

## 특허 정보를 이용한 유망기술 발굴 프로세스 연구 —요인분석과 텍스트마이닝 기법의 결합 접근—

송경태\* · 봉강호\*\* · 박재민\*\*\*

- |                      |                              |
|----------------------|------------------------------|
| I. 서론                | 1. 데이터 수집 및 선행 유망기술 선정       |
| II. 이론적 배경 및 선행연구 고찰 | 2. 탐색 대상 그룹 선정               |
| 1. 유망기술의 정의 및 특성     | 3. 선행 유망기술의 공통적 특징 추출        |
| 2. 기술예측              | 4. 탐색대상 그룹과 선행 유망기술의 상관관계 분석 |
| 3. 특허 데이터를 활용한 기술예측  | 5. 텍스트 전처리                   |
| III. 연구자료 및 프레임워크    | 6. 기술군 형성                    |
| 1. 연구자료              | 7. 텍스트마이닝 및 미래 유망기술 선정       |
| 2. 연구 프레임워크          |                              |
| IV. 실험 결과            | V. 결론                        |

\* 한국특허전략개발원 실장.

\*\* 건국대학교 기술경영학과 강사.

\*\*\* 건국대학교 기술경영학과 교수, 교신저자.

초록

본 연구에서는 통계적 정보와 서지적 데이터를 모두 활용하여 거시적 관점에서 미래 유망기술을 도출하는 방법을 제안한다. 기존 연구들은 대부분 특허 건수와 같은 단일의 통계적 데이터 또는 서지적 정보만을 활용해 유망기술을 도출하고 있다는 점에서 정밀성·예측력이 미흡하거나 상당한 비용이 수반된다는 한계가 있다. 본 연구에서 제안하는 방법은 우선 국제특허분류(IPC)에서 Sub-group으로 세분화되지 않은 Main-group을 데이터 수집 대상으로 설정하고, 통계적 데이터를 활용하는 주성분 요인분석 및 상관관계 분석을 통해 과거에 유망했던 기술분야들과 유사한 특징을 보이는 유망기술 분야 후보군을 도출한 후 특허의 서지적 데이터를 활용하는 텍스트마이닝 분석을 통해 최종적으로 유망기술을 확정한다는 점에서 선행연구의 방법과 차이가 있다. 본 연구에서는 IPC 기준 8개의 섹션 중 H섹션(전기분야)을 대상으로 분석하였다. 본 연구는 복수의 분석기법을 결합함으로써 유망기술 발굴의 효율성·객관성을 높일 수 있는 새로운 방법론을 제안한다는 점에서 학술적·실무적 의의가 있다고 하겠다.

주제어

유망기술, 기술예측, 특허분석, 국제특허분류, 주성분요인분석, 텍스트마이닝

## I. 서론

미래 성장동력 확보와 경쟁력 강화를 위한 유망기술 발굴의 노력은 개별 기업뿐만 아니라 국가차원에서도 지속적인 관심사이다. 새로운 성장동력 산업의 확보는 경제 성장의 한계를 극복하고 새로운 도약의 계기를 마련하기 위해 필수불가결하기 때문이다.

이러한 관점에서, 세계 주요국들은 미래 성장동력 확보를 위한 노력으로써 다양한 방법론을 적용한 유망기술 예측을 시도하고 있다. 전통적인 미래 유망기술 예측방법으로는 델파이 기법(Delphi), 계층적 분석방법(AHP), 시나리오(Scenario) 기법, 전문가 패널, 추세외삽법(Trend Extrapolation) 등이 있지만, 이러한 방법들은 관련 기술 전문가들의 주관적 견해가 개입된다는 한계점이 있다.<sup>1)</sup> 이로 인해 최근에는 서지학적 데이터인 논문과 특허정보를 활용한 데이터 기반 유망기술 예측방법이 주로 활용되고 있다. 특히 특허는 2020년 기준 전 세계적으로 4억 5천만 건이나 되는 방대한 양이 공개되어 있으며, 세상에 존재하는 최신 기술정보의 90% 이상이 담겨 있다는 점에서 유망기술 예측에 유용한 데이터로 간주되고 있다.<sup>2)</sup>

한편, 기존의 특허 데이터를 활용한 연구들은 통계적 또는 서지적 데이터 중 특정한 단일 정보를 이용하여 유망기술을 도출하는 방식이라는 점에서

1) 주재욱 외 7인, “데이터 기반 디지털 경제 미래예측 방법론 연구”, 경제·인문사회연구회, 2016, 36-53면.

2) Trappey, Charles V. et al., “Using patent data for technology forecasting: China RFID patent analysis”, *Advanced Engineering Informatics*, Vol.25 No.1(2011), pp. 53-64; Lee, Changyong et al., “Early identification of emerging technologies: A machine learning approach using multiple patent indicators”, *Technological Forecasting and Social Change*, Vol.127(2018), pp. 291-303; Yu, Xiang & Zhang, Ben, “Obtaining advantages from technology revolution: A patent roadmap for competition analysis and strategy planning”, *Technological Forecasting and Social Change*, Vol.145(2019), pp. 273-283; Choi, Youngjae et al., “Identifying emerging technologies to envision a future innovation ecosystem: A machine learning approach to patent data”, *Scientometrics*, Vol.126(2021), pp. 5431-5476.

한계가 있다. 예컨대 특허 수(특허 증가율)를 이용하는 시도<sup>3)</sup>는 일정 기간동안 누적된 데이터가 요구되나, 유망할 것으로 예측되는 기술의 특허가 실상 이미 상당한 규모로 출원되어 있어서, 곧 기술개발이 정체되는 성숙단계에 접어들 수 있거나 극히 단기적 유망성만 가질 가능성이 있다. 또한 특허 인용 정보를 활용한 네트워크 분석과 같은 서지적 분석기법을 활용하는 방법<sup>4)</sup>은 모든 기술분야를 대상으로 하는 경우 데이터 수집 및 분석에 천문학적 비용이 발생하므로, 연구목적에 따라 특정한 세부 기술분야를 지정하여 협소한 범위에서만 접근하는 경향이 있다.

본 연구에서는 통계적 정보와 서지적 데이터를 모두 활용하여 거시적 관점에서 미래 유망기술을 도출하는 방법을 제안하고자 한다. 통계적 분석기법을 적용하여 과거에 유망했던 기술분야들의 공통적 특징을 도출하고, 이들 기술분야들과 유사한 특징을 보이는 유망기술분야 후보군을 확정한다. 이후 서지적 분석기법을 적용하여 해당 기술분야에서 최종 유망기술을 발굴한다.

본 연구는 복수의 분석기법을 결합한 새로운 유망기술 예측방법을 제안하여 기존 기술예측 연구에서 제안된 방법론을 보완하고, 연구의 확장 가능성을 제공함으로써 연구분야 발전에 기여한다는 점에서 학문적 의의를 가진다. 실무적으로는 보다 효과적으로 유망기술을 예측하는 방법을 제공하고, 나아가 우리나라 과학기술 및 혁신 관련 정책수립을 지원하는 기초자료 창출에 기여한다는 점에서 의의가 있다고 하겠다.

---

3) 강희종 외 2인, “특허분석을 통한 유망융합기술의 예측”, 『기술혁신연구』, 제14권 제3호(2006), 93-116면; 김종찬 외 4인, “특허 키워드 시계열 분석을 통한 부상 기술 예측”, 『정보처리학회논문지. 소프트웨어 및 데이터 공학』, 제3권 제9호(2014), 355-360면.

4) Érdi, Péter et al., “Prediction of emerging technologies based on analysis of the US patent citation network”, *Scientometrics*, Vol.95(2013), pp.225-242; 주성현, “무인항공기 기술진화 탐색 및 유망기술 발굴 연구”, 『항공우주시스템공학회지』, 제13권 제6호(2019), 80-89면.

## II. 이론적 배경 및 선행연구 고찰

### 1. 유망기술의 정의 및 특성

유망기술은 용어적으로 미래기술(future technology), 잠재성이 있는 기술(promising technology), 신흥기술(emerging technology), 신기술(new technology), 돌파기술(breakthrough technology), 핵심기술(key technology) 등으로 명명되고 있으나, 아직까지 합의된 개념적 정의는 없다. 학자들은 유망기술을 연구의 목적 및 대상에 따라 다양하게 정의하였으며, 주요 내용을 정리하면 <표 1>과 같다.

<표 1> 선행연구의 유망기술 정의

| 연구자(연도)              | 정의   |
|----------------------|--|
| 김방룡 · 황성현(2009)      | 10년 이내에 시장에서 우위를 점할 것으로 기대되는 기술  |
| Cozzens et al.(2010) | 높은 잠재력이 있지만 아직 그 가치가 입증되지 않았거나 합의에 도달하지 않은 기술                                      |
| 감주식 외(2013)          | 최근 R&D 관심이 상승하고 있어 향후 5년 이내에 사회·경제적 파급효과가 큰 기술(분야)                                 |
| Small et al.(2014)   | 참신성 및 성장가능성에 대한 보편적 합의가 있는 기술  |
| Rotolo et al.(2015)  | 시간 경과에 따라 일정 수준의 일관성이 지속되고 사회·경제적으로 지대한 영향을 미칠 수 있으면서, 상대적으로 빠르게 성장하는 급진적이고 참신한 기술 |
| Jang et al.(2021)    | 사회적·기술적 발전에 상당한 영향을 미치고 미래에 빠르게 성장할 잠재력을 가진 기술                                     |

한편, Rotolo et al.(2015)<sup>5)</sup>은 유망기술에 초점을 둔 선행연구에서 제시한 유망기술의 개념정의를 대상으로 텍스트 분석을 실시하여, 그 결과를 토대로 유망기술의 주요 특성을 5가지로 정리하였다. 첫째, 유망기술은 급진적

5) Rotolo, Daniele et al., "What is an emerging technology?", *Research Policy*, Vol.44 No.10(2015), pp. 1827-1843.

혁신으로부터 파생된 불연속적 혁신(discontinuous innovation)의 형태를 보인다는 점에서 ‘급진적 참신성(radical novelty)’이라는 특성을 가진다고 하였다. 둘째, 유망기술은 동일 영역의 타 기술과 비교하여 ‘상대적으로 빠르게 성장(relatively fast growth)’하는 특성을 보인다고 하였다. 셋째, ‘일관성(coherence)’을 하나의 특성으로 제시하였는데, 실현 과정에 있는 유망기술이 원천기술(technological parents)로부터 일정 기간 이상 독립되어 있어야 별도의 정체성을 가진 기술로 간주될 수 있다는 점을 강조하였다. 넷째, 유망기술은 ‘두드러진 영향(prominent impact)’을 미치는 것으로, 하나 또는 그 이상 분야에 중대한 영향을 미치는 동시에 전체 경제·사회에 걸쳐 광범위하게 미치는 잠재적 영향도 포함한다고 하였다. 다섯째, 유망기술은 새로운 기술적 출현(technological emergence)이 비선형적·지속적으로 발생할 잠재력이 있고 그 최종 형태를 예측하기 어렵다는 의미에서 ‘불확실성 및 모호성(uncertainty and ambiguity)’이란 특성을 가진다고 하였다.

## 2. 기술예측

기술예측은 분석기법 측면에서 보면 정성적(qualitative) 및 정량적(quantitative) 방법으로 구분할 수 있다. 정성적 방법은 전문가들의 직관, 지식 및 판단력을 활용하는 기법으로서 델파이법, 시나리오법, 브레인스토밍, 전문가 패널, 설문조사 등이 일반적이다. 정량적 방법은 기술이 시계열적 일관성 또는 특정 패턴에 따라 규칙적으로 변화한다는 가정하에 기술 발전의 흐름을 예측하는 기법으로서 성장곡선, 추세외삽법(trend extrapolation) 등이 주로 활용되고 있다.

정성적 방법 중에서 가장 일반적으로 미래 예측에 활용되는 델파이법은 전문가 집단의 집중적, 반복적 설문조사를 통해 신뢰성이 높은 합의 결과를 도출하는 과정을 거치는 접근방법이다. Keller & von der Gracht(2014)<sup>6)</sup>의

6) Keller, Jonas & von der Gracht, Heiko A., "The influence of information and communication technology(ICT) on future foresight processes — Results from a

연구에서는 177명의 전문가를 활용하여 ICT(Information and Communication Technology) 분야의 미래 기술을 예측하는 연구를 수행하였다. 시나리오 기법은 과거와 현재의 흐름을 토대로 미래를 예측하는 기법으로, 미래에 발생 가능한 상황들을 예측하는 전문적 방법론이다.<sup>7)</sup> Wang et al.(2018)<sup>8)</sup>은 시나리오 기법을 활용하여 중국의 나노 발전기(nanogenerators) 기반 산업의 기술, 산업 동력, 생산 및 시장·애플리케이션 관점에서 미래 발전 로드맵(roadmap)을 제안하였다.

정량적 방법으로서 성장곡선 기술예측은 어느 특정한 기술성고가 유기체의 성장 과정과 유사할 것이라는 데 착안하여 기술변화 패턴을 예측하는 방법이다. 대표적으로 1980년대 매켄지 컨설팅의 리처드 포스터(Richard N. Foster)가 한계돌파의 경영전략(Innovation: The Attacker's Advantage)에서 소개한 S-curve이다. 신기술의 초기에는 기술발전 속도(생산성, 집적도 등)가 더디게 나타나고, 초기 투자기 및 시험기가 마무리되면 기술발전이 기하급수적으로 상승 곡선을 타게 된다. 이때 해당 기술에 대한 연구개발 투자가 최적기가 된다. 이후 성숙기에 접어들면 성장곡선은 완만해지고 혁신투자를 계속해도 기술의 한계에 도달하게 되고 기존 기술보다 효율적인 대체기술 개발에 대한 필요성이 증가한다. Priestley et al.(2020)<sup>9)</sup>은 1990년부터 2013년까지 웹 기술 관련 미국특허 20,493건으로 성장곡선을 추적하여 웹 기술이 S-curve를 따르는 것을 밝혀내고, 지속적 기술 성장을 위한 시사점을 제시하였다. 추세외삽법의 경우, 과거의 추세가 장래에도 그대로 지속될 것이라는 전제 아래 과거의 추세선을 연장하여 미래 일정 시점에서의 상황을 예

---

Delphi survey”, *Technological Forecasting and Social Change*, Vol.85(2014), pp. 81-92.

7) Schoemaker, Paul J. H., “Scenario planning: A tool for strategic thinking”, *MIT Sloan Management Review*, Vol.36 No.2(1995), p. 28.

8) Wang, Baicun et al., “Emerging nanogenerator technology in China: A review and forecast using integrating bibliometrics, patent analysis and technology roadmapping methods”, *Nano Energy*, Vol.46(2018), pp. 322-330.

9) Priestley, Maria et al., “Innovation on the web: the end of the S-curve?” *Internet Histories*, Vol.4 No.4(2020), pp. 390-412.

측하는 미래 예측 기법이다. 예컨대 Bodger & Tay(1986)<sup>10)</sup>의 연구에서는 뉴질랜드의 전력 소비 예측을 위해 추세외삽법을 활용한 바 있다.

### 3. 특허 데이터를 활용한 기술예측

특허 데이터를 활용한 기술의 분석과 계획 수립 및 기술예측에 대한 연구는 점차 활발해지고 있다. 이러한 연구들은 크게 특허추세분석(patent trend analysis), 특허 네트워크 분석(patent network analysis) 및 텍스트마이닝 분석(text mining analysis) 등의 기법을 활용한 연구로 구분할 수 있다.

먼저, 특허 추세분석은 특허 증가율이나 키워드의 변화 등을 토대로 미래의 기술변화를 예측하는 방식이다. 대표적으로 한국특허전략개발원은 특허출원 증가율, 점유율, 시장성, 발전단계 등을 고려하여 부상기술 및 유망기술을 선정하고 있다. 강희종 외(2006)의 연구에서는 특허 건수 및 특허 증가율에 의해 결정되는 유망지수(PI: promise index)와 해당 기술군이 다른 분류체계하에서 서로 다른 영역으로 분류되는 정도로서 IPC 분류체계상의 총 분류수를 근거로 하는 융합지수를 활용하여 유망기술군을 도출하는 방법을 제안하였다.

둘째, 특허 네트워크 분석은 특허의 발명자, 출원인, 패밀리 정보, 인용 정보 등을 활용하여 각 특허 간의 연결 관계를 통해 최근 유망기술을 도출하는 분석기법이다. Erdi et al.(2013)은 특허 네트워크 분석 및 클러스터링 분석을 통해 유망기술 발굴을 위한 기술의 성장, 수축, 통합, 분할, 발생 및 소멸의 진화를 예측하였다. 주성현(2019)의 연구에서는 네트워크 분석을 통해 무인항공기산업 분야의 기술진화 및 확산패턴을 탐색하고 유망기술을 도출하였다.

셋째, 텍스트마이닝은 텍스트 데이터에서 자연어처리(natural language processing)에 기반하여 키워드를 추출하는 방법으로, 유망기술 발굴을 위한

---

10) Bodger, P. S. & H.S, Tay, "Trend extrapolation in long-term forecasting: An investigation using New Zealand electricity consumption data", *Technological Forecasting and Social Change*, Vol.30 No.2(1986), pp. 167-188.

연구에서도 활발하게 활용되고 있다. Trappey et al.(2011)은 특허 문서에서 핵심 구절(key phrase)을 추출하여 상관도를 측정하고, 특허 콘텐츠 클러스터링 방법을 통해 중국 RFID 분야의 유망기술을 도출하였다. 김종찬 외(2014)는 텍스트마이닝을 통해 핵심 키워드를 추출하고, 이들 키워드의 연도별 출현빈도로 시계열 분석을 실시하는 방법으로 미국 탄소 복합소재 분야의 유망기술을 도출하는 연구를 수행하였다. 서원철(2021)<sup>11)</sup>은 텍스트 데이터를 분석하는 기법인 토픽모델링과 연관규칙마이닝을 활용해 3D프린팅 분야에서 미래 유망할 것으로 보이는 기술테마를 도출하였다.

유망기술 발굴을 위해 국제특허분류(IPC: International Patent Classification) 코드를 활용한 연구도 있다. 김요한(2016)<sup>12)</sup>의 연구에서는 1차적으로 IPC H 섹션(전기분야)에 위치한 서브그룹(sub-group)이 아닌 전체 메인그룹(main-group)에서 연도별 출원 증가율이 높은 분류를 유망기술 후보군으로 선정하고, 선정된 유망기술 후보군에 대해 특허 동향분석 및 환경분석을 실시한 결과를 토대로 최종 유망기술을 도출하였다.

### Ⅲ. 연구자료 및 프레임워크

#### 1. 연구자료

각 특허에는 고유한 IPC 코드가 부여되어 있다. IPC는 세계 공통적인 특허분류 체계로서 <표 2>와 같이 섹션(section), 클래스(class), 서브클래스(sub-class), 메인그룹(main-group) 및 서브그룹(sub-group)이라는 계층적 구조를 이루고 있으며, 기술분류 코드는 약 7만여 개에 이르고 있다. 통상적으

11) 서원철, “기술테마 분석을 통한 기술기회발굴 연구: 3D 프린팅 기술사례를 중심으로”, 『지식재산연구』, 제16권 제2호(2021), 205-248면.

12) 김요한, “국제특허분류를 이용한 유망기술 도출에 관한 연구”, 홍익대학교대학원, 석사, 2016, 11-13면.

로 기존 IPC 코드로 분류되어 있지 않는 신기술이 포함된 특허가 출원되었을 때는 먼저 메인그룹으로만 분류하게 된다. 이후 이와 유사한 기술의 특허 출원이 활발해져서 별도의 IPC 코드를 부여할 정도의 규모가 되면, 서브그룹으로 세분화되는 방식이 일반적이다. Kang(2012)<sup>13)</sup>은 신규 IPC 분류를 만드는 절차를 다룬 연구에서 특허의 양에 따른 메인그룹의 세분화와 함께 기술을 명확하게 구분하기 어렵고 분류의 어려움이 있는 경우에 메인그룹을 처리하는 과정을 설명하고 있다. 이와 같은 점에 착안하여, 본 연구에서는 서브그룹으로 세분화되지 않은 메인그룹에는 특허출원이 많지 않은 초기의 신기술이 포함되어 있을 가능성이 높다는 전제하에, 메인그룹의 특허를 대상으로 미래 유망기술을 발굴해 보고자 한다. 결과적으로, 본 연구에서는 IPC 기준 8개의 섹션 중에서 첨단 기술이 많이 포함된 H섹션(전기분야)을 분석 대상으로 선정하였다.

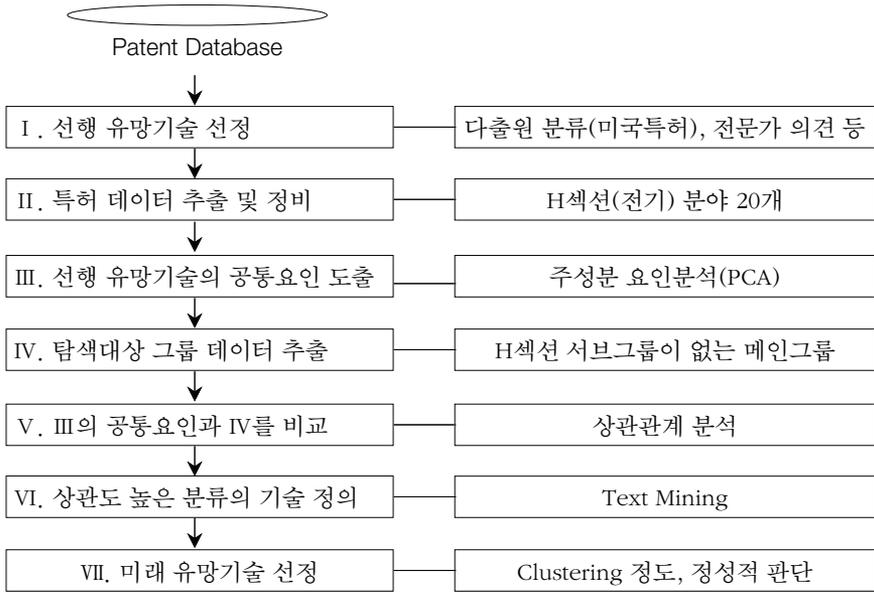
〈표 2〉 국제특허분류 체계(2019)

| 구분 | Section | Class | Sub-Class | Main-group | Sub-group |        |
|----|---------|-------|-----------|------------|-----------|--------|
| 개수 | 8       | 129   | 639       | 7,483      | 67,020    |        |
| 예시 | 표기      | F     | 16        | K          | 1/00      | 1/02   |
|    | 내용      | 기계공학  | 기계요소      | 밸브         | 리프트 밸브    | 나사 스핀들 |

## 2. 연구 프레임워크

본 연구에서는 선행 유망기술 선정, 선행 유망기술의 공통적 특징 추출, 유망기술 후보군 도출, 최종 유망기술 선정 등 절차로 유망기술을 도출한다. 전반적 프레임워크는 〈그림 1〉과 같다.

13) Kang, Hyo Y., "Science inside law: The making of a new patent class in the international patent classification", *Science in Context*, Vol.25 No.4(2012), pp. 551-594.



〈그림 1〉 연구 프레임워크

먼저, H섹션 중에서 과거 유망했던 ‘선행 유망기술’을 선정한다. 메인그룹 또는 서브그룹 중에서 미국 특허청(USPTO)에 출원된 특허<sup>14)</sup>가 1만 건 이상인 특허분류를 1차로 추출한다. 그리고 일반적으로 널리 알려진 기술과 한국특허전략개발원의 특허분류 전문가 및 기술전문가 3인을 대상으로 델파이 조사를 통해 공통적으로 유망기술로 추천된 기술분류들을 최종 선행 유망기술로 확정한다.

둘째, 비교분석 대상으로서 선행 유망기술의 특허출원 건수(number of patents), 피인용 수(number of citations), 발명자 수(number of inventors) 및 청구항 수(number of claims) 등 4개 항목의 공통적 추세를 생성하기 위해 주성분 요인분석(PCA: principal component analysis)을 실시한다. 개별 특허 서지 데이터로부터 알 수 있는 정보는 출원, 등록, 청구항, 출원인, 발명자, 심사청

14) 미국은 세계 최대의 시장으로서 중요한 특허들이 집결되는 국가라는 점에서 미국 특허 데이터를 활용한 경우 그만큼 기술의 트렌드 변화를 잘 파악할 수 있다는 장점이 있다.

구, 인용도, 권리유지·소멸, 패밀리 정보 등이 있다. 특허 데이터를 이용한 기술혁신 특성에 관한 연구는 이와 같은 정보들을 다양하게 단일·복합 지표로 응용하여, 활동성, 권리성, 기술성, 파급성 등의 관점에서 특성을 분석한다.<sup>15)</sup> 본 연구에서는 산출에 시간 소요가 많은 복합지표는 배제하고, 이들에 가지 특성별로 대표적인 단일 지표를 활용하였다. 첫째, 활동성에 대해서는 가장 대표적이고 광범위하게 활용되고 있는 특허출원 건수를 이용하였으며, 이는 연구개발 활동이 얼마나 활발하게 이루어지고 있는지 분석하기에 적합하다. 둘째, 권리성 부분은 청구항 수를 활용하였다. 특허 청구항은 특허받은 발명에서 보호받고자 하는 사항을 모두 기재한 것으로, 복수의 발명을 하는 경우라도 개개의 발명마다 청구항으로 구분하여 기재하기 때문에 청구항 수는 특허출원에 포함되어 있는 발명의 수로 볼 수 있다. 또한 청구항 수는 특허 보호의 법적 범위와 복잡성을 나타내고, 획기적인 돌파형 기술 특허는 더 높은 청구항 수를 포함하는 것이 일반적이다.<sup>16)</sup> 셋째, 기술성 측면은 피인용 수를 활용하였다. 이는 학술저널의 영향력 지수(impact factor)와 유사하게 기술의 우수성과 영향력을 나타내며, 기술계적 분석<sup>17)</sup>뿐 아니라 신기술 예측<sup>18)</sup>에도 사용된다. 넷째, 발명자 수는 파급성 측면에서 선정하였다. 발명자의 지식은 일종의 암묵지(tacit knowledge)라고 볼 수 있으며, 특허에 포함된 발명자 정보로 암묵지의 흐름을 분석하거나 발명자의 생산성에 따른 지식의 경쟁력 등을 판단할 수 있다.<sup>19)</sup>

15) 특허청·한국지식재산연구원, “지식재산 경쟁력 및 특성지표 개발: 지식재산 포트폴리오의 특성 분석지표 개발”, 한국지식재산연구원, 2012, 40면.

16) Moore, Kimberly A., “Worthless patents”, *Berkeley Technology Law Journal*, Vol. 20 No. 4(2005), p. 1532.

17) Wong, Chan-Yuan & Wang, Lili, “Trajectories of science and technology and their co-evolution in BRICS: Insights from publication and patent analysis”, *Journal of Informetrics*, Vol. 9 No. 1(2015), pp. 90-101.

18) Breitzman, Anthony & Thomas, Patrick, “The emerging clusters model: A tool for identifying emerging technologies across multiple patent systems”, *Research Policy*, Vol. 44 No. 1(2015), pp. 195-205.

19) 특허청·한국지식재산연구원, “지식재산 경쟁력 및 특성지표 개발: 지식재산 포트폴리오의 특성 분석지표 개발”, 한국지식재산연구원, 2012, 52면.

한편, 주성분 요인분석은 여러 항목의 공통요인을 도출하여 집단화하는 방법으로, 잠재변수(latent variable) 또는 요인 점수(factor score)를 이용한 새로운 지표 생성에 활용된다. 즉, 일반적으로 주성분 요인분석은 동일한 개념을 측정하는 여러 응답(값)을 단일의 응답(값)으로 축소하는 데 활용된다는 것이다. Dai et al.(2021)<sup>20)</sup>의 연구에서는 20개 국가의 경제정책불확실성지수(the Economic Policy Uncertainty Indices)를 하나의 차원으로 축소한 ‘글로벌 경제정책불확실성지수(Global Economic Policy Uncertainty: GEPUI)’ 생성을 위해 주성분 요인분석을 활용하였다. Radovanović et al.(2018)<sup>21)</sup>의 경우, 유럽 28개국의 전기 요금, 1인당 전기 사용량, 에너지 의존도, 국가신용등급 등 9개 항목의 통계값(2004-2013년 기간)를 활용해 여러 에너지 안보 및 경제적 특징 지표를 단일 차원으로 축소한 ‘에너지 안보의 지리경제적 지표(Geo-economic Index of Energy Security)’를 생성하고, 이 지표의 값을 종속변수로 활용한 계량경제학적 분석을 통해 시사점을 도출한 바 있다. 본 연구에서는 앞서 기술한 4가지 특징 지표별로 20개 선행 유망기술의 값을 단일의 값으로 축소하고 선행 유망기술의 공통적 추세를 생성하기 위해 주성분 요인분석을 실시하며, 주성분 요인분석은 민인식(2018)<sup>22)</sup>을 참고하여 통계패키지인 Stata 16.1을 이용한다. 여기서 4가지 특징 지표는 단기적 예측 및 중기적 예측에 활용하기 위하여 각각 특허출원이 최고점에 도달한 시점의 3년 전, 5년 전 시점을 기준으로 과거 20년 데이터로 분석하여 생성한다.<sup>23)</sup>

20) Dai, Peng-Fei et al., “A global economic policy uncertainty index from principal component analysis”, *Finance Research Letters*, Vol.40(2021), Article No.101686.

21) Radovanović, Mirjana et al., “Geo-economic approach to energy security measurement – principal component analysis”, *Renewable and Sustainable Energy Reviews*, Vol.82 No.2(2018), pp.1691-1700.

22) 동 연구에서는 주성분 요인분석을 통해 국가통계포털(KOSIS)에서 제공하는 100대 통계지표 중 56개 지표를 이용하여 우리나라를 대표하는 하나의 지표를 생성하는 예를 제시한 바 있다; 민인식, “Stata를 이용한 KOSIS 100대 지표 웹크롤링”, 『The Korean Journal of Stata』, 제5권 제2호(2018), 29-38면.

23) 예를 들어, 선행 유망기술 A의 특허출원이 2019년에 최고점에 도달하였다면 단기적 예측지표 생성에는 1997~2016년 기간의 데이터를, 중기적 예측지표 생성에는 1995~2014년 기간의 데이터를 활용한다는 것이다.

셋째, 유망기술 후보군 도출을 위해 H섹션 내에서 서브그룹이 없는 메인 그룹의 특허 중 특허출원 건수, 피인용 수, 발명자 수 및 청구항 수 등 4개 항목에 대해 선행 유망기술과 높은 유사성을 보이는 분류를 확인한다. 여기서 앞선 주성분 요인분석을 통해 생성된 선행 유망기술의 공통적 특징 지표와의 유사성을 검토하기 위하여 두 계량형 변수 간 선형 상관관계를 나타내는 피어슨 상관계수(Pearson correlation coefficient) 추정을 실시한다. 결과적으로 특허출원 건수, 피인용 수, 발명자 수 및 청구항 수 등 4개 항목 모두 선행 유망기술 그룹과 5% 유의수준에서 양(+)의 상관성을 보이는 분류를 유망기술 후보군으로 선정한다.<sup>24)</sup>

넷째, 최종 유망기술 선정을 위하여 유망기술 후보군을 대상으로 한 키워드 추출 및 텍스트마이닝을 실시한다. 텍스트마이닝을 위해 본 연구에서는 대표적인 군집화 알고리즘 중에서 K-means 클러스터링을 사용한다. 클러스터(cluster)란 비슷한 특성들을 갖는 데이터들의 집합을 의미하며, 클러스터링(clustering)은 전체 데이터 내에서 서로 유사한 데이터들은 같은 그룹으로, 서로 유사하지 않은 데이터는 다른 그룹으로 분리하는 과정이다. 각 클러스터는 하나의 중심을 가지며, 각 개체는 가장 가까운 중심에 할당된다. 그리고 같은 중심에 할당된 개체들이 모여 하나의 클러스터를 형성하게 된다. 이를 유망기술 후보 메인그룹 각각에 적용하면, 하나의 메인그룹 내에서 서로 유사한 기술끼리 묶이게 되며, 이렇게 함께 묶인 그룹이 정성적 판단으로 어떤 기술인지 확인한다. 한편, K-means 클러스터링을 실시할 때 사전에 설정한 클러스터 수(k)에 따라 결과가 다르게 나타나기 때문에 클러스터 수 설정에 주의가 필요하다. 본 연구에서는 과학적 방법으로 최적의 클러스터 수를

24) 한 익명의 심사자는 본 연구에서 유망기술 후보군을 선정하는 과정에서 활용한 상관관계 분석이 두 변수 간의 '선형적' 관계를 파악하는 방법이므로, 탐색대상 그룹의 특허출원이 선행 유망기술보다 지수함수적으로 급증하는 경우 선행 유망기술과의 유사성을 파악하지 못할 수 있음을 지적한 바 있음을 밝힌다. 단지 특허출원이 폭발적으로 증가하고 있는 경우라면 이미 메인그룹 분류에서 서브그룹 분류로 분화하였을 가능성이 높다고 판단되며, 본 연구에서는 서브그룹으로 분화하지 않은 메인그룹만을 탐색대상으로 선정한다는 점에 유의할 필요가 있다고 하겠다.

설정하기 위하여 실루엣 계수(silhouette coefficient)를 도출하는 기법을 활용한다. 실루엣 계수는 한 클러스터 안에 있는 데이터들이 다른 클러스터와 비교하여 얼마나 비슷한가를 나타낸다. 즉, 실루엣 계수 개수는 하나의 메인그룹 내에 포함된 기술군의 개수를 의미한다. 또한 클러스터 안의 거리가 짧을수록, 다른 클러스터와의 거리가 멀수록 클러스터링이 잘 되어 있다고 볼 수 있다.  $i$ 번째 데이터  $x(i)$ 에 대한 실루엣 계수  $s(i)$ 는 아래와 같은 식으로 정의된다.

$$s(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]}$$

여기서  $a(i)$ 는 군집의 응집도(cohesion)이며, 데이터  $x(i)$ 와 동일한 클러스터 내의 나머지 데이터들과의 평균 거리를 의미한다.  $b(i)$ 는 클러스터 간의 분리도(separation)이며, 데이터  $x(i)$ 와 가장 근접한 클러스터 내의 모든 데이터들과의 평균 거리를 의미한다. 실루엣 계수  $s(i)$ 는 -1부터 1 사이의 값을 가지게 된다. 만일 클러스터의 개수가 최적화되어 있는 경우,  $b(i)$ 의 값은 크고  $a(i)$ 의 값은 작아져  $s(i)$  값이 1에 가까운 숫자가 된다.

## IV. 실험 결과

### 1. 데이터 수집 및 선행 유망기술 선정

본 연구에서 사용한 선행 유망기술 데이터는 20개 특허 분류별로 특허출원이 최고치를 기록한 연도를 기준으로 하여 각각 과거 25년간 미국에 출원된 특허 데이터를 활용하였으며, 수집된 특허 건수는 총 392,859이다. 또한 탐색 대상인 68개의 메인그룹은 1998년부터 2017년까지 20년간의 미국 특허 데이터를 활용하였으며, 수집된 특허 건수는 총 68,254건이다.

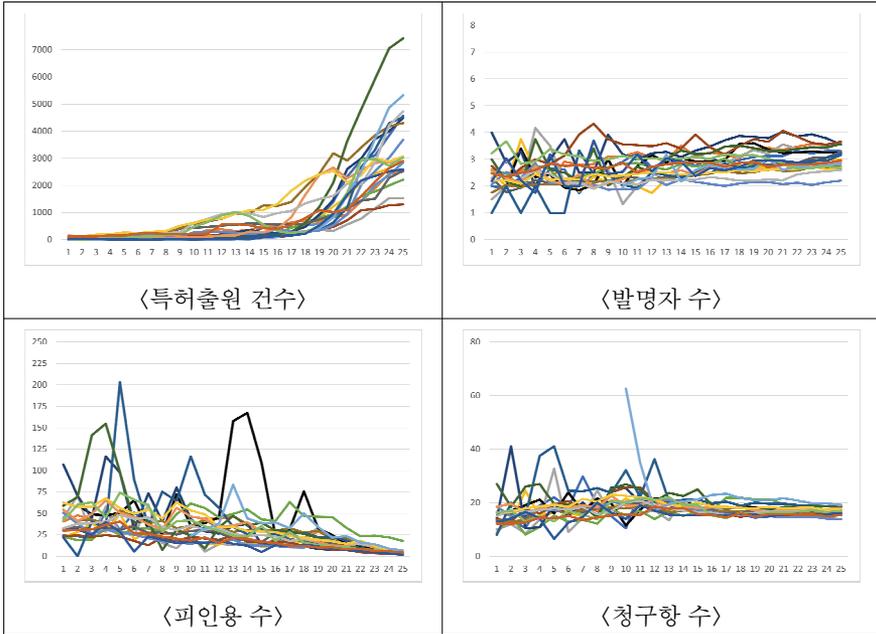
선행 유망기술 선정 대상은 H섹션의 메인그룹 또는 서브그룹 중에서 미국

특허청(USPTO)에 출원된 특허가 1만 건 이상인 특허분류를 1차로 추출하였다. 그 결과 1차 추출된 후보군은 55개 분류였으며, H01L 분야가 19개, H04N은 6개 순이다. 이 중에서 한국특허전략개발원의 특허분류 전문가와 기술 전문가 등 3명의 추천 기술 및 디지털카메라, LED 등 일반적으로 널리 알려진 과거 유망기술을 중심으로 총 20개의 분류를 선정하였다. 선행 유망 기술 선정 기준은 후보군 중에서 특허출원량이 최고점에 도달하여 성숙기의 모습을 보이고, 특허출원량이 상위권인 분류만을 대상으로 하였다. 정확한 결과 도출을 위해 특정 기술분야에 편중되지 않도록 1차 후보군의 분류 분포를 고려하여 적절히 배분하였다.

한편, 전체 출원량은 3만 건 이상으로 많으나 현재도 지속적으로 출원이 증가하고 있는 ‘디지털 통신 에러 검출 및 방지 기술(H04L 1/00)’ 같은 분류는 아직 유망성을 고려하기 위한 전체 패턴이 완성되지 않았다고 판단하여 선행 유망기술에서 제외하였다. 또한 특허출원량이 최고점을 기록한 시점이 오래된 경우(데이터 수집 시점 기준으로 10년 이상 경과)에도 시의성 측면에서 미래 유망기술 예측에 활용하기는 적절하지 않다고 판단하여 제외하였다. 예를 들어, 이동통신 관련 기술(H04J 3/00, H04J 7/20, H04B 1/06) 등은 2000년대 중반에 특허출원이 가장 많았던 기술로, 현 시점에 미래 유망기술을 예측하는 데 있어서 이들 기술의 부상 패턴은 다소 시의성이 미흡하다고 판단하였다.

최종 수집된 선행 유망기술 20개 분류의 출원 건수, 발명자 수, 피인용 수 및 청구항 수 등 특징 지표 항목에 대한 시계열적 패턴은 다음 <그림 2>와 같다.<sup>25)</sup>

25) 전술한 바와 같이, 단기적 예측 및 중기적 예측을 위한 지표 생성에는 선행 유망기술의 특허출원이 최고점에 도달한 시점의 3년 전, 5년 전 시점을 기준으로 과거 20년 데이터를 활용한다. 따라서 실제 선행 유망기술의 데이터는 특허출원이 최고점에 도달한 시점을 기준으로 과거 25년 치를 수집하였으며, <그림 2>와 <그림 3>은 초기 수집한 데이터로 도식화한 것임을 밝힌다.



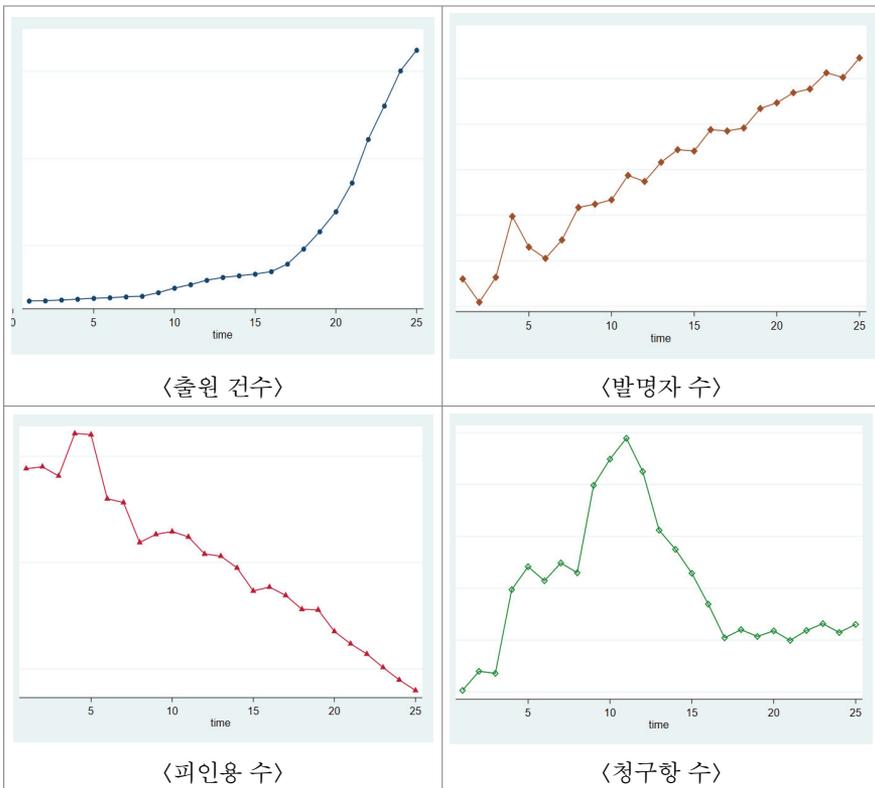
〈그림 2〉 선행 유망기술 그룹의 시계열 패턴

## 2. 탐색 대상 그룹 선정

전술한 바와 같이, 본 연구에서는 아직 서브그룹으로 세분화되지 않은 메인그룹에 특허출원이 많지 않은 초기의 신기술이 포함되어 있을 가능성이 높다는 전제하에 메인그룹의 특허를 대상으로 미래 유망기술을 발굴해 보고자 한다. 탐색 대상 그룹은 H섹션에서 서브그룹이 없는 메인그룹 125개 중 20년간 전체 특허 건수 합계가 100개가 되지 않는, 즉 특허출원이 극히 미미한 57개 분류는 제외한 68개 분류로 선정하였다. 탐색 대상 그룹의 특허 수는 최소 100건에서 최대 6,874건이다.

### 3. 선행 유망기술의 공통적 특징 추출

다음으로, 선행 유망기술 특허의 출원 건수, 피인용 수, 발명자 수, 청구항 수 등 4개 특징 지표의 공통적 추세를 생성하기 위해 주성분 요인분석을 실시하였다. 다음 <그림 3>은 주성분 요인분석을 통해 선행 유망기술의 특허 출원이 최고점에 도달한 시점을 기준으로 과거 25년 기간의 특징별 추이를 나타낸 것이다.



<그림 3> 선행 유망기술 그룹의 항목별 추이(주성분 요인분석 결과)

첫째, 특허출원 건수는 최고점 부근을 기준으로 과거 5~6년 전부터 급증하는 추세를 보였다. 둘째, 선행 유망기술의 발명자 수는 지속적·선형적으

로 증가하는 추세를 보였다. 이는 유망기술 분야에 연구개발 투자가 집중되어 개발에 참여하는 연구인력이 늘어나고, 더불어 많은 연구자가 참여하는 공동연구가 활발한 결과일 것으로 사료된다. 셋째, 피인용 수는 특허출원량이 최고점인 시기에서 약 20년 이전 시점에 매우 높은 모습이다가 이후 지속적으로 감소하는 추세를 보였다. 이는 기술개발의 초기에는 중요한 원천특허가 출원되므로 이들 특허를 인용하는 횟수가 많고, 이후에는 기술의 상용화를 위한 다양한 애플리케이션에 관한 많은 응용특허가 주로 출원되는 현상을 나타낸 것으로 사료된다. 넷째, 청구항 수는 특허출원량이 최고점인 시점에서 약 10년 정도 이전에 가장 높은 모습을 보였다. 이는 선행 유망기술에 상용화를 위한 초기 기술의 난제를 해결한 혁신성과 시장성이 높은 특허가 많다는 의미로 볼 수 있다. 일반적으로 새로 개발한 기술의 특허를 출원할 때 다양한 실시 예에 대해 권리를 확보하고자 특허권 범위의 청구항의 개수를 늘리는 경향이 있기 때문이다.

#### 4. 탐색대상 그룹과 선행 유망기술의 상관관계 분석

유망기술 후보군을 선정하는 과정으로서, 4개 특징 지표에 대해 선행 유망기술의 공통적 추세와 각 탐색대상 그룹의 유사성을 확인하기 위한 상관관계 분석을 실시하였다. 세부적으로, 선행 유망기술의 특허출원이 최고점에 도달한 시점의 3년 전과 5년 전 시점을 기준으로 과거 20년 기간에 대한 공통적 추세를 비교군으로 하여 각 탐색대상 그룹과 분석하였다. 여기서 3년 전 시점을 기준으로 하여 비교한 경우 단기적 예측, 5년 전 시점을 기준으로 하여 비교한 경우 중기적 예측이라 할 수 있다.

먼저, 단기적 예측 결과로서 각 탐색대상 그룹에 대해 선행 유망기술의 특허출원이 최고점에 도달한 시점의 3년 전 시점을 기준으로 과거 20년 기간에 대한 공통적 추세와 상관관계 분석한 결과는 <표 3>과 같다. 분석 결과, H04J 11/00과 H05K 9/00의 두 개 메인그룹이 4가지 항목 모두 선행 유망기술 그룹과 5% 수준에서 통계적으로 유의한 양(+)의 상관관계를 가지는 것으

로 관측되었다.

〈표 3〉 상관관계 분석 결과(단기-최고점 기준 과거 3년)

| 특허분류       | 출원 건수   | 발명자 수   | 청구항 수   | 피인용 수   |
|------------|---------|---------|---------|---------|
| H01F13/00  | .641**  | .194    | -.080   | .536*   |
| H01H39/00  | .703*** | .105    | -.095   | .433    |
| H01H49/00  | .702*** | .794*** | -.104   | -.106   |
| H01H57/00  | -.108   | .053    | .410    | .802*** |
| H01H59/00  | .822*** | .093    | .054    | .919*** |
| H01H65/00  | -.169   | .333    | .252    | .150    |
| H01M14/00  | .056    | .306    | .042    | .787*** |
| H01M16/00  | .725*** | .629**  | .115    | .732*** |
| H01P11/00  | .358    | .410    | .577**  | .489*   |
| H01Q17/00  | .423    | .096    | .643**  | .829*** |
| H01Q23/00  | .803*** | .065    | .452*   | .795*** |
| H01R29/00  | -.504*  | .838*** | .415    | .739*** |
| H01R101/00 | .866*** | .797*** | -.555*  | -.660** |
| H01R103/00 | .956*** | .621**  | -.014   | -.227   |
| H01R105/00 | .934*** | .617**  | -.525*  | -.456*  |
| H01R107/00 | .949*** | .837*** | -.424   | -.481*  |
| H01S4/00   | -.313   | .659**  | .286    | .612**  |
| H01T23/00  | .072    | .062    | .246    | .551*   |
| H02G13/00  | .720*** | .194    | -.080   | .536*   |
| H02H11/00  | .571**  | .337    | .008    | .007    |
| H02J4/00   | .418    | .244    | .084    | .688*** |
| H02J5/00   | .597**  | -.235   | .530*   | .379    |
| H02J11/00  | .749    | .175    | .148    | .369    |
| H02J13/00  | .916*** | -.292   | .316    | .125    |
| H02J17/00  | .356    | -.138   | .324    | .254    |
| H02K53/00  | .571**  | .821*** | .407    | .573    |
| H02M11/00  | -.095   | .099    | -.299   | .875*** |
| H02N3/00   | .095    | .639**  | .112    | .765*** |
| H02N6/00   | -.184   | .568**  | .818*** | .588**  |
| H02N10/00  | -.220   | .440    | .081    | .920*** |
| H02N11/00  | .393    | .306    | -.061   | .565*   |
| H02N13/00  | .406    | .102    | .152    | .394    |

|            |         |         |        |         |
|------------|---------|---------|--------|---------|
| H02P4/00   | .753*** | -.148   | .188   | .753**  |
| H02P31/00  | .699*** | -.551*  | .051   | .589*   |
| H02S99/00  | .955*** | .016    | .382   | .553    |
| H03B17/00  | .774*** | -.287   | .610** | .464    |
| H03B27/00  | -.233   | .398    | .158   | .855*** |
| H03B28/00  | .124    | .241    | .669** | .661**  |
| H03B29/00  | -.532*  | .656**  | .086   | .518*   |
| H03D13/00  | -.608** | .451*   | -.124  | .797*** |
| H03F99/00  | .054    | .303    | .369   | .481    |
| H03H19/00  | .939*** | -.012   | -.169  | .877*** |
| H03H21/00  | .833*** | -.370   | .326   | .746*** |
| H03M9/00   | .331    | .147    | -.242  | .951*** |
| H04B11/00  | .852*** | -.040   | -.166  | .517*   |
| H04H5/00   | -.601** | .067    | .052   | .316    |
| H04H9/00   | -.505*  | .019    | .446   | .714**  |
| H04J4/00   | -.462*  | .717*** | .552*  | .814*** |
| H04J9/00   | -.434   | .515*   | -.116  | .865*** |
| H04J11/00  | .660**  | .963*** | .545*  | .932*** |
| H04J15/00  | -.615** | .580*   | .169   | .888*** |
| H04J99/00  | .289    | -.318   | .574*  | .203    |
| H04K3/00   | .404    | .500*   | .101   | .668**  |
| H04N15/00  | -.444   | .169    | .245   | .661**  |
| H04N101/00 | .465*   | .238    | .607*  | .694**  |
| H04R29/00  | .911*** | .839*** | .192   | .822*** |
| H04R31/00  | .692*** | -.399   | -.284  | .005    |
| H04S3/00   | .972*** | -.047   | .080   | .259    |
| H04S7/00   | .976*** | .650**  | -.197  | .523*   |
| H04W56/00  | .892*** | .018    | .319   | .788*** |
| H04W64/00  | .878*** | .333    | -.011  | .598*   |
| H05B11/00  | -.366   | -.159   | .020   | .777*** |
| H05G2/00   | .779*** | .371    | .122   | .687*** |
| H05H9/00   | .022    | .042    | .217   | .759*** |
| H05K9/00   | .852*** | .829*** | .592** | .918*** |
| H05K10/00  | -.414   | -.341   | .123   | .751*   |

주: \*, \*\*, \*\*\*는 각각 5%, 1%, 0.1% 수준에서 유의함을 의미

둘째, 중기적 예측 결과로서 각 탐색대상 그룹에 대해 선행 유망기술의 특

허출원이 최고점에 도달한 시점의 5년 전 시점을 기준으로 과거 20년 기간에 대한 공통적 추세와 상관관계 분석한 결과는 <표 4>와 같다. 분석 결과, 유일하게 H05K 9/00 분류만이 4가지 항목 모두 선행 유망기술 그룹과 5% 수준에서 통계적으로 유의한 양(+)의 상관관계를 가지는 것으로 나타났다. 특히, H05K 9/00 분류는 앞선 단기적 예측에서도 유의한 상관관계가 확인된 바 있다.

<표 4> 상관관계 분석 결과(중기-최고점 기준 과거 5년)

| 특허분류       | 출원 건수   | 발명자 수   | 청구항 수  | 피인용 수    |
|------------|---------|---------|--------|----------|
| H01F13/00  | .652**  | .339    | .051   | .561**   |
| H01H39/00  | .720*** | -.034   | -.122  | .521*    |
| H01H49/00  | .702*** | .776*** | .451*  | -.104    |
| H01H57/00  | -.065   | .111    | .138   | .708***  |
| H01H59/00  | .848*** | -.144   | .308   | .816***  |
| H01H65/00  | -.133   | .275    | .455*  | .054     |
| H01M14/00  | .136    | .090    | -.242  | .763***  |
| H01M16/00  | .742*** | .754*** | -.116  | .763***  |
| H01P11/00  | .430    | .344    | .355   | .548*    |
| H01Q17/00  | .446*   | -.092   | .347   | .777***  |
| H01Q23/00  | .812*** | -.263   | .467*  | .546*    |
| H01R29/00  | -.528*  | .869*** | .599** | .777***  |
| H01R101/00 | .853*** | .765*** | -.073  | -.678*** |
| H01R103/00 | .947*** | .591**  | .495*  | -.208    |
| H01R105/00 | .934*** | .661**  | -.065  | -.486*   |
| H01R107/00 | .934*** | .812*** | .101   | -.541*   |
| H01S4/00   | -.281   | .554*   | .301   | .718***  |
| H01T23/00  | .090    | .075    | -.102  | .705***  |
| H02G13/00  | .753*** | .339    | .051   | .561**   |
| H02H11/00  | .595**  | .572**  | .589** | -.157    |
| H02J4/00   | .470*   | .106    | .152   | .584**   |
| H02J5/00   | .601**  | -.247   | .236   | .399     |
| H02J11/00  | .733*** | .545*   | .316   | .055     |
| H02J13/00  | .923*** | .372    | .268   | -.053    |

|            |         |          |        |         |
|------------|---------|----------|--------|---------|
| H02J17/00  | .398    | .271     | .595** | .107    |
| H02K53/00  | .629**  | .859***  | .298   | -.183   |
| H02M11/00  | -.026   | .225     | -.296  | .824*** |
| H02N3/00   | .097    | .686***  | -.148  | .763*** |
| H02N6/00   | -.182   | .491*    | .462*  | .684*** |
| H02N10/00  | -.177   | .410     | -.291  | .914*** |
| H02N11/00  | .468*   | .487*    | .236   | .483*   |
| H02N13/00  | .393    | .049     | -.264  | .417    |
| H02P4/00   | .759*** | .586**   | .469*  | .260    |
| H02P31/00  | .729*** | .398     | .526*  | -.238   |
| H02S99/00  | .939*** | .737***  | -.278  | -.569** |
| H03B17/00  | .809*** | .061     | .318   | .411    |
| H03B27/00  | -.267   | .449*    | .270   | .807*** |
| H03B28/00  | .176    | .702***  | .538*  | .458*   |
| H03B29/00  | -.532*  | .688***  | -.029  | .526*   |
| H03D13/00  | -.591** | .430     | -.099  | .890*** |
| H03F99/00  | .145    | .753***  | .521*  | -.177   |
| H03H19/00  | .930*** | .371     | .033   | .321    |
| H03H21/00  | .840*** | -.231    | .060   | .455*   |
| H03M9/00   | .352    | .195     | -.196  | .920*** |
| H04B11/00  | .855*** | -.079    | -.223  | .452*   |
| H04H5/00   | -.651** | -.760*** | -.185  | .686*** |
| H04H9/00   | -.548*  | -.532*   | .421   | .809*** |
| H04J4/00   | -.400   | .667**   | .097   | .897*** |
| H04J9/00   | -.423   | .436     | -.239  | .851*** |
| H04J11/00  | .661**  | .896***  | .227   | .956*** |
| H04J15/00  | -.667** | -.253    | .241   | .967*** |
| H04J99/00  | .360    | .684***  | .400   | -.379   |
| H04K3/00   | .461*   | .560**   | -.253  | .684*** |
| H04N15/00  | -.379   | .083     | .102   | .788*** |
| H04N101/00 | .493*   | .845***  | .434   | -.346   |
| H04R29/00  | .918*** | .912***  | .106   | .865*** |
| H04R31/00  | .708*** | -.173    | -.175  | -.052   |
| H04S3/00   | .966*** | .246     | .279   | .057    |
| H04S7/00   | .969*** | .825***  | .533*  | -.345   |
| H04W56/00  | .894*** | .817***  | .505*  | -.162   |

|           |         |         |       |         |
|-----------|---------|---------|-------|---------|
| H04W64/00 | .889*** | .731*** | .172  | .108    |
| H05B11/00 | -.377   | -.085   | .020  | .801*** |
| H05G2/00  | .777*** | .487*   | .094  | .470*   |
| H05H9/00  | .028    | .038    | .539* | .724*** |
| H05K9/00  | .870*** | .777*** | .493* | .925*** |
| H05K10/00 | -.431   | -.237   | -.332 | .657**  |

주: \*, \*\*, \*\*\*는 각각 5%, 1%, 0.1% 수준에서 유의함을 의미

결과적으로, H04J 11/00과 H05K 9/00 등 2개 특허분류는 선행 유망기술들의 공통적 패턴을 따라 부상할 가능성이 높다고 예상되는바, 상관관계 분석 결과를 근거로 이들 2개 메인그룹을 미래 유망기술 특허분류로 선정하였다.

## 5. 텍스트 전처리

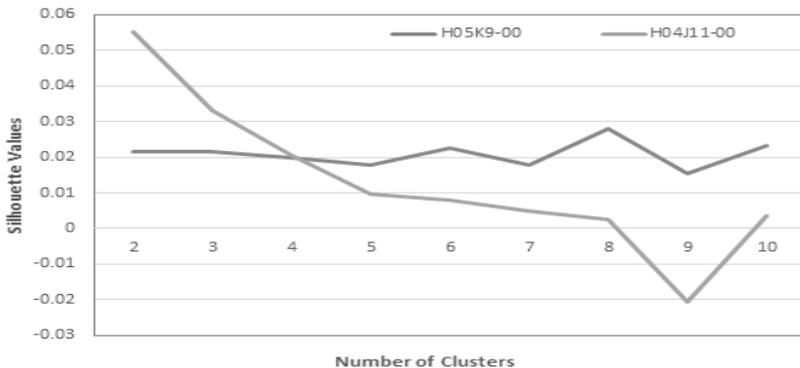
앞서 미래 유망기술 특허분류로 선정된 2개 메인그룹에 포함되어 있는 기술을 확인하기 위한 과정을 진행한다. 먼저, 텍스트마이닝 분석을 위해 텍스트 전처리 작업을 실시하였다. 우선 기술 분류별 문서들의 텍스트마이닝 분석을 위해 특허의 제목, 요약문, 대표 청구항을 분석 대상으로 선정하였다. 문서의 대문자를 소문자로 변환하였고, the, a, and 등 텍스트 분석에 불필요한 불용어를 제거하는 텍스트 전처리 작업을 수행하였다. 다음으로 문서의 중요한 단어들을 선별하기 위해 TF-IDF 가중치를 적용하였다. TF-IDF 가중치는 문서 집합 내에 특정 단어가 얼마나 중요한지를 나타내는 통계적 수치이다. 여기서 TF는 Term Frequency의 약어으로써 문서 집합 내에 등장한 단어의 빈도수를 나타내고, IDF(Inverse Document Frequency)는 특정 단어가 출현한 문서들의 역수를 나타낸다. 이러한 TF 값과 IDF 값을 곱함으로써 문서 집합 내 특정 단어의 중요도를 산출할 수 있다. 본 연구에서 TF-IDF 가중치 임계값은 0.4로 선정하였다. 그 결과 다음 <표 5>와 같이 H05K 9/00의 중요 단어 수는 557개, H04J 11/00은 647개로 확인되었다.

〈표 5〉 유망기술 특허분류에 대한 텍스트 전처리 결과

| 특허분류       | 전체 단어 수 | TF-IDF $\geq 0.4$ |
|------------|---------|-------------------|
| H05K 9/00  | 18,595  | 557               |
| H04J 11/00 | 20,008  | 647               |

## 6. 기술군 형성

선정된 중요 단어를 기준으로 Word2Vec 분석을 수행하였다. Word2Vec은 단어의 의미를 벡터화할 수 있는 신경망 기반의 word embedding 방법들 중 하나로서 크게 CBOW(continuous bag of words)와 skip-gram 방식으로 구분된다. CBOW 모델은 문장 내 주변 단어로부터 타깃 단어를 예측하는 신경망이고, 반대로 skip-gram은 타깃 단어로부터 주변 단어를 예측하는 방식의 신경망이다. 본 연구에서는 타깃 단어로 주변 단어를 예측할 수 있도록 skip-gram으로 설정하였으며, 1회에 학습하는 단어의 개수는 5로 설정하였다. 각각의 단어를 밀집 벡터로 변환시켜 줌으로써 임베딩된 단어들을 클러스터링에 적용하였으며, 임베딩된 단어들을 K-means 클러스터링에 적용하기 위한 최적의 클러스터 수(k)는 실루엣 점수(silhouette score)를 기준으로 선정하였다. K-means 클러스터링 결과는 〈그림 4〉와 같으며, 결과적으로 특



〈그림 4〉 유망기술 후보군의 K-means 클러스터링 결과

허분류 H04J 11/00와 H05K 9/00의 최적 클러스터 수는 각각 2개, 8개로 확인되었다.

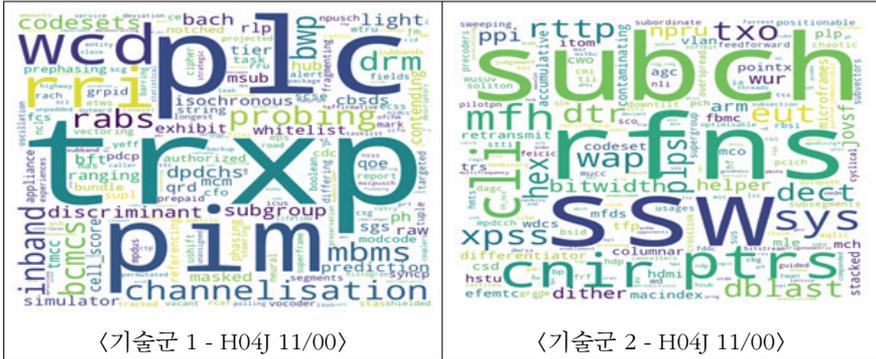
## 7. 텍스트마이닝 및 미래 유망기술 선정

유망기술 분류로 선정된 H04J 11/00과 H05K 9/00에 어떠한 세부 기술이 포함되어 있는지를 파악하기 위하여 텍스트마이닝을 실시하였다. 다음으로, 텍스트마이닝 결과에 나타난 핵심 키워드를 바탕으로 각 분류의 특허에 포함된 구체적 기술 내용을 정성적으로 확인하였다. 텍스트마이닝 결과에서 기술군이 두 개 이상으로 구분된 경우, 이들 기술군이 상호 독립적인 기술인지 또는 단일 기술분야의 세부기술로서 그룹화할 수 있는지 검토하였다.

특허분류 H04J 11/00 기술의 IPC상 정의는 다중 통신의 직교 다중화 시스템으로 통신 관련 기술에 해당한다. 기술군 1은 PLC(Power Line Communication), PIM(Passive Intermodulation), TRXP(Transmission Reception Point) 등의 키워드를 포함하며, 통신 신호의 전송 기술, PIM에 의해 발생하는 신호 간섭 저감 기술 등 통신 신호의 품질 향상을 위한 기술이 해당된다. 기술군 2는 직교 다중분할통신(OFDM: Orthogonal Frequency Division Multiplexing) 기술로 정의되며, 이는 전통적인 통신기술로 알려져 있다. 텍스트마이닝 결과 값의 상위 words에는 포함되지 않았으나 최근 수년간의 특허를 살펴보면 5G 및 IoT(Internet of Thing)와 같은 키워드가 상당수로 확인된다. 특허분류 H04J 11/00에 대한 텍스트마이닝 결과를 종합하면 ‘차세대 통신 신호 전송 및 품질 향상 기술’을 핵심적인 세부 유망기술로 볼 수 있다. 이들 특허의 주요 출원인으로서 LG전자(636건), QUALCOMM(562건), 삼성전자(461건), ERICSSON(317건) 등 통신 관련 글로벌 기업들이 활발하게 기술개발 및 특허 출원을 하고 있고, 일본 신에너지·산업기술 종합개발기구(NEDO)도 지난 2021년 차세대 유망 정보통신 기술로 5G 정보통신 전송 기술을 공표<sup>26)</sup>한 바

26) 新エネルギー・産業技術総合開発機構, “次世代に期待される情報通信技術”, 新エネルギー・産業技術総合開発機構, 2021.

있다는 점에서도 ‘차세대 통신 신호 전송 및 품질 향상 기술’이 유망기술로서 사회 · 경제에 상당한 영향력을 미칠 가능성은 상당히 높다고 판단된다.



<그림 5> H04J 11/00의 기술군별 워드 클라우드

<표 6> H04J 11/00의 기술군별 텍스트마이닝 결과값(상위 10개 워드)

| 기술군 1          |          | 기술군 2 |          |
|----------------|----------|-------|----------|
| words          | 점수       | words | 점수       |
| trxp           | 0.96255  | subch | 0.988259 |
| plc            | 0.95983  | ssw   | 0.960611 |
| pim            | 0.914409 | rfns  | 0.942411 |
| wcd            | 0.910401 | ptrs  | 0.940133 |
| rri            | 0.905161 | cnir  | 0.911241 |
| channelisation | 0.887401 | cli   | 0.910607 |
| probing        | 0.851859 | sys   | 0.907248 |
| bwp            | 0.84215  | mfh   | 0.896731 |
| mbms           | 0.841924 | xpss  | 0.892838 |
| inband         | 0.836098 | txo   | 0.88512  |

특허분류 H05K 9/00는 전장 또는 자장에 대한 장치 또는 부품의 차폐에 관한 것으로, 전자기와 차폐 기술이 주를 이루고 있다. 기술군 1은 earbud, fot photoconductor 등의 키워드를 포함하고 있으며, 초소형 헤드폰 기술로

정의된다. 기술군 2는 boss, pack, digitizer, reser 등의 키워드를 포함하고 있으며, 전자기파 차폐를 위한 구조로 정의된다. 기술군 3은 nanomesh, imd, blackend, harmonizer, ecat, ecn, softmagnetic 등의 키워드를 포함하고 있으며, 전자기파 필터로 정의된다. 기술군 4는 restraint, knuckle, contactor, sidepanel 등으로 구성되어 전자기기 등의 전자기파 흡수 접촉 단자, 기술군 5는 indicator, principle, qc(quantum computing), modularized 등으로 구성되어 컴퓨터 회로 기판의 전자기파 차폐 회로 및 칩, 기술군 6은 assemble, facial, slickline, cru, fia 등으로 구성되어 케이블 구조, 기술군 7은 forearm, plumbing, sas, microelectrode, earphone 등으로 구성되어 기기간 연결 기술, 기술군 8은 magnetometer, trimmable, icpt, undulated, cloak, protuberance 등으로 구성되어 전자기파 측정 기술로 정의된다. 한편, ‘전자기파 차폐 기술’은 자동차 통신, 5G 이동통신, 전자 기기(스마트폰, 웨어러블, 노트북 등)의 증가, 전기차 무선 충전 및 발전 전력의 무선 전송 등 사용처가 다양하며 전자기파에 의한 안정성 문제는 해결해야 할 핵심 과제로 알려져 있다. APPLE(116건), 3M(76건), MURATA MANUFACTURING(63건) 등 글로벌 기업들이 이들 특허의 주요 출원인으로 확인되며, 최근까지도 활발하게 특허출원을 하고 있다. 국립전파연구원(2013)<sup>27)</sup>의 경우 2025년까지의 미래사회에 필요한 전파기술에 대한 수요를 예측하고, 이를 실현하기 위한 핵심 요소기술로 ‘기기간 전자파 영향분석·평가·보호 기술’을 선정하기도 하였다. 결론적으로, 각종 기기의 무선화에 따라 다양한 방면에 필요한 ‘전자기파 차폐 기술’은 미래 글로벌 사회·경제에 상당한 영향력을 미칠 가능성이 상당히 높다고 판단된다.

27) 국립전파연구원, “2025년 미래전파 기술수요 예측조사”, 2013.



| 기술군 1           |      | 기술군 2         |      | 기술군 3          |      | 기술군 4           |      |
|-----------------|------|---------------|------|----------------|------|-----------------|------|
| words           | 점수   | words         | 점수   | words          | 점수   | words           | 점수   |
| pwa             | 0.90 | twinax        | 0.82 | ecn            | 0.69 | pedal           | 0.75 |
| obscuration     | 0.89 | earthing      | 0.81 | softmagnetic   | 0.68 | gratings        | 0.70 |
| glove           | 0.86 | osd           | 0.80 | pouchy         | 0.64 | elbow           | 0.69 |
| sliced          | 0.86 | induction     | 0.79 | harmonizer     | 0.58 | semicylindrical | 0.69 |
| ddi             | 0.85 | strengthening | 0.75 | micropattern   | 0.57 | aerial          | 0.67 |
| trailer         | 0.85 | pcd           | 0.75 | reticular      | 0.56 | carriable       | 0.66 |
| 기술군 5           |      | 기술군 6         |      | 기술군 7          |      | 기술군 8           |      |
| words           | 점수   | words         | 점수   | words          | 점수   | words           | 점수   |
| principle       | 0.86 | assemble      | 0.96 | forearm        | 0.86 | magnetometer    | 0.83 |
| indicator       | 0.84 | facial        | 0.87 | sas            | 0.80 | icpt            | 0.83 |
| qc              | 0.83 | cru           | 0.84 | microelectrode | 0.80 | trimmable       | 0.83 |
| modularized     | 0.78 | slickline     | 0.80 | earphone       | 0.78 | undulated       | 0.82 |
| tree            | 0.74 | fia           | 0.80 | plumbing       | 0.77 | cloak           | 0.81 |
| important       | 0.68 | tgvs          | 0.79 | resealable     | 0.71 | protuberance    | 0.80 |
| bristle         | 0.63 | sock          | 0.77 | wiper          | 0.71 | rooftop         | 0.80 |
| atca            | 0.60 | vise          | 0.73 | score          | 0.70 | condensers      | 0.77 |
| bristles        | 0.58 | workslate     | 0.69 | turf           | 0.70 | ltcc            | 0.74 |
| instrumentation | 0.54 | wireline      | 0.65 | suppressive    | 0.70 | weighing        | 0.72 |

## V. 결론

본 연구에서는 요인분석 및 텍스트마이닝 분석을 통해 특허정보의 통계적 및 서지적 데이터를 공히 활용하는 미래 유망기술 도출 방법을 제안하였다. 본 연구에서 선정된 H섹션 내 미래 유망기술은 ‘차세대 통신 신호 전송 및 품질 향상 기술’과 ‘전자기파 차폐 기술’이며 이를 정리하면 아래 <표 8>과 같다.

〈표 8〉 H섹션 내 미래 유망기술 선정 결과

| 연 번 | 유망기술                    |
|-----|-------------------------|
| 1   | 차세대 통신 신호 전송 및 품질 향상 기술 |
| 2   | 전자기파 차폐 기술              |

앞서 논의한 바와 같이, 선행연구들은 대부분 특허건수와 같은 단일의 통계적 데이터 또는 서지적 정보만을 활용해 기술의 유망성을 판단하고 있다는 점에서 정밀성·예측력이 미흡하거나 상당한 비용이 수반된다는 한계가 있다. 본 연구에서 제안한 방법은 서브그룹으로 세분화되지 않은 메인그룹에 유망기술이 포함되어 있을 것이라는 전제에서 이 같은 메인그룹을 데이터 수집 대상으로 설정하고, 통계적 데이터를 활용하는 주성분 요인분석 및 상관관계 분석을 통해 과거에 유망했던 기술분야들과 유사한 특징을 보이는 유망기술분야 후보군을 도출한 후 특허의 서지적 데이터를 활용하는 텍스트 마이닝 분석을 통해 최종적으로 유망기술을 확정한다는 점에서 선행연구로부터 제안된 방법과 차이가 있다. 본 연구는 복수의 분석기법을 결합함으로써 미래 유망기술 발굴의 효율성·객관성을 높일 수 있는 새로운 방법론을 제안한다는 점에서 학문적·실무적 기여를 찾을 수 있다고 하겠다.

한편, 전술한 연구의 의의에도 불구하고, 본 연구에서는 IPC의 메인그룹 중 H섹션(전기)만을 대상으로 분석한 결과에 근거하고 있다는 점에서 한계가 있음을 밝혀 둔다. 후속 연구에서는 기계, 바이오 등 다른 분야의 데이터를 수집·분석하여 제안된 방법론을 검증할 필요가 있을 것이다. 아울러 본 연구에서는 선행 유망기술의 공통적 특징 도출을 위해 출원 건수, 발명자 수, 피인용 수 및 청구항 수 등 4개 항목을 기준으로 삼았으나, 유망기술의 특징을 보다 효과적으로 포착해낼 수 있는 지표에 대한 추가적 연구도 필요하다고 사료된다.

참고문헌

〈학술지(국내 및 동양)〉

- 감주식 외 2인, “특허정보 기반의 바이오 기술개발 트렌드 분석 및 유망기술분야 도출에 관한 연구”, 『기술혁신연구』, 제21권 제2호(2013).
- 강희종 외 2인, “특허분석을 통한 유망융합기술의 예측”, 『기술혁신연구』, 제14권 제3호(2006).
- 김방룡·황성현, “특허 정보를 활용한 IT 유망기술 도출에 관한 연구”, 『한국통신학회 논문지』, 제34권 제10호(2009).
- 김종찬 외 4인, “특허 키워드 시계열 분석을 통한 부상 기술 예측”, 『정보처리학회논문지. 소프트웨어 및 데이터 공학』, 제3권 제9호(2014).
- 민인식, “Stata를 이용한 KOSIS 100대 지표 웹크롤링”, 『The Korean Journal of Stata』, 제5권 제2호(2018).
- 서원철, “기술테마 분석을 통한 기술기회발굴 연구: 3D 프린팅 기술사례를 중심으로”, 『지식재산연구』, 제16권 제2호(2021).
- 주성현, “무인항공기 기술진화 탐색 및 유망기술 발굴 연구”, 『항공우주시스템공학회지』, 제13권 제6호(2019).

〈학술지(서양)〉

- Bodger, P. S. & H.S, Tay, “Trend extrapolation in long-term forecasting: An investigation using New Zealand electricity consumption data”, *Technological Forecasting and Social Change*, Vol.30 No.2(1986).
- Breitzman, Anthony & Thomas, Patrick, “The emerging clusters model: A tool for identifying emerging technologies across multiple patent systems”, *Research Policy*, Vol.44 No.1(2015).
- Choi, Youngjae et al., “Identifying emerging technologies to envision a future innovation ecosystem: A machine learning approach to patent data”, *Scientometrics*, Vol.126(2021).
- Cozzens, Susan et al., “Emerging technologies: quantitative identification and measurement”, *Technology Analysis and Strategic Management*, Vol.22 No.3(2010).
- Dai, Peng-Fei et al., “A global economic policy uncertainty index from principal component analysis”, *Finance Research Letters*, Vol.40(2021).

- Érdi, Péter et al., “Prediction of emerging technologies based on analysis of the US patent citation network”, *Scientometrics*, Vol.95(2013).
- Jang, Wooseok et al., “Identifying emerging technologies using expert opinions on the future: A topic modeling and fuzzy clustering approach”, *Scientometrics*, Vol.126(2021).
- Kang, Hyo Y., “Science inside law: The making of a new patent class in the international patent classification”, *Science in Context*, Vol.25 No.4(2012).
- Keller, Jonas & von der Gracht, Heiko A., “The influence of information and communication technology(ICT) on future foresight processes — Results from a Delphi survey”, *Technological Forecasting and Social Change*, Vol.85(2014).
- Lee, Changyong et al., “Early identification of emerging technologies: A machine learning approach using multiple patent indicators”, *Technological Forecasting and Social Change*, Vol.127(2018).
- Moore, Kimberly A., “Worthless patents”, *Berkeley Technology Law Journal*, Vol.20 No.4(2005).
- Priestley, Maria et al., “Innovation on the web: the end of the S-curve?” *Internet Histories*, Vol.4 No.4(2020).
- Radovanović, Mirjana et al., “Geo-economic approach to energy security measurement — principal component analysis”, *Renewable and Sustainable Energy Reviews*, Vol.82 No.2(2018).
- Rotolo, Daniele et al., “What is an emerging technology?”, *Research Policy*, Vol.44 No.10(2015).
- Schoemaker, Paul J. H., “Scenario planning: A tool for strategic thinking”, *MIT Sloan Management Review*, Vol.36 No.2(1995).
- Small, Henry et al., “Identifying emerging topics in science and technology”, *Research Policy*, Vol.43 No.8(2014).
- Trappey, Charles V. et al., “Using patent data for technology forecasting: China RFID patent analysis”, *Advanced Engineering Informatics*, Vol.25 No.1(2011).
- Wang, Baicun et al., “Emerging nanogenerator technology in China: A review and forecast using integrating bibliometrics, patent analysis and technology roadmapping methods”, *Nano Energy*, Vol.46(2018).
- Wong, Chan-Yuan & Wang, Lili, “Trajectories of science and technology and their co-evolution in BRICS: Insights from publication and patent analysis”, *Journal of*

*Informetrics*, Vol.9 No.1(2015).

Yu, Xiang & Zhang, Ben, "Obtaining advantages from technology revolution: A patent roadmap for competition analysis and strategy planning", *Technological Forecasting and Social Change*, Vol.145(2019).

〈학위논문(국내 및 동양)〉

김요한, "국제특허분류를 이용한 유망기술 도출에 관한 연구", 홍익대학교, 석사, 2016.

〈연구보고서〉

국립전파연구원, "2025년 미래전파 기술수요 예측조사", 국립전파연구원, 2013.

주재욱 외 7인, "데이터 기반 디지털 경제 미래예측 방법론 연구", 경제·인문사회연구회, 2016.

특허청·한국지식재산연구원, "지식재산 경쟁력 및 특성지표 개발: 지식재산 포트폴리오의 특성 분석지표 개발", 한국지식재산연구원, 2012.

新エネルギー・産業技術総合開発機構, "次世代に期待される情報通信技術", 新エネルギー・産業技術総合開発機構, 2021.

## A Study on the Process of Identifying Emerging Technology Using Patent Data — A Joint Approach of Factor Analysis and Text Mining Methods —

Song, Kyoungtae; Bong, Kangho & Park, Jaemin

This study proposes a new process of identifying emerging technologies from a macro perspective using statistical and bibliographic data. Since most of the existing studies identify emerging technologies using only a single statistical data or bibliographic information such as number of patents, there is a limitation in that precision and prediction power is insufficient or costs considerable. The proposed process in this study differs from that of previous studies in that the main-group not subdivided into sub-group is set as the target of data collection, and the candidate group for emerging technology similar to those of the past is identified through factor analysis and correlation analysis. In this study, we analyze H-section(electric field) based on International Patent Classification(IPC). This study contributes to the methodology for identifying emerging technologies, proposing a new process that can increase the efficiency and objectivity by combining multiple analysis techniques.

Keyword

emerging technology, technology forecasting, patent analysis, IPC, principal component factor analysis, text mining