

## 기술토픽 간 기술융합기회 발굴 연구\*

### — 웨어러블 기술 사례를 중심으로 —

서원철\*

- |                                  |                        |
|----------------------------------|------------------------|
| I. 서론                            | 2. 기술토픽 간 링크예측         |
| II. 관련 연구                        | 3. 기술토픽 간 융합기회 발굴      |
| 1. 특허분석 기반의 기술융합기회 발굴            | IV. 사례 분석              |
| 2. 토픽모델링                         | 1. 웨어러블 기술 관련 기술토픽 생성  |
| 3. 링크예측                          | 2. 기술토픽 네트워크 구축 및 링크예측 |
| III. 기술토픽 분석을 통한 기술융합기회<br>발굴 방안 | 3. 기술융합기회 발굴모델 개발      |
| 1. 기술토픽 생성                       | V. 논의 사항               |
|                                  | VI. 결론                 |

\* 이 논문은 부경대학교 자율창의학술연구비(2021년)에 의하여 연구되었음.

\*\* 부경대학교 산업및데이터공학과 산업데이터공학융합전공 교수.

초록

기술융합은 서로 다른 분야의 기술지식을 융합적으로 활용하여 혁신을 선도하고 산업에 새로운 가치를 창출해 내는 기술적 진보를 촉진한다. 따라서 기술융합을 통해 새로운 혁신을 실현할 수 있는 잠재적 기술기회를 발굴해 낼 수 있도록 함으로써 기술기회를 선점하고 경쟁우위를 갖출 수 있도록 지원할 수 있다. 특허데이터를 활용하여 기술융합기회를 발굴하기 위한 연구는 다양한 관점에서 수행되어 왔지만, 이들은 주로 기술적 의미가 모호한 기술클래스를 활용함으로써 기술융합기회가 지니는 기술적 함의를 구체적으로 도출하지 못하는 한계를 지니고 있다. 또한, 소수의 링크예측 알고리즘들을 개별적으로만 적용함으로써 인해 이들 각각이 지니는 개별적 장점을 효과적으로 결합하여 활용하지 못하는 문제를 유발하기도 한다. 이에 본 연구에서는 기술 구체성이 높은 기술토픽 단위에서 융합 네트워크를 구축하고 다양한 링크예측 알고리즘들을 복합적으로 적용하여 기술융합기회를 발굴하는 지도학습 모델을 개발한다. 그리고 본 연구의 타당성을 살펴보기 위하여 개발된 모델이 미래 융합 가능성을 높은 확률로 예측한 기술토픽들에 대해 살펴보고 기술융합기회로서의 의미에 대해서도 논의한다. 본 연구는 구체화된 기술토픽 간 융합을 통해 실행 가능한 새로운 기술기회를 선점할 수 있도록 지원하는 새로운 형태의 지도학습 기반 기술융합기회 발굴모델을 제시한다는 측면에서 기여점을 지닌다. 이를 통해 국내 기업들의 취약한 R&D 기획 역량을 증진하고 지속가능한 성장을 위한 경쟁우위를 확보하도록 지원할 수 있을 것으로 기대된다.

주제어

기술융합, 지도학습, 토픽모델링, 링크예측, 웨어러블

## I. 서론

서로 다른 기술지식을 융합적으로 활용하여 새로운 기술을 창출해 내는 기술융합은 단선적인 기술개발의 한계를 극복하고 혁신적인 제품 및 서비스의 지속적 파생을 가능하게 한다는 점에서 그 중요성이 매우 높게 인식되고 있다.<sup>1)2)3)</sup> 서로 단절된 기술 도메인 간 연계를 통해 새로운 기술지식을 생성해냄으로써 혁신을 선도하고 산업발전을 도모하는 구실을 할 뿐만 아니라 새로운 가치를 창출해 내는 기술적 진보를 이끌어 낸다.<sup>4)5)</sup> 따라서 기업이 기술융합을 통해 새로운 혁신을 실현할 수 있는 잠재적 기술기회를 발굴해 낼 수 있도록 지원하는 체계를 구축하는 것은 해당 기업이 기술기회를 선점하고 지속가능한 성장을 위한 경쟁우위를 갖출 수 있도록 하는 훌륭한 도구가 된다. 기술융합을 통해 실현 가능한 잠재적 기술기회를 발굴해 내기 위해서는 과거에는 존재하지 않았던 기술 속성 간 융합을 예측하기 위한 접근이 요구된다. 이는 새로운 경쟁우위를 선점할 수 있도록 하고 기술융합과 관련한 기술적 변화를 경쟁기업보다 빠르게 포착해 낼 수 있도록 하기 때문이다.

정량적 데이터를 활용하여 특정 기술분야의 트렌드를 분석하고 잠재적 기술기회를 발굴하기 위한 연구는 매우 활발하게 진행되어 오고 있다. 이들은

- 
- 1) 정오뜸 외 2인, “기술적 파급효과 도출을 통한 기술융합 분석 연구—나노바이오 융합 기술 사례를 중심으로”, 『지식재산연구』, 제10권 제4호(2015), 255-290면.
  - 2) 김은정 · 최희진, “토픽모델링과 네트워크분석을 활용한 헬스케어 분야의 핵심기술과 기술융합 분석 연구—특허정보를 중심으로”, 『한국정보통신학회논문지』, 제26권 제5호(2022), 763-778면.
  - 3) Feng, Sida et al., “The technology convergence of electric vehicles: Exploring promising and potential technology convergence relationships and topics”, *Journal of Cleaner Production*, Vol.260(2020), pp.120992.
  - 4) 조용래 · 김의석, “특허 네트워크와 전략지표 분석을 통한 기업 기술융합 전략 연구”, 『지식재산연구』, 제9권 제4호(2014), 191-221면.
  - 5) Ko, Namuk et al., “Analyzing interdisciplinarity of technology fusion using knowledge flows of patents”, *Expert Systems with Applications*, Vol.41 No.4(2014), pp.1955-1963.

주로 R&D의 성과물인 특허데이터를 이용하는데,<sup>6)7)8)</sup> 특허가 발명을 통해 산출된 기술적 신뢰도가 높은 결과물일 뿐만 아니라 급격하게 변화하는 기술발전의 트렌드를 효과적으로 반영할 수 있는 구조화된 지식의 원천으로서 인식되기 때문이다.<sup>9)10)11)12)</sup> 이러한 특허데이터 기반의 기술융합기회 발굴 연구는 일반적으로 특허분류코드인 International Patent Classification(IPC)과 Cooperative Patent Classification(CPC)으로 대표되는 기술클래스 단위에서 이들 간 융합기회를 탐색한다. 각 기술클래스에 대한 특허분류코드의 정의를 바탕으로 관련된 기술 속성을 도출하고 이들 간 융합을 통해 파생 가능한 기회를 정성적으로 분석한다. 이들은 정형화된 기술클래스 데이터를 활용하기 때문에 정량적 분석과정이 수월한 장점이 있다. 하지만 기술클래스는 상당히 넓은 범위의 기술 속성을 포괄하기 때문에 이를 통해 도출해 낸 기술융합기회의 의미가 구체적이지 못하고 상당히 모호한 한계가 존재한다. 본 연구에서는 보다 명확한 기술적 함의를 내포할 수 있도록 기술토픽을 정의하고 이들 간 융합기회를 발굴함으로써 기존연구의 한계를 극복하고자 한다. 또한, 특허데이터 기반의 기존연구는 기술클래스 간 네트워크상 존재하는 기존의 노드 간 링크정보를 기반으로 미래 생성 가능한 새로운 링크를 예

6) 정재민 외 2인, “비즈니스 기회 발굴을 위한 문제-해결방법 기반의 특허분석 방법”, 『지식재산연구』, 제15권 제2호(2020), 187-222면.

7) Kim, Jeeun & Lee, Sungjoo, “Forecasting and identifying multi-technology convergence based on patent data: The case of IT and BT industries in 2020”, *Scientometrics*, Vol.111 No.1(2017), pp.47-65.

8) Lee, Changyong et al., “Anticipating multi-technology convergence: a machine learning approach using patent information”, *Scientometrics*, Vol.126 No.3(2021), pp.1867-1896.

9) 김경수, “국제특허분류를 활용한 기술융합 연결망 구조 및 패턴 분석 : 해양수산 산업 육성 분야의 특허를 중심으로”, 『과학기술정책』, 제3권 제2호(2020), 159-186면.

10) 전상규, “특허 네트워크 분석을 통한 기술융합 및 융합기술의 확산 연구-디지털 데이터 처리 기술 중심으로”, 『지식재산연구』, 제16권 제4호(2021), 161-202면.

11) Choi, Sungchul et al., “A Supervised Learning-Based Approach to Anticipating Potential Technology Convergence”, *IEEE Access*, Vol.10(2022), pp.19284-19300.

12) Seo, Wonchul, “A patent-based approach to identifying potential technology opportunities realizable from a firm’s internal capabilities”, *Computers & Industrial Engineering*, Vol.171(2022), pp.108395.

측하는 접근을 취한다. 그들은 주로 다수의 링크예측 알고리즘을 함께 활용하여 기술클래스 간 링크생성 가능성을 측정하고 미리 정의한 임계치보다 큰 가능성을 지닌 것들만을 잠재적 기술융합기회로서 지정한다.<sup>13)</sup> 이러한 접근은 기존 링크로부터 미래의 잠재적 링크를 발굴하고 새롭게 생성될 것으로 여겨지는 융합기회를 발굴해 낼 수 있는 장점이 있다. 하지만 소수의 링크예측 알고리즘들을 개별적으로만 적용함으로써 인하여 다양한 관점의 링크예측 알고리즘 각각이 지니는 개별적 장점을 효과적으로 결합하여 활용하지 못한다는 단점도 존재한다. 이에 본 연구에서는 링크예측 결과를 효과적으로 결합할 수 있도록 지도학습 기반의 접근법을 제안하고자 한다. 이를 통해 구체화된 기술토픽 간 융합을 통해 실행 가능한 새로운 기술기회를 선점할 수 있도록 지원하는 새로운 형태의 지도학습 기반 기술융합기회 발굴모델을 제시한다는 측면에서 기여점을 지닌다.

우선, 특정 기술분야 관련 특허데이터를 수집하고 토픽모델링을 활용하여 기술적 함의를 내재적으로 포괄하고 있는 기술토픽을 생성한다. 기술토픽은 기술의 합리적 실시를 가능하게 하는 구체화된 기술 속성의 집합으로서 기술융합을 분석하기 위한 기본 단위가 된다. 토픽모델링 적용을 통해 특허와 기술토픽 간 정량적 연관관계를 생성할 수 있는데 이를 활용하여 기술토픽 네트워크를 구축하고 링크예측을 실행함으로써 기술토픽 간 잠재적 링크 생성 가능성을 측정한다. 그리고 다양한 지도학습 알고리즘을 활용하여 기술융합기회를 예측하는 모델을 학습하고 성능을 측정한 후, 우수한 성능을 나타내는 모델을 중심으로 논의를 수행한다. 또한, 본 연구의 타당성을 살펴보기 위하여 본 연구의 모델이 미래 융합 가능성을 높은 확률로 예측한 기술토픽 쌍에 대해 이들의 실현 가능성을 정성적으로 탐색한다. 본 연구는 구체화된 기술토픽 간 융합을 통해 실행 가능한 새로운 기술기회를 선점할 수 있도록 함으로써 국내 기업들의 취약한 R&D 기획역량을 증진하고 지속가능한

13) Kim, Tae San & Sohn, So Young, "Machine-learning-based deep semantic analysis approach for forecasting new technology convergence", *Technological Forecasting and Social Change*, Vol.157(2020), pp.120095.

성장을 위한 경쟁우위를 확보하도록 지원할 수 있을 것으로 기대된다.

## II. 관련 연구

### 1. 특허분석 기반의 기술융합기회 발굴

특허데이터를 활용하여 기술융합 트렌드를 분석하고 기술기회를 발굴하기 위한 기존 연구들은 주로 특허 간 인용관계 분석 또는 특허공동분류 분석을 통해 기술클래스 간 네트워크를 구축하고 다양한 네트워크 분석 기법을 적용하여 의미있는 기술적 함의를 도출하는 형태로 수행되어 왔다.<sup>14)</sup> <sup>15)</sup><sup>16)</sup><sup>17)</sup> 다양한 기술 도메인을 대상으로 산업 수준 또는 기술클래스 수준의 기술융합 트렌드를 파악하고 기술융합 패턴을 규명하거나 해당 분야의 발전 방향에 대한 논의를 수행한다.<sup>18)</sup> 또한, 기술클래스 네트워크에 대한 다양한 특성값을 도출하여 기술융합 수준을 평가하고 기술융합의 파급성을 정량화함으로써 기술융합 발생에 있어 핵심적 역할을 수행하는 기술클래스를 규명한다.<sup>19)</sup><sup>20)</sup><sup>21)</sup> 이들은 기업이 기술융합 트렌드로부터 새로운 기술기회를 발굴해 낼 수 있도록 지원한다. 기술융합의 특성을 보다 세밀하게 살펴볼 수

---

14) 전상규, 앞의 글(주 10), 161-202면.

15) 김은정·최희진, 앞의 글(주 2), 763-778면.

16) Kim, Tae San & Sohn, So Young, *op. cit.*, pp.120095.

17) Choi, Sungchul et al., *op. cit.*, pp.19284-19300.

18) 박수현 외 3인, “5개국 바이오헬스 산업의 기술융합과 트렌드 분석 : 특허 동시분류분석과 텍스트마이닝을 활용하여”, 『한국융합학회논문지』, 제12권 제4호(2021), 9-21면.

19) 이민정 외 2인, “국제특허분류(IPC) 공동출현 네트워크분석을 통해 살펴본 기업의 지식융합 특성이 혁신성과에 미치는 영향—전자통신산업 중소기업을 중심으로”, 『지식재산연구』, 제13권 제1호(2018), 245-284면.

20) 이현민 외 2인, “융합기술 개발전략 기획을 위한 특허 인용 네트워크 기반의 분석 방법론—스마트공장 ICT 기술을 중심으로”, 『한국산학기술학회 논문지』, 제19권 제1호(2018), 34-47면.

21) 김경수, 앞의 글(주 9), 159-186면.

있도록 다양한 정량지표를 활용하여 산업 간 융합을 촉진하기 위한 정책적 시사점을 도출하고 기술융합을 위한 정형화된 프로세스를 정립하기 위한 연구도 수행되었다.<sup>22)23)</sup> 이들은 현재의 기술융합 트렌드를 파악하고 혁신을 유발하기 위한 의미 있는 시사점을 제시한다는 긍정적 의미가 있지만, 기술환경의 변화에 능동적으로 대응하기 위한 구체적 실행방안을 제시하지 못한다는 한계가 있다. 이를 위하여 새로운 기술융합의 발현 여부를 예측하기 위한 연구가 수행되었다.<sup>24)25)</sup> 기술융합기회를 적시에 포착하도록 함으로써 기업의 기술혁신 효율성을 증진할 수 있도록 지원하는 장점이 있다.<sup>26)</sup> 하지만 이들은 기술융합기회가 지니는 구체적 기술적 함의를 도출하지 못하고 의미가 모호한 기술클래스만을 활용하거나 기술융합기회 예측을 위해 활용하는 다양한 지표들을 효과적으로 결합하여 활용하지 못하는 한계를 지닌다. 따라서 본 연구에서는 명확한 기술적 함의를 내포하는 기술토픽을 정의하고 이들 간 융합기회를 발굴할 수 있도록 지도학습 기반의 예측모델을 개발하고자 한다. 이를 통해 구체화된 기술토픽 간 융합을 통해 실행 가능한 새로운 기술기회를 선점할 수 있도록 지원하는 새로운 형태의 지도학습 기반 기술융합기회 발굴모델을 제시한다는 측면에서 기여점을 지닌다.

---

22) Choi, Jae Young et al., "A study on diffusion pattern of technology convergence: Patent analysis for Korea", *Sustainability*, Vol.7 No.9(2015), pp.11546-11569.

23) Kose, Toshihiro & Sakata, Ichiro, "Identifying technology convergence in the field of robotics research", *Technological Forecasting and Social Change*, Vol.146(2019), pp.751-766.

24) Lee, Changyong et al., op. cit., pp.1867-1896.

25) Aaldering, Lukas Jan et al., "Competition or collaboration?—analysis of technological knowledge ecosystem within the field of alternative powertrain systems: a patent-based approach", *Journal of Cleaner production*, Vol.212(2019), pp.362-371.

26) Tang, Ying et al., "A study on dynamic patterns of technology convergence with IPC co-occurrence-based analysis: The case of 3D printing", *Sustainability*, Vol.12 No.7(2020), pp.2655.

## 2. 토픽모델링

토픽모델링은 전체 문서들의 집합으로부터 숨겨진 의미 구조를 도출해 내기 위한 텍스트마이닝 기법으로서 문서 내 단어가 출현한 빈도 및 연관성을 바탕으로 의미가 유사한 단어들을 클러스터링하고 이를 통해 토픽을 추론한다.<sup>27)28)</sup> Latent Dirichlet Allocation(LDA)<sup>29)</sup>는 토픽모델링을 실행하기 위한 가장 대표적인 알고리즘으로 모든 문서가 디리클레 분포를 따르는 토픽들의 확률적 혼합으로 되어 있음을 가정하고 각 토픽을 추론 가능한 단어들의 확률분포로서 표현한다.<sup>30)31)</sup> 각 토픽을 구성하는 단어들의 집합에 대한 확률분포를 생성함으로써 단어들이 각 토픽에 기여하는 수준을 정량화할 수 있기 때문에 문서-토픽-단어 간 관계를 체계적으로 묘사할 수 있다.<sup>32)</sup> 또한, 비구조화된 텍스트 중심의 문서에 내재되어 있는 암묵지 형태의 정보를 형식지의 정보로 추출함으로써 토픽의 기술적 함의를 더욱 명확히 파악할 수 있도록 하는 특징이 있다.

LDA가 문서의 토픽 및 트렌드를 파악하는 효과적인 수단으로 인식되면서 이를 특허문서에 적용함으로써 특정 기술분야의 트렌드를 분석하고 R&D 방향을 기획하기 위한 연구들이 다양한 관점에서 수행되었다. 스마트팩토리 분야 기술 트렌드를 파악하고 핫 & 콜드 기술토픽을 규명함으로써 미래

27) Jelodar, Hamed et al., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey", *Multimedia Tools and Applications*, Vol.78 No.11 (2019), pp.15169-15211.

28) 음수민 외 4인, "LDA 기반의 토픽모델링을 이용한 철도차량용 무선급전시스템 연구 동향 분석", 『대한산업공학회지』, 제45권 제4호(2019), 284-301면.

29) Blei, David et al., "Latent Dirichlet allocation", *Journal of Machine Learning Research*, Vol.3(2003), pp.993-1022.

30) 오승현 외 2인, "특허의 토픽 모델링을 활용한 증강현실 기술 모니터링", 『대한산업공학회지』, 제43권 제3호(2017), 213-228면.

31) 이금실·이인주, "LDA 토픽 모델링 기법을 적용한 관광분야의 가상현실(Virtual Reality) 연구 동향분석", 『관광레저연구』, 제33권 제10호(2021), 85-105면.

32) 서원철, "기술테마 분석을 통한 기술기회발굴 연구—3D 프린팅 기술 사례를 중심으로", 『지식재산연구』, 제16권 제2호(2021), 205-248면.

R&D 방향을 설계하는 연구,<sup>33)</sup> 기술토픽들 간 의미론적 연관관계를 파악하고 유망한 기술의 획득을 위한 기술 로드맵을 구축하는 연구,<sup>34)</sup> 기업의 내부적 보유역량 및 기술기회를 기술토픽으로 모델링하고 토픽 간 연관규칙마이닝을 통해 해당 역량을 활용하여 실행 가능한 실질적 기술기회를 발굴하는 연구<sup>35)</sup> 등이 대표적이다. 기존의 기술융합기회 발굴 연구들은 주로 기술클래스 단위에서 기술적 내용을 정의하고 있어, 기술 속성을 구체적으로 파악하기 어려운 한계가 있다. 이에 본 연구는 구체화된 기술적 함의를 내포하고 있는 기술토픽을 활용함으로써 보다 실효성 있는 기술융합기회 발굴모형을 개발하고자 한다. 특정 기술분야 특허텍스트 정보에 LDA를 적용하여 기술토픽을 추출한 후, 각각의 확률분포를 나타내는 문서-토픽 매트릭스 및 토픽-단어 매트릭스를 생성한다. 우선, 문서-토픽 매트릭스를 통해 특허문서와 기술토픽 간 확률적 연관관계를 파악하고 충분한 수준의 확률관계가 존재하는 경우, 해당 연도 구간 내 관련 기술토픽들 간 융합이 존재한다고 판단한다. 이를 통해 기술융합기회 발굴모형의 학습 및 성능 평가를 위한 클래스 정보를 생성할 수 있다. 그리고 토픽-단어 매트릭스를 통해 각 기술토픽에 내재된 기술 속성을 정의한다. 이는 본 연구의 합리성 및 타당성을 확인하는 방안으로서 마지막 연도 구간 특허데이터로부터 발굴된 새로운 기술융합기회의 기술적 함의를 정성적으로 분석하는 데 활용된다. 본 연구는 문서에 내포된 핵심적 단어들을 중심으로 관련 토픽을 추출하는 토픽모델링을 기술트렌드 분석 분야에 접목하여 새로운 형태의 지도학습 기반 기술융합기회 발굴모형을 개발하는 데 활용할 수 있음을 입증함으로써 해당 분야 학문적·실무적 연구 범위를 크게 확장할 수 있다는 측면에서 기여점을 지닌다.

33) Hussain, Adnan et al., "Analyzing technological trends of smart factory using topic modeling", *Asian Journal of Innovation and Policy*, Vol.10 No.3(2021), pp.380-403.

34) Ma, Tingting et al., "Combining topic modeling and SAO semantic analysis to identify technological opportunities of emerging technologies", *Technological Forecasting and Social Change*, Vol.173(2021), pp.121159.

35) Seo, Wonchul, *op. cit.*, pp.108395.

### 3. 링크예측

링크예측은 복잡한 네트워크상에서 노드 및 링크의 특성에 대한 분석을 바탕으로 가까운 미래에 생성될 것으로 여겨지는 새로운 링크를 예측하는 기법이다.<sup>36)37)</sup> 현재 네트워크 상 연결되어 있지 않은 노드들의 모든 쌍에 대해 미래 연결될 가능성을 다양한 관점에서 산출한다.<sup>38)39)</sup> 링크예측 알고리즘들은 크게 로컬(local), 글로벌(global), 준로컬(quasi-local) 방법으로 구분된다.<sup>40)41)</sup> 로컬 방법은 각 노드들의 이웃노드에 대한 정보만을 활용하여 링크 생성 가능성을 탐색하며,<sup>42)</sup> 대부분 common neighbors(cn) 알고리즘으로부터 파생된 것들이다. 이들은 두 객체가 공동의 이웃을 많이 가질수록 그들이 연결될 가능성이 높을 것이라 가정하고 예측을 수행한다. 반면, 글로벌 방법은 모든 노드를 구성하는 전체 네트워크의 속성을 종합적 관점에서 분해하여 예측을 수행하고, 준로컬 방법은 로컬과 글로벌의 특성이 혼합된 형태를 지닌다.<sup>43)</sup>

다양한 객체들 간 미래 융합가능성을 탐색하는 링크예측을 활용하여 기술 융합 트렌드를 분석하기 위한 연구들 역시 넓은 범위에서 수행되었다. 특히

36) 장동호·이충목, “GAN을 이용한 비지도학습 기반의 새로운 링크 예측 알고리즘”, 『대한산업공학회지』, 제46권 제3호(2020), 245-255면.

37) Liu, Yingjie et al., “Link prediction algorithm based on the initial information contribution of nodes”, *Information Sciences*, Vol.608(2022), pp.1591-1616.

38) Cai, Lei et al., “Line graph neural networks for link prediction”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.44 No.9(2022), pp.5103-5113.

39) Ma, Jing et al., “Identifying translational indicators and technology opportunities for nanomedical research using tech mining: The case of gold nanostructures”, *Technological Forecasting and Social Change*, Vol.146(2019), pp.767-775.

40) Lü, Linyuan & Zhou, Tao, “Link prediction in complex networks: A survey”, *Physica A: Statistical Mechanics and its Applications*, Vol.390 No.6(2011), pp.1150-1170.

41) Choi, Sungchul et al., *op. cit.*, pp.19284-19300.

42) Kumar, Ajay et al., “Link prediction techniques, applications, and performance: A survey”, *Physica A: Statistical Mechanics and its Applications*, Vol.553(2020), pp.124289.

43) Kumar, Ajay et al., *op. cit.*, pp.124289.

서지 및 텍스트 정보와 IPC 간 링크예측 정보를 활용하여 기술클래스 단위에서의 미래 기술융합 예측모델을 개발하기 위한 연구,<sup>44)</sup> IPC 간 링크예측, 기술적 파급효과 및 기술적 유사도 정보를 결합하여 잠재적 기술융합기회를 발굴하기 위한 연구<sup>45)</sup> 등이 대표적이다. 하지만 이들은 기술 속성을 구체적으로 파악하기 어려운 기술클래스를 활용할 뿐만 아니라 소수의 링크예측 알고리즘들을 개별적으로만 적용함으로 인해 이들 각각이 지니는 개별적 장점을 효과적으로 결합하여 활용하지 못하는 한계를 지니고 있다.<sup>46)47)48)</sup> 따라서 본 연구에서는 기술클래스에 비해 기술적 구체성이 높은 기술토픽 단위에서의 융합 네트워크를 구축하고 다양한 링크예측 알고리즘들을 복합적으로 적용하여 기술융합기회를 발굴할 수 있도록 지도학습 기반의 접근법을 개발하고자 한다. 이는 소수의 링크예측 알고리즘들을 개별적으로만 적용함으로 인해 다양한 관점의 링크예측 알고리즘 각각이 지니는 개별적 장점을 효과적으로 결합하여 활용하지 못한다는 기존연구의 한계를 극복할 수 있다는 측면에서 기여점을 지닌다.

### Ⅲ. 기술토픽 분석을 통한 기술융합기회 발굴 방안

기술 속성을 구체적으로 표현하는 기술토픽 간 링크예측을 바탕으로 기술융합기회를 발굴하기 위한 본 연구는 [그림 1]과 같이 1) 특정 기술분야 관련 특허데이터 수집 및 기술토픽 생성, 2) 기술토픽 네트워크 구축 및 다양한 링

44) Kim, Tae San & Sohn, So Young, *op. cit.*, pp.120095.

45) Choi, Sungchul et al., *op. cit.*, pp.19284-19300.

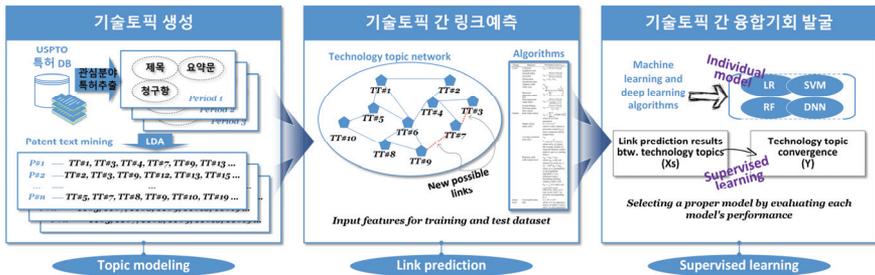
46) Lee, Changyong et al., "Novelty-focused patent mapping for technology opportunity analysis", *Technological Forecasting and Social Change*, Vol.90(2015), pp.355-365.

47) Park, Hyochul et al., "Regional innovation capability from a technology-oriented perspective: An analysis at industry level", *Computers in Industry*, Vol.129(2021), pp.103441.

48) Choi, Sungchul et al., *op. cit.*, pp.19284-19300.

크예측 알고리즘을 활용한 기술토픽 간 연계관계 확장, 3) 머신러닝 및 딥러닝 기법의 적용을 통해 기술토픽 간 융합기회를 발굴하는 지도학습 모델 개발의 3단계로 이루어진다. 추가적으로 본 연구를 통해 발굴된 기술융합기회의 타당성을 살펴보기 위한 논의를 수행한다.

[그림 1] 기술토픽 분석을 통한 기술융합기회 발굴 연구 프레임워크



## 1. 기술토픽 생성

기술토픽 분석을 통해 기술융합기회를 발굴하기 위한 본 연구는 기본적으로 특허데이터를 활용하기 때문에 첫 번째 단계에서는 특정 기술분야 관련 특허를 수집한다. 본 연구를 통해 발굴된 기술융합기회의 의미를 보다 구체적으로 파악할 수 있도록 기술융합이 활발하게 발생하고 있는 특정 기술분야의 최근 특허를 중심으로 수집을 진행한다. 특허출원의 절대적 규모가 큰 Intellectual Property(IP) 5 중, 미국은 출원량이 많을 뿐만 아니라 전체 출원인 중 해외 출원인의 비중이 상당히 큰 국가이기 때문에 미국특허에 대한 분석은 글로벌 기업들의 기술융합 활동을 종합적으로 분석하기에 적합하다고 할 수 있다.<sup>49)</sup> 따라서 본 연구에서도 United States Patent and Trademark Office(USPTO)에 출원된 특허를 활용하여 기술토픽을 생성하고 이들 간 기술융합기회를 탐색하고자 한다. 이를 위해서는 특허문서 내

49) 서원철, 앞의 글(주 32), 205-248면.

텍스트 정보가 필수적으로 요구되기 때문에 텍스트 정보를 중심으로 특허데이터를 수집한다.

수집된 특허텍스트 정보에 LDA를 적용하여 기술 속성을 내재적으로 함축하고 있는 기술토픽을 생성한다. 특허문서에는 제목, 요약문, 청구항, 배경기술, 발명의 내용 등 다양한 종류의 텍스트 정보가 포함되어 있다. 이 중, 배경기술은 해당 기술 관련 주변기술에 대한 내용을 포함하고 발명의 상세한 설명은 방대한 규모의 텍스트를 지니고 있어 기술 속성에 대한 핵심내용의 효과적 추출이 어려운 문제가 있다. 반면 제목, 요약문 및 대표 청구항은 해당 기술을 구성하는 핵심 구성요소를 중심으로 간결하게 기술되어 있기 때문에 기술 속성을 명확하게 함축하는 기술토픽을 추출하는 데 있어 상대적으로 적합하다. 따라서 본 연구에서는 제목, 요약문 및 대표 청구항을 활용하여 코퍼스(corpus)를 구축하고 불용어 처리 및 어간 추출을 통한 동의어 통합 등의 텍스트 전처리 과정을 진행하여 코퍼스를 정제한다. 이렇게 정제된 코퍼스에 LDA를 그대로 적용할 경우, 특허문서에 등장하는 빈도수가 매우 미미한 단어들까지 활용함으로써 기술적 의미가 모호한 토픽까지도 추출할 가능성이 크다. 따라서 본 연구에서는 충분한 빈도수를 보이는 단어들만을 선별한 후, 이들만을 활용하여 토픽을 생성함으로써 기술적 함의가 충분히 내재될 수 있도록 한다. LDA 적용을 위해서는 생성할 토픽의 수를 사전에 지정하여야 한다. 토픽의 수는 토픽모델링 결과의 품질에 영향을 미친다. 적정 수준의 토픽 수를 결정하기 위한 기준으로는 토픽모델의 품질수준에 대한 가시적 가이드를 제공하는 perplexity와 토픽생성 결과에 대한 해석 용이성을 나타내는 coherence 등이 있다.<sup>50)</sup> 이 중, 본 연구에서는 명확한 기술적 함의를 내포하고 있어 품질이 우수하다고 여길 수 있는 기술토픽을 생성할 수 있도록 perplexity를 활용한다. Perplexity 값이 낮을수록 토픽모델의 품질수준이 우수함을 의미하기 때문에 토픽의 수를 달리하며 perplexity 값을 측정하고, 이 값의 변화가 완만해지는 지점을 나타내는 토픽의 수를 최종

50) O'callaghan, Derek et al., "An analysis of the coherence of descriptors in topic modeling", *Expert Systems with Applications*, Vol.42 No.13(2015), pp.5645-5657.

적으로 생성할 토픽의 수로 설정한다.

본 연구는 기술융합기회를 발굴하기 위한 지도학습 모델을 개발하는 것을 목적으로 한다. 이를 위해서는 모델 학습을 위한 인풋 피처 및 기술융합 발현 여부를 나타내는 클래스 정보가 필요하다. 따라서 특허데이터를 3개의 연도 구간으로 구분하고, 첫 번째 구간의 인풋 피처를 활용하여 두 번째 구간에서의 기술융합 발현 여부를 예측하는 모델을 학습하고 두 번째 구간의 인풋 피처를 활용하여 세 번째 구간에서의 기술융합 발현 여부에 대한 예측 성능을 측정하는 접근을 취하고자 한다.

## 2. 기술토픽 간 링크예측

LDA의 적용을 통해 최종적으로 특허문서와 토픽 간 정량적 연관관계를 생성할 수 있는데, 해당 관계는 상당히 큰 편차를 보이는 것이 일반적이다. 만약, 이를 모두 포함하게 되면 매우 미미한 관계까지도 포괄하는 기술토픽 네트워크가 생성되게 된다. 이에 본 연구에서는 적정 임계값 이하는 제거하고 특허-토픽 관계를 정립한다. 이는 특허에 내재된 발명에 대한 기술적 사상이 명확한 의미를 지니는 기술토픽의 형태로 정형화될 수 있도록 한다. 하나의 특허와 연관된 기술토픽들은 서로 간 융합을 바탕으로 해당 특허발명의 창출에 기여하기 때문에,<sup>51)</sup> 특허-토픽 관계를 활용함으로써 기술토픽 네트워크를 생성할 수 있다. 네트워크에 존재하는 기술토픽 간 연관관계는 각 토픽과 연계된 기술지식들이 융합적인 형태로 새로운 발명에 얼마나 활발히 응용되는지를 나타낸다. 특정 연도 구간의 네트워크에서 링크로 연계된 기술토픽들은 그들 사이에 이미 기술융합이 발현되어 해당 시기에 특허발명이 창출되었음을 의미하고, 연계되지 않은 토픽들은 다음 연도 구간에서 기술융합의 발현 가능성에 대한 탐색이 필요함을 의미한다. 따라서 본 연구에서는 링크예측 결과에 대한 지도학습을 활용하여 현재는 존재하지 않지만 다

51) Choi, Sungchul et al., *op. cit.*, pp.19284-19300.

음 구간에서 나타나게 될 것으로 기대되는 기술토픽 간 융합기회를 발굴해 내는 모델을 개발하고자 한다. 하나의 네트워크 내 노드 간 근접성의 정도를 계산하는 링크예측 알고리즘은 그 종류가 매우 다양하다.<sup>52)</sup> 이들은 모두 두 노드 간 근접성의 정도가 높을수록 다음 연도 구간에는 그들 간 링크가 형성될 가능성이 크다고 판단한다. 링크예측 알고리즘은 <표 1>과 같이 로컬, 글로벌, 준로컬 방법으로 구분되고, 본 연구에서는 이들을 모두 지도학습을 위한 인풋 피처로 활용한다. 3개의 연도 구간에 대해 개별적으로 기술토픽 네트워크를 생성하고 각 네트워크에서 연계되지 않은 기술토픽들에 대해 링크예측을 수행한다. 각 링크예측 결과를 인풋으로 활용하여 다음 연도 구간에서 새로 등장할 것으로 여겨지는 기술융합 기회를 예측하는 모델을 개발한다. 이때, 각 특허데이터에 LDA가 독립적으로 적용되기 때문에 산출된 기술토픽들은 기본적으로 서로 연계관계가 존재하지 않게 된다. 하지만 3개의 구간에 걸쳐서 기술토픽 간 융합기회를 탐색하기 위해서는 이들을 연계시킬 수 있어야 한다. 따라서 본 연구에서는 각 토픽과 관련된 단어의 유사성을 통해 토픽 간 유사도를 측정하고 가장 유사도가 높은 토픽들을 도출하는 형태로 구간별 토픽 간 연계관계를 형성한다. 물론, 하나의 토픽은 다른 두 개 이상의 토픽과 동일한 개수의 동일한 단어를 공유할 수 있기 때문에 동일한 단어의 개수를 기반으로 연계성을 생성하게 되면 동일한 수준의 유사도를 갖는 관계가 나타날 수밖에 없다. 이에 본 연구에서는 각 단어의 출현빈도에 따라 단어의 중요도를 산출하고 이를 가중치로 활용함으로써 이러한 문제를 해결한다. 이러한 과정을 통해 최종적으로 각 구간에 존재하는 기술토픽들 모두 3구간에 걸쳐 1:1:1의 매핑관계를 지닐 수 있게 된다.

---

52) Ma, Jing et al., *op. cit.*, pp.767-775.

<표 1> 링크예측 알고리즘 ( $I(x)$ 와  $k_x$ 는 노드  $x$ 의 이웃 노드 집합과 연결된 노드의 수를 의미함)

구분	알고리즘	산출기준	참조문헌
Local	Common neighbors(cn)	$s_{xy} =  I(x) \cap I(y) $	Newman (2001)
	Salton Index(cos)	$s_{xy} = \frac{ I(x) \cap I(y) }{\sqrt{k_x \times k_y}}$	-
	Jaccard Index(jaccard)	$s_{xy} = \frac{ I(x) \cap I(y) }{ I(x) \cup I(y) }$	Jaccard (1912)
	Sørensen Index(sor)	$s_{xy} = \frac{2 I(x) \cap I(y) }{k_x + k_y}$	Sørensen (1948)
	Hub Promoted Index(hpi)	$s_{xy} = \frac{ I(x) \cap I(y) }{\min\{k_x, k_y\}}$	Ravasz et al. (2002)
	Hub Depressed Index(hdi)	$s_{xy} = \frac{ I(x) \cap I(y) }{\max\{k_x, k_y\}}$	
	Leicht-Holme-Newman Index (lhn_local)	$s_{xy} = \frac{ I(x) \cap I(y) }{k_x \times k_y}$	Leicht et al. (2006)
	Preferential Attachment(pa)	$s_{xy} = k_x \times k_y$	Barabási & Albert (1999)
	Adamic-Adar Index(aa)	$s_{xy} = \sum_{z \in I(x) \cap I(y)} \frac{1}{\log k_z}$	Adamic & Adar (2001)
	Resource Allocation Index(ra)	$s_{xy} = \sum_{z \in I(x) \cap I(y)} \frac{1}{k_z}$	Zhou et al. (2009)
Global	Katz Index(katz)	$S = (I - \beta A)^{-1} - I$ , $A$ 는 네트워크에 대한 인접행렬, $\beta$ 는 자유 파라미터	Katz (1953)
	Leicht-Holme-Newman Index, global version(lhi_global)	$S = D^{-1} \left( I - \frac{\emptyset A}{\lambda_1} \right)^{-1} D^{-1}$ , $\lambda_1$ 은 $A$ 의 고유값 최대치, $\emptyset$ 는 자유 파라미터	Leicht et al. (2006)
	Average Commute Time(act)	$s_{xy} = \frac{1}{m(x,y) + m(y,x)}$ , $m(x,y)$ 는 노드 $x$ 에서 노드 $y$ 로 이동하기 위한 평균 단계 수	Klein & Randić (1993)
	Normalized Average Commute Time(act_n)	$s_{xy} = \frac{1}{m(x,y)\pi_y + m(y,x)\pi_x}$ , $\pi_x = \frac{k_x}{\sum_y k_y}$	
	Random Walk with Restart(rwr)	$s_{xy} = q_{xy} + q_{yx}$ , $q_{xy}$ 는 벡터 $q_x = (1 - \alpha)P^T q_x + \alpha e_x$ 의 $y$ 번째 요소, $\alpha$ 는 PageRank 알고리즘의	Brin & Page (1998)

		<p>확률, <math>P</math>는 노드 <math>x, y</math>가 연결된 경우 <math>P_{xy} = \frac{1}{k_x}</math>의 인접행렬</p>	
	Matrix Forest Index(mf)	<p><math>S = (I + L)^{-1}</math>,  <math>L</math>은 네트워크에 대한 라플라시안 행렬</p>	Chebotarev & Shamis (2006)
Quasi-local	Local Path Index(lp)	<p><math>S = A^2 + \epsilon A^3</math>,  <math>\epsilon</math>는 자유 파라미터</p>	Zhou et al. (2009)

### 3. 기술토픽 간 융합기회 발굴

연도 구간 1에서 연계되지 않은 토픽들에 대한 링크예측 결과를 인풋 피쳐로 하고 해당 토픽들이 연도 구간 2에서 연계되어 있는지의 여부를 클래스 정보로 하여 기술융합 기회를 예측하는 모델을 학습한다. 기본적으로 링크예측 결과는 알고리즘에 따라 값의 범위가 매우 다를 수밖에 없다. 따라서 본 연구에서는 이들에 대한 정규화를 진행한 후, 모델 학습을 실행한다. 예측 또는 분류문제를 위해 다양한 종류의 머신러닝 및 딥러닝 알고리즘들이 활용되어 오고 있는데, 이들은 각각의 특징을 지니고 있기 때문에 본 연구에서는 여러 가지 알고리즘을 활용하여 모델 학습 및 성능 측정을 실행하고 우수한 성능을 나타내는 모델을 중심으로 논의를 수행하고자 한다. Logistic Regression(LR)은 간단하고 직관적으로 실행 가능하다는 장점으로 인해 널리 사용되는 확률모델로서 여러 독립변수의 선형 결합을 바탕으로 종속변수에 대한 발생 가능성을 예측한다. Support Vector Machine(SVM)은 훈련 데이터 셋으로부터 클래스 값 간 간격을 최대화할 수 있는 공간상의 초평면(hyperplane)을 찾는 것을 목적으로 한다. 하나의 모델을 활용하는 것에 비해, 여러 기본 모델을 복합적으로 활용하는 앙상블 모델은 더 나은 예측 성능을 보이는 것으로 알려져 있다.<sup>53)</sup> 따라서 Decision Tree 기반의 앙상블 모

53) Xu, Xiujuan et al., "MM-UrbanFAC: Urban functional area classification model based on multimodal machine learning", *IEEE Transactions on Intelligent Transportation Systems*, Vol.23 No.7(2021), pp.8488-8497.

델인 Random Forest(RF)도 본 연구에서 활용한다. 가장 기본적인 형태의 딥러닝 알고리즘인 Deep Neural Network(DNN) 학습 또한 이용된다. 학습 과정을 조절하기 위해서는 하이퍼파라미터 설정이 필요하므로 본 연구에서는 교차검증 기반의 그리드 서치를 활용한다.

기술융합기회 발굴모델에 대한 학습은 연도 구간 1의 링크예측 결과와 연도 구간 2의 기술융합 발현 여부에 대한 클래스 정보를 활용하여 이루어진다. 이렇게 훈련된 모델의 성능 측정에는 연도 구간 2의 링크예측 결과와 연도 구간 3의 클래스 정보가 활용된다. 본 연구는 잠재적인 미래 기술융합기회에 대한 예측도 수행하고자 하기 때문에, 연도 구간 3의 링크예측 결과를 학습된 모델에 입력하여 미래의 기술융합에 대한 예측을 수행하고 잠재적 기술융합기회에 대한 논의도 수행한다.

## IV. 사례 분석

### 1. 웨어러블 기술 관련 기술토픽 생성

본 연구에서 제안하는 기술융합기회 발굴모델을 개발하기 위해서는 특정 기술분야 관련 최근 특허데이터가 필요하다. 이를 위하여 본 연구에서는 웨어러블을 사례 분석 기술분야로 선정하였다. 웨어러블은 인체에 착용할 수 있을 정도로 가볍고 작은 휴대용 전자기기로서 많은 도구와 센서를 장착하여 착용한 사람의 다양한 감정 패턴을 인식할 수 있도록 한다. 웨어러블 디바이스의 광범위한 활용 도메인에 의해 다양한 종류의 기술들을 융합한 혁신적인 제품들이 활발하게 파생되고 있다. 따라서 이러한 웨어러블 기술분야를 대상으로 사례 분석을 시행하여 본 연구의 기술융합기회 발굴 방안에 대한 타당성 및 실제적 활용성을 탐색하고자 한다. Song et al.(2017)은 웨어러블과 밀접한 기술영역으로 signal transmission and telecommunications 와 medical equipment를 제시하고 웨어러블 관련 특허데이터를 추출할 수

있도록 각각의 기술영역을 표현할 수 있는 IPC를 제시하고 있다.<sup>54)</sup> 따라서 위 두 기술영역 관련 IPC가 동시에 부여된 특허를 추출함으로써 웨어러블 관련 기술융합을 탐색할 수 있는 특허데이터를 수집할 수 있다. <표 2>는 본 연구의 사례 분석을 위해 수집된 USPTO에 등록된 웨어러블 관련 특허의 연도별 수를 나타낸다. 웨어러블에 대한 기술적 관심이 높아지는 추세에 따라 출원 및 등록되고 있는 특허의 양 또한 지속적으로 증가하고 있음을 알 수 있다.

<표 2> USPTO에 등록된 웨어러블 기술 관련 특허 수

연도 구간	등록연도	특허 수
연도 구간 1	2012	197
	2013	175
	2014	341
연도 구간 2	2015	734
	2016	855
	2017	1,036
연도 구간 3	2018	1,107
	2019	1,471
	2020	1,481
Total		7,397

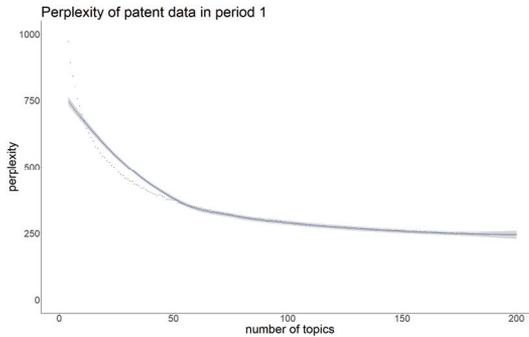
각 연도 구간의 특허텍스트 데이터에 LDA를 적용하여 기술도픽을 생성한다. 제목, 요약문, 대표 청구항을 활용하여 코퍼스를 구축하고 기술적 의미가 명확한 단어들을 나타낼 수 있도록 unigram뿐만 아니라 bigram까지 포함하여 코퍼스를 정제한다. 연도 구간 1에서는 총 51,756개의 단어로 구성된 코퍼스가 구축되었는데, 이를 모두 활용하게 되면 특허문서 내 출현 빈도수가 매우 미미한 것들까지 포함된 토픽이 생성되는 문제가 있다. 따라서 기술

54) Song, Chie Hoon et al., "Anticipation of converging technology areas — A refined approach for the identification of attractive fields of innovation", *Technological Forecasting and Social Change*, Vol.116(2017), pp.98-115.

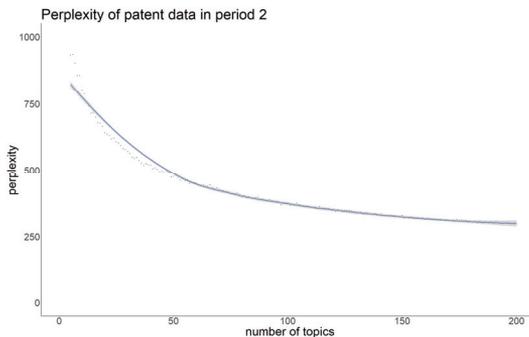
적 함의가 충분히 내재될 수 있도록 단어의 출현빈도가 충분한 단어만을 포함하도록 한다. 출현빈도에 따라 정렬한 후, 누적 비율을 살펴본 결과 상위 10%의 단어들만 전체 출현빈도의 약 90% 이상을 차지하고 있어 총 단어 수의 10% 수준에 해당하는 5,000개의 단어만을 선택하고 최종적으로 이들만을 활용하여 기술토픽을 생성하도록 한다. 연도 구간 2 및 3의 특허데이터에 대해서도 동일한 기준에 따라 코퍼스를 정제하고 각각의 기술토픽을 생성한다.

생성할 토픽의 개수를 결정하기 위해 perplexity를 측정한다. 토픽의 개수를 200개까지 늘려가며 perplexity 값을 측정한 결과 [그림 2]와 같이 대략 60개부터 비슷한 수준의 값을 나타내는 것을 확인할 수 있다. Perplexity 값

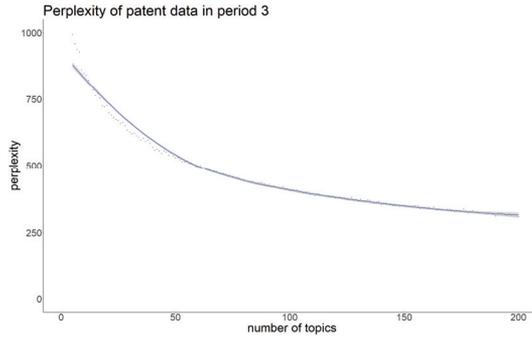
[그림 2] 기술토픽 개수의 증가에 따른 perplexity 값 변화 (a) 연도 구간 1, (b) 연도 구간 2, (c) 연도 구간 3



(a)



(b)



(c)

은 일반적으로 토픽의 개수가 늘어남에 따라 자연스럽게 줄어드는 특징이 있기 때문에 60개 이상부터는 비슷한 토픽모델 품질수준을 확보할 수 있음을 알 수 있다. 하지만 60개는 기술융합기회를 탐색하기에 충분하다고 보기 어렵고, 너무 많은 수의 토픽을 생성하게 될 경우에는 의미가 중복되는 토픽들이 다수 생성될 우려가 있다. 따라서 본 연구에서는 기술융합의 특징을 명확히 나타낼 수 있도록 할 뿐만 아니라 중복의 우려도 줄일 수 있도록 최종적으로 100개의 토픽을 생성하기로 한다.

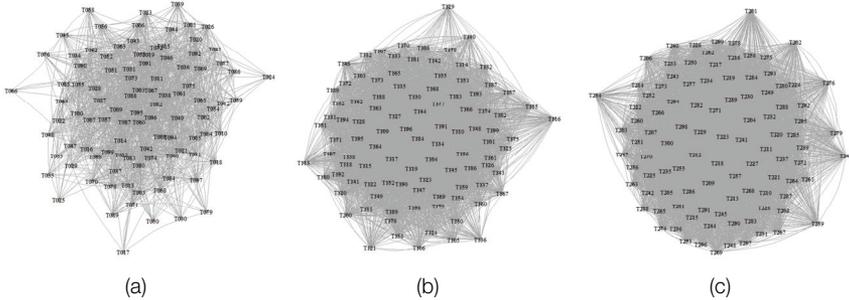
LDA의 적용을 통해 연도 구간 1에서는 713개의 특허문서와 기술토픽 100개 간 확률적 연관관계가 생성되고, 연도 구간 2에서는 2,625개의 특허문서와 기술토픽 100개, 연도 구간 3에서는 4,059개의 특허문서와 기술토픽 100개 간 연관관계가 생성된다. 각 기술토픽이 명확한 기술적 함의를 내재하도록 하기 위해서는 이러한 연관관계를 바탕으로 기술토픽을 특허에 매핑하는 과정이 요구된다. 하지만 연관관계의 수준도 매우 큰 편차를 나타내기 때문에 0이 아닌 모든 관계를 활용하는 것은 상당히 미미한 관계까지도 포괄해 버리는 문제가 있다. 이에 본 연구에서는 기술토픽의 의미는 명확하게 표현하되, 중복될 수 있는 토픽의 생성을 최대한 방지할 수 있도록 각 연도 구간에서의 평균수준을 활용하여 기술토픽과 특허문서를 매핑하도록 한다.

## 2. 기술토픽 네트워크 구축 및 링크예측

기술토픽과 특허 간 정제된 매핑관계를 활용하여 기술토픽 네트워크를 구축한다. 하나의 특허와 연관관계를 지니고 있는 기술토픽들을 모두 추출하고 이들 간 링크를 형성한 후 이를 누적함으로써 가중치가 있는 네트워크를 구성할 수 있다. 하지만 모든 가중치를 활용하여 네트워크를 구축하게 될 경우, 미래 연결가능성을 탐색해야 할 대상인 현재의 네트워크에서는 연결되지 않은 기술토픽 쌍들이 거의 존재하지 않게 되는 문제가 있다. 따라서 적정 수준의 임계치를 설정하고 이에 미치지 못하는 가중치를 지닌 링크는 끊어 내기 위한 기준을 마련할 필요가 있다. 너무 낮은 수준의 임계치 활용은 거의 모든 기술토픽에서 이미 융합이 이루어지고 있어 예측을 위한 대상 자체를 줄이는 문제가 있고, 너무 높은 수준의 임계치 활용은 거의 대부분의 기술토픽들 사이에서 융합이 발생하지 못하게 됨을 나타내는 문제가 있다. 따라서 본 연구에서는 충분한 수준의 기술융합기회 탐색이 가능하도록 네트워크상 링크 가중치의 평균값을 임계치로 활용한다.

[그림 3]은 이러한 과정을 거쳐 구축된 3개 연도 구간에서의 기술토픽 네트워크를 나타낸다. 각 네트워크에서 링크예측을 수행함으로써 현재의 구간에서는 연결되지 않은 기술토픽 쌍이 다음 구간에서는 연결될 가능성을 탐색한다. 링크예측은 기본적으로 서로 연결되지 않은 서브 네트워크가 존재할 때 각 서브 네트워크의 노드들에 대해서는 연결 가능성을 탐색하지 않는 특징이 있다. 따라서 본 연구의 기술토픽 네트워크에 대해서도 분리된 서브 네트워크가 존재하는지 확인하는 것이 필요하다. 모든 연도 구간에서의 네트워크에서 이와 같이 격리된 기술토픽은 없는 것으로 확인되었다. 따라서 각 네트워크에 링크예측을 적용함으로써 기술토픽 간 링크가 다음 연도 구간에 생성될 가능성을 정량화할 수 있다.

[그림 3] 기술토픽 네트워크: (a) 연도 구간 1, (b) 연도 구간 2, (c) 연도 구간 3



3개의 네트워크에 존재하는 노드는 모두 각 연도 구간의 특허데이터로부터 생성된 기술토픽이다. 활용하는 특허데이터가 연도 구간별로 다르기 때문에 기술토픽 또한 서로 연계성이 없는 상태로 생성된다. 본 연구는 각 연도 구간을 거치며 변화하는 기술토픽 간 융합 양상을 예측하는 모델 개발을 목적으로 하기 때문에, 각 연도 구간의 기술토픽을 의미적 유사성에 기반하여 연계시킬 수 있어야 한다. LDA 적용을 통해 생성된 토픽-단어 매트릭스는 각 토픽의 기술적 의미를 유추할 수 있도록 한다. 따라서 각 연도 구간의 토픽에 대해 연계된 단어가 얼마나 유사한지를 측정하게 되면, 기술토픽 간 의미적 연계성을 파악할 수 있다. 이를 위해 우선 각 토픽 생성에 기여하는 수준 상위 100개의 단어에 대한 토픽-단어 매트릭스를 추출한다. 너무 작은 수의 단어를 활용할 경우, 토픽 간 단어 유사성이 거의 나타나지 않게 되고, 너무 많은 수의 단어를 활용할 경우, 모든 토픽에 있어 유사성이 차별화되지 않은 문제가 있어 본 연구에서는 적정 수준의 유사성을 표현할 수 있도록 상위 100개의 단어를 활용하기로 한다. 연도 구간 1의 기술토픽 100개와 연도 구간 2의 기술토픽 100개에 대해 총 10,000가지 경우에 대한 단어 유사성을 살펴보고 연도 구간 1의 기술토픽을 중심으로 최대의 유사성을 나타내는 연도 구간 2의 기술토픽을 1개씩 연계시키는 방법으로 기술토픽 간 1:1의 의미적 연계관계를 생성한다. 연도 구간 2의 기술토픽 100개와 연도 구간 3의 기술토픽 100개에 대해서도 동일한 방법을 적용함으로써 최종적으로 3

개의 연도 구간에 존재하는 각 100개의 기술토픽에 대해 1:1:1의 연계관계를 생성한다.

다양한 링크예측 알고리즘을 활용하여 연도 구간 1에서 연결되지 않은 모든 기술토픽 쌍에 대해 이들이 연도 구간 2에서 연결될 가능성을 정량화한다. <표 3>은 측정된 링크예측 가능성에 대한 기술통계량을 나타낸다. 각 알고리즘별로 값이 클수록 다음 연도 구간에서 기술토픽 간 융합 가능성이 높아질 수 있다는 긍정적 함의를 유추할 수 있다. <표 3>에서 확인할 수 있는 것처럼 알고리즘에 따라 측정된 링크예측 값의 편차가 매우 크기 때문에 이들을 기술토픽 간 융합기회를 발굴하는 지도학습 모델을 위한 인풋 피처로 활용할 수 있도록 정규화를 통한 피처 스케일링을 먼저 실행한다.

<표 3> 연도 구간 1 내 미연결 기술토픽 쌍에 대한 링크예측 수행 결과 기술통계량

알고리즘	Min.	Q1	Median	Mean	Q3	Max.
cn	1	8	14	14.687	20	55
cos	0.0357	0.2987	0.4009	0.3945	0.4950	0.7828
jaccard	0.0169	0.1633	0.2407	0.2445	0.3182	0.6429
sor	0.0333	0.2807	0.3881	0.3805	0.4828	0.7826
hpi	0.0192	0.2206	0.3269	0.3240	0.4222	0.7660
hdi	0.0370	0.3810	0.5000	0.4975	0.6154	1.0000
lhn_local	0.0013	0.0095	0.0113	0.0115	0.0132	0.0317
pa	48	775	1,222	1,297	1,710	5,100
aa	0.2316	2.1833	3.6636	3.9230	5.2835	15.113
ra	0.0133	0.1920	0.3226	0.3559	0.4836	1.4854
katz	0.0000	0.0000	0.0000	0.0004	0.0010	0.0011
lhi_global	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
act	0.0035	0.0148	0.0206	0.0207	0.0259	0.0525
act_n	1.6178	2.0461	2,1979	2,2202	2,3715	3.5138
rwr	0.0012	0.0080	0.0106	0.0140	0.0221	0.0496
mf	0.0092	0.0098	0.0099	0.0099	0.0100	0.0127
lp	0.1500	11,965	19,560	20,647	27,460	76,350

### 3. 기술융합기회 발굴모델 개발

링크예측 결과를 활용하여 기술토픽 간 융합을 예측하기 위한 지도학습 모델을 개발한다. Random Forest와 같이 Tree 기반의 알고리즘에서는 피쳐 스케일링이 필수는 아니지만, 본 연구에서는 다양한 알고리즘을 활용하여 모델 학습 및 성능 측정을 실행하고 우수한 성능을 나타내는 모델을 중심으로 논의를 수행하고자 하기 때문에 모델 간 일관성을 유지할 수 있도록 피쳐 스케일링 후 학습을 진행한다. 모델 학습을 위해서는 알고리즘별로 다양한 하이퍼파라미터를 설정하는 것이 필요하다. 최적의 파라미터를 활용한 모델 학습이 이루어질 수 있도록 본 연구에서는 교차검증 기반의 그리드 서치를 이용한다. <표 4>는 각 알고리즘의 그리드 서치를 위해 활용한 파라미터 값 범위와 최종적으로 선택된 값을 나타낸다. 기본적으로 본 연구에서의 학습 모델은 비선형 분류를 시행해야 하기 때문에 SVM의 커널함수는 비선형 분류를 위해 많이 사용하는 RBF를 기준으로 그리드 서치를 수행한다.

<표 4> 하이퍼파라미터 설정을 위한 그리드 서치 기준

모델	파라미터	값 범위	선택된 값
Logistic Regression(LR)	Cost	{ 0.001, 0.01, 0.1, 1, 10, 100, 1,000 }	0.001
	Penalty	{ 11, 12 }	12
Support Vector Machine(SVM)	Cost	{ 0.001, 0.01, 0.1, 1, 10, 100, 1,000 }	10
	Gamma	{ 0.0005, 0.001, 0.01, 0.1, 1, 10, 100 }	10
Random Forest(RF)	트리 개수	{ 50, 100, 200, 500 }	500
	트리 최대 깊이	{ 4, 6, 8, 10, 12 }	12
	최대 피쳐 개수	{ 'auto', 'sqrt', 'log2' }	'auto'
	불순도 기준	{ 'gini', 'entropy' }	'gini'
Deep Neural Network(DNN)	배치 크기	{ 32, 64, 80, 100 }	80
	에포크	{ 50, 100 }	100
	옵티마이저	{ 'SGD', 'RMSprop', 'Adagrad', 'Adadelta', 'Adam', 'Adamax', 'Nadam' }	'Adam'

선택된 최적의 하이퍼파라미터 조합을 각 알고리즘에 설정하여 기술토픽

간 융합기회를 예측하는 특성이 서로 다른 4개의 지도학습 모델을 생성한다. 이들은 연도 구간 1에서의 특허데이터로부터 추출한 인풋 피처를 활용하여 연도 구간 2에서 기술융합이 일어날 가능성을 학습한다. 생성된 모델의 성능을 측정하기 위하여 연도 구간 2에서의 인풋 피처를 활용하여 연도 구간 3에서의 기술융합 기회를 예측하고 실제 발생 여부와 비교한다. 각 연도 구간에서의 기술융합 발생 여부는 상당히 불균형(imbalanced)하기 때문에 불균형 데이터셋에 대한 예측모델의 성능을 높일 수 있도록 본 연구에서는 Synthetic Minority Over-sampling Technique(SMOTE)를 활용한다. SMOTE는 데이터 수가 충분하지 않은 불균형 데이터셋의 오버 샘플링을 위한 기법으로 통계분포에 따라 임의의 데이터 샘플을 생성함으로써 모델에 대한 학습이 균형적으로 이루어질 수 있도록 한다.<sup>55)</sup>

〈표 5〉는 각 모델에 대한 성능 측정 결과를 나타낸다. 4가지 모델 중, SVM을 활용한 모델은 상당히 우수한 성능을 보이는 것으로 나타났다. 따라서 본 연구에서 개발된 기술토픽 간 융합기회를 발굴하는 SVM 예측모델은 충분한 합리성을 지닌 것으로 판단된다. 하지만 모든 기술토픽 간 융합기회를 예측해 내는 것보다 더욱 중요한 것은 주요한 융합기회를 빠르게 발굴해 내고 기업이 이를 선점할 수 있도록 지원하는 것이다. 따라서 본 연구의 실제적 활용성을 확인하기 위해서는 본 연구의 모델이 미래 융합 가능성을 높은 확률로 예측한 기술토픽 쌍에 대해 이들의 실현 가능성을 정성적으로 탐색해 볼 수 있어야 한다. 이러한 논의를 다음 장에서 수행하고자 한다.

〈표 5〉 개별 모델의 기술토픽 간 융합기회 예측성능

모델	Accuracy	Precision	Recall	F1 score
LR	0.4109	0.9142	0.3816	0.5385
SVM	0.8752	0.9048	0.9626	0.9328
RF	0.3787	0.9163	0.3411	0.4972
DNN	0.4544	0.9259	0.4283	0.5857

55) Chawla, Nitesh V. et al., "SMOTE: synthetic minority over-sampling technique", *Journal of Artificial Intelligence Research*, Vol.16(2002), pp.321-357.

## V. 논의 사항

본 연구에서 개발한 기술융합기회 발굴모델의 타당성을 살펴보기 위하여 SVM 모델에 연도 구간 3의 특허데이터를 적용함으로써 미래 기술토픽 간 융합기회를 탐색한다. <표 6>은 미래 기술융합이 발생될 것으로 예측된 기술토픽 쌍들 중 가장 확률이 높은 것들에 대한 기술토픽 내용, 주요 키워드 및 기술적 함의를 나타낸다. T213과 T255는 멀티활동 플랫폼 디바이스 및 극자외선(EUV: extreme ultraviolet) 생성 시스템 관련 기술토픽들로 EUV 공정을 적용한 운동 등 외부 활동을 위한 웨어러블 디바이스용 반도체 및 프로세서 설계와 밀접한 관련이 있다. EUV 공정은 반도체 칩 내 작고 미세한 회로소자 수십억 개를 형성하는 노광 단계에서 극자외선 파장을 가진 광원을 활용함으로써 성능 및 전력효율을 향상하는 역할을 한다.<sup>56)</sup> 이에, 삼성전자, TSMC, 인텔 등 선두권 파운드리 업체들은 경쟁적으로 5-3나노 공정 양산에 돌입하고 있으며, 고성능 컴퓨터 및 웨어러블 디바이스 등 첨단 IT 기기 개발을 위해 EUV 공정 도입에 앞장서고 있다. 고사양에 대한 시장 요구가 높아지고 있는 웨어러블 디바이스에서는 프로세서의 높은 성능 및 저전력 구현이 필수적이고, EUV 공정은 이를 실현할 수 있는 기반으로 인식되고 있다. T203은 광간섭 토모그래피(OCT: Optical Coherence Tomography) 시스템 관련 기술토픽으로 T203과 T213의 융합은 OCT 기술을 활용한 원격 외부 활동 모니터링 및 헬스케어를 위한 웨어러블 디바이스와 밀접한 관련이 있다. OCT 기술은 광을 활용함으로써 비접촉 방식으로 사물의 단층 영상을 획득할 수 있도록 하는 3차원 영상화 기술로서, 안과 진료 분야에서부터 점차 확대되어 제조업 분야로 응용 영역이 확대되고 있다.<sup>57)</sup> 웨어러블 디바이스를 이용한 원격 의료정보 모니터링은 원격 헬스케어 실행을 위한 핵심 요소로

56) 유종우 외 2인, “반도체 파운드리 EUV 있는 가치 상승”, 한국투자증권, 2020, 27-44면.

57) 김형태 외 6인, “적층구조 내부 비파괴 검사용 광간섭 토모그래피 기술 개발에 관한 연구”, 한국생산기술연구원, 2017, 7-15면.

서, OCT 기술과의 융합은 모빌리티 기반의 디지털 헬스케어 서비스를 가능하게 하는 기반이 된다. 특히 2022년 현시점에서, 미래 성장동력 산업의 한 축으로서 디지털 헬스케어 서비스 산업을 육성하기 위한 노력이 정부 차원에서 매우 활발하게 이루어지고 있다는 것을 고려할 때, 본 융합기회가 지니는 잠재적 가능성은 상당히 크다고 판단된다. T245와 T280은 디바이스 에너지 공급 및 웨어러블 디스플레이 관련 기술토픽들로 소형화, 경량화 및 저전력화에 대한 수요가 높은 웨어러블 디스플레이에 전력을 공급하기 위한 웨어러블 마이크로그리드와 밀접한 관련이 있다. 마이크로그리드는 소규모 스마트그리드로서 작은 단위의 전력 공동체를 통해 전기수요를 충당할 수 있도록 구축한 전력 공급 시스템으로 중앙집중형 전력 공급방식에서 벗어나 분산 에너지를 활용하는 자급자족이 가능한 전력체계를 의미한다.<sup>58)</sup> 신체의 활동으로부터 에너지를 얻는 웨어러블 발전기는 웨어러블 디스플레이에 대한 전력 공급의 문제를 해결할 수 있는 하나의 방안으로서 주목받고 있다. 대표적인 예로, Yin et al.(2021)은 사람의 땀으로부터 에너지를 얻는 바이오연료전지와 운동에너지를 전기로 변환하는 마찰 발전기를 개발함으로써 안정적이고 독립적인 웨어러블 에너지 마이크로그리드 환경을 구축하기 위한 기반을 마련하였다.<sup>59)</sup> 이처럼 본 연구를 통해 도출된 웨어러블 관련 미래 기술융합기회들은 다양한 관점에서 이미 시도되고 있거나 가시적 성과가 창출되고 있는 것들이기 때문에 본 연구는 충분한 수준의 합리성 및 타당성을 지니고 있다고 판단된다.

58) 민현동, “스마트그리드/마이크로그리드”, 한국IR협의회, 2019, 2-8면.

59) Yin, Lu et al., “A self-sustainable wearable multi-modular E-textile bioenergy microgrid system”, *Nature Communications*, Vol.12 No.1(2021), pp.1-12.

〈표 6〉 본 연구의 모델을 통해 도출된 기술융합기회 도출 사례

no.	기술토픽	내용(주요 키워드)	기술적 함의
1	T213	멀티활동 플랫폼 디바이스 (health data, remote monitor, exercise, activity)	EUV 공정을 적용한 운동 등 외부 활동을 위한 웨어러블 디바이스용 반도체 및 프로세서 설계
	T255	극자외선 생성 시스템 (extreme ultraviolet, laser beam, optical amplification, light generation)	
2	T203	광간섭 토모그래피 시스템 (tomography, xray image, optical coherence, radiation detector)	OCT 기술을 활용한 원격 외부 활동 모니터링 및 헬스케어용 웨어러블 디바이스
	T213	멀티활동 플랫폼 디바이스 (health data, remote monitor, exercise, activity)	
3	T245	디바이스 에너지 공급 (energy transfer, wireless energy, electromagnetic energy, energy storage)	소형화, 경량화 및 저전력화에 대한 수요가 높은 웨어러블 디스플레이에 대해 전력을 공급하기 위한 웨어러블 마이크로그리드
	T280	웨어러블 디스플레이 (display screen, touch screen, visual, graphic interface)	

## VI. 결론

기술융합은 서로 다른 분야의 기술지식을 융합적으로 활용하여 혁신을 선도하고 산업에 새로운 가치를 창출해 내는 기술적 진보를 촉진한다. 따라서 기업이 기술융합을 통해 실행 가능한 잠재적 기술기회를 미리 발굴할 수 있도록 하는 것은 해당 기업이 기술기회를 선점하고 지속가능한 성장을 위한 경쟁우위를 갖추는 데 매우 중요한 역할을 한다. 특허데이터를 활용하여 이러한 기술융합기회를 발굴하기 위한 연구가 활발하게 수행되어 왔다. 하지만 이들은 기술적 의미가 모호한 기술클래스를 활용함으로써 기술융합기회가 지니는 기술적 함의를 구체적으로 도출하지 못할 뿐만 아니라 소수의 링크예측 알고리즘들을 개별적으로만 적용함으로써 이들 각각이 지니는 개별적 장점을 효과적으로 결합하여 활용하지 못하는 한계를 지니고 있다. 따

라서 본 연구에서는 기술클래스에 비해 기술 구체성이 높은 기술토픽 단위에서 융합 네트워크를 구축하고 다양한 링크예측 알고리즘들을 복합적으로 적용하여 기술융합기회를 발굴하는 지도학습 기반의 접근법을 개발하였다. 본 연구는 구체화된 기술토픽 간 융합을 통해 실행 가능한 새로운 기술기회를 선점할 수 있도록 함으로써 국내 기업들의 취약한 R&D 기획역량을 증진하고 지속가능한 성장을 위한 경쟁우위를 확보하도록 지원할 수 있을 것으로 기대된다.

이러한 기여점에도 불구하고 본 연구에는 여러 가지 한계점이 있어 다양한 추후 연구가 진행될 필요가 있다. 우선, 본 연구는 오직 두 기술토픽 간 융합기회만을 탐색하였다. 기술융합은 단지 두 토픽 사이에서뿐만 아니라 여러 기술토픽이 함께 복합적으로 활용되어 발생할 수 있다. 따라서 보다 넓은 범위에서 융합기회를 발굴하기 위한 방안을 제시하기 위한 연구가 수행되어야 한다. 각 연도 구간의 기술토픽 간 연계관계를 구축하기 위해 본 연구는 각 토픽과 관련된 단어 유사성만을 고려하는 단순한 방안을 활용하였다. 서로 다른 연도 구간에 존재하는 기술토픽들의 연계관계를 정립하는 것은 기술융합기회 발굴모델의 품질 및 성능에 직접적 영향을 미친다. 따라서 모델의 실제적 활용성을 증진하기 위해서는 기술토픽 간 연계성을 형식적 및 의미적 관점에서 정밀하게 측정할 수 있는 방안이 적용되어야 한다. 마지막으로 본 연구는 기술토픽 간 링크예측 결과만을 활용하여 융합기회 발굴을 위한 모델 학습을 진행하였다. 이는 다양한 링크예측 결과를 효과적으로 결합하여 융합기회를 예측해 내는 장점이 있지만, 기술토픽 간 융합에 영향을 미치는 요인을 링크 생성 가능성이라는 매우 한정적인 관점에서만 고려한다는 한계 또한 지니고 있다. 기술토픽과 관련된 속성들을 폭넓게 도출하고 융합기회에 대한 영향 수준을 측정한 후, 이들을 포괄적으로 활용하기 위한 후속연구가 수행되어야 한다.

## 참고문헌

## 〈학술지(국내 및 동양)〉

- 김경수, “국제특허분류를 활용한 기술융합 연결망 구조 및 패턴 분석—해양수산 산업 육성 분야의 특허를 중심으로”, 『과학기술정책』, 제3권 제2호(2020).
- 김은정·최희진, “토피코모델링과 네트워크분석을 활용한 헬스케어 분야의 핵심기술과 기술융합 분석 연구—특허정보를 중심으로”, 『한국정보통신학회논문지』, 제26권 제5호(2022).
- 박수현 외 3인, “5개국 바이오헬스 산업의 기술융합과 트렌드 분석—특허 동시분류분석과 텍스트마이닝을 활용하여”, 『한국융합학회논문지』, 제12권 제4호(2021).
- 서원철, “기술태마 분석을 통한 기술기회발굴 연구—3D 프린팅 기술 사례를 중심으로”, 『지식재산연구』, 제16권 제2호(2021).
- 오승현 외 2인, “특허의 토피코모델링을 활용한 증강현실 기술 모니터링”, 『대한산업공학회지』, 제43권 제3호(2017).
- 음수민 외 4인, “LDA 기반의 토피코모델링을 이용한 철도차량용 무선급전시스템 연구 동향 분석”, 『대한산업공학회지』, 제45권 제4호(2019).
- 이금실·이인주, “LDA 토피코 모델링 기법을 적용한 관광분야의 가상현실(Virtual Reality) 연구 동향분석”, 『관광레저연구』, 제33권 제10호(2021).
- 이민정 외 2인, “국제특허분류(IPC) 공동출현 네트워크분석을 통해 살펴본 기업의 지식융합 특성이 혁신성과에 미치는 영향: 전자통신산업 중소기업을 중심으로”, 『지식재산연구』, 제13권 제1호(2018).
- 이현민 외 2인, “융합기술 개발전략 기획을 위한 특허 인용 네트워크 기반의 분석 방법론—스마트공장 ICT 기술을 중심으로”, 『한국산학기술학회 논문지』, 제19권 제1호(2018).
- 장동호·이충목, “GAN을 이용한 비지도학습 기반의 새로운 링크 예측 알고리즘”, 『대한산업공학회지』, 제46권 제3호(2020).
- 전상규, “특허 네트워크 분석을 통한 기술융합 및 융합기술의 확산 연구—디지털 데이터 처리 기술 중심으로”, 『지식재산연구』, 제16권 제4호(2021).
- 정으뜸 외 2인, “기술적 과급효과 도출을 통한 기술융합 분석 연구—나노바이오 융합 기술 사례를 중심으로”, 『지식재산연구』, 제10권 제4호(2015).
- 정재민 외 2인, “비즈니스 기회 발굴을 위한 문제—해결방법 기반의 특허분석 방법”, 『지식재산연구』, 제15권 제2호(2020).
- 조용래·김의석, “특허 네트워크와 전략지표 분석을 통한 기업 기술융합 전략 연구”,

『지식재산연구』, 제9권 제4호(2014).

〈학술지(서양)〉

- Aaldering, Lukas Jan et al., “Competition or collaboration?—analysis of technological knowledge ecosystem within the field of alternative powertrain systems: a patent-based approach”, *Journal of Cleaner production*, Vol.212 (2019).
- Adamic, Lada A. & Adar, Eytan, “Friends and neighbors on the web”, *Social Networks*, Vol.25 No.3(2003).
- Barabási, Albert-László & Albert, Réka, “Emergence of scaling in random networks”, *Science*, Vol.286 No.5439(1999).
- Blei, David et al., “Latent Dirichlet allocation”, *Journal of Machine Learning Research*, Vol.3(2003).
- Brin, Sergey & Page, Lawrence, “The anatomy of a large-scale hypertextual web search engine”, *Computer Networks and ISDN Systems*, Vol.30 No.1-7(1998).
- Cai, Lei et al., “Line graph neural networks for link prediction”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.44 No.9(2022).
- Chawla, Nitesh V. et al., “SMOTE: synthetic minority over-sampling technique”, *Journal of Artificial Intelligence Research*, Vol.16(2002).
- Chebotarev, Pavel & Shamis, Elena, “The matrix-forest theorem and measuring relations in small social groups.” *Automation and Remote Control*, Vol.58 No.9(1997).
- Choi, Jae Young et al., “A study on diffusion pattern of technology convergence: Patent analysis for Korea”, *Sustainability*, Vol.7 No.9(2015).
- Choi, Sungchul et al., “A supervised learning-based approach to anticipating potential technology convergence”, *IEEE Access*, Vol.10(2022).
- Feng, Sida et al., “The technology convergence of electric vehicles: Exploring promising and potential technology convergence relationships and topics”, *Journal of Cleaner Production*, Vol.260(2020).
- Hussain, Adnan et al., “Analyzing technological trends of smart factory using topic modeling”, *Asian Journal of Innovation and Policy*, Vol.10 No.3(2021).
- Jaccard, Paul, “The distribution of the flora in the alpine zone”, *New Phytologist*, Vol.11 No.2(1912).

- Jelodar, Hamed et al., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey", *Multimedia Tools and Applications*, Vol.78 No.11(2019).
- Katz, Leo, "A new status index derived from sociometric analysis", *Psychometrika*, Vol.18 No.1(1953).
- Kim, Jeeun & Lee, Sungjoo, "Forecasting and identifying multi-technology convergence based on patent data: The case of IT and BT industries in 2020", *Scientometrics*, Vol.111 No.1(2017).
- Kim, Tae San & Sohn, So Young, "Machine-learning-based deep semantic analysis approach for forecasting new technology convergence", *Technological Forecasting and Social Change*, Vol.157(2020).
- Klein, Douglas J., & Randić, Milan, "Resistance distance", *Journal of Mathematical Chemistry*, Vol.12 No.1(1993).
- Ko, Namuk et al., "Analyzing interdisciplinarity of technology fusion using knowledge flows of patents", *Expert Systems with Applications*, Vol.41 No.4 (2014).
- Kose, Toshihiro & Sakata, Ichiro, "Identifying technology convergence in the field of robotics research", *Technological Forecasting and Social Change*, Vol.146 (2019).
- Kumar, Ajay et al., "Link prediction techniques, applications, and performance: A survey", *Physica A: Statistical Mechanics and its Applications*, Vol.553(2020).
- Lee, Changyong et al., "Novelty-focused patent mapping for technology opportunity analysis", *Technological Forecasting and Social Change*, Vol.90 (2015).
- Lee, Changyong et al., "Anticipating multi-technology convergence: a machine learning approach using patent information", *Scientometrics*, Vol.126 No.3 (2021).
- Leicht, Elizabeth A. et al., "Vertex similarity in networks", *Physical Review E*, Vol.73 No.2(2006).
- Liu, Yingjie et al., "Link prediction algorithm based on the initial information contribution of nodes", *Information Sciences*, Vol.608(2022).
- Lü, Linyuan & Zhou, Tao, "Link prediction in complex networks: A survey", *Physica A: Statistical Mechanics and its Applications*, Vol.390 No.6(2011).
- Ma, Jing et al., "Identifying translational indicators and technology opportunities for nanomedical research using tech mining: The case of gold nanostructures",

- Technological Forecasting and Social Change*, Vol.146(2019).
- Ma, Tingting et al., “Combining topic modeling and SAO semantic analysis to identify technological opportunities of emerging technologies”, *Technological Forecasting and Social Change*, Vol.173(2021).
- Newman, Mark EJ., “Clustering and preferential attachment in growing networks”, *Physical Review E*, Vol.64 No.2(2001).
- O’callaghan, Derek et al., “An analysis of the coherence of descriptors in topic modeling”, *Expert Systems with Applications*, Vol.42 No.13(2015).
- Park, Hyochul et al., “Regional innovation capability from a technology-oriented perspective: An analysis at industry level”, *Computers in Industry*, Vol.129 (2021).
- Ravasz, Erzsébet et al., “Hierarchical organization of modularity in metabolic networks”, *Science*, Vol.297 No.5586(2002).
- Seo, Wonchul. “A patent-based approach to identifying potential technology opportunities realizable from a firm’s internal capabilities”, *Computers & Industrial Engineering*, Vol.171(2022).
- Song, Chie Hoon et al., “Anticipation of converging technology areas — A refined approach for the identification of attractive fields of innovation”, *Technological Forecasting and Social Change*, Vol.116(2017).
- Sørensen, Thorvald A., “A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons”, *Biologiske Skrifter*, Vol.5 (1948).
- Tang, Ying et al., “A study on dynamic patterns of technology convergence with IPC co-occurrence-based analysis: The case of 3D printing”, *Sustainability*, Vol.12 No.7(2020).
- Xu, Xiujian et al., “MM-UrbansFAC: Urban functional area classification model based on multimodal machine learning”, *IEEE Transactions on Intelligent Transportation Systems*, Vol.23 No.7(2021).
- Yin, Lu et al., “A self-sustainable wearable multi-modular E-textile bioenergy microgrid system”, *Nature Communications*, Vol.12 No.1(2021).
- Zhou, Tao et al., “Predicting missing links via local information”, *The European Physical Journal B*, Vol.71 No.4(2009).

## Discovering Technology Convergence Opportunities among Technology Topics —The Case of Wearable Technology—

Seo, Wonchul

Technology convergence promotes technological advances that can create new values in industries by merging technological knowledge from different industrial fields. Anticipating potential technological opportunities that enable new innovations through technology convergence can facilitate seizing technology opportunities and gaining competitive advantage. Although extensive research has been conducted to identify technological convergence opportunities using patent data, these studies mainly rely on technology classes whose technical meaning is quite ambiguous, thereby not being able to specifically derive the technical implications of identified opportunities. Therefore, this study proposes a supervised learning model that builds a convergence network of technology topics with high technological specificity and anticipates potential technology convergence opportunities by applying various link prediction algorithms. In addition, to explore the applicability of the proposed model, we discuss the technological implications of the predicted technology topic pairs that show high probability of future convergence. We expect that this study will help to find new technological prospects that can be exploited through technology convergence. Moreover, this study can assist relevant firms in

strengthening their underdeveloped R&D planning capabilities and securing competitive advantage for sustainable growth.

Keyword

Technology convergence, Supervised learning, Topic modeling, Link prediction, Wearable