

# 토픽모델링을 활용한 과학기술동향 및 예측에 관한 연구

## (A Study on Science Technology Trend and Prediction Using Topic Modeling)

박 주 섭<sup>1)</sup>, 홍 순 구<sup>2)\*</sup>, 김 종 원<sup>3)</sup>

(Park Ju Seop, Hong Soon-Goo, and Kim Jong-Weon)

**요 약** 기업이나 정부에서는 연구나 기술 동향을 파악하고 예측하기 위해 주로 델파이 기법이 활용되어 왔다. 이 기법은 많은 시간과 비용이 소요되는 단점이 있기에 본 논문에서는 LDA 토픽모델링 기법을 활용하여 과학기술의 동향 및 예측에 관한 연구를 실시하였다. 이를 위해 미국 특허 문서 중 AI(Artificial Intelligence) 초록을 대상으로 LDA 토픽모델링 기법을 활용하여 20개의 AI 세부기술을 추출하였다. 도출된 세부기술에 대해 핵심기술을 파악하고, 연도별 비중 추이 분석을 통하여 Hot 기술과 Cold기술을 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 논문에서 제시한 방법론은 사회문제나 지역혁신, 경영 등 다양한 분야에서 동향분석이나 정책 도출 또는 기술 수요 예측에 활용되어 질 수 있을 것이다.

**핵심주제어** : 과학기술예측, LDA 토픽모델링, AI(Artificial Intelligence), 특허분석

**Abstract** Companies and Governments have Mainly used the Delphi Technique to Understand Research or Technology Trends. Because this Technique has the Disadvantage of Consuming a Large Amount of Time and Money, this Study Attempted to Understand and Predict Science and Technology Trends using the Topic Modeling Technique Latent Dirichlet Allocation (LDA). To this end, 20 Specific Artificial Intelligence (AI) Technologies were Extracted From the Abstracts of the US Patent Documents on AI. With Regard to the Extracted Specific Technologies, Core Technologies were Identified, and then these were Divided into Hot and Cold Technologies though a Trend Analysis on their Annual Proportions. Text/Word Searching, Computer Management, Programming Syntax, Network Administration, Multimedia, and Wireless Network Technology were Derived From Hot Technologies. These Technologies are Key Technologies that are

---

\* Corresponding Author : shong@dau.ac.kr

이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015S1A3A2046781)

Manuscript received May 30, 2017 / revised Jun 16, 2017 / accepted Aug 17, 2017

1) 동아대학교 경영정보학과, 제1저자

2) 동아대학교 경영정보학과, 교신저자

3) 동의대학교 정보경영학부

Actively Studied in the Field of AI in Recent Years. The Methodology Suggested in this Study may be used to Analyze Trends, Derive Policies, or Predict Technical Demands in Various Fields such as Social Issues, Regional Innovation, and Management.

**Key Words :** Science and Technology Predictions, Latent Dirichlet Allocation (LDA), Topic Modeling, Artificial Intelligence (AI), Patent Analysis

## 1. 서 론

과학기술은 국가 경쟁력의 핵심요소이다. 세계 각국은 경제 성장을 위해 과학기술의 핵심 과제를 선정하여 지속적인 R&D 투자를 하고 있다. 우리나라에서도 R&D 투자가 2016년을 기준으로 19조원을 넘었고, 이는 1인당 GDP 대비 세계 1 위이다. 총금액으로 보면 R&D 투자가 경제규모에 비해 큰 편이다. IMD(International Institute for Management Development) 보고서[9]에 의하면 우리나라 기초과학 수준이 30위권으로 R&D 투자에 비해 실적이 부실하다. 핫 이슈가 발생하면 임기응변으로 밀어주기식 투자 형태가 반복되고 있는 것도 부실의 원인이 되고 있어 합리적인 R&D 투자에 대한 연구가 필요한 시점이다. 제대로 된 과학기술 정책을 장기적으로 수립하기 위해서는 과학기술 예측이 선행되어야 한다. 이를 바탕으로 미래 유망기술을 예측하여 선택하고 자원을 집중적으로 투입함으로써 큰 성과를 이뤄낼 수 있다.

유망 기술을 예측하는 방법은 델파이 기법을 이용하여 전문가들의 의견을 집약하는 방식이 주로 활용되어 왔다[13]. 델파이 기법은 절차의 복잡성과 시간 및 경제적 비용 증가, 소수 전문가의 과도한 영향력 발휘에 의한 왜곡 등 많은 단점이 있다[14]. 본 연구에서는 토픽모델링 기법을 활용하여 델파이 기법의 단점을 극복하고자 한다.

토픽모델링은 기존 정성적 분석의 한계를 극복하고 대량의 문서들에서 숨겨진 주제를 찾아내기 위하여 제시된 통계적 방법론으로 한 개의 문서를 한 개의 토픽으로만 할당하는 일반적인 접근법과는 달리 한 개의 문서에 여러 개의 토픽이 존재할 수 있으므로 융합성격의 기술 문서를 분

석하는 데 유용하다[3]. 토픽모델링 중 가장 대표적인 방법론은 LDA(Latent Dirichlet Allocation) 기법으로 산업공학, 문헌정보학, 전산학 등 기술 예측 연구에 활발히 활용되고 있다.

본 연구의 목적은 LDA 토픽모델링을 활용하여 과학기술동향 및 예측을 위한 분석 방법론을 제시하는 것이다. 이를 위해 기술동향 및 예측 분석에 널리 활용되고 있는 AI 특히 초록을 이용하여 LDA 토픽모델링으로 실증분석을 실시한다.

## 2. 이론적 배경

토픽모델링은 여러 문서들 중에서 주제를 찾아내고, 유사한 단어들끼리 군집화하는 기법으로 문서의 주제를 찾는 데 사용된다. 토픽모델링 기법중 하나인 LDA(Latent Dirichlet Allocation) 분석은 광범위하고 비정형적인 문서 집합에 잠재되어 있는 주제들을 발견하기 위한 통계적 알고리즘으로, 간단하고 보편적으로 사용되고 있는 토픽 모델링 기법이다[10]. LDA 알고리즘은 문서내 단어들을 대상으로 문서 구조와 보이지 않는 변수를 추론하고, 문서 집합내의 단어들에 대한 디리클레(dirichlet) 분포를 통하여 지정한 수만큼의 토픽을 생성한다.

Fig.1은 잠재 디리클레 할당을 위한 과정을 형식적으로 표현한 것으로, 관측 값  $W_{d,n}$ 을 통하여 잠재변수인  $\theta_d$ ,  $Z_{d,n}$ 와  $\beta_k$ 를 찾는 것이 목적이다.  $\theta_d$ 의 파라미터 값은  $\alpha$ 이고  $\theta_d$ 의 분포로 디리클레 분포를 따른다. 문서d의 n번째 단어에 대한 토픽  $Z_{d,n}$ 은  $\theta_d$ 를 통해 얻어지고, 단어  $W_{d,n}$ 은  $Z_{d,n}$ 와 전체 토픽  $\beta_{1:k}$ 로부터 얻을 수 있다.

이러한 LDA 토픽모델링은 단어 사이의 동시

발생 빈도를 바탕으로 은닉 변수를 학습하는 비감독 학습법을 사용하여 문서의 토픽을 얻는 것이 특징이며, 새로운 문서에 대한 경험적 접근 없이도 쉽게 확률을 적용시킬 수 있는 것이 장점이다. <Fig. 1>에서는 LDA 토픽 모델링의 기본적인 개념을 잘 보여준다. 그 과정은 첫 번째로 맨 오른쪽 도표와 같이 주제에 대한 분포를 선택하고, 두 번째로 문서의 개별 단어마다 주제를 할당한 뒤, 그에 상응하는 주제에서 단어를 선택하는 것이다.

토픽모델링을 바탕으로 선행연구들을 살펴보면 학술분야, 기술분야, 사회분야로 세 가지로 분류될 수 있다.

첫째로, 학술분야에서 보면 Park et al.[2]은 국내 문헌정보학 분야의 연구동향을 살펴보기 위해 주요 학술지의 논문 초록을 수집하여 LDA기반의 토픽모델링 실험을 수행하였다. 기존의 양적 분석과 내용분석이 연구자의 주관적 관점에서 진행되는 단점을 보완하였다는 특징이 있다. Jeong et al.[6]은 국내 산업공학 분야의 연구 주제를 탐색하기 위해 대한산업공학회 발간 학술지에 게재된 논문 초록 정보를 바탕으로, LDA 토픽모델링을 적용하여 산업공학의 연구 주제를 식별하였으며, 산업공학 분야 내의 세부 연구 주제들의 내용 및 분포를 파악하고 이들의 동적 변화를 분

석하였다는 데 의의가 있다. Kim et al.[1]은 국내의 주요 산업공학 학술지의 영문 초록 정보를 이용하여 국내 산업경영공학의 동향 분석을 수행하였으며, 산업공학의 전체적인 흐름 및 변화를 살펴보았다는 데 의의가 있다.

둘째로, 기술분야에서 보면 Seo et al.[3]은 특허 데이터를 이용하여 핀테크 기술의 최신 트렌드를 도출하고, 핀테크 기술 간 연계 구조를 분석하기 위해 금융 분야의 BM(business method) 특허인 미국 특허청 705/35 클래스에 등록된 특허 7000여 건을 수집하여 20개의 핀테크 하부기술에 대한 동향분석을 하였다. 핀테크 산업에 대한 연구가 주로 거시적 측면을 다루어 왔으나, 이 논문은 핀테크 기술에 대한 학술적 연구를 하였다는 데 의의가 있다. Jeong et al.[7]은 증강현실기술의 특허들과 토픽모델링 기법을 활용하여, 증강현실기술의 주요 기술토픽을 추출하고 추출된 토픽의 기술발전 동향 및 기업별 잠재적 특허 경쟁전략을 분석하였다.

셋째로, 사회분야에서 Jin et al.[8]은 트위터에서 상품명에 관한 키워드를 추출한 후, 동시출현 단어분석을 이용하여 네트워크로 표현하였으며 시계열 분석과 LDA 토픽모델링 분석을 실시하였고, 이 연구방법을 활용하면 소비자 의견을 신속히 파악하고 즉각적 의사결정을 지원할 수 있

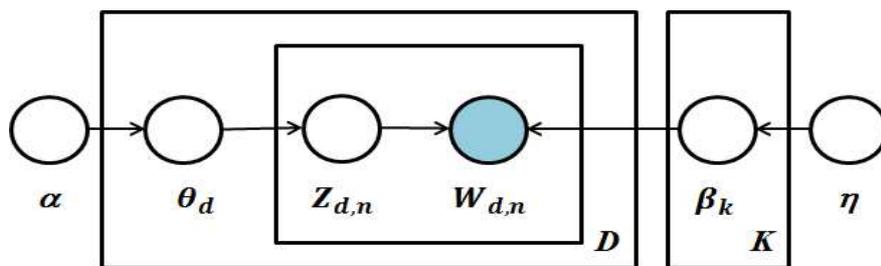


Fig. 1 Document creation process of LDA[11]

- D : 문서집합
- d : 개별문서
- k : 토픽의 수
- $\beta_k$  : k번째 토픽의 단어 분포의 확률.
- $\theta_d$  : 문서d에 대한 토픽 비율
- $Z_{d,n}$  : 문서d의 n번째 단어에 대한 토픽할당.
- $W_{d,n}$  : 문서d에서 관찰되어진 n번째 단어
- $\alpha$  : 양의 K-벡터 모수

는 역할도 가능하다. Lee et al.[4]은 트윗 데이터를 이용하여 트렌드를 추출하는 방법으로 LDA 토픽 모델을 적용하였으며 리트윗과 댓글을 이용한 그래프를 생성하고, 유사한 사용자를 그룹핑하여 데이터를 표현함에 따라 LDA 토픽 모델에 맞는 데이터마이닝 기법을 제안하였다. Jeong et al.[5]은 온라인 서점의 소개 글을 크롤링하였고, 본격문학과 장르소설 분야에서 1155권의 소개글과 서명, 저자, 출판년도 등의 메타데이터를 수집하여 분석 실시하였으며, 많은 양의 소설책으로부터 자동으로 토픽 키워드를 추출하고 이 토픽별로 소설책을 보여줘 이야기별로 소설을 탐색할 수 있는 방안을 제시하였다.

### 3. 연구방법론

#### 3.1 분석의 틀

본 연구의 목적을 달성하기 위해 2장에서 이론적 배경을 바탕으로 과학기술동향 및 예측을 위한 분석 방법론의 절차는 <Fig. 2>와 같다.

#### 3.2 분석 개요

특허는 일정기간 동안 발명자의 기술에 대한 독점적 권리가 보장되는 지식재산이다. 보통 특허 기술이 상용화 되는 데는 많은 시간이 소요되는데, 이는 특허 자체가 미래 기술의 의미를 내포하고 있다고 볼 수 있다. 따라서 특허 분석을 통해 미래기술에 대한 효과적인 예측이 가능하다. 특히 미국은 AI(Artificial Intelligence) 분야에서 탁월한 기술을 보유하고 있어 본 논문에서는 미국 AI 특허를 분석하여 AI 기술을 예측하

고자 하였다. 이에 따라 AI 분야의 기술예측을 위해 미국 특허 DB의 초록을 분석대상으로 하였다. 주제어인 ‘Artificial Intelligence’를 키워드로 2000년 이후 미국 특허 DB에서 특허명, 특허출원일자, 특허초록 등 14,187개의 특허문서를 연단위로 추출하였다. 수집된 특허문서 중 초록들을 대상으로 전처리 작업을 수행하여 표제어(Lemma)를 추출하였다. 이후 표제어에 대해 LDA 토픽모델링 기법을 활용하여 AI 기술의 토픽을 추출한 후, 이를 AI 세부 기술로 정의한다. 추출된 AI 세부 기술의 비중도 분석 및 Hot/Cold 토픽 분석을 통해 기술 흐름을 예측하였다.

#### 3.3 데이터 수집 및 전처리

기술동향을 분석하기 위해 수집된 14,187개 (2000년~2016년)의 미국 특허 초록을 연도별 분류 작업을 하였다. 연도별 초록 개수는 <Fig. 3>에서 보는 바와 같이 2000년 이후 특허가 연 200개 이상 출원되어 활발한 연구가 이루어지기 시작하여 2016년 1810개에 이른다. 미국 특허 데이터를 특허정보 검색서비스인 키프리스(Kipris)에서 키워드 검색 후 온라인다운로드 서비스를 통해 발명의 명칭, 출원 번호와 일자, 등록 번호와 일자, IPC(국제특허분류), 초록 등을 가져왔다. 특허 초록의 전처리 과정은 <Fig. 4>와 같다.

특허 초록을 단어 단위로 나누기 위해 공백을 기준으로 단위를 분리하였다. 문서 내에 나타난 불필요한 특수문자, 관사, 조사 등의 불용어는 삭제하고 각 키워드들에 품사를 할당한 뒤에 명사들만 선택하였다. 도출된 키워드들의 레이블명을 정하기 위해 표제어선정(Lemmatization)을 하였다.

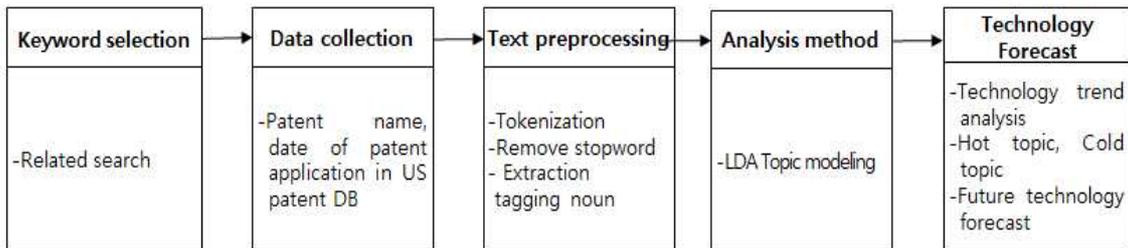


Fig. 2 Research framework

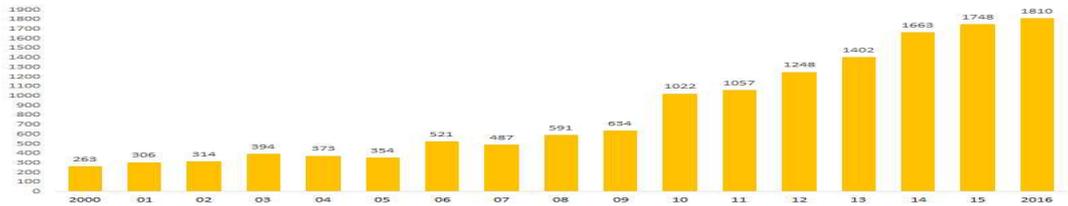


Fig. 3 Number of AI technology patents by year(2000~2016)



Fig. 4 Text preprocessing

### 3.4 AI(Artificial Intelligence) 기술 추세 분석

#### 3.4.1 AI(Artificial Intelligence) 세부 기술 추출

AI 세부 기술을 추출하기 위해 전처리과정을 끝낸 특허 초록을 대상으로 오픈소스인 LDA 토픽모델링 프로그램을 사용하여 토픽을 추출하였다. 이 과정에서 적절한 토픽수와 샘플링 반복횟수는 연구자가 결과를 효과적으로 해석할 수 있는 수준에서 결정할 수 있다[12]. 본 연구에서는 토픽의 수를 20개, 샘플링 반복 횟수를 1,000회로 수행하였고 LDA 프로그램에 사용된  $\alpha$ ,  $\beta$ 는 기본 값을 사용하였다.

AI 세부 기술로 추출된 20개의 토픽과 토픽내 확률 값이 높은 단어는 아래 <Table 1>에 있다. 토픽의 레이블명은 확률 값이 높은 단어들을 기본으로 추론하여 정의하였다. 토픽의 주제는 로봇공학, 비디오, 프로그래밍, 컴퓨터 관리, 문제해결 모델링, 객체인식, 게임, 가상화, 미적분학 등 다양한 AI 세부 기술을 포함하고 있다. 정의된 20개의 세부 기술들을 대상으로 추후 분석을 실시하였다.

#### 3.4.2 AI(Artificial Intelligence) 세부 기술 비중도

AI 분야의 핵심 기술을 파악하기 위해 특허

초록을 대상으로 2000년부터 2016년까지 세부 기술별 점유율(비중)을 분석한 결과는 <Table 2>와 같다. AI 분야의 핵심 세부 기술로 [T14] Problem solving method, [T7] SQL, [T5] Computer management, [T2] Text/Word searching 기술 등의 순으로 높은 비중을 보였다. 반면에 [T16] Game, [T6] Programming syntax, [T10] General networking, [T4] Programming 기술 등의 순으로 낮은 비중을 보였다.

#### 3.4.3 Hot/Cold AI(Artificial Intelligence) 세부 기술

미국 특허 초록을 연단위로 분류한 후 개별 기술(토픽) 비중(값)을 분석한다. 즉 전 기간에 대한 연도별 비중 추이를 분석하여, 개별 기술(토픽)의 값이 상승추세를 보이는 Hot 기술과 하락 추세를 보이는 Cold 기술을 도출한다. 개별 기술(토픽)의 추세를 판단하는 기준은 선형 회귀분석의 회귀계수 값을 사용한다. 독립변수는 연도로 하고, 종속변수는 개별 기술(토픽)의 연도별 비중 평균값을 사용하여 단순회귀 분석을 수행한다. 유의 수준은 0.05에서 통계적으로 유의하다고 판단한다. 유의미한 개별 기술(토픽)을 대상으로, 회귀계수가 (+)면 Hot 기술, (-)면 Cold 기술로 나타낸다. 개별 기술에 대한 회귀 분석 결과는 <Table 3>과 같다. 개별 기술의 연도별 추이를 보면 <Fig. 5>와 같이 6개의 Hot 기술과 <Fig.

Table 1 20 topics and keywords in topics

[T1] Robotics	[T2] Text/Word searching	[T3] Video	[T4] Programming	[T5] Computer management
sensor power control robot fluid light system	user search context input display be interest	display frame motion screen camera video color	code processor program array element data address	product task system item tool work computer
[T6] Programming syntax	[T7] SQL	[T8] Text preprocessing	[T9] Network speed adm.	[T10] General networking
first second portion set parameter section end	data system input store event be set	document text speech web word file search	time level real high traffic line speed	network node neural data packet protocol computer
[T11] VLSI	[T12] Network administration	[T13] Object recognition	[T14] Problem solving method	[T15] Multimedia
signal input unit output circuit control pattern	server agent event client call request session	object region light subject scene interest detector	model rule state system process problem decision	content media audio video user music stream
[T16] Game	[T17] Virtual computer	[T18] Wireless network	[T19] Calculus	[T20] Health
game character layer player card risk account	virtual environment system real control world user	wireless access cell network radio station base	set target test vector point matrix graph	patient health treatment central blood condition person

6>과 같이 9개의 Cold 기술이 도출되었다. 6개의 Hot 토픽은 [T2]Text/Word searching, [T5] Computer management, [T6] Programming syntax, [T12]Network administration, [T15]Multimedia 그리고 [T18]Wireless network 기술이 있다.

최근 AI 산업에서 활발하게 연구되는 분야는 자율주행 자동차, 비서 서비스, 금융, 쇼핑, 사진이나 동영상 서비스 등이 있다. 자율주행 자동차

개발에서는 [T5]Computer management, [T6] Programming syntax, [T12]Network administration, [T15]Multimedia 그리고 [T18]Wireless network 기술이 주를 이룬다. 비서 서비스 분야에서는 [T2] Text/Word searching, [T5]Computer management, [T15]Multimedia 기술이 주를 이룬다. 금융 산업에서는 이상 금융거래 탐지시스템인 FDS(Fraud Dection System)와 투자자들의 거래 및 습성을

Table 2 Ranking of AI technology

AI Technology Topics	Weight	Rank
[T1] Robotics	5.84%	5
[T2] Text/Word searching	6.15%	4
[T3] Video	4.48%	14
[T4] Programming	4.03%	17
[T5] Computer management	6.25%	3
[T6] Programming syntax	3.67%	19
[T7] SQL	6.45%	2
[T8] Text preprocessing	5.17%	8
[T9] Network speed administration	4.69%	13
[T10] General networking	4.00%	18
[T11] VLSI	4.21%	16
[T12] Network administration	5.57%	6
[T13] Object recognition	4.83%	11
[T14] Problem solving method	6.61%	1
[T15] Multimedia	4.81%	12
[T16] Game	3.24%	20
[T17] Virtual computer	5.09%	10
[T18] Wireless network	5.14%	9
[T19] Calculus	5.48%	7
[T20] Health	4.31%	15

Table 3 Regression coefficient and significance level by AI technology

No	AI topic	Regression coefficient	P-value	Hot/Cold
1	[T1]	-0.422	0.027	Cold
2	[T2]	0.696	0	Hot
3	[T3]	0.164	0.357	X
4	[T4]	-0.733	0.001	Cold
5	[T5]	0.299	0.043	Hot
6	[T6]	0.333	0.029	Hot
7	[T7]	0.228	0.099	X
8	[T8]	0.061	0.691	X
9	[T9]	-0.466	0.003	Cold
10	[T10]	-0.333	0.023	Cold
11	[T11]	-0.740	0	Cold
12	[T12]	0.917	0	Hot
13	[T13]	-0.588	0.002	Cold
14	[T14]	-0.578	0.002	Cold
15	[T15]	0.887	0	Hot
16	[T16]	-0.027	0.695	X
17	[T17]	-0.498	0.003	Cold
18	[T18]	1.189	0	Hot
19	[T19]	-0.282	0.043	Cold
20	[T20]	-0.108	0.555	X

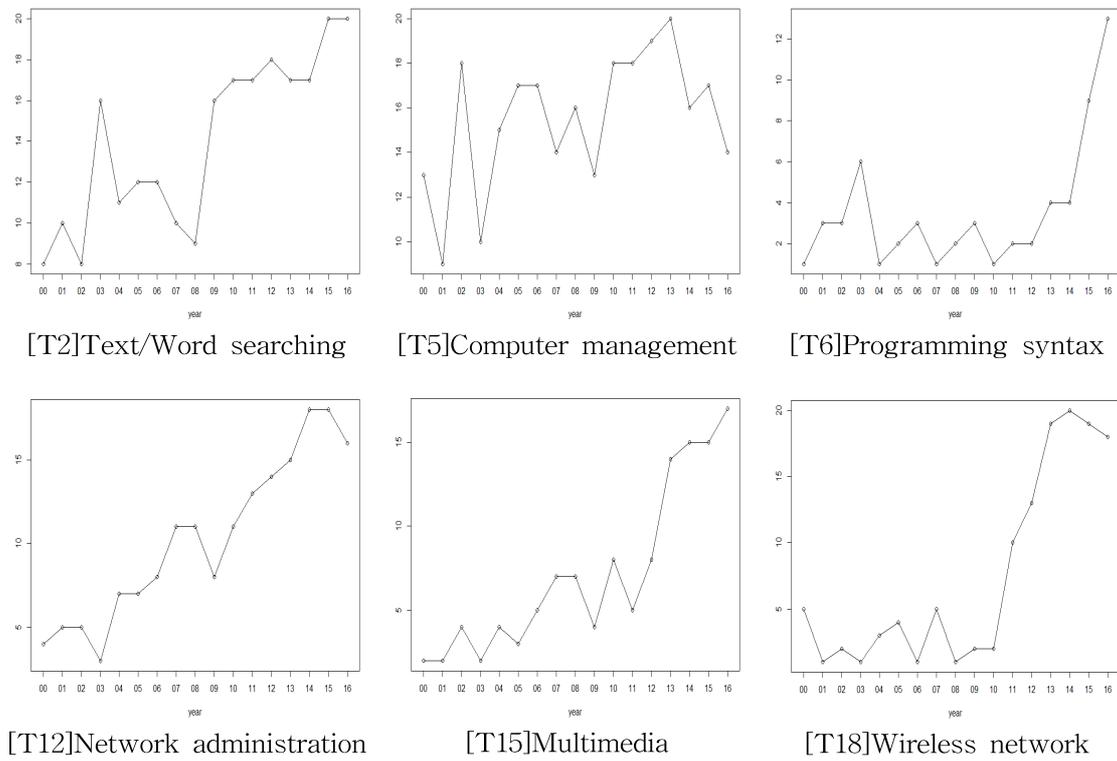


Fig. 5 Hot AI yearly trend by technology

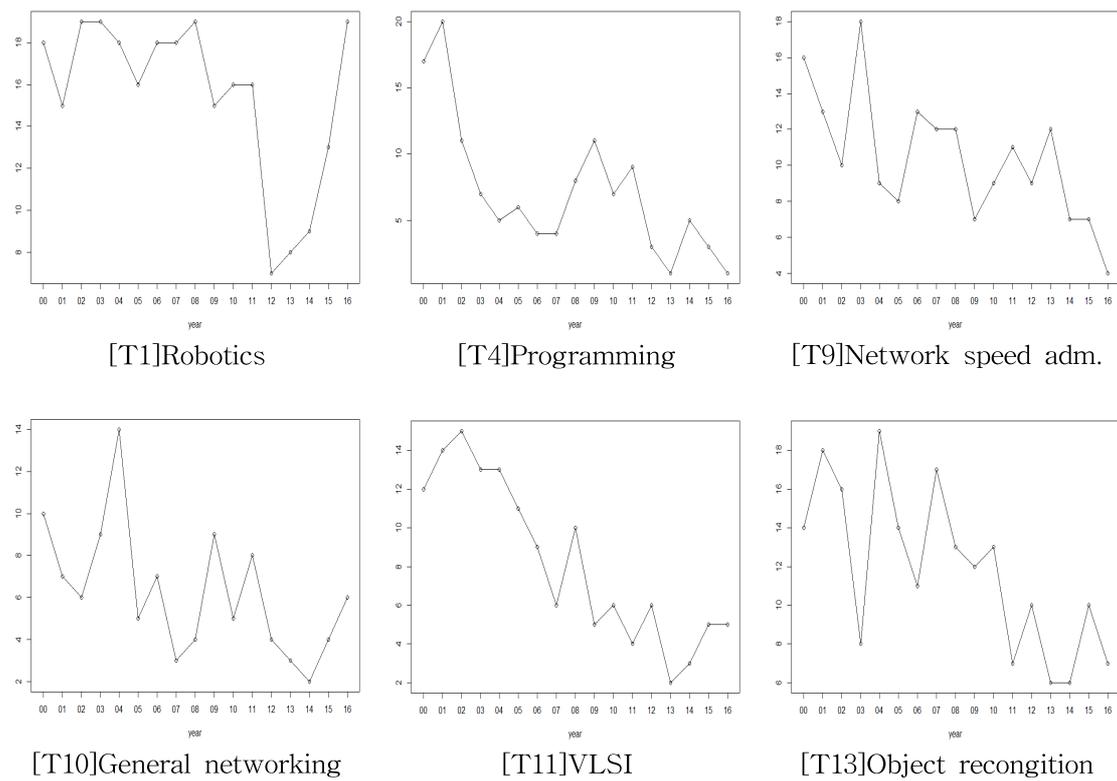


Fig. 6 Cold AI yearly trend by technology

분석하여 투자 방향을 제시해주는 로봇 어드바이저(Robot advisor)가 관심받는 분야이다. 이 두 분야는 [T5]Computer management, [T12]Network administration, [T15]Multimedia 기술이 주를 이룬다. 쇼핑 분야에서는 상품이나 영화, 드라마 등을 추천하기 위해 [T2]Text/Word searching 기술과 [T15]Multimedia 기술이 필요하다. 사진이나 동영상 분야에서는 패턴인식을 적용시키기 위해 [T2]Text/Word searching, [T6]Programming syntax, [T15]Multimedia 기술 등이 필요하다.

#### 4. 결론

본 연구는 LDA 토픽모델링을 활용하여 과학 기술동향 및 예측을 위한 분석 방법론을 제시하고 특히 데이터를 대상으로 실증분석을 실시하였다. LDA 토픽모델링 알고리즘을 이용하여 AI 세부 기술을 추출한 후 전체 기간(2000~2016년) 동안 핵심 기술을 도출하였으며, Hot/Cold AI(Artificial Intelligence) 세부 기술을 도출하였다. 텍스트 탐색, 컴퓨터 관리, 프로그래밍 구문, 네트워크 관리, 멀티미디어, 무선 네트워크 기술 등이 Hot 기술로 도출되었다. 이런 기술들은 최근 AI 산업에서 활발하게 연구되는 자율주행 자동차, 비서 서비스, 금융, 쇼핑, 사진이나 동영상 서비스 분야등에서는 핵심 기술들이다. 반면에 프로그래밍, 망관리, 일반 네트워킹, 대규모 집적 회로, 객체인식, 문제해결, 가상 컴퓨터, 미적분학 기술 등은 Cold 기술로 간주되었다. 1990년대 후반부터 머신러닝 개념이 소개되면서 수많은 데이터를 분석하여 인공지능 시스템 스스로 학습하는 형태로 진화했고, 2006년 이후 부터는 딥러닝으로 발전하게 된다. 그로 인해 Cold 기술은 머신러닝에 필요한 기술들과 많은 연관성이 있다. 특히 머신러닝에 사용되는 많은 데이터의 특징들을 추출하고 해석하는 기술이 요구된다. 즉 프로그래밍, 객체인식, 문제해결 그리고 미적분학 기술 등이 많이 사용된다. 딥러닝은 자동음성인식, 영상인식, 자연어처리 분야 등에서 활발한 연구가 진행 중이고, 이런 분야의 핵심 기술은 텍스트/단어 검색, 프로그래밍 구문, 멀티미디어 기술 등을 들 수 있다.

본 연구를 통해 과학기술 정책 수립을 위해 빅

데이터 분석 방법을 제시하였고, 후속 연구를 유발하였다는 학문적 기여도가 있다. 현재 기술 예측을 위한 방법은 정성적 방법과 정량적 방법으로 나뉘어진다. 델파이 기법과 같은 정성적 방법은 시간과 비용이 많이 소요되어 개인 연구에 제약이 따른다. 본 논문에서는 기술 예측을 위한 정량적 방법론을 제시하고, 미국 AI 특허 초록을 이용하여 실증 분석을 하였다는데 의의가 있다. 또한 최근 각광을 받고 있는 AI 산업 기술에 대한 초기의 학술 연구로서 의의가 있다고 할 수 있다. AI 기술 동향 및 예측 기법은 텍스트 마이닝을 활용하여 미래 기술을 예측하는 방법이다. 이런 방법론은 사회문제나 지역혁신, 경영 등 다양한 분야에서의 동향이나 예측에 활용이 가능할 것으로 생각된다. 실무적으로는 실제 행정기관의 과학정책 수립에 본 연구결과가 활용될 수 있을 것으로 기대된다. 그럼에도 불구하고 본 연구 방법이 과학 기술 예측에 있어서 최선의 방법인가에 대한 의문점이 있을 수 있다. 향후 연구과제로는 첫째, 기업이나 개인에게 맞는 텍스트 마이닝을 활용한 기술동향 및 예측을 위한 방법론이 개발된다면 예측 정확성을 높이고, 시간과 비용이 훨씬 줄어드는 효과가 발생할 것이다. 둘째, 특허 초록을 5년이나 10년 주기로 나누어서 분석한다면 기술 예측에 있어 좀 더 의미 있는 결과를 유추할 수 있을 것이다.

#### References

- [1] Kim, S. K. and Jang, S. Y., "A Study on the Research Trends in Domestic Industrial and Management Engineering using Topic Modeling", Journal of the Korea Management Engineers Society, Vol. 21, No. 3, pp. 71-95, 2016.
- [2] Park, J. H. and Song, M., "A Study on the Research Trends in Library & Information Science in Korea using Topic Modeling", Korea Society for Information Management, Vol. 30, No. 1, pp. 7-32, 2013.
- [3] Seo, S. H. and Lee, H. Y., "Fintech Trend Analysis using Topic Modeling of BM

Patents”, Proceedings of the Korean Industrial Engineering Society Fall Conference, pp. 471-480, 2015.

[4] Lee, S. Y. and Lee, K. M., “Trend Extraction using Topic Model Based on Reply Graph”, Korean Institute of Intelligent Systems, Vol. 24, No. 2, pp. 99-100, 2014.

[5] Jeong, D. M., “Study of Topic Classification and Keywords for Online Fiction Search”, Department of Transdisciplinary Studies, The Graduate School Seoul National University, 2015.

[6] Jeong, B. K. and Lee, H. Y., “Research Topics in Industrial Engineering 2001~2015”, Journal of the Korean Institute of Industrial Engineers, Vol. 42, No. 6, pp. 421-431, 2016.

[7] Jeong, B. K., Kim, J. W. and Yoon, J. H., “Patent-Based Competitive Intelligence Analysis of Augmented Reality Technology : Application of Topic Modeling”, Proceedings of the Korean Industrial Engineering Society Fall Conference, pp. 2262-2267, 2015.

[8] Jin, S. A., Heo, C. E., Jeong, Y. K. and Song, M., “Topic-Network based Topic Shift Detection on Twitter”, Korea Society for Information Management, Vol. 30, No. 1, pp. 285-302, 2013.

[9] “IMD World Talent Report 2015”, The IMD World Competitiveness Center, 2015.

[10] Blei, D. M., Ng, A. Y. and Jordan, M. I., “Latent Dirichlet Allocation”, The Journal of Machine Learning Research, Vol. 3, pp. 993-1022, 2003.

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

[12] Song, M. and Kim, S. Y., “Detecting the Knowledge Structure of Bioinformatics by Mining Full-Text Collections”, Scientometrics, Vol. 96, No. 1, pp. 183-201, 2013.

[13] Choi, Y. S., Baek, S. C. and Kwon, H. I., “Study on City Revitalization Plan Via U-City Project”, The Journal of Internet Electronic Commerce Research, Vol. 8, No. 3, pp. 183-209, 2008.

[14] Suh, C. K., Kim E. J. and Lee Y. S., “The Development of Delphi Support Decision System on the Web”, The Journal of Information Systems, Vol. 10, No. 1, pp. 5-25, 2001.



**박 주 섭** (Park Ju Seop)

- 울산대학교 전산학과 공학사
- 부산대학교 전산학과 이학석사
- 동아대학교 경영정보학과 박사과정
- 관심분야 : 빅데이터, 텍스트마이닝, 기술예측, 지역혁신, Co-creation



**홍 순 구** (Hong Soon-Goo)

- 종신회원
- 영남대학교 경영학과 경영학박사
- Univ. of Nebraska-Lincoln 경영학석사
- Univ. of Nebraska-Lincoln 경영학박사
- 동아대학교 경영대학 경영정보학과 교수
- 관심분야 : 중소기업의 정보화, 웹 접근성, 정보시스템 평가, RFID, Co-creation



**김 종 원** (Kim Jong-Weon)

- 종신회원
- 인하대학교 경영학과 경영학사
- Univ. of Nebraska-Lincoln, 경영학석사
- Univ. of Nebraska-Lincoln, 경영학박사
- 동의대학교 상경대학 경영정보학과 교수
- 관심분야 : SCM, ERP, BSC, CSR, CSV, 서비스 품질, 시스템 품질