

# 디지털 헬스케어 기술 구조 분석: Fractional Assignment Vs. Discrete Assignment

이서영<sup>1</sup> · 안승섭<sup>2</sup> · 이학연<sup>3\*</sup>

<sup>1</sup>서울과학기술대학교 IT정책전문대학원 / (주)웍스 기술평가센터

<sup>2</sup>서울과학기술대학교 데이터사이언스학과 / (주)컬리 물류전략팀

<sup>3</sup>서울과학기술대학교 산업공학과

## Technological Structure Analysis of Digital Healthcare: Fractional Assignment Vs. Discrete Assignment

Seoyoung Lee<sup>1</sup> · Seungseop Ahn<sup>2</sup> · Hakyeon Lee<sup>3</sup>

<sup>1</sup>The Graduate School of Public Policy and Information Technology, Seoul National University of Science and Technology / Technology Evaluation Center, WIPS Co., Ltd.

<sup>2</sup>Department of Data Science, Seoul National University of Science and Technology / Logistics strategy office, Kurly InC.

<sup>3</sup>Department of Industrial Engineering, Seoul National University of Science and Technology

This study identifies the technology structure of the digital healthcare field using patent data and compares the characteristics between techniques in the technical structure analysis. 2,064 patents obtained from USPTO, and analyzed by LDA to extract 15 technology topics, and documents classified as Doc2vec performed 15 technology groups through Spherical K-means clustering. As a result, LDA derived component technology of digital healthcare such as implantable medical devices, but Doc2ve-based clustering was confirmed promising technology of digital healthcare, including blockchain, telehealth service. This study are expected to be useful for establishing technology strategies and commercializing technologies for digital healthcare companies in the future.

**Keywords:** Digital Healthcare, Technology landscape, LDA Topic Modelling, Doc2vec, Fractional Assignment, Discrete Assignment

### 1. 서론

헬스케어(healthcare)는 질병의 치료를 주목적으로 하는 기존 의료 서비스와 질병의 예방 및 관리 영역을 모두 포함하는 전반적인 건강관리 서비스를 의미한다. 디지털 기술이 등장한

이후 헬스케어 산업에서도 디지털 기술을 활용하여 의료 서비스의 질을 개선하는 작업이 지속적으로 이루어져 왔다. 그러나 개인 스스로가 건강을 관리 및 예방하는 방향으로 헬스케어 패러다임이 변화함에 따라(Jeong *et al.*, 2016), 헬스케어 산업과 정보통신기술이 융합된 ‘디지털 헬스케어(digital health-

이 논문은 제1저자 이서영의 석사학위논문을 발췌 및 수정하여 작성한 것임. 이 연구는 한국연구재단의 이공분야기초연구사업(기본연구)의 지원을 받아 수행되었음(NRF-2021R1F1A1045787).

\* 연락저자 : 이학연 교수, 서울시 노원구 공릉로 232 서울과학기술대학교 프론티어관 607호, Tel : 02-970-6469, Fax : 02-974-2849,

E-mail : hylee@seoultech.ac.kr

2022년 9월 19일 접수; 2022년 10월 30일 수정본 접수; 2022년 11월 21일 게재 확정.

care)'가 주목받고 있다.

산업의 매력도가 제고됨에 따라 시장조사기관인 Frost & Sullivan에서는 세계 디지털 헬스케어 시장이 연평균 12% 수준으로 성장하여 2017년에 1,470억 달러에서 2023년에는 2,200억 달러로 확대될 것으로 전망한 바 있다. 특히 헬스케어 시장에서도 기술 활용을 통한 효율성 향상을 추진하고 있으며, 의료 서비스의 디지털화에 우호적인 규제 환경이 조성된 점도 시장 성장을 견인하고 있는 것으로 파악된다(Frost and Sullivan, 2019). 이처럼 디지털 헬스케어 기술이 기업들에게 새로운 수익창출 기회를 제공함에 따라 과거부터 헬스케어 산업에서 관련 사업을 영위하던 의료, 제약, 생명공학 기업은 물론 Apple, Google, Microsoft, Samsung을 비롯한 주요 ICT 기업에 이르는 다양한 산업군의 업체들이 디지털 헬스케어 시장에 진입하고 있다(Lupton, 2014). ICT 기업 및 헬스케어 기업들은 병원과 중형으로 연계하여 새로운 헬스케어 산업 생태계를 구축하고 있으며, 기업 뿐 아니라 미국, 중국 등 주요 국가들도 헬스케어 시장을 선점하기 위한 정책을 잇달아 내놓으면서 디지털 헬스케어 시장의 주도권 경쟁이 전개되고 있다(Lee, 2019).

헬스케어 시장을 둘러싼 경쟁이 심화되고 있는 가운데, 시장에서 경쟁우위를 확보하기 위해서는 디지털 헬스케어 기술에 대한 전략적인 접근이 요구된다. 즉, 기업이 연구개발을 수행하기에 앞서 자사의 기술경쟁력에 기초하여 발전 가능한 기술 영역을 알고 있다면 연구개발의 성공 확률을 높일 수 있으며, 신규 사업을 추진하고 있는 기업의 경우에도 산업을 구성하는 핵심 기술 및 기술 구조에 대한 이해를 토대로 사업전략을 수립할 수 있다. 그러나 지금까지 수행된 디지털 헬스케어 기술 분석 연구들은 키워드 분석, 동향 분석을 통해 단순히 기술의 현황을 보여주거나(Kim *et al.*, 2018; Kim and Kim, 2019) 사물인터넷(Lee *et al.*, 2019), 웨어러블 기기(Mück *et al.*, 2019)와 같이 특정 ICT 기술에 초점을 맞추어 헬스케어 분야를 파악하였다. 이에 따라 디지털 헬스케어를 구성하는 기술 전반을 탐색하여 기술 구조를 파악하고 핵심 기술을 도출하는 연구는 미진한 상황이다.

한편, 기술 문서로부터 기술 구조를 파악하려는 시도는 다양한 기술 도메인을 대상으로 활발히 이루어져 왔다(Seong and Lee, 2020). 과거에는 동시인용분석(co-citation analysis), 동시출현단어분석(co-word analysis), 동시분류분석(co-classification analysis) 등 전통적인 계량서지분석(bibliometric analysis) 기법이 주로 활용되어 왔으나, 최근에는 기계학습 기반의 텍스트 임베딩 기법의 활용도가 높아지고 있다(Kang and Lee, 2018). 이러한 기법들은 기술 문서로부터 세부 기술을 도출하는 방식에 따라 크게 비율 할당(fractional assignment) 방식과 이산 할당(discrete assignment) 방식으로 구분될 수 있다(Lee and Kang, 2018). 비율 할당 방식은 한 문서 내에 여러 가지 세부 기술이 존재한다고 가정하고, 각 문서로부터 세부기술의 비율을 추정하여 전체 문서집합(corpus) 내에서 각 세부기술의 비중을 추

출하는 방법이다. 반면 이산 할당은 하나의 문서가 하나의 기술만을 나타낸다고 가정하고, 기술의 유사성에 따라 각 문서를 군집화하여 세부기술을 추출하는 방식이다. 두 가지 접근 방식 모두 장단점을 가지고 있으나, 기술 구조 분석의 결과 및 해석 측면에서 구체적인 특성 비교가 이루어진 바 없어, 연구자들이 분석 목적에 따라 어떤 접근 방식을 취해야 하는지를 결정하기가 쉽지 않은 상황이다.

이에 본 연구는 특히 문서를 대상으로 비율 할당 방식과 이산 할당 방식을 모두 활용하여 디지털 헬스케어 기술 구조를 분석한다. 디지털 헬스케어 관련 특허 데이터를 대상으로 대표적인 비율 할당 방식인 LDA(latent dirichlet allocation) 토픽 모델링 기법과 이산 할당 방식인 Doc2vec 기반 클러스터 분석을 동시에 수행하여 핵심 기술을 도출하고, 기술 구조 분석 방식 간의 차이를 탐색한다.

이후 본 논문의 구성은 다음과 같다. 제2장에서는 디지털 헬스케어 기술 및 기술 구조 분석 관련 선행연구를 살펴보고, 제3장에서는 본 연구에서 사용한 연구방법을 설명한다. 제4장에서는 제시된 연구방법을 통해 디지털 헬스케어 기술 구조와 핵심 기술을 도출하고 제5장에서는 기법 간 비교를 통해 차이점을 분석한다. 마지막으로 제6장에서는 연구의 의의, 한계점 및 향후 연구방향을 포함한 결론을 제시한다.

## 2. 선행 연구

### 2.1 디지털 헬스케어 기술 분석

디지털 헬스케어의 개념은 학계 및 산업계에서 명확하게 정의하지 않아 연구자에 따라서 디지털 헬스케어와 스마트 헬스케어, e-health, m-health, u-health 등을 혼재하여 표기하고 있다(Kim and Kim, 2019). WHO를 비롯한 FDA, OECD 등의 국제 기구에서는 디지털 헬스(digital health)를 공식적으로 사용하고 있으나, 다수의 선행연구에서 넓은 범위에서 헬스케어 기술과 ICT 기술이 결합한 형태를 아우르는 개념으로 디지털 헬스케어를 사용하고 있다(Lee, 2014; Kim *et al.*, 2018; Kim *et al.*, 2021).

한편, 디지털 헬스케어 기술을 분석한 연구들은 주로 특허 데이터를 활용하여 기술에 대한 실증적인 접근이 이루어지고 있다. Kim *et al.*(2018)은 디지털 헬스케어 기술을 크게 건강정보 측정 기술, 건강관리 플랫폼 기술, 건강관리 원격서비스 기술로 구분하고 주요 4개국(미국, 일본, 유럽, 중국) 대비 한국의 기술경쟁력을 비교·분석하였다. 또한 Kim and Kim(2019)은 빈도 분석을 통해 구간별 상위 키워드를 추출하여 디지털 헬스케어 기술의 변화 추이를 분석하였다. 아울러 ICT 기술을 활용한 세부 기술 주제에 초점을 맞추어 헬스케어 기술을 분석한 연구들도 존재한다. Lee *et al.*(2019)은 Girvan-Newman 모델을 사용하여 네트워크 분석을 수행함으로써 헬스케어와 IoT가 융합된 서비스 분야의 유망기술을 제시하였다. Mück *et*

al.(2019)은 웨어러블 디바이스를 중심으로 헬스케어 특허를 분석하여 전자산업의 기업들이 관련 기술 개발을 주도하고 있다는 점을 확인하였다. 이 밖에도 Dantu *et al.*(2019)은 논문 데이터를 토픽모델링으로 분석하여 IoT 기술을 바탕으로 한 기술 주제를 확인하였으며, Farahani *et al.*(2018)과 같이 사례분석을 통해 최신 헬스케어 기술을 파악한 연구도 존재한다.

그러나 디지털 헬스케어 기술 전반을 대상으로 한 기존 연구들은 기술의 트렌드 변화나 기술경쟁력을 확인하는 정도에 그치고 있다. 뿐만 아니라 특정 기술에 초점을 맞추어 헬스케어 기술을 분석한 연구들도 디지털 헬스케어 산업의 일부 기술만을 분석하여 산업 전체의 기술 구조를 파악하기는 어렵다는 한계가 있다. 따라서 본 연구는 특허 데이터를 활용하여 실증적인 관점에서 디지털 헬스케어 분야 전반의 기술 구조를 제시하는 것을 목적으로 한다.

### 2.2 기술 구조 분석 기법

헬스케어 분야는 산업 내 디지털 전환(digital transformation)이 빠르게 진행되고 있어 기술범위 정의에 어려움이 존재하므로 기술 구조 분석을 위해서는 다각적인 접근이 필요하다. 기술 구조 분석 기법은 기술 문서의 집합으로부터 세부 기술을 추출하는 방식에 따라 크게 비율 할당(fractional assignment)과 이산 할당(discrete assignment)의 두 가지 유형으로 구분될 수 있다.

먼저 비율 할당은 문서에 포함된 복수의 주제를 비율에 따라 추출함으로써 기술 문서에 내재된 다양한 주제를 확인할 수 있는 방법으로, 토픽모델링 기법이 대표적이다(Lee and Kang, 2018). 토픽모델링은 다량의 비정형적인 문서 집합에 잠재된 주제를 자동으로 탐색하는 통계 모델이며(Blei *et al.*, 2003), LSI(latent semantic indexing), pLSI(probabilistic latent semantic indexing), LDA(latent dirichlet allocation)를 포함한 다양한 기법이 있다. LSI는 SVD(singular value decomposition)를 이용하여 문서-단어 행렬(Document-Term Matrix)의 차원을 축소함으로써 문서 혹은 단어 간 잠재적 의미를 발견하는 방법이다(Deerwester *et al.*, 1990). LSI는 구현이 용이하다는 장점이

있으나 새로운 문서가 추가되면 연산을 다시 수행해야 하므로 확장이 어렵다는 한계가 존재한다. 이에 따라 등장한 pLSI은 문서를 토픽들의 조합으로 가정하여 문서의 잠재의미를 분석하며, 확률 조합을 통해 단어의 출현 빈도를 대체한다(Hofmann, 1999). pLSI와 LDA는 확률 모델을 이용하여 문서의 주제를 파악한다는 점에서 상호 유사성이 있으나 pLSI는 말뭉치(corpus)의 크기에 따라 매개변수의 수가 선형적으로 증가한다는 단점이 존재한다(Blei *et al.*, 2003). LDA는 이러한 pLSI의 한계를 보완한 알고리즘으로서 비구조적인 텍스트 집합에서 의미 있는 주제를 찾아내는 생성 확률(generative probabilistic) 모델이며, 문서나 단어 등에 기초하여 문서 구조를 비롯한 숨겨진 변수의 추론을 목적으로 한다(Blei, 2012).

LDA의 모델링 작업은 문서의 작성 과정과 유사한 메커니즘으로 구현되어 LDA는 문서 내 단어들을 이용하여 주제별 단어를 할당하고 각각의 문서가 어떤 주제를 포함할 확률을 파라미터로 설정한다(Blei *et al.*, 2003; Park and Song, 2013). <Figure 2>는 LDA의 문서생성 프로세스를 나타낸 것이다.  $W_{d,i}$ 는 문서에서 관측되는 변수로서 d번째 문서의 i번째 단어를 의미하며,  $Z_{d,i}$ 는 각 단어에 대한 주제 할당에 해당한다.  $\theta_d$ 는 d번째 문서의 토픽 비율이고  $\phi_k$ 는 k번째 토픽에 대한 단어 분포를 말한다. 더불어 문서 전체에 적용되는 파라미터  $\alpha$ 와  $\beta$ 는  $\alpha$ 에 따라 각 문서의 토픽 비율( $\theta$ )이 결정되면  $\beta$ 로 인해 토픽별 단어 분포( $\phi_k$ )가 정해진다. 따라서 LDA는 문서 집합에서 각 토픽별 단어 분포( $\phi_k$ )가 정해진 이후 문서별 토픽 비율( $\theta_d$ )이 결정되어 각각의 단어는 토픽의 할당( $Z_{d,i}$ )에 따라 선택되는 과정을 거치게 되며, 최종적으로 문서 집합의 토픽( $\phi$ ), 문서별 토픽 비율( $\theta$ ), 각 토픽에 포함되는 단어 분포( $Z$ )를 추정한다.

즉, LDA는 사전에 정의된 코드 또는 의미 범주를 지정하지 않아도 연구자가 토픽 수를 정한 만큼 말뭉치에서 토픽을 추출하므로(Mohr and Bogdanov, 2013) 거대 문서 집합에서 다수의 토픽을 탐색하는 것이 용이하여 텍스트의 의미 구조를 파악하기에 적합하다(Blei *et al.*, 2003; Blei, 2012). 나아가, LDA 알고리즘은 데이터의 차원을 축소하기에 유용할 뿐 아니라 해석이 가능하면서 의미적으로는 일관성 있는 토픽을 생성할 수 있으므로 다수의 텍스트 분석에서 LDA 토픽모델링을 활용하고 있다(Mimno and McCallum, 2008). 이에 따라 물류 기술(Choi and Song, 2018), 블록체인(Shahid, 2020), 태양광 발전(Venugopalan and Rai, 2015), 클라우드 컴퓨팅(Kang and Lee, 2018), 인공지능(Lee and Lee, 2019) 등 다양한 분야의 기술 분석에 LDA 토픽모델링이 유용하게 활용되었다.

반면, 이산 할당은 하나의 문서가 하나의 기술만을 포함한다고 가정하고, 문서의 관계를 정량적으로 측정한 후, 이를 바탕으로 개별 문서를 군집화하여 세부기술을 추출하는 방식이다. 이때, 기술 문서 간의 관계는 인용, 동시인용, 동시출현단어, 동시분류, 동시저술 등 다양한 계량서지 지표를 통해 측정하는 방식과 문서의 의미적 유사성(semantic similarity)을 통해 측정하는 방식으로 구분될 수 있다. 최근 텍스트 분석 기법의

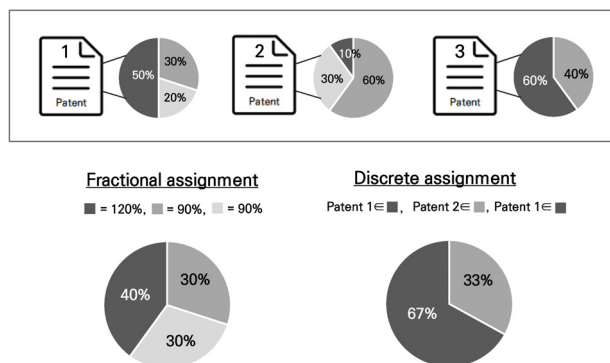


Figure 1. Fractional Assignment Vs. Discrete Assignment

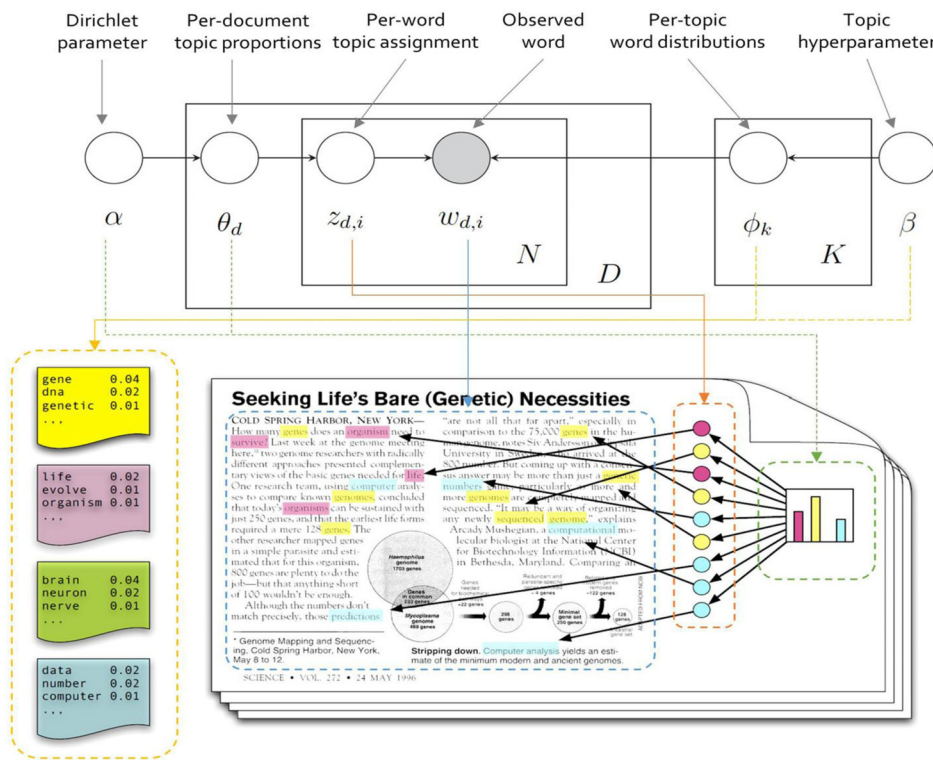


Figure 2. LDA Document Generation Process (Blei *et al.*, 2003; Blei, 2012)

급속한 발전으로 인해 다양한 기계학습 기반 문서 유사도 측정 기법들이 활용되고 있으며, 이 중 문서 단위의 유사도 측정에는 Doc2Vec 기법이 주로 활용되고 있다.

Doc2Vec은 문서, 문장과 같은 가변 길이의 텍스트(variable-length pieces of texts)에 대한 벡터 표현을 학습하는 알고리즘이다. Word2vec을 확장한 개념인 Doc2Vec은 Word2vec과 기본 구조가 유사한 바, Doc2Vec과 Word2vec 모두 분포 가설(distributional hypothesis)에 기초하여 비슷한 의미의 단어가 유사한 맥락에서 나타남을 가정하고 있다(Radu *et al.*, 2020). 다만, Word2vec이 단어 임베딩 모델인 것과 달리 문헌 임베딩 모델인 Doc2vec은 문서 단위를 벡터로 표현한다는 점에서 차이가 존재한다.

문서 내 단어들을 예측하기 위해 Doc2Vec은 인공신경망(artificial neural network)을 통해 훈련되고, 훈련이 끝나면 문서의 벡터 표현으로 각 가중치들을 사용한다. Doc2vec가 등장하기 전까지 텍스트에 대한 일반적인 벡터 표현은 Bag-of-words였으나, Bag-of-words는 단어의 순서 및 의미를 무시한다는 한계가 있었다. 반면에 Doc2vec은 문서 내에 존재하는 같은 단어도 순서가 달라지면 다른 벡터를 생성함으로써 벡터 공간상에서 유사한 의미를 가진 문서들이 가깝도록 위치시키는 바, Bag-of-words 모델보다 Doc2Vec이 향상된 성능을 보인다는 점이 실험으로 증명되었다(Le and Mikolov, 2014). 즉, Doc2vec 알고리즘은 문서를 대표할 수 있을 정도로 식별성이 높은 자질은 상대적으로 낮은 차원에 나타내어 자질 크기 때문에 분류 알고리즘의 학습이 저하되는 문제를 해결(Yuk and Song, 2018)한 것

으로 평가된다. 더욱이 Doc2Vec은 문서 전체를 고려하여 학습이 이루어져 기존 텍스트마이닝 기법인 TF-IDF, Text Rank 대비 향상된 단어 벡터를 획득할 수 있으므로 키워드 추출 효과를 개선한다(Wang *et al.*, 2020).

Doc2vec에서 문서가 포함하는 단어들을 예측하는 방식으로 DBOW(Distributed bag of words) 모델과 DM(Distributed memory) 모델이 존재하며, 두 모델의 구조는 <Figure 3>과 같다. 먼저 DBOW 모델은 어떠한 문서가 주어지면 그 문서가 포함하고 있는 단어들을 예측하는 알고리즘이며, 주어진 문서에서 단어를 임의로 추출하여 해당 문서에서 특정 단어가 도출될 확률을 최대한 높이는 방식으로 훈련이 이루어진다. DBOW 모델은 단어 벡터를 저장하지 않기 때문에 DM 모델에 비해 저장되는 데이터 양이 적다는 장점이 있다. 반면 word2vec의 CBOW 모델과 유사한 DM 모델은 문서 매트릭스(D)가 추가된 것 외에는 CBOW 모델과 동일한 구조를 갖는다. 이로 인해 문서가 포함하고 있는 단어를 예측하는 훈련이 종료될 경우, DM 모델은 훈련의 결과로서 의미 및 순서를 포함한 D와 W 매트릭스를 획득하게 된다(Le and Mikolov, 2014). 다만 일반적으로 DBOW 모델 대비 DM 모델의 성능이 더 우수하다고 알려져 있으므로 본 연구에서는 DM 모델을 활용하여 분석을 수행하였다.

한편, Doc2vec은 클러스터링 기법 등과 결합하여 대량의 문서 집합을 효과적으로 그룹화 할 수 있다. 이에 따라 측정된 벡터의 유사도에 기초하여 문서를 군집화함으로써 기술의 변화 추이를 파악하거나 유망기술을 발굴하는데 Doc2vec을 활용한



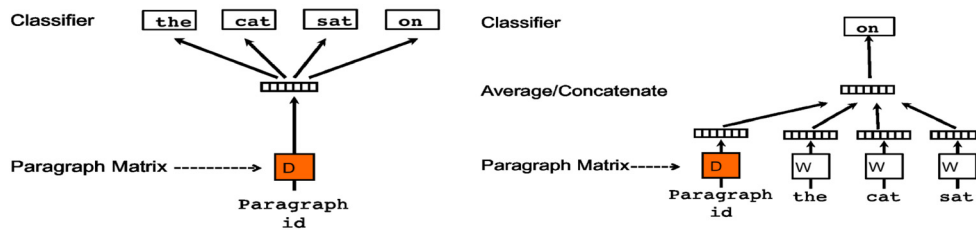


Figure 3. DBOW Model Vs. DM Model (Le and Mikolov, 2014)

연구들도 존재한다. 구체적으로 Doc2Vec을 활용한 클러스터 분석을 통해 유사 기술 그룹을 분류한 연구(Kim *et al.*, 2020)가 수행되었으며, Doc2vec을 이용하여 태양광 발전 기술의 변화 추이를 파악하거나(Trappey *et al.*, 2019) ICT R&D 분야의 유망기술을 발굴(Heo *et al.*, 2018)에 활용된 바 있다.

결과적으로 LDA 토픽모델링은 논문, 특허를 비롯한 각종 데이터를 토대로 특정 분야의 기술동향 탐색, 기술 구조 분석 등을 수행하는 연구가 진행되었다. 이처럼 LDA 토픽모델링은 대량의 문서 집합에서 기술 주제를 효과적으로 도출할 수 있으므로 본 연구에서는 LDA 토픽모델링을 활용하여 디지털 헬스케어의 기술 토픽을 파악하였다. 또한 Doc2vec은 클러스터 분석과 결합하여 문서 집합을 용이하게 그룹화 할 수 있어 유사 기술 그룹을 탐색하는데 이용되고 있는 것으로 확인된다. 따라서 본 연구에서는 Doc2vec 기반의 클러스터 분석을 토대로 디지털 헬스케어의 기술 클러스터를 탐색하였다.

### 3. 연구방법

#### 3.1 연구 프레임워크

본 연구의 프레임워크는 <Figure 4>와 같이 4단계로 구성된다. 1단계인 데이터의 수집 및 전처리 과정은 미국 특허청으로부터 디지털 헬스케어 관련 특허를 추출한 이후 전처리 과정을

거친다. 2단계에서는 전처리를 수행한 특허 데이터에 대해 LDA 토픽모델링으로 기술 토픽을 도출한다. 3단계는 전처리를 거친 데이터를 Doc2vec으로 임베딩하고 산출된 문서 벡터에 대해 Spherical k-means로 클러스터링하여 기술 클러스터를 탐색한다. 4단계는 LDA 기반의 기술 토픽과 Doc2vec 기반의 기술 클러스터의 결과를 비교함으로써 기법 간 차이점을 도출한다.

#### 3.2 데이터 수집 및 전처리

특허 검색서비스인 WINTELIPS를 이용하여 2011년부터 2020년까지 미국 특허청(USPTO)에 출원된 특허 중 디지털 헬스케어와 관련된 키워드가 제목이나 초록에 포함하고 있는 특허를 추출하였다. 특허 검색을 위한 검색식은 한국특허전략개발원의 특허기술동향조사 보고서(2018)를 바탕으로 디지털 헬스케어에 포함되는 개념인 smart healthcare, ubiquitous healthcare, mobile healthcare를 추가하여 <Table 1>과 같이 작성하여 특허를 추출하였으며, 중복 데이터를 제외하고 총 2,064건의 특허를 이용하여 데이터베이스를 구성하였다. <Figure 5>는 수집한 디지털 헬스케어 특허의 연도별 출원 추이를 나타내며 2014년을 기점으로 큰 폭의 증가세를 보이고 있다는 점을 확인할 수 있다. 더욱이 출원 특허가 공개될 때까지 약 1년 6개월이 소요되는 점을 고려했을 때 디지털 헬스케어 분야의 기술개발이 활발하다는 것을 알 수 있다.

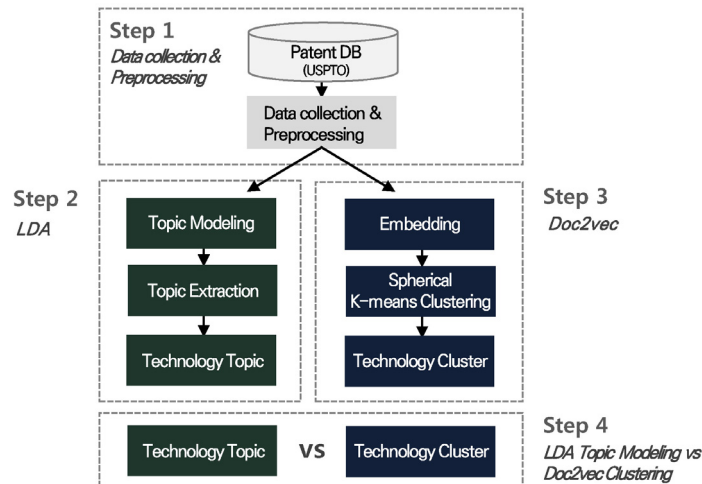


Figure 4. Research Framework

Table 1. Query for Keyword Search

Keyword	((digital* smart* ubiquitous* mobile*) and (healthcare wellness* fitness* clinic* medical*)).ti.ab. and (@AD>=20110101)
---------	---

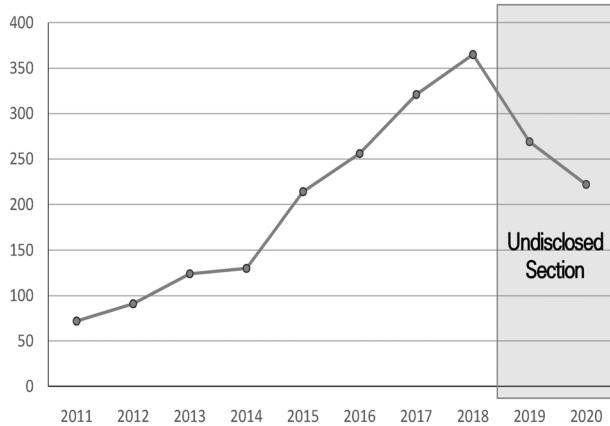


Figure 5. Annual Changes in the Numbers of Digital Healthcare Patents

특히 데이터에서 추출한 초록 텍스트에 대해 전처리를 수행하였다. 문장 부호 및 숫자를 제거한 이후 토큰화를 위해 띄어쓰기를 기준으로 텍스트를 단어 단위로 분리하였다. 분리된 단어 중 관사 및 전치사를 포함한 불용어(stopwords)를 제거한 후, 마지막으로 표제어 추출(lemmatization) 단계를 수행하였다. 표제어 추출은 문장에 사용된 단어들의 표제어(lemma)를 추출하는 방법으로 단어의 어간을 추출하는 스템밍(stemming)과 달리 문장 내에서 사용된 단어의 품사를 비롯한 의미적 정보를 고려하여 표제어를 도출하였다.

#### 4. 디지털 헬스케어 핵심 기술 탐색

##### 4.1 LDA 기반 기술 토픽

전처리 작업을 수행한 특허 초록을 대상으로 LDA를 수행하여 디지털 헬스케어 기술의 세부 토픽을 도출하였다. LDA 수행을 위해 R 패키지인 topicmodels를 수정하여 활용하였고(Hornik and Grün, 2011), 추론 방법은 깰스 샘플링(Gibbs sampling)을 적용하였다. 샘플링 반복횟수는 1,000번,  $\alpha$ 값은 기본값을 적용하였으며,  $\beta$ 값 역시 기본값에 해당하는 0.1을 사용하였다. 토픽 수는 10개, 15개, 20개, 30개로 추출하였으며, 이 중 15개일 때 설명력이 가장 뛰어난 것으로 판단되어 최종 토픽수는 15개로 설정하였다. 추출된 토픽의 주요 단어와 상위 비중 특허 문서의 내용을 참조하여, 각 토픽의 네이밍을 수행하였다. <Table 2>는 LDA 알고리즘을 통해 추출된 15개의 기술 토픽과 토픽별 핵심 단어를 나타낸 것이며, 토픽의 번호는 토픽 비중이 높은 것부터 내림차순으로 부여하였다.

15개의 기술 토픽은 원격진단, 의료용 장비, 바이오 센서 등 디지털 헬스케어의 다양한 세부 분야를 모두 포괄하고 있는 것으로 나타났다. 가장 비중이 높은 [T-01] IoT 기술은 인터넷에 기반을 두고 모든 사물을 연결하고 있어 기술의 활용범위가 점차 넓어지고 있으며, 인구 고령화, 만성 질환 환자의 증가 등의 영향을 받아 헬스케어 산업에서 IoT 기술이 차지하는 비중도 확대될 전망이다(Lee, 2015). 뿐만 아니라 디바이스 및 애플리케이션으로부터 수집된 개인의 건강정보는 IoT 기술을 통해 의료진에게 전송되어 능동적인 헬스케어 서비스를 제공할 수 있도록 지원하고 있으므로 [T-01]은 [T-03] Mobile healthcare application, [T-12] Data security, [T-15] Wearable device를 포함한 기술 토픽 전반과 높은 연관성을 가진다. 또한 [T-15] Wearable device는 형태에 따라 휴대형(portable), 부착형

Table 2. Technology Topics of Digital Healthcare

Technology Topics	Relevant word	Proportion
[T-01] IoT	communication, internet, security, system, packet	9.13%
[T-02] Business intelligence	system, support, business, intelligent, path	7.60%
[T-03] Mobile healthcare application	service, network, scheme, connectivity, dvertisement	7.53%
[T-04] Home healthcare support	wireless, service, home, support, report	7.25%
[T-05] Portable device	device, interface, wireless, portable, handheld	6.98%
[T-06] Telediagnosis	patient, database, diagnostic, biometric, clinician	6.72%
[T-07] Smart emergency application	user, emergency, analysis, application, pattern	6.31%
[T-08] Medical imaging	image, target, capture, camera, storage	6.30%
[T-09] Healthcare consulting service	identifier, entity, message, management, response	6.20%
[T-10] Real-time locating system	information, location, indicate, feedback, search	6.18%
[T-11] Biosensor & Monitoring system	monitor, parameters, track, alarm, sensor	6.16%
[T-12] Data security	data, server, encrypt, protocol, program	6.14%
[T-13] Medical supplies & equipment	attach, wheel, vehicle, RFID, container	5.93%
[T-14] Implantable medical device	signal, antenna, control, beacon, pulse	5.78%
[T-15] Wearable device	wearable, sensor, measurement, sign, application	5.77%

(attachable), 이식/복용형(eatable)으로 분류되며(Jeong, 2014), 초기 모델인 휴대형에서 기술 발전을 거쳐 이식/복용형에 이를 것으로 예상된다. 다만 아직까지 헬스케어 분야에서 손목시계 형태의 휴대형 디바이스 개발이 활발한 바, [T-05] Portable device이 상위 개념인 [T-15] Wearable device 보다 비중이 높은 것으로 분석된다.

나아가, LDA 기반의 기술 토픽에서는 의료산업에서 일반적으로 사용 중인 의료장비와 관련된 기술 주제가 확인되어 디지털 방식의 의료용 기기인 [T-13] Medical supplies & equipment, 이식용 장비에 대한 토픽인 [T-14] Implantable medical device가 도출되었다. 또한 응급 및 재난상황에 대한 긴급 알람에 대한 기술인 [T-07] Smart emergency application, 사용자의 위치를 실시간으로 추적하는 기술과 관련된 [T-10] Real-time locating system과 같이 각종 산업군에서 활발하게 사용되고 있는 기술 토픽도 도출되었다.

#### 4.2 Doc2vec 기반 기술 클러스터

전처리를 거친 특허 초록에 대하여 Doc2vec 기법 중 DM 모델을 활용하여 각 문서를 벡터화하였다. 파이썬의 gensim 패키지를 사용하여 PV-DM 알고리즘을 활용하였고, 벡터의 차원 수(vector size)는 10, 윈도우 크기(window size)는 15, 초기 학습률 값(learning rate)은 기본값인 0.0025를 사용하였다. 도출된 문서별 벡터값을 바탕으로 Spherical K-Means 분석을 통해 특허 문서를 군집화하여 기술 클러스터를 도출하였다. 기술 클러스터의 수는 LDA 토픽모델링의 산출 결과와 비교하기 위해 15개로 설정하였고, 각 클러스터에 포함된 특허 기술의 내용을 고려하여 클러스터의 명칭을 결정하였다. <Table 3>은 Doc2vec에 기반한 기술 클러스터의 현황과 클러스터별 비중

을 나타낸 것이다. 클러스터의 번호 역시 비중이 높은 것부터 내림차순으로 부여하였다.

[C-01] IoT in healthcare, [C-12] AI in medical imaging, [C-15] Healthcare virtual assistant, [C-09] Blockchain in healthcare와 같은 기술 클러스터를 통해 헬스케어 산업 내에서 디지털 전환이 활발하게 진행되고 있음을 확인할 수 있다. 특히 IoT 기술은 원격으로 건강관리 시스템을 제공할 수 있을 뿐 아니라 생체 신호의 측정 및 모니터링, 의료 데이터의 저장·전송·보안과 같이 디지털 헬스케어 기술 전반과 연결되어 핵심 기술에 해당하는 것으로 분석된다. [C-15] Healthcare virtual assistant은 현재는 비중이 가장 낮으나 점차 의료기관의 인적·물적 자원 관리에 활용되어(Tian *et al.*, 2019) 기존의 의료 서비스를 개선할 것으로 전망된다.

더불어 헬스케어 데이터가 개인 맞춤형 의료 서비스, 원격 의료를 제공하기 위한 기초 자료로 사용되면서 데이터 수집 및 관리는 물론 보안 기술의 중요성이 높아지고 있으며, 관련 분야의 기술 개발도 활발한 상황이다(Agbo *et al.*, 2019). 이에 따라 개인의 헬스케어 데이터의 관리 및 제어, 정보보안 향상에 활용되는 [C-09] Blockchain in healthcare가 헬스케어 산업의 주요 기술로 추출되었다. 나아가, 블록체인 기술은 개인의 의료 데이터를 통합적으로 관리하는 [C-08] Personal health record (PHR) platform, 생체 인증을 포함한 데이터 보안 기술인 [C-13] Data security in healthcare와도 밀접한 상관관계를 가지고 있는 것으로 파악된다.

이 밖에도 Doc2vec 기반의 기술 클러스터에서는 디지털 기술에 기반하여 비대면 의료 서비스를 제공하는 기술인 [C-10] Telehealth service와 [C-14] Remote patient monitoring이 도출되었다. 의료 비용을 절감하고 의료 서비스를 개선하기 위해 헬스케어 산업에서 원격 의료 서비스를 확대하기 위한 움직임이

Table 3. Technology Clusters of Digital Healthcare

Technology Clusters		Proportion
[C-01]	IoT in healthcare	17.20%
[C-02]	Smart emergency medical system	12.16%
[C-03]	Mobile healthcare management system	9.88%
[C-04]	Biosignal measurement system	6.30%
[C-05]	Portable medical device	5.86%
[C-06]	Personalized healthcare	5.81%
[C-07]	Wearable medical device	5.43%
[C-08]	Personal health record (PHR) platform	5.18%
[C-09]	Blockchain in healthcare	5.14%
[C-10]	Telehealth service	4.99%
[C-11]	Digital diagnostic system	4.94%
[C-12]	AI in medical imaging	4.51%
[C-13]	Data security in healthcare	4.31%
[C-14]	Remote patient monitoring	4.31%
[C-15]	Healthcare virtual assistant	3.97%

이어지고 있어(Tuckson *et al.*, 2017), 향후 헬스케어 산업에서 관련 시장이 크게 성장할 전망이다.

### 5. 디지털 헬스케어 핵심 기술 비교

4장에서는 비율 할당 방식인 LDA(latent dirichlet allocation) 토픽모델링 기법과 이산 할당 방식인 Doc2vec 기반 클러스터 분석을 동시에 수행하여 디지털 헬스케어의 핵심 기술을 도출하였다. 두 가지 서로 다른 접근법을 통해 도출된 핵심 기술의 목록을 살펴보면, 동일하거나 유사한 기술이 동시에 나타나기도 하지만, 서로 다른 기술들이 포함되어 있음을 확인할 수 있다. 두 가지 방식을 통해 도출된 핵심 기술 목록을 비교하기 위해서는 표준화된 분류 체계를 활용하는 것이 효과적이나 기술의 범위 및 분류가 정립되지 않은 디지털 헬스케어 분야는 관점에 따라 기술 분류가 다양한 상황이다(Kim *et al.*, 2018). 이에 본 연구에서는 산업연구원(2017)의 스마트 헬스케어 산업 분류체계를 활용하여 비교를 수행한다.

산업연구원(2017)은 스마트 헬스케어 산업의 제품 및 서비스를 하드웨어(hardware), 소프트웨어(software), 서비스(service)의 크게 세 가지로 분류하고 각각을 다시 두 개의 세부 유형으로 분류하여 총 6개로 구분하였다. 보다 구체적으로, 하드웨어는 제품·서비스 일체형, 단품 형태 제품(products and services all-in-one, single-piece product), 부품(component)의 두 가지 유형, 소프트웨어는 의료·건강 관리 콘텐츠(medical·helathcare contents)와 미들웨어, 플랫폼, 통신 네트워크(middleware, platform communication network), 서비스는 진단 서비스(diagnosis

service)와 건강관리 서비스(healthcare service)로 세분된다. <Table 4>는 본 연구에서 두 가지 방식으로 도출된 디지털 헬스케어 핵심 기술을 6개의 유형별로 매칭한 결과를 나타낸 것이다.

디지털 헬스케어 분야에서 IoT가 핵심 기술을 담당함에 따라 LDA 기반의 기술 토픽과 Doc2vec 기반의 기술 클러스터 모두 미들웨어 및 통신 네트워크 기술의 비중이 가장 높은 것으로 나타났다. 또한 IoT 기술에 기초한 진단서비스와 건강관리 서비스도 다수를 차지하고 있는바, 수집된 개인의 헬스케어 데이터를 인터넷으로 연결하여 새로운 의료 서비스를 제공하려는 시도가 늘어나고 있는 것으로 분석된다. LDA 기반의 기술 토픽과 Doc2vec 기반의 기술 클러스터에서 모두 IoT 기술이 디지털 헬스케어의 핵심 기술로 파악되지만, 기술 분류별 세부적인 특성과 기술의 분포 현황을 바탕으로 LDA 토픽모델링과 Doc2vec 기반 클러스터링의 차이점을 확인할 수 있다.

LDA 기반의 기술 토픽에서는 [T-13] Medical supplies & equipment, [T-14] Implantable medical device과 같이 아날로그 의료 산업에서 사용되었던 의료용 장비 및 의료기기에 대한 핵심 기술이 포함되었다. 또한 하드웨어, 소프트웨어, 서비스에 관한 기술이 각각 5개씩 제시되어 대분류 수준에서는 기술별 비중이 동일하게 나타난다. 즉, 비율 할당 접근법은 각 특허 문서에 포함된 주제 전반을 추출하게 되므로, 특정 분야를 구성하고 있는 기반 기술이 핵심 기술로 도출되며, 새로운 기술 분야와 함께 해당 산업의 근간을 이루고 전통적인 기술 분야가 동시에 추출된다. 따라서 LDA에 기반한 비율할당 방식은 특정 기술 분야의 요소 기술(component technology)을 핵심 기술로써 추출하는데 효과적이다.

반면에 Doc2vec 기반의 기술 클러스터는 [C-15] Healthcare

**Table 4.** Classifications of Digital Healthcare Technology Topics and Clusters

Type		Technology Topics based LDA	Technology Clusters based Doc2vec
Hardware	Products and Services all-in-one, Single-piece product	[T-05] Portable device [T-15] Wearable device [T-08] Medical imaging [T-14] Implantable medical device	[C-05] Portable medical device [C-07] Wearable device [C-12] AI in medical imaging
	Component	[T-13] Medical supplies & equipment	-
Software	Medical · Healthcare Contents	[T-03] Mobile healthcare application	[C-03] Mobile healthcare management system
	Middleware, Platform, Communication network	[T-01] IoT [T-12] Data security [T-02] Business intelligence [T-04] Home healthcare support	[C-01] IoT in healthcare [C-13] Data security in healthcare [C-09] Blockchain in healthcare [C-08] Personal health record(PHR) platform
Service	Diagnosis service	[T-11] Biosensor & Monitoring system [T-06] Tlediagnosis	[C-04] Biosignal measurement system [C-14] Remote patient monitoring [C-11] Digital diagnostic system
	Healthcare service	[T-07] Smart emergency application [T-09] Healthcare consulting service [T-10] Real-time locating system	[C-02] Smart emergency medical system [C-06] Personalized healthcare [C-15] Healthcare virtual assistant [C-10] Telehealth service



virtual assistant, [C-09] Blockchain in healthcare와 같이 디지털 전환에 따라 새롭게 대두되는 헬스케어 산업의 유망 분야를 포함하고 있다. 또한 [C-10] Telehealth service 및 [C-14] Remote patient monitoring 처럼 최근 연구개발이 활발한 비대면 의료 서비스에 대한 기술 클러스터도 발견되었다. 이처럼 Doc2vec 기반의 클러스터링에서는 디지털 전환으로 인해 새롭게 대두되고 있는 소프트웨어 및 서비스 분야를 중심으로 기술 주제가 도출되었는데, 이는 Doc2vec 기반 클러스터링이 각 특허 문서를 특허의 주요 목적에 따라 하나의 군집에만 할당하기 때문이다. 즉, Doc2vec 기반 이산 할당 방식에서는 각 특허가 담고 있는 최신 개발 기술만이 표면적으로 드러나게 되므로, 이는 특정 기술 분야의 유망기술(emerging technology)을 핵심 기술로써 추출하는데 효과적이다.

## 6. 결론

본 연구는 미국 특허청(USPTO)에서 디지털 헬스케어 관련 특허 2,064건을 수집하여 초록 데이터에 대하여 LDA 토픽모델링과 Doc2vec 기반 클러스터링을 수행함으로써 디지털 헬스케어 분야의 기술 구조를 분석하고 핵심 기술을 파악하였다. 또한 각각 15개씩 도출된 LDA 기반 기술토픽과 Doc2vec 기반 기술 클러스터에 대한 비교를 통해 기술 구조의 분석 방법으로 LDA 토픽모델링과 Doc2vec 기반 클러스터링의 특성 및 차이를 확인하였다. 이에 비유 할당 접근방법에 기초한 LDA 토픽모델링은 특허 문서에 내재된 다양한 주제를 추출하여 하드웨어, 소프트웨어, 서비스의 기술을 포괄적으로 제시하였다. 즉, LDA 토픽모델링을 통해 디지털 헬스케어 산업 전반의 기술 주제를 확인함으로써 디지털 헬스케어 산업을 구성하는 요소기술(component technology)을 탐색할 수 있었다. 반면, 각 특허 문서별로 하나의 주제만을 할당하는 이산 할당 접근방법인 Doc2vec 기반의 클러스터링은 디지털 헬스케어 산업에서 제품 개발이 활발한 소프트웨어 및 서비스 분야 위주로 기술 클러스터가 분석되었다. Doc2vec 기반 클러스터링은 문서 내에 높은 비중을 차지하는 기술 주제를 중심으로 특허의 군집화가 이루어져 디지털 헬스케어 산업에서 기업들이 기술 개발에 초점을 맞추고 있는 유망기술(emerging technology)을 확인할 수 있었다.

헬스케어 산업의 매력도가 제고되어 생명공학, 의료, 제약 기업을 넘어 ICT 기업 등 다양한 산업군의 기업들이 시장에 참여함에 따라 경쟁이 심화되고 있으므로, 기업이 경쟁우위 확보를 위해서는 기술개발에 앞서 산업을 구성하는 기술 구조에 대한 이해가 선행되어야 한다. 이에 본 연구는 특허 데이터를 활용함으로써 실증적 관점에서 디지털 헬스케어 산업의 기술 구조를 탐색하고 핵심 기술을 파악하였다. 본 연구를 통해 파악된 디지털 헬스케어 분야의 기술 구조와 핵심 기술 분류 체계는 정부의 관련 정책 수립 및 기업의 기술 전략 수립에 기본

프레임워크로서 활용될 수 있을 것으로 기대된다.

다만 본 연구는 몇 가지 한계점이 존재하며 향후 연구에서 다루어질 필요가 있다. 첫째, 특허 데이터는 출원 후 공개될 때까지 일반적으로 1년 6개월이 소요되므로 최신 기술을 반영하지 못하였다. 빠르게 변화하고 있는 디지털 헬스케어 기술을 분석하기 위해서는 오픈소스, 논문 등의 데이터를 활용한 분석이 필요하다. 둘째, ‘Digital healthcare’, ‘Mobile healthcare’와 같이 포괄적인 키워드로 이루어진 검색식을 사용하여 특허를 수집함에 따라 디지털 헬스케어 기술 전체를 포함하지 못하였다는 한계가 존재한다. 따라서 본 연구에서 도출된 기술 주제를 키워드로 사용하여 세부요소별 기술 분석을 수행해볼 수 있다. 셋째, 지난 10년 동안 출원된 디지털 헬스케어 분야의 특허를 정태적으로 분석하여 기술 구조의 변화 양상을 살펴보지 못하였다. 세부 기술의 변화 추이를 살펴봄으로써 유망기술뿐 아니라 회복기로 돌아선 기술 등을 확인할 수 있으므로 향후 연구에서는 디지털 헬스케어 분야의 기술 구조에 대한 동적 변화를 살펴볼 필요가 있다.

## 참고문헌

- Agbo, C. C., Mahmoud, Q. H., and Eklund, J. M. (2019), Blockchain Technology in Healthcare: A Systematic Review, *Healthcare*, **7**(2), 56
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), Latent Dirichlet Allocation, *The Journal of Machine Learning Research*, **3**, 993-1022.
- Blei, D. M. (2012), Probabilistic Topic Models, *Communications of the ACM*, **55**(4), 77-84
- Choi, D. H. and Song, B. M. (2018), Exploring Technological Trends in Logistics: Topic Modeling-based Patent Analysis, *Sustainability*, **10**(8), 2810
- Dantu, R., Dissanayake, I., and Nerur, S. (2019), Exploratory Analysis of Internet of Things (IoT) in Healthcare: A Topic Modeling Approach, *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990), Indexing by Latent Semantic Analysis, *Journal of the American Society for Information Science*, **41**(6), 391-407.
- Farahani, B., Firouzi, F., Chang, V., Badaroglu, M., Constant, N., and Mankodiya, K. (2018), Towards fog-driven IoT eHealth: Promises and Challenges of IoT in Medicine and Healthcare, *Future Generation Computer Systems*, **78**, 659-676
- Frost & Sullivan (2019), Global Digital Health Outlook, 2020.
- Heo, Y. S., Kim, K. H., and Kang, J. S. (2018), Research on the Process of Searching Emerging Technology: The Case of the Result of Exploring Emerging Technology in ICT R&D Area, *Proceedings of the Korea Technology Innovation Society Conference*, 525-545.
- Hofmann, T. (1999), Probabilistic Latent Semantic Indexing, In *Proceedings of the 22nd annual international ACM SIGIR*.
- Hornik, K. and Grün, B. (2011), Topicmodels: An R Package for Fitting Topic Models, *Journal of Statistical Software*, **40**(13), 1-3.
- Jeong, H. S. (2014), Trends and Prospects of Healthcare Wearable Devices, *Health Industry Brief*, **115**, 1-20.

- Jeong, I. Y., Kim, S. K., Lee, D. E., and Lee Y. H. (2016), Emerging Healthcare Innovations Driven by Data and Its Policy Implications, *Science and Technology Policy Institute*, 1-204.
- Kang, J. H. and Lee, H. Y. (2018), Analyzing the Technological Structure of Cloud Computing Based on Patent Information, *Korean Institute of Industrial Engineers*, **44**(1), 69-81.
- Kim, D. S., Jo, S. H., Lee, J. S., Kim, M. S., and Kim, N. H. (2018), A Study on the Competitive Analysis of Digital Healthcare in Korea through Patent Analysis, *Journal of Digital Policy & Management*, **16**(9), 229-237.
- Kim, H. J. and Kim, H. K. (2019), Analysis of Research Trends in Digital Healthcare based on Patent Data, *Korea Technology Innovation Society*, 2154-2162
- Kim, H. S., Kwon, I. H., and Cha, W. C. (2021), Future and Development Direction of Digital Healthcare, *Healthcare Informatics Research*, **27**(2), 95-101
- Kim, J. Y., Yoon, J. H., Park, E. J., and Choi, S. C. (2020), Patent Document Clustering with Deep Embeddings, *Scientometrics*, 563-577.
- Korea Institute for Industrial Economics & Trade (2017), New Growth Engines, Smart Healthcare Industries in the Era of the 4th Industrial Revolution, *i-KIET Industrial Economy Issues*, **12**, 1-12.
- Korea Intellectual Property Strategy Agency, Korean Intellectual Property Office, Ministry of Health and Welfare and Korea Health Industry Development Institute (2018), The Ministry of Health and Welfare's Health Care Technology R&D Project: A Report on Patent Technology Trends using Blockchain Technology in the Healthcare Sector, *Korea Intellectual Property Strategy Agency*, 1-79.
- Le, Q. and Mikolov, T. (2014), Distributed Representations of Sentences and Documents, In *International Conference on Machine Learning*, 1188-1196.
- Lee, J. S. (2014), Digital Healthcare Platforms and Trends in Major Companies, *Health Industry Brief*, **140**, 1-12.
- Lee, B. M. (2015), Registration Protocol for Health IoT Platform to Share the use of Medical Devices, *International Journal of Bio-Science and Bio-Technology*, **7**(4), 1-10
- Lee, H. Y. and Kang, P. S. (2018), Identifying Core Topics in Technology and Innovation Management Studies: A Topic Model Approach, *The Journal of Technology Transfer*, **43**(5), 1291-1317.
- Lee, J. Y. (2019), Trends and Implications of Digital Healthcare, *Issue Report*, **3**, 1-14
- Lee, W. J. and Lee, H. Y. (2019), A Technology Landscape of Artificial Intelligence: Technological Structure and Firms' Competitive Advantages, *Journal of Korea Technology Innovation Society*, **22**(3), 340-361.
- Lee, S. G., Choi, J. W., and Sawng, Y. W. (2019), Foresight of Promising Technologies for Healthcare-IoT Convergence Service by Patent Analysis, *Journal of Scientific & Industrial Research*, **78**(8), 489-494.
- Lupton, D. (2014), Beyond Techno-utopia: Critical Approaches to Digital Health Technologies, *Societies*, **4**(4), 706-711.
- Mimno, D. and McCallum, A. (2008), Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression, In *Uncertainty in Artificial Intelligence*, **24**, 411-418.
- Mohr, J. W. and Bogdanov, P. (2013), Introduction—Topic Models: What they are and why they Matter, *Poetics*, **41**(6), 545-569.
- Mück, J. E., Unal, B., Butt, H., and Yetise, A. K. (2019), Market and Patent Analyses of Wearables in Medicine, *Trends in Biotechnology*, **37**(6), 563-566.
- Park, J. H. and Song, M. (2013), A Study on the Research Trends in Library & Information Science in Korea using Topic Modeling, *Journal of the Korean Society for Information Management*, **30**(1), 7-32.
- Radu, R. G., Rădulescu, I. M., Trucă, C. O., Apostol, E. S., and Mocanu, M. (2020), Clustering Documents using the Document to Vector Model for Dimensionality Reduction, In *2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*, 1-6.
- Seong, K. and Lee, H. (2020), Exploring the Interdisciplinary Structure of Information and Communication Technologies: A Co-Classification Analysis of Research Proposals, *Korean Institute of Industrial Engineers*, **46**(5), 491-505.
- Shahid, M. N. (2020), A Cross-disciplinary Review of Blockchain Research Trends and Methodologies: Topic Modeling Approach, In *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 4053-4059.
- Tian, S., Yang, W., Grange, J. M. Le., Wang, P., Huang, W., and Ye, Z. (2019), Smart Healthcare: Making Medical Care More Intelligent, *Global Health Journal*, **3**(3), 62-65.
- Trappey, A. J., Chen, P. P., Trappey, C. V., and Ma, L. (2019), A Machine Learning Approach for Solar Power Technology Review and Patent Evolution Analysis, *Applied Sciences*, **9**(7), 1478.
- Tuckson, R. V., Edmunds, M., and Hodgkins, M. L. (2017), Telehealth, *New England Journal of Medicine*, **377**(16), 1585-1592.
- Venugopalan, S. and Rai, V. (2015), Topic Based Classification and Pattern Identification in Patents, *Technological Forecasting and Social Change, Technological Forecasting and Social Change*, **94**, 236-250.
- Wang, W., Li, X., and Yu, S. (2020), Chinese Text Keyword Extraction Based on Doc2vec And TextRank, In *2020 Chinese Control And Decision Conference (CCDC)*, 369-373.
- Yuk, J. H. and Song, M. (2018), A Study of Research on Methods of Automated Biomedical Document Classification using Topic Modeling and Deep Learning, *Journal of the Korean Society for Information Management*, **35**(2), 63-88.

## 저자소개

**이서영** : 숙명여자대학교 행정학과에서 학사학위를 취득하고, 서울과학기술대학교 산업정보시스템전공에서 석사학위를 받았으며, 현재 (주)웹스 기술평가센터에서 기술가치평가 업무를 수행하고 있다. 주요 연구분야는 기술경영, 기술평가이다.

**안승섭** : 서울과학기술대학교 경영학과에서 학사학위를 취득하고 서울과학기술대학교 데이터사이언스학과에서 석사학위를 취득하였으며, 현재 (주)컬리 물류전략팀에 재직 중이다. 연구 분야는 로지스틱스 애널리틱스, 비즈니스 애널리틱스 등이다.

**이학연** : 서울대학교 산업공학과에서 학사학위를 취득하였으며, 동 대학원에서 박사학위를 받았다. 현재 서울과학기술대학교 산업공학과 교수로 재직 중이다. 주요 연구 분야는 기술예측, 이노베이션 애널리틱스, 디지털 혁신 전략 등이다.