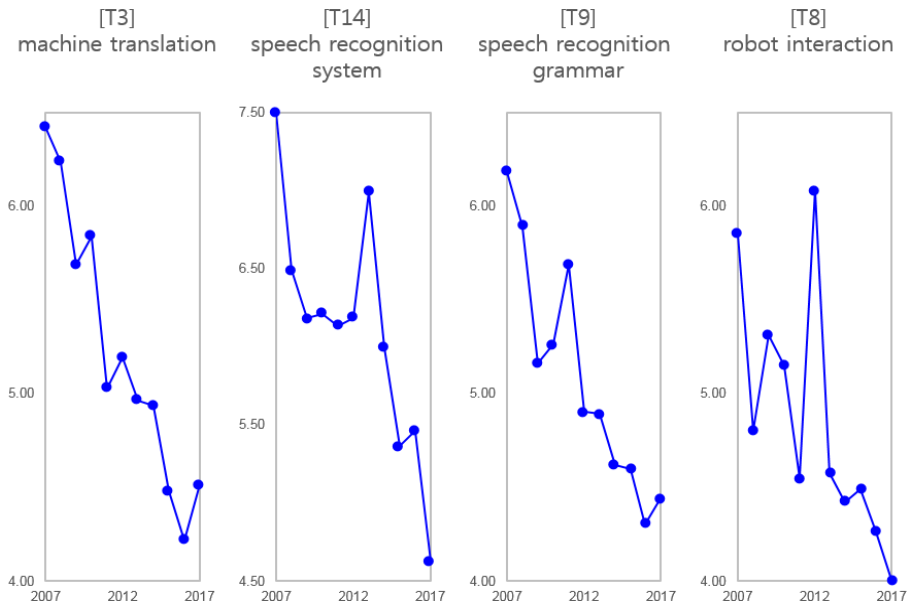
인공지능 기술개발 트렌드를 살펴보기 위해 기술토픽의 연도별 비중 변화를 분석하였다. 연도별 비중이 증가하는 토픽은 성장하는 기술이며, 연도별 비중이 감소하는 토픽은 쇠퇴하는 기술이라 할 수 있다. 본 연구에서는 회귀분석을 바탕으로 각 토픽의 성장과 쇠퇴를 분석하였다. 각 토픽에 대한 토픽 비율을 종속변수로, 연도를 독립변수로 하는 선형회귀 분석을 수행하였다(Griffiths and Steyvers, 2004). 산출된 회귀계수 기울기(slope)의 양수 또는 음수 여부와 유의수준 p값(p-value)을 통해 유망(hot) 토픽과 쇠퇴(cold) 토픽을 선정하였다. 유의수준 0.05에서 기울기가 양수인 토픽을 유망 토픽, 음수인 토픽을 쇠퇴 토픽으로 정의하였다.

(그림 5)와 (그림 6)은 각각 유망 토픽과 쇠퇴 토픽의 연도별 비중 변화를 나타낸 것이다. 로봇,



(그림 5) 유망 토픽의 비중 변화

자율주행 자동차 등에 관심이 높아짐에 따라 [T11] object detection, [T15] device control와 같은 시각이해와 동작제어 기술 성장이 두드러졌으며 [T3] machine translation, [T9] speech recognition grammar, [T14] speech recognition system와 같은 언어이해와 음성처리 기술은 성숙기에 접어들었음을 알 수 있다. 대용량 데이터 처리에 필수적인 [T10] search technique, [T16] identification method와 같은 데이터분석 기술은 유망한 기술로 분류되었다. [T20] predictive analytics와 같은 머신러닝 기술은 최근 한 해 동안 급격한 증가를 보여 기업들이 중점적으로 연구개발하고 있음을 알 수 있다. 요약하면 업계에서 인공지능 기술 분야 중 최근 언어이해, 음성처리와 같은 분야보다는 데이터분석, 시각처리, 동작제어, 머신러닝 분야에 좀 더 집중하고 있는 것으로 보인다.



(그림 6) 쇠퇴 토픽의 비중 변화

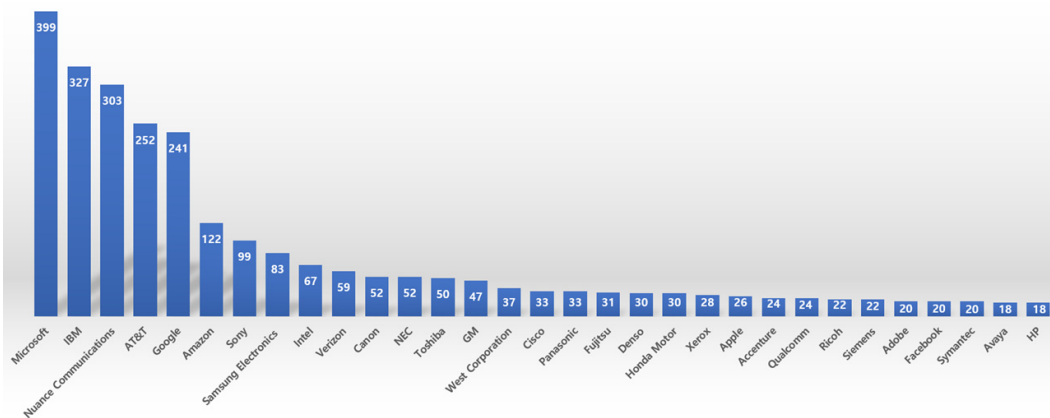
V. 인공지능 기술역량 분석

치열한 시장 경쟁에서 앞서기 위해 기업들은 특허 확보에 심혈을 기울이고 있다. 따라서 각 기업이 보유한 특허를 분석함으로써 기업들의 기술 역량과 기술 전략을 파악할 수 있다. 본 장에서는 IV장에서 도출된 인공지능 기술 분야별로 주요 글로벌 IT 기업의 인공지능 기술역량

을 분석한다.

글로벌 IT 기업의 인공지능 특허 보유현황은 (그림 7)과 같다. Microsoft가 1위이며 그 뒤를 IBM이 따르고 있다. 3위 Nuance Communication은 미국 시장에서 인지도가 높은 음성인식 엔진 개발 소프트웨어 회사이다. 이 기업은 Apple 인공지능 솔루션인 Siri의 음성인식 엔진을 개발한 회사로 알려져 있다. 4위는 통신 기업인 AT&T이고 5위는 Google이 차지했다. 6위는 Amazon, 7위는 Sony, 8위는 한국의 삼성전자이다. 기업별 기술역량 분석에는 인공지능 특허 보유 상위 기업과 시가총액, 인지도가 높은 Apple과 Facebook을 포함하여 10개 기업을 분석하였다.

기업별 인공지능관련 기술역량 분석은 두 가지 측면에서 진행하였다. 하나는 기업별 토픽비중 분석을 통해 기업별로 어떤 기술을 중점적으로 다루는지 살펴봤다. 다른 하나는 기업의 토픽별 특허 보유수를 검토하여 분야별로 기술력이 우수한 기업을 도출하였다.



(그림 7) 글로벌 IT기업별 인공지능관련 특허 수

1. 기업별 중점 기술 분석

토픽 비중을 통해 <표 2>와 같이 기업별로 다루는 중점 기술을 분석하였다. 기업별 강점 기술은 <표 3>과 같다. 기업의 강점 기술로 선정된 토픽 비중 상위 3개 기술에 대해서는 세부 분야를 병기하였다.

Microsoft의 경우 기계학습 분야의 [T12] pattern recognition, [T20] predictive analytics 와 데이터분석 분야의 [T2] parameter optimization 토픽 비중이 크다. 최근 Azure 클라우드 서비스를 강화하고 있는 Microsoft는 인공지능 기술 중 [T12] pattern recognition, [T20] predictive

〈표 2〉 글로벌 IT기업의 토픽 비중

	Language understanding			Speech technology						Computer vision		Data analysis				Motion control		Machine learning		
	[T1]	[T3]	[T7]	[T4]	[T9]	[T13]	[T14]	[T18]	[T19]	[T11]	[T16]	[T2]	[T5]	[T6]	[T10]	[T17]	[T8]	[T15]	[T12]	[T20]
Microsoft	5.01	4.69	4.9	4.28	4.62	4.72	4.64	4.67	4.67	4.36	4.13	6.40	5.30	4.17	5.65	3.93	5.28	4.85	7.12	6.60
IBM	6.15	5.09	9.79	4.33	3.96	4.46	4.19	4.71	4.43	3.41	5.59	5.04	4.26	4.36	6.10	4.68	4.70	4.69	4.65	5.41
Nuance Comm.	4.79	5.31	4.08	5.85	6.78	4.67	8.10	5.26	5.09	2.90	4.71	4.52	4.96	7.02	4.89	5.02	4.48	4.31	3.90	3.36
AT&T	3.95	5.26	4.04	10.66	7.88	4.87	6.54	7.34	3.65	3.29	5.01	4.03	5.71	3.96	3.77	4.17	3.73	4.89	3.62	3.63
Google	5.43	4.79	3.94	5.00	4.01	4.90	5.70	4.77	5.13	4.85	5.77	5.24	5.58	4.6	6.23	4.92	3.56	6.07	4.14	5.35
Amazon	4.21	3.5	4.87	4.29	5.64	5.09	4.84	4.76	7.94	4.42	4.46	4.97	5.52	5.57	3.91	3.9	3.98	6.47	5.97	5.68
Sony	3.65	6.64	4.11	3.89	3.77	6.92	6.38	3.46	3.96	8.44	4.42	5.36	4.53	6.73	3.92	5.92	4.66	5.16	4.2	3.87
Samsung Electronics	4.35	5.78	3.05	4.63	4.21	4.57	9.21	3.94	5.87	7.77	4.48	4.72	4.15	5.13	3.93	6.86	5.48	5.43	3.23	3.20
Apple	3.94	7.34	3.83	3.64	4.64	3.07	6.30	5.86	7.93	2.83	4.71	4.08	3.77	5.17	4.11	5.08	3.17	13.49	3.48	3.56
Facebook	12.65	5.27	5.37	4.15	3.26	3.06	2.65	4.25	2.79	4.89	7.46	3.82	6.23	3.85	4.69	3.35	4.37	3.89	8.05	5.96

〈표 3〉 글로벌 IT기업의 인공지능관련 기술 강점

Company	Major topics	Major sectors
Microsoft	[T12] pattern recognition	machine learning
	[T20] predictive analytics	
	[T2] parameter optimization	data analysis
IBM	[T7] natural language understanding	language understanding
	[T1] text processing	
	[T10] searching technique	data analysis
Nuance Comm.	[T14] speech recognition system	speech technology
	[T9] speech recognition grammar	
	[T6] recognition result evaluation	data analysis
AT&T	[T4] spoken dialog system	speech technology
	[T9] speech recognition grammar	
	[T18] voice service system	
Google	[T10] searching technique	data analysis
	[T16] identification method	computer vision
	[T15] device control	motion control
Amazon	[T19] signal processing	speech technology
	[T15] device control	motion control
	[T12] pattern recognition	machine learning
Sony	[T11] object detection	computer vision
	[T13] utterance feature extracting	language understanding
	[T6] recognition result evaluation	data analysis
Samsung Electronics	[T14] speech recognition system	speech technology
	[T11] object detection	computer vision
	[T17] candidate generation	data analysis
Apple	[T15] device control	motion control
	[T19] signal processing	speech technology
	[T3] machine translation	language understanding
Facebook	[T1] text processing	language understanding
	[T12] pattern recognition	machine learning
	[T16] identification method	data analysis

analytics 기술 비중이 커 기계학습 분야가 중점기술이라 할 수 있다.

IBM의 경우 언어이해 분야의 [T7] natural language understanding, [T1] text processing 과 데이터분석 분야의 [T10] search technique 토픽 비중이 크다. IBM은 최근 힘을 쏟고 있는 AI서비스 Watson의 기반기술인 언어이해 분야의 자연어처리 기술이 중점기술임을 알 수 있다. IBM의 두드러진 보유 기술은 [T7] natural language understanding을 포함한 언어이해 분야라 할 수 있다.

Google의 경우 데이터분석 분야의 [T10] search technique와 동작제어 분야의 [T15] device control, 그리고 시각이해 분야의 [T16] identification method 토픽 비중이 크다. 세계 최고의 검색기술을 자랑하는 Google은 검색기술에 강점이 있다는 것을 [T10] search technique 기술 비중이 가장 큰다는 것에서 확인할 수 있다. Google은 다양한 인공지능 기술을 보유하고 있는데, 특히 검색과 더불어 동작제어 분야의 [T15] device control, 시각이해 분야의 [T11] object detection 기술은 자율주행 자동차와 홈 네트워크 서비스 기술과 관련된 것으로 보인다.

Amazon의 경우 토픽 비중을 볼 때 중점 보유기술은 음성처리 분야의 [T19] signal processing, 동작제어 분야의 [T15] device control, 기계학습 분야의 [T12] pattern recognition이다. Amazon은 인공지능 서비스 알렉사(Alexa), 인공지능 스피커 에코(Echo), 그리고 이와 관련된 홈 네트워크 서비스에서 미국 시장 선두이다. 이런 점은 음성처리와 동작제어 분야 기술 비중에서 잘 나타나 있다. 또한 글로벌 시장에서 앞선 Amazon은 클라우드 서비스에서 기계학습 분야 지원을 강화하고 있는데 Amazon의 기계학습 기술 보유비중도 상당함을 알 수 있다.

삼성전자는 음성처리 분야의 [T14] speech recognition system, 시각이해 분야의 [T11] object detection 토픽이 중점기술로 나타났다. Sony는 시각이해 분야의 [T11] object detection, 언어이해 분야의 [T13] utterance feature extraction 토픽이 중점 보유기술이었다. TV와 같은 영상 기기 시장에서 글로벌 선두인 삼성전자와 Sony는 시각이해의 [T11] object detection 토픽과 같은 영상관련 인공지능 기술에 모두 강점을 보였다. 삼성전자는 스마트폰과 같은 스마트 기기 사업, Sony는 디지털 카메라와 같은 영상 사업도 활발하다. 스마트 기기는 카메라와 같은 영상 기기를 포함한다. 하지만 삼성전자의 가장 큰 강점기술은 음성처리 분야이고 Sony의 경우는 시각이해 분야여서 서로 차이가 있음을 알 수 있다.

Apple은 동작제어 분야의 [T15] device control, 음성처리 분야의 [T19] signal processing, 언어이해 분야의 [T3] machine translation 토픽에 강점을 나타냈다. 이중 가장 비중이 큰 중점 보유 기술은 동작제어 분야로 나타났다. 인공지능 비서 또는 스피커 서비스에 글로벌 강자인 Amazon, Google, Apple, 세 개 회사 모두 동작제어 분야의 [T15] device control이 중점기술로 나타났다. 이는 홈 네트워크 관련 기술에 중점을 두고 있음을 알 수 있다.

Facebook은 언어이해 분야의 [T1] text processing, 기계학습 분야의 [T12] pattern recognition, 데이터분석 분야의 [T16] identification method 토픽 기술을 확보하고 있음을 알 수 있다. 특히 텍스트 서비스가 많은 Facebook은 다른 기업과는 다르게 문자 기반의 [T1] text conversion 기술이 상대적으로 큰 비중을 차지했다.

그 밖에 음성인식 엔진으로 유명한 Nuance Communication은 [T14] speech recognition system 등 음성처리 분야에 강점을 보였고 통신 기업인 AT&T도 음성처리 분야인 [T4] spoken dialog system 등 음성처리 분야에 강점을 나타냈다.

2. 분야별 우수 기업 분석

토픽에 따른 기업별 인공지능 특허 보유수는 <표 4>와 같으며, 이를 통해 각 토픽별 우수 기업을 도출할 수 있다. 언어이해 분야의 [T7] natural language understanding 기술은 2위인 Microsoft보다 3배나 많을 정도로 IBM이 압도적으로 특허를 많이 보유하고 있다. IBM이 자연어이해 분야의 최강자임을 확인할 수 있다. 음성처리 분야의 [T4] spoken dialog system 기술은 AT&T가 2위인 Nuance Communication 보다 3배나 많이 갖고 있다. 오랜 동안 대화형 시스템 연구를 수행한 AT&T는 이 분야에 많은 특허를 보유하고 있고 강자라 할 수 있다. 기계학습 분야의 [T12] pattern recognition 기술은 Microsoft가 2위인 IBM보다 4배나 많은 특허를 보유하고 있어 이 분야 기술력이 우수함을 알 수 있다.

세부 기술 분야별 우수기업은 <표 5>와 같다. 기술 분야별 우수기업은 해당 분야의 특허수를

<표 4> 글로벌 IT기업의 토픽에 따른 특허 수

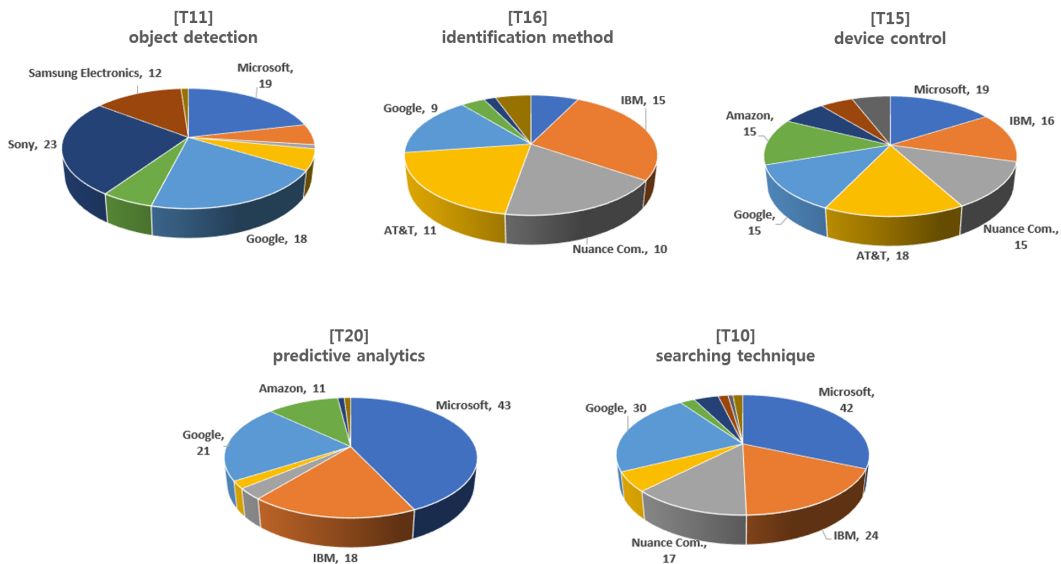
	Language understanding			Speech technology						Computer vision		Data analysis					Motion control		Machine learning	
	[T1]	[T3]	[T7]	[T4]	[T9]	[T13]	[T14]	[T18]	[T19]	[T11]	[T16]	[T2]	[T5]	[T6]	[T10]	[T17]	[T8]	[T15]	[T12]	[T20]
Microsoft	28	23	26	4	11	12	13	15	19	19	4	41	12	5	42	5	12	19	46	43
IBM	25	28	75	6	6	10	11	19	12	5	15	17	7	7	24	5	10	16	11	18
Nuance Comm.	10	27	11	19	31	18	29	16	24	1	10	14	15	23	17	9	9	15	2	3
AT&T	6	13	5	60	40	10	4	37	6	5	11	6	11	1	7	3	1	18	6	2
Google	16	11	11	12	3	9	13	14	14	18	9	14	9	3	30	10	2	15	7	21
Amazon	3	3	1	0	10	6	4	4	21	5	2	5	11	4	3	0	2	15	12	11
Sony	1	15	1	2	0	11	4	0	3	23	1	8	3	6	5	1	3	8	3	1
Samsung Electronics	4	11	0	2	1	2	18	2	7	12	0	3	4	2	2	5	2	6	0	0
Apple	1	4	0	0	0	0	2	2	7	0	0	0	0	0	1	2	0	7	0	0
Facebook	7	2	1	0	0	0	0	0	0	1	3	0	0	0	2	0	0	0	3	1

합쳐 특허수가 많은 1위부터 3위까지의 업체를 우수기업으로 표기하였다. 인공지능 기술 분야별 우수기업을 살펴보면 Microsoft가 데이터분석, 동작제어, 기계학습 분야에서 모두 1위를 차지했다. 특히 기계학습 분야는 2위 IBM에 두 배 이상 앞선다. 앞에서 살펴본 바와 같이 자연어 이해 기술이 우수한 IBM은 언어이해 분야에 선두를 차지했다. 음성처리 분야는 오랜 기간 음성 기술을 연구한 AT&T가 가장 우수했으며 시각이해 분야에선 비교적 신생기업인 Google이 선두를 차지했다. 특히 시각이해 분야에서만 2위를 차지한 Sony가 눈에 띈다.

〈표 5〉 인공지능 분야별 우수 기업

AI Sector	Leading companies		
Language understanding	IBM	Microsoft	Nuance Communications
Speech technology	AT&T	Nuance Communications	Microsoft
Computer vision	Google	Sony	Microsoft
Data analysis	Microsoft	Nuance Communications	Google
Motion control	Microsoft	IBM	Nuance Communications
Machine learning	Microsoft	IBM	Google

인공지능 기술은 발전 속도가 무척 빠르며 시장에서 새로운 기술 확보를 위해 기업 간에 치열한 경쟁을 하고 있는 분야다. 따라서 인공지능 분야에서 최근 성장하고 있는 유망 토픽에 대해서 추가적인 분석을 수행하였다. (그림 8)은 다섯 개의 유망 토픽에 대한 주요 기업별 인공



(그림 8) 유망 토픽에 대한 글로벌 IT기업의 특허 수

지능 특허 수를 나타낸 것이다.

가장 눈에 띄는 것은 Google이다. Google은 5개 토픽, 모든 유망 기술 보유에서 상위권을 차지했다. 이 회사는 ‘AI 혜택을 모두에게(Bringing the benefits of AI to everyone)’ 구호아래 최근 AI 기술에 집중하고 있음을 알 수 있다. Google은 대부분의 유망 기술에서 선두 주자라 할 수 있을 것이다. 또한 Microsoft와 IBM도 4개 토픽에서 상위권에 위치하여 인공지능 유망 기술에서 강자임을 알 수 있다. 특히 Microsoft는 기계학습 분야의 [T20] predictive analytics에서 압도적 선두였다. IBM은 유망 토픽 중 데이터분석 분야의 [T16] identification method에서 가장 많은 특허를 보유하고 있다. 시각이해 분야의 [T11] object detection에서는 Sony와 삼성전자가 상당 비중을 차지했는데 두 기업 모두 TV, 카메라 같은 영상 기기 분야의 선두 업체라는 공통점이 있다. [T10] searching technique에서는 예상을 깨고 Google보다 Microsoft가 더 많은 특허를 보유하고 있었다. 동작제어 분야의 [T15] device control에서는 인공지능 비서와 스피커 서비스를 제공하는 Amazon, Google, Apple 이 같은 비중을 차지하고 있다. 단, 음성인식 엔진을 Apple에 제공하는 Nuance Communications는 Apple로 간주했다.

요약하면 인공지능 유망 토픽 기술에서는 글로벌 시가총액 상위 IT기업인 Microsoft, Google, Amazon, 삼성전자, IBM 등이 선전하고 있다. 반면 Apple과 Facebook은 시가총액 규모나 명성에 비해 상대적으로 낮은 비중을 차지하고 있다. 오히려 영상 기기 등을 생산하는 Sony, 통신 기업인 AT&T, 음성인식 엔진 개발 기업인 Nuance Communications 등이 선전하고 있는 것으로 나타났다.

VI. 결 론

본 연구에서는 미국 특허청에서 인공지능 관련 기업 특허 2,589건을 수집하였으며 제목과 초록 데이터를 활용하여 토픽 모델링을 수행하여 인공지능 분야의 세부 기술을 추출하여 동 분야의 기술구조를 정의하였다. 또한 각 토픽의 연도별 변화 추이를 분석하여 5개 유망토픽과 4개 쇠퇴 토픽을 도출하였다. 글로벌 IT기업의 토픽별 기술 비중과 보유 특허수를 바탕으로 기업별 인공지능 기술역량을 분석하고, 인공지능 분야별 우수 기업을 도출하였다.

인공지능 기술은 지난 십여 년간 급속한 성장을 하고 있다. 글로벌 IT 기업들은 새로운 성장 동력이자 경쟁력의 핵심으로 인공지능 연구개발에 매진하고 있다. 이런 환경 속에서 산업정책과 기술전략 수립을 위해서는 인공지능 기술 구조 및 기업 기술역량 파악이 요구된다. 본 연구는 특허 데이터를 활용한 토픽 모델링 분석을 통해 인공지능 기술구조를 도출하고, 글로벌 IT

기업의 기술역량을 분석함으로써, 인공지능 기술별 우수 기업과 기업별 강점 분야를 실증적으로 도출하였다. 본 연구를 통해 도출된 인공지능 기술구조와 기업별 기술역량 정보는 인공지능 기술을 개발하는 공급 기업뿐만 아니라 이를 활용하는 수요 기업들의 기술기획 및 전략 수립에도 매우 유용하게 활용될 수 있다.

그러나 본 연구는 향후연구가 필요한 몇 가지 한계점도 지니고 있다. 첫째, 기술 역량을 분석함에 있어 특허의 건수만을 고려하였기에 특허의 질적인 측면을 반영하지 못 하였다. 피 인용수, 패밀리 특허 수 등 특허의 품질 지표들을 활용하여 기업별 기술역량을 측정할 필요가 있다. 둘째, 인공지능 기술이 급속히 발전함에 따라 기업에서 수행하는 선행 연구가 특허가 아닌 논문 형식으로 발표되는 경우가 많아지고 있다. 따라서 기업의 기술 역량을 분석하는데 특허 데이터와 더불어 논문 데이터도 함께 살펴볼 필요가 있다. 셋째, 인공지능 특허를 많이 보유한 글로벌 IT기업 뿐 아니라 소규모 기업도 분석할 필요가 있다. 새로운 기술을 근간으로 창업하여 성장하는 신생기업이나 영향력이 있는 벤처기업 등을 포함하여 연구의 범위를 확장하는 것이 필요하다.

참고문헌

- 강전학·이학연 (2018), “특허 정보를 활용한 클라우드 컴퓨팅 기술 구조 분석”, 「대한산업공학 회지」, 44(1): 69-81.
- 김봉선·김언수 (2014), “특허기술의 특성과 가치의 관계”, 「전략경영연구」, 17(3): 163-181.
- 김태경·최희련·이홍철 (2016), “토픽 모델링을 이용한 핀테크 기술 동향 분석”, 「한국산학기술 학회논문지」, 17(11): 670-681.
- 박재용 (2018), “특허정보를 이용한 인공지능 기술 동향 분석”, 「한국컴퓨터정보학회논문지」, 23(4): 9-16.
- 박주섭·홍순구·김종원 (2017), “토픽모델링을 활용한 과학기술동향 및 예측에 관한 연구”, 「한국산업정보학회논문지」, 22(4): 19-28.
- 임지연 (2016), “글로벌 인공지능 SW 기술 개발 동향”, 「한국차세대컴퓨팅학회논문지」, 12(4): 33-46.
- 정명석·이주연 (2018), “Latent Dirichlet Allocation (LDA) 모델 기반의 인공지능(A.I.) 기술 관련 연구 활동 및 동향 분석”, 「한국산업정보학회논문지」, 23(3): 87-95.
- 정보권·이학연 (2016), “국내 산업공학 연구 주제 2001~2015”, 「대한산업공학회지」, 42(6):

421-431.

- Andrzejewski, D., Mulhern, A., Liblit, B. and Zhu, X. (2007), "Statistical Debugging Using Latent Topic Models", *Proceedings of European Conference on Machine Learning*, 6-17.
- Anusuya, M. A. and Katti, S. K. (2009), "Speech Recognition by Machine: A Review", *International Journal of Computer Science and Information Security*, 6(3): 181-205.
- Bates, M. (1995), "Models of Natural Language Understanding", *PNAS*, 92: 9977-9982.
- Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer.
- Blei, D. M. (2012), "Probabilistic Topic Models", *Communications of the ACM*, 55(4): 77-84.
- Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003), "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, 3: 993-1022.
- Britannica (2019), "Encyclopedia Britannica Article: Artificial Intelligence", <https://www.britannica.com/technology/artificial-intelligence> (20 March 2019).
- Brynjolfsson, E., Rock, D. and Syverson, C. (2017), "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics", *NBER Working Paper*.
- Cincera, M. (1997), "Patents, R&D, and Technological Spillovers at the Firm Level: Some Evidence from Econometric Count Models for Panel Data", *Applied Econometrics*, 12: 265-280.
- Drucker, P. F. (1995), *Managing in a Time of Great Change*, Truman Talley Books: New York.
- French, C. S. (1996), *Data Processing and Information Technology*, Thomson.
- Fujii, H. and Managi, S. (2018), "Trends and Priority Shifts in Artificial Intelligence Technology Invention: A Global Patent Analysis", *Economic Analysis and Policy*, 58: 60-69.
- Gaikwad, S. K., Gawali, B. W. and Yannawar, P. (2010), "A Review on Speech Recognition Technique", *International Journal of Computer Applications*, 10(3): 16-24.
- Griffiths, T. L. and Steyvers, M. (2004), "Finding Scientific Topics", *Proceedings of the National Academy of Sciences*, 101: 5228-5235.
- Hornik, K. and Grün, B. (2011), "Topicmodels: An R Package for Fitting Topic Models", *Journal of Statistical Software*, 40(13): 1-30.

- IITP (2016), *ICT Long Term Technology Road Map 2022*.
- Kim, J., Jun, S., Jang, D. and Park, S. (2018), “Sustainable Technology Analysis of Artificial Intelligence Using Bayesian and Social Network Models”, *Sustainability*, 10(1): 155.
- Lee, H. and Kang, P. (2018), “Identifying Core Topics in Technology and Innovation Management Studies: a Topic Model Approach”, *Technology Transfer*, 43(5): 1291-1317.
- Lee, H., Seo, H. and Geum, Y. (2018), “Uncovering the Topic Landscape of Product-Service System Research: from Sustainability to Value Creation”, *Sustainability*, 10: 911.
- Maresch, D., Fink, M. and Harms, R. (2016), “When Patents Matter: The Impact of Competition and Patent Age on the Performance Contribution of Intellectual Property Rights Protection”, *Technovation*, 57: 14-20.
- Mejia, A. (2019), “Language Understanding”, <https://azure.microsoft.com/en-us/services/cognitive-services/language-understanding-intelligent-service/> (9 March 2019).
- Russell, S., Dewey, D. and Tegmark, M. (2015), “Research Priorities for Robust and Beneficial Artificial Intelligence”, *AI Magazine*, 36(4): 105-114.
- Samuel, A. L. (1959), “Some Studies in Machine Learning Using the Game of Checkers”, *IBM Journal of Research and Development*, 3(3): 210-229.
- Szeliski, R. (2010), *Computer Vision: Algorithms and Applications*, Springer.
- Winograd, T. (1972), “Understanding Natural Language”, *Cognitive Psychology*, 3(1): 1-191.
- Witten, I. H., Frank, E., Hall, M. A. and Pal, C. J. (2016), *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann.

이왕재

연세대학교에서 공학 학사, 석사학위를 취득하였으며, 현재 서울과학기술대학교 IT정책전문대학원 박사 과정에 있다. 삼성전자 연구소인 삼성리서치의 수석 연구원으로 재직 중이며 관심분야는 인공지능, 빅데이터, 기술전략 등이다.

이학연

서울대학교 산업공학과에서 학사학위를 취득하였으며, 동대학원에서 박사학위를 받았다. 현재 서울과학기술대학교 글로벌융합산업공학과 부교수로 재직 중이다. 주요 연구 분야는 기술예측, 기술전략, 연구개발 성과평가 등이다.