

A Study on Big Data Analysis of Related Patents in Smart Factories Using Topic Models and ChatGPT

Sang-Gook Kim · Minyoung Yun · Taehoon Kwon · Jung Sun Lim[†]

Korea Institute of Science and Technology Information

토픽 모형과 ChatGPT를 활용한 스마트팩토리 연관 특허 빅데이터 분석에 관한 연구

김상국 · 윤민영 · 권태훈 · 임정선[†]

한국과학기술정보연구원

In this study, we propose a novel approach to analyze big data related to patents in the field of smart factories, utilizing the Latent Dirichlet Allocation (LDA) topic modeling method and the generative artificial intelligence technology, ChatGPT. Our method includes extracting valuable insights from a large data-set of associated patents using LDA to identify latent topics and their corresponding patent documents. Additionally, we validate the suitability of the topics generated using generative AI technology and review the results with domain experts. We also employ the powerful big data analysis tool, KNIME, to preprocess and visualize the patent data, facilitating a better understanding of the global patent landscape and enabling a comparative analysis with the domestic patent environment. In order to explore quantitative and qualitative comparative advantages at this juncture, we have selected six indicators for conducting a quantitative analysis. Consequently, our approach allows us to explore the distinctive characteristics and investment directions of individual countries in the context of research and development and commercialization, based on a global-scale patent analysis in the field of smart factories.

We anticipate that our findings, based on the analysis of global patent data in the field of smart factories, will serve as vital guidance for determining individual countries' directions in research and development investment. Furthermore, we propose a novel utilization of ChatGPT as a tool for validating the suitability of selected topics for policy makers who must choose topics across various scientific and technological domains.

Keywords : Smart Factory, Patent Big Data Analysis, Latent Dirichlet Allocation, ChatGPT, Global Technology Development Trends

1. 서 론

현대 산업 환경은 끊임없는 변화와 혁신의 과정에 놓여 있으며, 이러한 환경에서 디지털 기술은 기업의 경쟁력을

형성하고 결정하는 주요 요소 중 하나로 부상하고 있다. 디지털 전환(Digital Transformation, DT)은 기업이 디지털 기술과 탄탄한 비즈니스 모델을 결합하여 새로운 가치를 창출하는 프로세스로, 이는 제품 디자인, 공학, 생산 및 수명주기 관리와 같은 다양한 산업 분야에서 혁신을 가져오고 있다. 특히, 스마트팩토리 분야에서의 DT는 더욱 중요하며, 글로벌 산업 경쟁의 핵심 요소로 주목받고 있다. 스

Received 26 September 2023; Finally Revised 20 October 2023;
Accepted 24 October 2023

[†] Corresponding Author : jsunnylim@kisti.re.kr

마트팩토리는 생산 및 공정을 자동화하고 최적화하기 위해 디지털 기술을 활용하는 개념으로, 제조업체들이 경쟁력을 유지하고 향상하는 데 필수적이다.

본 논문은 스마트팩토리 분야에서 DT의 연구 동향과 글로벌 기술 개발 추세, 그리고 국가 간의 차이를 이해하기 위한 새로운 접근 방법을 제안한다. 이 방법은 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 토픽 모형 분석 방법과 생성형 인공지능 기술인 ChatGPT를 결합하여 대규모 연관 특허 빅데이터를 분석하고, DT와 관련된 주요 주제를 도출한다. 본 연구에서는 LDA를 사용하여 연관 특허 문서에서 숨겨진 주제를 식별하고, 이를 통해 스마트팩토리 분야의 기술 발전 동향을 분석한다. 더 나아가, ChatGPT를 활용하여 도출된 주제의 의미를 추론하고 주제의 적합성을 검증하며, 전문가들의 검토를 통해 선별된 주제의 검증된 결과를 확정한다. 과학기술 문헌인 특허 문서를 대상으로 LDA를 분석할 때 토픽을 구성하는 단어들은 과학기술 분야의 전문용어를 구성하는 경우가 일반적이다. 이때 비전문가의 경우에 LDA의 토픽을 구성하고 있는 전문용어(단어)를 근거로 토픽의 의미를 이해하고 설명하는데 어려움이 있을 수 있으며, 본 연구에서는 이 같은 문제를 해결하기 위해 ChatGPT의 의미추론 엔진을 도입하였다. 또한 토픽에 대한 의미 추론 엔진으로써의 ChatGPT 활용 가능성을 검증하기 위해 스마트팩토리 분야를 대상으로 해당 전문가를 통해 토픽별 의미추론 결과를 검토하였다.

마지막으로 스마트팩토리 DT 관련 주요 주제들의 국가간 특허 환경과 특허 수준을 양적 및 질적 수준 측면에서 차이를 비교분석 한다. 양적 수준 분석을 위해 주제별로, 특허 출원 발생빈도, 피인용 수, 인용 수, 패밀리-특허 건수를 분석하였고, 질적 수준 분석을 위해 특허 청구항 수와 독립항수에 대한 비교분석을 수행하였다.

우리는 글로벌 특허 빅데이터를 대상으로 자료를 수집하기 위하여 GPASS(<https://gpas.kisti.re.kr>) 검색 시스템을 활용하였으며, LDA 분석 및 주제별 시각화를 위해 강력한 빅데이터 분석 도구인 Kontanz Information Miner (KNIME)를 활용하였다.

이 방법을 통해 스마트팩토리 분야에서의 국가별 기술 개발 주제의 수준을 이해하고 기술사업화 방향에 대한 특징을 탐색한다. 특히 스마트팩토리의 DT 분야에서 국가 기술 개발 주제의 차이를 확인하여 향후 기술사업화 가능성과 후속 기술 개발이 요구되는 분야 탐색을 수행한다. 결과적으로, 우리의 연구는 스마트팩토리 분야의 글로벌 특허 데이터 분석을 통해 해당 국가들의 DT 연구개발 전략 수립에 중요한 지침을 제공할 것으로 기대되며, 특히 과학기술 정책입안자나 기술 기획 전문가들에게 실무적인 측면에서 큰 기여가 예상된다.

2. 선행연구

2.1 특허데이터 기반 기술 추세 예측

Nam and Choi[16]는 자동차 반도체의 신기술 도출을 위한 연구를 통해 미국 등록 특허 문서를 수집하고 해당 특허의 IPC를 기반으로 기술 추세를 분석하였다. 이 연구에서는 자동차 반도체의 주요기술을 토픽모델링을 통해 분석하고, 토픽 모형을 통해 신기술과 같은 기술 경로를 조사하였다. Wang et al.[19]은 토픽 진화의 확률과정 분석을 통해 특허 텍스트로부터 기술 추세를 추적하고 예측하였다. 이를 통해 석탄 슬러리 파이프라인 운송 기술 개발 과정에서의 어려움을 이해하고 기술의 동향과 개발 특성을 파악하여 연구개발(R&D) 효율성과 지속 가능성을 향상시키고자 하였다. 결과적으로 이 연구에서는 잠재 디리클레 분포(Latent Dirichlet Distribution) 방법을 사용하여 특허 텍스트에서 잠재적인 기술 주제를 추출하고, 주제 간의 코사인 유사성을 계산하기 위해 Word2vec 기반의 주제 단어 벡터 모델을 적용하였다. Hidden Markov Model(HMM)을 도입하여 이중 확률과정을 묘사할 수 있는 HMM 기반의 토픽 진화 추세 모델을 구축하고 이를 사용하여 이 분야의 기술 진화 추세를 분석하고 예측하였다. Lee et al.[14]은 이상치를 사용하는 기계 학습 알고리즘을 기반으로 한 기술 추세 스크리닝 프레임워크를 제안하였다. 이 방법은 먼저, 데이터 세트를 시간별로 훈련 및 테스트 세트로 분할하여 Doc2Vec 모델을 훈련하고, 훈련된 모델을 사용하여 특허 문서를 전처리한다. IQR, 3-시그마 규칙 및 Isolation Forest 알고리즘을 사용하여 추출된 이상치 문서에 대한 투표(voting)를 통해 최종 이상치 문서를 선택한다. 마지막으로 토픽 모형을 통해 추출된 이상치 문서의 기술 주제를 식별한다. 이 연구에서는 드론에 관한 특허데이터를 분석하였으며, 그 결과 드론 관련 하드웨어와 시스템 기술에 관한 누적 연구에도 불구하고 자율 비행 분야에 대한 연구 부족이 있음을 확인하였다.

2.2 특허 및 논문 데이터 기반 기술 추세 예측

Wang et al.[20]은 특허 분석과 과학 논문 분석을 결합하여 기술 추세를 예측하고 기술 기회를 발견하기 위한 프레임워크를 제안하였다. 먼저, 텍스트마이닝을 사용하여 특허와 과학 논문에 포함된 지식을 추출한다. 그런 다음, 자연어 처리를 통해 데이터를 정리하고 코딩한다. LDA 토픽 모형을 사용하여 주제를 클러스터링하고 다차원 계층 구조 다이어그램을 작성한다. 결과를 비교분석하고 전문가 지식과 결합하여 기술 로드맵을 작성하여 미래 개발 추세와 기술 기회를 예측한다. 마지막으로 냉동 컨테이너 기술을

사용하여 유효성을 검증하였다. Park et al.[17]은 LDA 주제 모델링 기술을 사용하는 과학기술 추세 예측 시스템을 개발하기 위한 연구를 수행하였다. 이 연구에서는 미국의 인공지능(AI) 관련 특허에 관한 13,618개의 초록을 분석하여 결과를 AI 주제 내의 관련 단어의 빈도 변화에 기반하여 검증을 시도하였다. AI 주제의 추세 분석 결과, 11가지 기술 중 8가지가 일치하였으며, 이 연구는 간단하고 저비용의 기술 예측을 가능하게 하는 엔진 개발을 돕는 기술 예측 및 주제 모델링 기술에 대한 기초를 제공하였다.

2.3 전문가 견해 및 텍스트마이닝 기반 기술 추세 연구

특허와 과학 논문을 분석하는 데 텍스트마이닝 기법을 사용하여 기술 정보를 추출하는 것은 기술 추세 예측과 기술 기회 발견을 위해 중요하다.

기술 추세를 식별하는 것은 기업이 경쟁력을 갖고 미래 기술 추세를 파악하여 이를 이용하는 데 있어 중요한 성공 요소가 될 수 있다. 기업은 항상 미래 제품과 서비스를 위한 계획을 세우기 위해 기술 예측에 사용되는 도구를 찾는 것이 중요하다. Daim et al.[17]은 전문가 판단과 함께 텍스트마이닝 기술을 사용하여 Software as a Service(SaaS) 사례 연구에서 가까운 미래 기술 진화 추세를 감지하고 분석하였다. 이 사례에서는 장기적인 기술 개발 추세를 과학과 기술 사이의 격차를 분석함으로써 예측하였다. Li et al.[15]은 특허데이터를 사용하여 나라, 기관, 기술 분야 및 주제 관점에서 체계적으로 기술 개발 추세를 모니터링하였다. 그런 다음 데이터마이닝 접근법, 사회연결망분석을 적용하여 협력 네트워크 기관과 발명가를 탐색하였다. 또한 Order Analysis를 적용하여 보다 포괄적이고 객관적인 협력 관계, 파트너 및 중심성을 제시하였다. Ampornphan and Tongngam[2]는 데이터마이닝 방법과 사회연결망분석을 적용하여 특허데이터에서 신흥 기술에 대한 지식을 얻고 정보성 있는 기술 추세를 찾고자 했다. 이러한 기법은 K-means 클러스터링, 텍스트마이닝 및 연관 규칙 마이닝 방법을 포함하고 있으며, 분석된 특허데이터에는 국제 특허 분류(IPC) 코드와 특허 제목이 포함된다. 연관 규칙 마이닝은 특허데이터 사이의 연관 관계를 찾기 위해 적용되었고, 이후 사회연결망분석(SNA)과 결합하여 기술 추세를 더 자세히 분석하였다. SNA는 가장 영향력 있는 기술을 탐색하고 다양한 네트워크 레이아웃에서 데이터를 시각화하는 데 사용되었으며, 결과는 신흥 기술 클러스터, 의미 있는 패턴 및 네트워크 구조를 보여주었으며 기술 및 발명의 발전에 대한 정보를 제안하였다. Govindarajan et al.[8]은 복잡한 제조 공정에서의 인간 중심 사이버물리시스템을 위한 몰입형 기술에 대한 포괄적인 개요를 제공하였다. 몰

입적 세계의 사용자가 몰입감을 느끼며 사이버 세계와 연결될 수 있게 하는 가상 현실, 증강 현실, 뇌-컴퓨터 인터페이스 및 뇌-기계 인터페이스와 같은 복잡한 시스템들이 제조 시스템 개선의 잠재력을 갖고 나타나고 있다는 것을 밝혔다. Industry 4.0은 지능적인 제조를 촉진하기 위한 제4차 산업 혁명을 위한 모든 기술, 표준 및 프레임워크를 포함하고 있고, 산업용 몰입형 기술은 Industry 4.0의 인간-기계 인터페이스 컨텍스트에서 스마트 제조 혁신을 위해 사용될 것으로 예측하였다. 이 연구는 문헌의 철저한 검토, 도메인 온톨로지 구축, 특허 메타 추세 통계 분석 및 기술 기능 매트릭스를 활용한 데이터마이닝 분석을 제공하였으며, LDA 모델을 사용하여 기술 및 기능 개발 추세를 강조하였다. IEEE 및 IET 데이터베이스에서 얻은 179개 참고 자료와 2,672개 특허를 체계적으로 분석하여 현재의 동향을 파악하였다. 이 연구는 복잡한 제조 공정에서의 고급 인간 중심 사이버물리시스템 개발을 위한 필수적인 기초를 제공하고 있다. Jung and Chang[10]은 온라인 뉴스 기사에서 스마트팩토리와 관련된 추세와 이슈를 분석하기 위해 텍스트마이닝을 활용하였으며, 이를 통해 기술 중심 주제의 이동과 정책 중심 기사의 감소하는 경향을 확인하였다.

2.4 토플모델링 기반 기술 추세 연구

Lee et al.[13]은 고급 제조에서의 DT를 탐구하며, 제품 디자인, 공학, 생산 및 수명주기 관리에 미치는 영향을 분석하였다. 이 연구에서는 기계 학습 기반의 토플모델링을 활용하여 99편의 논문을 분석하고, DT와 관련된 여섯 가지 주요 주제를 식별하였다. Yang et al.[22]은 스마트팩토리의 연구 동향을 해외와 국내를 대표하는 사례를 조사하면서, 잠재 의미 분석(LSA)을 활용하여 주 연구 방향을 확인하고 한국과 국제 연구 동향을 비교분석하였다. Abdirad and Krishnan[1]은 공급망 관리(SCM)에서의 Industry 4.0의 역할에 중점을 두며, 체계적인 문헌 고찰을 수행하였다. 이때 연구를 탐색적 대 확증적, 질적 대 양적, 관리 수준 대 공정/기술 수준으로 분류하고 토플모델링(topic modeling) 기술을 활용하여 주제 군집(topic cluster)을 추출하였다. Cao and Fei-Fei[4]은 한국의 스마트팩토리 연구 논문을 중점적으로 다루며, LSA를 활용하여 스마트팩토리 연구 동향을 다섯 가지 주제로 분류하였다. 이를 통해 연구 경향을 파악하고 연구 지형에 대한 통찰을 제공하였다. Wang and Hsu[17]는 스마트 제조 분야에서 기술 동향을 탐구하기 위해 특허데이터를 활용한 주제 기반의 특허 분석 방법을 제시하였다. 스마트 제조가 제조 산업의 생산 기술과 비즈니스 모델에 혁명적인 변화를 가져올 수 있으며, 이 연구는 스마트 제조 분야에서의 기술적 주제를 식

별하고 기술 동향을 조사하며 잠재적 기회를 탐구하였다. 여기서는 LDA 주제 모델링 기술을 활용하여 잠재적인 기술 주제를 추출하고, 일반화된 선형 혼합 모델(Generalized Linear Mixed Model, GLMM)을 사용하여 주제의 상대적 등장 수준을 분석하였다. 또한 주제 가치 및 주제 경쟁 분석을 개발하여 각 주제의 잠재적 가치를 평가하고 경쟁 기업의 기술적 위치를 파악하였다. 이 연구 결과에서는 수집한 특허데이터에서 총 14개의 주제를 추출하였으며, 스마트 연결, 사이버-물리 시스템(CPS), 제조 데이터 분석 및 분말 침착성 가공 제조 등 몇 가지 빠른 성장과 고가치 주제가 식별되었다. 또한, 일부 선도 기업은 다양한 기술 주제에 관한 광범위한 연구 및 개발을 수행하고 있으며, 다른 기업은 몇 가지 기술 주제에 집중하고 있음을 확인하였다. 이 연구에서 개발된 방법론은 기업들이 스마트 제조 분야에서 중요한 기술 주제를 식별하여 연구 및 개발 투자 결정을 내릴 수 있도록 도움을 줄 것으로 기대하였고, 기업은 주제의 등장 위치에 따라 적절한 기술 전략을 선택할 수 있다. 이 연구는 또한 기술적 기회를 탐구하기 위한 다중 가치 평가 프레임워크를 제공하여 스마트 제조 기술 랜드스케이프를 보다 견고하게 이해하는 데 유용한 정보를 제공하였다. Choi and Song[6]은 미국 특허청(USPTO)에서 등록된 물류 관련 특허를 사용하여 주제 모델링 기반 접근 방법의 기술 추세 분석 방법을 제안하였다. 제안된 접근 방법의 핵심은 LDA이며 특허 내용의 기술 주제를 식별한다. LDA에 의해 식별된 주제는 파일 수준 및 기업 수준의 추세에 대해 추가 조사하였다. Kwon et al.[12]은 물류산업에서 디지털화를 위한 기술 개발 전략을 수립하기 위하여 특허데이터를 사용하여 기술 전략 수립 시 기회와 영역을 식별하고자 하였다. 이전 연구에서는 주로 전문가 인터뷰 방법에 의존하였으며, 토픽모델링을 기반으로 한 특허 분석 연구도 기술 추세 파악만을 위해 사용되었다. 이 연구는 기술 추세 파악 단계로의 확장을 제안하는 새로운 프레임워크를 제안하였으며, Word2Vec 알고리즘을 사용하여 추세를 반영하는 특허 검색식, 특허데이터의 LDA 클러스터링을 통한 물류 기술의 변화 예측, 실험 방법을 통한 공백 기술 도출을 조사하였다. 제안된 프레임워크는 물류 기업에서 기술 변화를 예측하고 유망한 기술을 도출하기 위한 목적으로 사용될 것으로 예상하였다. Kim et al.[11]은 의미론적 특허 주제 분석을 기반으로 한 특허 개발지도(PDM) 생성을 제안하였다. 이 방법은 1) 특허를 수집하고 전처리하는 단계, 2) 각 특허를 용어 벡터로 구조화하는 단계, 3) LDA를 적용하여 특허의 기술 분류를 식별하는 단계, 마지막으로 4) 의미론적 특허 유사성과 인용을 기반으로 한 민감성 분석을 통해 특허 간의 개발 경로를 시각화하는 단계로 구성된다. 이 방법은 3D 프린팅 기술과 관련된 특허를 사용하여, PDM 생성의 정량화에

기여하면서 특허 간의 기술 내용에 대한 연속성 관계와 각 특허의 기술 분류에 대한 분류 방법을 제안하였다.

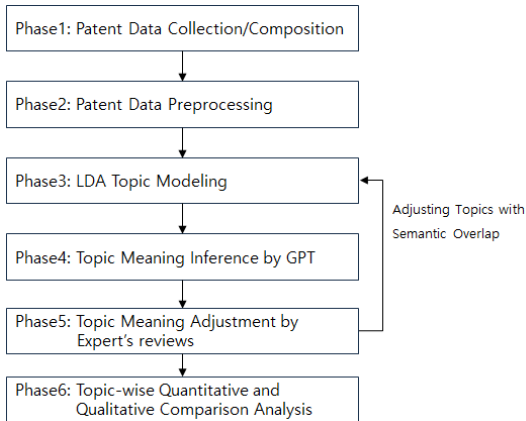
이처럼 기술 추세를 예측하기 위해, 특허 문헌과 과학기술 문헌 등 다양한 증거 데이터를 활용하여 전문가 의견과 통계적 방법, 그리고 텍스트마이닝 기법에 이르기까지 다양한 기법을 활용하여 기술 추세를 탐색해 왔다. 특히 최근 연구 동향은 텍스트 데이터를 대상으로 데이터마이닝이 가능한 토픽모델링을 주도적으로 활용해 왔다. 하지만 토픽모델링에서 최적의 토픽 수를 결정하는 방식의 결과물들은 관련 주제의 의미를 해석하고 이해하는 데 상당히 어려움이 있는 게 사실이다. 특히 핵심어 중심의 토픽 분석 결과가 기술 추세 분석에 있어서 해당 기술 분야의 비전문가가 이해하는 데에는 많은 어려움이 있는 게 사실이다. 본 연구에서는 이러한 문제를 해결하기 위하여 생성형 인공지능 ChatGPT 모형을 활용하였으며, 이 모형을 통해 산출된 최적의 토픽에 대한 의미추론과 토픽 간의 중복적인 의미해석의 가능성이 있지 않은지에 대한 검증 수단으로써 활용하였다. 마지막으로 해당 분야의 전문가 견해를 통해 최종적인 주제 수를 확정하고, 이 주제에 대한 의미를 부여할 수 있도록 보조하였다.

3. 연구방법론

3.1 연구방법론 설계

글로벌 기술 추세를 탐색하기 위해 토픽 모형과 생성형 인공지능 추론 모형 활용을 위한 연관 특허 빅데이터 분석 체계를 아래 <Figure 1>과 같이 설계하였다. 먼저 글로벌 특허 자료를 수집하고 분석을 위한 데이터를 구성하기 위하여 특허 검색 시스템을 선정하였으며, 한국과학기술정보연구원에서 구축하여 운용 및 관리 중인 ‘글로벌 특허 분석 서비스 시스템’(GPASS)을 활용하였으며, 검색 적용 기간은 최근 6년(2017~2023)을 반영하였다. GPASS는 1940년부터 현재까지 101개 관계국의 특허청 자료를 구축 운용하고 있으며, 2023년 10월 기준 147,705,746건의 특허 빅데이터를 보유하고 있으며, 현재까지도 매 2주 마다 특허 자료를 갱신하고 있다.

스마트팩토리 분야 연관 특허를 검색하기 위하여 검색어와 검색식을 작성하고, 특허 검색 필드 중에서 특허 제목, 초록, 특허 핵심어를 대상으로 데이터를 구성한다. 특허데이터의 정제 및 처리를 위해 구두점, 숫자, 불용어, POS Lemmatizing 등 정제를 위한 사전 처리를 수행하여 LDA 토픽 모형 분석을 위한 데이터 구성을 완료한다.



<Figure 1> Research Framework

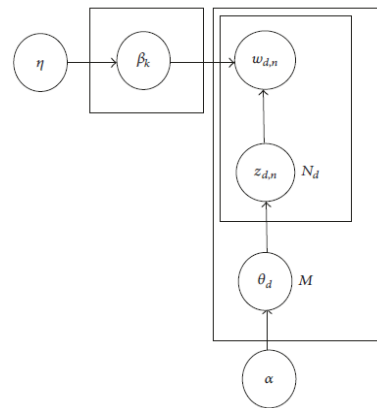
텍스트마이닝 기법 중 가장 흔히 사용되는 토픽모델링(topic modelling)은 대량의 문서(document) 집단에서 잠재적으로 의미가 존재하는 토픽(topic)을 찾아내는 방법론이다. 결과적으로, 문서는 토픽의 혼합물이고 토픽은 단어(word)의 혼합물이기때문에, 이러한 문서들의 비지도 분류를 위한 방법으로 사용된다. LDA는 초기 텍스트 데이터 모델링을 위해 소개되었으며, 문서 내의 단어를 활용한다. 생성된 토픽은 문서 내 단어의 확률분포를 기반으로 하며, 전통적인 단어 빈도와 비교하여 생성된다. 이 모델의 가정은 단어가 강력한 의미 정보를 가지고 있고, 유사한 주제를 다루는 문서는 비슷한 단어 그룹을 사용할 것이라는 사실이다. 따라서 문서는 숨겨진 토픽에 대한 확률분포로 간주하며, 토픽은 단어에 대한 확률분포로 간주한다.

<그림 2>에서처럼, LDA를 수행할 때 문서집단에 k개의 토픽이 존재할지 사용자가 가정해야 한다. 이후 각 문서집

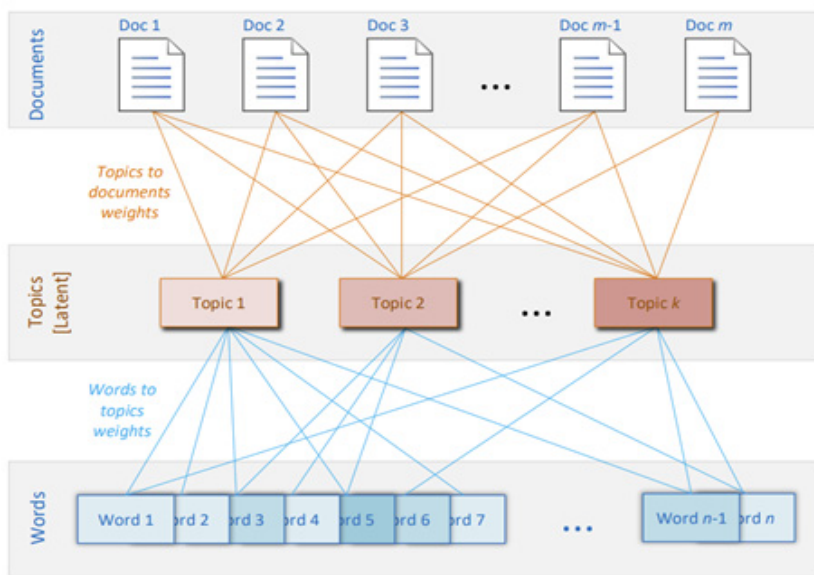
단에 K개의 토픽별 발생확률이 존재하게 되고, 이러한 각 토픽을 구성하는 단어가 확률적으로 존재하게 된다.

결과적으로 각 토픽을 구성하는 단어의 발생확률을 근거로, 토픽에 대한 해석과 설명을 할 수 있다. 하지만 과학기술 문헌의 경우에 토픽을 구성하는 단어, 즉 과학기술 전문용어의 발생빈도를 토대로 일반인이 해당 토픽에 대한 해석과 설명이 어려울 수 있다.

본 연구에서 각 토픽은 어휘 집합 V 내의 단어로 특성화되며, 각 단어별 토픽 분포로 표현된다. 토픽들의 혼합물은 특허 문서를 나타내는 데 사용되며, 각 토픽별 특허 단어분포를 이용한다. 어휘 집합 V에 |V|개의 단어가 있다고 가정하고, 특허 문서 D 내에 K개의 토픽과 M개의 특허 문서가 있다고 가정한다. $N_{d=1, \dots, M}$ 특허 문서 D 내에 d번째 특허의 단어 수를 나타내며, 디리클레 분포의 사전 매개변수 벡터는 알파(α)와 에타(η)이다.



<Figure 3> LDA Flowchart



<Figure 2> Relationship in LDA Model: Documents, Topics, Words

토픽별 특허 분포에서 $z_{d,n=1,\dots,N_d} \in \{1,\dots,K\}$ 는 d번째 특허 내의 n번째 단어가 속할 수 있는 토픽을 나타내며, $z_{d,n}$ 는 디리클레($\theta_d; \alpha$)로 부터 추출된 매개변수 벡터 $\theta_d \in (0,1)^K$ 를 따르는 다항분포 확률변수이다. 단어 구조가 주제를 나타내므로, 단어별 토픽 분포가 필요하다. 각 단어별 토픽 분포를 위해, $w_{d,n=1,\dots,N_d} \in V$ 를 d번째 특허 문서의 제 n번째 단어로 가정하고, 단어 $w_{d,n}$ 은 특정 토픽인 $z_{d,n}$ 에 할당될 수 있다. 토픽은 단어로 구성되므로, $w_{d,n} | z_{d,n}$ 를 다항분포 ($\beta_{z_{d,n}=k}$)를 따르는 확률변수로 가정하며, 여기서 $\beta_k \in (0,1)^{|V|}$ 는 디리클레 ($\beta_{k,\eta}$) 분포에서 추출된다. 단어별 토픽 및 토픽별 특허 분포의 결합분포는 아래와 같이 설명된다.

$$P(w_d, z_d, \theta_d, B; \alpha, \eta) = \prod_{j=1}^{N_d} \prod_{k=1}^K P(w_{d,j} | \beta_{z_{d,j}=k}) \text{Dirichlet}(\beta_k; \eta) P(z_{d,j} | \theta_d) \text{Dirichlet}(\theta_d; \alpha) \quad (1)$$

여기서 B는 $K \times |V|$ 차원의 행렬로, 행벡터는 $\beta_{k=1,\dots,K}$ 에 의해 구성된다. 더 나아가, $\theta_1, \dots, \theta_M$ 과 B를 통합함으로써, (1)의 결합분포는 $(w, z; \alpha, \eta)$ 로 유도된다. 그런 다음 사전 매개변수 벡터 알파(α)와 에타(η)는 경험적인 값을 사용하여 근사화하며, 최종 특허 토픽 할당이 결정된다. 예를 들어, N 개의 단어를 가진 특허에 잠재 토픽 k가 포함되어 있는지를 결정하기 위해 식 (2)가 사용된다. $\epsilon \in [0,1]$ 범위 내의 모든 ϵ 에 대해, 다음과 같이 정의된다.

$$I_{z=k}(w) = \begin{cases} 1, & \sum_{n=1}^N P(w_n, z_n = k; \alpha, \eta) > \epsilon \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

LDA 토픽 모형의 파라미터는 토픽 수(k), 토픽에 관해 설명할 핵심어(keyword) 개수, 알파(α), 그리고 에타(η)가 존재한다. 여기서 알파는 문서에서 각 토픽의 이전 가중치를 설정하며, 에타는 토픽에서 각 핵심어의 이전 가중치를 설정하게 된다. 작은 알파 값(0.1)은 희박한 토픽 분포를 생성하며, 각 문서에 대해 덜 두드러진 토픽을 나타낸다. 또한 작은 에타 값(0.001)은 희박한 핵심어(keyword) 분포를 생성하며, 이것은 각 토픽을 설명하는데 덜 두드러진 핵심어를 말한다. 최적의 사용자 정의 파라미터(hyper parameter)를 선택하기 위한 경험적인 레시피는 기본적으로 존재하지 않으며, 가시화를 위해서는 토픽 수를 관리할 수 있어야 한다(1~100개 사이). 경험적으로 알파 값은 50/토픽 수로 결정할 수 있으며, 본 연구에서는 이를 근거로 활용하였으며, 토픽에 관해 설명할 핵심어 수를 20개, 에타 값은 0.01을 적용하였다. 최적의 토픽 수를 결정하기 위해서는 Griffiths and Steyvers [9]가 제안한 perplexity index를

활용하며, 이때 아래 식 (3)을 근거로 line plot 분석을 수행한다.

$$2^{-\log P(w | T)} \text{ for each } k=2, \dots, 80 \quad (3)$$

ChatGPT ver 4.0을 활용하여 앞서 LDA 분석을 통해 도출된 주제별 의미를 추론하고, 중복적인 의미가 발생하면 최적 주제 수(k)를 조정하는 단계를 수행하여 LDA 토픽 모형 재분석을 수행한다. 이렇게 정리된 주제들을 대상으로 다시 전문가 의견을 반영하여 주제별 추론 의미의 적절성을 검토한다. 최종 확정된 주제들을 대상으로 국가별 양적 및 질적 수준을 비교 분석한다. 이때 토픽별 등록 특허의 발생빈도, 인용(backward citation) 수/피인용(forward citation) 수, 패밀리특허 건수, 청구항 수, 독립항 수의 지표를 활용하여 국가별로 수준 차이를 비교 검토한다.

3.2 특허 분석 데이터 구성

스마트팩토리 연관 특허를 검색하기 위해서 한국과학기술정보연구원 (KISTI)에서 제공하고 있는 GPASS 검색 시스템을 활용하였다. 이때 검색 기간은 2017년부터 2023년까지 최근 6년 동안에 등록된 글로벌 특허데이터를 대상으로 하였다. 연관 특허 검색을 위한 핵심어와 검색식은 아래 <Table 1>과 같이 구성하였다.

<Table 1> The Search Keywords and Search Query

Type	Description
Keywords	smart factory, smart-factory, smart manufacturing, smart-manufacturing, smart manufacture, smart-manufacture, smart production, smart-production, intelligent manufacturing, intelligent-manufacturing, digital factory, digital-factory
Search query	AU=(CA OR CN OR DE OR EP OR FR OR GB OR JP OR KR OR US OR WO) MF_ALIVE=(Y) DTYPE=(P) PUBL_TYPE=(A OR G) (TI=("smart factory" OR smart-factory OR "smart manufacturing" OR smart-manufacturing OR "smart manufacture" OR smart-manufacture OR "smart production" OR smart-production OR "intelligent manufacturing" OR intelligent-manufacturing OR "digital factory" OR "digital-factory") OR AB=("smart factory" OR smart-factory OR "smart manufacturing" OR smart-manufacturing OR "smart manufacture" OR smart-manufacture OR "smart production" OR smart-production OR "intelligent manufacturing" OR intelligent-manufacturing OR "digital factory" OR "digital-factory"))

위의 검색어 및 검색식을 근거로 스마트팩토리 부문의 글로벌 연관 특허를 검색한 결과, 총 4,353건이 분석되었으며, 미확인 1건이 포함되어 있어 이를 제외하였다. 주요 국가별로 검색 건수는 아래 <Table 2>와 같다.

<Table 2> The Count and Relative Ratio by Authority

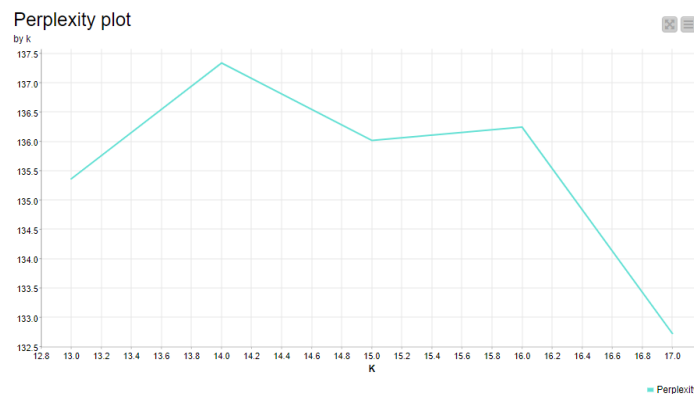
Authority	Name	Search results	Relative ratio
CN	China	3,928	90.3%
JP	Japan	3	0.1%
KR	South Korea	335	7.7%
US	United States	36	0.8%
WO	WIPO	50	1.1%

3.3 특허 분석 데이터 전처리

특허 문헌의 텍스트 정보를 전처리하기 위해, KNIME에서 제공하는 다양한 분석 노드(node)를 활용하였으며, String to Document(OpenNLP English Word Tokenizer 포함), Punctuation Erasure, Number Filter, N Chars Filter, Stop Word Filter, Stanford Tagger(Tagger model: English left 3 words 사용), Standford Lemmatizer, Case Converter 노드를 사용하였다.

Row ID	Document
Row1	"accommodation rack facilitate protection component smart manufacturing
Row2	"edge device interworking heterogeneous legacy protocols smart factory
Row3	"identification system automatic transplantation vertical smart farm
Row4	"system monitoring smart factory technology disclose smart factory
Row5	"system smart learn factory invention relate smart learning factory smr
Row6	"manage system heterogenous types multi mobile manipulators smart factory
Row7	"unmanned feeding intelligent device flat plate type workpiece inver
Row8	"intelligent manufacturing flexible production line packaging line inventi
Row9	"dimensional complex element form system method base series parallel robot
Row10	"automatic drilling device base intelligsystem manufacturing invention di
Row11	"fixture tool system invention relate technical field fixture particular fixture
Row12	"robot intelligent manufacturing method invention relate robot intelligent me
Row13	"quarter bend intelligent clamp rotary cut device invention belong techni
Row14	"compressor housing intelligent manufacturing seam automatic detection system

<Figure 4> The Case of Patent Document Preprocessing



<Figure 5> The Result of Perplexity Plot

<Table 3> The 1st Result of Keywords for 15 Optimal Topics

Topic id	Keywords
topic_00	production, line, intelligent, platform, system, manufacturing, design, training, simulation, technology, invention, teaching, process, method, practical, control, cost, industrial, comprise, digital
topic_01	datum, manufacturing, system, intelligent, module, platform, digital, service, management, cloud, information, equipment, industrial, method, model, base, layer, comprise, invention, acquisition

3.4 LDA 토픽 모형 분석

KNIME 도구에서 LDA 토픽모델링을 분석하기 위해서는 1) 데이터 입력, 2) 데이터 사전 처리, 3) 토픽 탐색, 4) 토픽별 핵심어 그룹핑, 마지막으로 5) 가시화 단계를 수행한다. KNIME에서 제공하는 Topic-Extractor (Parallel LDA) 노드를 사용하여 사전 처리된 특허 문서의 주제 목록을 작성한다. 이때 최적의 주제 수를 결정하기 위하여 LDA 사용자 정의 파라미터인 알파 값은 50/주제 개수, 에타(베타)값은 0.01, 주제 수는 2개에서부터 80개 범위를 선정하여 perplexity index를 적용해 식 (3)을 근거로 line plot을 수행하였다.

따라서 본 연구에서는 최적의 토픽 수를 15개로 선정하고, 토픽별 핵심어를 20개 썩 도출하였다. 1차로 분석된 최적의 토픽 15개의 결과는 아래 <Table 3>과 같이 도출되었다.

3.5 GPT 기반 LDA 주제의 의미추론

앞서 분석된 최적의 토픽 15개에 대한 핵심어를 근거로, 생성형 인공지능 ChatGPT ver 4.0 (유료 버전)에게 다음과 같은 질의를 던져 토픽별 의미를 추론하였다.

질의예시: “스마트팩토리 분야에서 [주제별 핵심어 리스트]와 같은 핵심어가 동일 주제로 분석되었다. 이 주제에 대한 제목과 해설을 작성해주세요.”

상기 질의 내용을 15개의 주제를 대상으로 의미 추론한 결과, topic_00과 topic_01이 동일 의미로 추론되어 토픽 수 14개로 다시 수정하여 LDA 토픽모델링 분석을 수행하였으며, 그 결과 14개 토픽별 핵심어는 아래 <Table 4>와 같이 도출되었다.

<Table 3> The 1st Result of Keywords for 15 Optimal Topics(Continued)

Topic id	Keywords
topic_02	factory, smart, datum, system, information, unit, device, server, invention, management, iot, process, provide, facility, base, method, include, collect, sensor, receive
topic_03	cut, manufacturing, spray, device, wire, intelligent, steel, assembly, cleaning, mechanism, equipment, roller, wind, pipe, paint, bend, dry, cloth, comprise, machine
topic_04	control, module, system, unit, intelligent, manufacturing, device, power, signal, comprise, processing, invention, connected, vehicle, detection, display, datum, terminal, information, controller
topic_05	device, body, air, box, intelligent, water, manufacturing, liquid, heat, pipe, filter, treatment, material, storage, valve, invention, arrange, seal, dust, gas
topic_06	rod, connected, intelligent, manufacturing, fixedly, block, base, plate, air, drive, move, body, wheel, device, magnetic, rotate, arrange, spring, shell, relate
topic_07	plate, device, arrange, manufacturing, intelligent, support, fix, frame, mechanism, mount, base, top, bottom, fixedly, connect, body, slide, upper, box, install
topic_08	machine, robot, welding, machining, tool, arm, mechanical, intelligent, manufacturing, position, workpiece, control, printing, clamp, laser, comprise, base, industrial, device, head
topic_09	production, product, manufacturing, process, system, management, intelligent, information, method, material, equipment, module, workshop, datum, accord, quality, time, monitoring, invention, scheduling
topic_10	rod, rotate, connect, manufacturing, block, connected, plate, intelligent, drive, shaft, device, fixedly, gear, wall, slide, movable, shell, inner, outer, movably
topic_11	assembly, convey, device, mechanism, material, intelligent, automatic, manufacturing, arrange, feed, feeding, production, assemble, comprise, belt, equipment, packaging, transfer, product, invention
topic_12	une, pcb, material, composite, des, aluminum, invention, pour, preparation, process, method, metal, syst, alloy, hub, les, communication, cell, intelligente, dispositif
topic_13	method, model, datum, manufacturing, intelligent, base, obtain, process, accord, step, detection, evaluation, image, invention, fault, feature, target, parameter, equipment, time
topic_14	equipment, manufacturing, intelligent, main, rotate, grind, drive, mold, metal, press, piece, block, body, move, rod, push, polish, die, plastic, speed

<Table 4> The 2nd Result of Keywords for 14 Optimal Topics

Topic id	Keywords
topic_00	device, cut, power, board, wire, manufacturing, circuit, intelligent, wind, chip, electric, automatic, bend, cabinet, alarm, lead, invention, comprise, mechanism, control
topic_01	material, welding, manufacturing, steel, packaging, intelligent, device, raw, automatic, phone, film, invention, process, bin, glue, form, control, field, mobile, feed
topic_02	factory, smart, datum, system, information, unit, device, process, management, server, base, provide, invention, method, iot, facility, control, include, receive, sensor
topic_03	spray, printing, grind, manufacturing, intelligent, mold, device, metal, paint, form, die, polish, surface, dry, coating, screening, equipment, plastic, invention, molding
topic_04	robot, system, unit, control, intelligent, manufacturing, platform, industrial, vehicle, station, training, teaching, invention, agv, comprise, plc, line, practical, warehouse, provide
topic_05	production, product, line, process, intelligent, manufacturing, material, system, method, accord, invention, automatic, quality, improve, efficiency, cost, design, battery, comprise, scheduling
topic_06	method, datum, model, manufacturing, intelligent, base, process, information, equipment, digital, accord, obtain, time, invention, step, system, parameter, evaluation, fault, feature
topic_07	une, glass, pcb, des, pour, communication, syst, les, intelligente, dispositif, invention, sur, donn, fabrication, est, smart, proc, device, tin, usine
topic_08	equipment, manufacturing, intelligent, main, body, drive, rotate, rod, block, protection, move, mount, spring, prevent, connected, field, force, technical, movably, push
topic_09	module, system, intelligent, datum, manufacturing, management, platform, control, information, equipment, comprise, industrial, acquisition, cloud, service, terminal, invention, processing, operation, base
topic_10	mechanism, assembly, device, arrange, intelligent, manufacturing, convey, plate, frame, equipment, comprise, clamp, support, base, table, mount, drive, automatic, move, feed
topic_11	detection, machine, tool, machining, control, device, intelligent, position, detect, manufacturing, camera, workpiece, inspection, image, invention, sensor, laser, method, comprise, numerical
topic_12	rod, plate, manufacturing, intelligent, connected, fixedly, connect, block, rotate, slide, device, drive, fix, base, shaft, arrange, support, wall, gear, surface
topic_13	air, body, box, device, intelligent, water, pipe, cleaning, manufacturing, arrange, liquid, dust, filter, heat, storage, bag, comprise, invention, cavity, valve

상기 14개의 토픽을 대상으로 ChatGPT를 활용하여 의미를 다시 추론한 결과, 아래 <Table 5>와 같이 분석되었다.

‘기계 구성요소를 활용한 제품의 조립 및 제조를 위한 지능형 자동화 시스템의 설계 및 구현’ 기술 분야가 가장 많은 866건으로 확인되었고, ‘제품을 조립하고 제조하기 위한 지능형 자동화 시스템의 설계 및 구현’ 608건, ‘제조 공정을 관리하고 제어하기 위한 지능형 자동화 시스템의 설계 및 구현’ 482건, ‘생산 공정을 최적화하고 제품 품질을 향상시키기 위한 데이터 기반 접근 방식 및 지능형 제조 시스템의 사용’ 433건, ‘IoT 기술 및 데이터 기반 의사

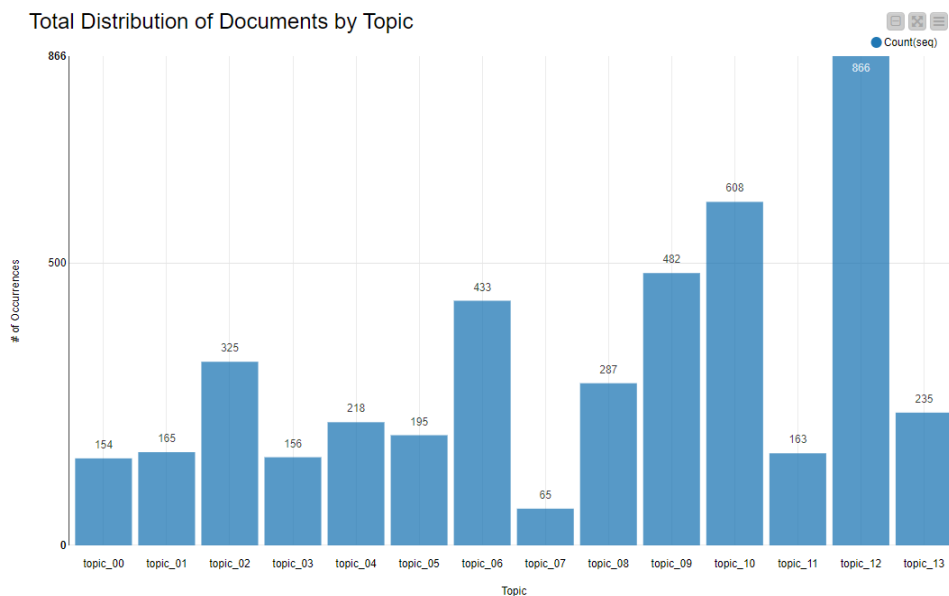
결정’ 325건 순으로 분석되었다.

3.6 전문가 견해 기반 주제의 검토 및 조정

앞서 분석된 스마트팩토리 분야의 14개 기술 주제별 제목과 내용에 대해 전문가 3인(연구계 1인, 학계 2인)을 대상으로 검토를 의뢰하였다. 그 결과 topic_01의 경우에 “모바일 장치 관련 산업”으로 의미를 제한하는 것이 더 적절하다는 의견과 topic_03의 경우에 “기초 및 원천 부문”으로 의미 제한을, topic_04에서는 “생산 공정 최적화”란 용

<Table 5> The 2nd Topic Semantic Inference Result Based on ChatGPT ver 4.0

Topic id	Topic	Counts
topic_00	Design and implementation of intelligent automation manufacturing systems and devices for electronic component production and assembly.	154 (3.5%)
topic_01	Design and implementation of automated manufacturing systems and equipment for material processing and packaging.	165 (3.8%)
topic_02	IoT (Internet of Things) technology and data-based decision-making.	325 (7.5%)
topic_03	Design and implementation of advanced manufacturing processes and equipment for material processing and finishing.	156 (3.6%)
topic_04	Design and implementation of intelligent automation manufacturing systems and equipment for optimizing production processes.	218 (5.0%)
topic_05	Design and implementation of intelligent automation manufacturing systems and processes for optimizing production efficiency, enhancing product quality, and reducing costs.	195 (4.5%)
topic_06	Use of data-driven approaches and intelligent manufacturing systems to optimize production processes and improve product quality.	433 (9.9%)
topic_07	Development and implementation of advanced manufacturing technology and systems.	65 (1.5%)
topic_08	Design and implementation of intelligent and automated manufacturing equipment.	287 (6.6%)
topic_09	Design and implementation of intelligent automation systems for managing and controlling manufacturing processes.	482 (11.1%)
topic_10	Design and implementation of intelligent automation systems for assembling and manufacturing products.	608 (14.0%)
topic_11	Design and implementation of intelligent automation systems for detecting and controlling machining processes.	163 (3.7%)
topic_12	Design and implementation of intelligent automation systems for assembling and manufacturing products using mechanical components.	866 (19.9%)
topic_13	Design and implementation of intelligent automation systems for managing and controlling fluids and gases.	235 (5.4%)



<Figure 6> The Total Distribution of Documents by Topic

어보다는 “생산 공정 유연화”가 타당하다는 의견을 접수하였다. 이외에도 topic_05에서는 “제품 관점”의 의미 제한을, topic_06에서는 “설비 관점”의 의미 제한이 적절하며, topic_10과 topic_12는 유사한 주제로 판단된다는 의견을 접수하였다. 결과적으로 상기 의견을 수렴하여 최종적으로 13개의 주제로 분류하였으며 (topic_10과 topic_12 통합), 의미를 제약하는 형태로 주제를 조정하였으며 그 결과는 아래 <Table 6>과 같다.

결과적으로, ‘기계 구성요소를 활용한 제품 조립 및 제조 지능형 자동화 시스템 개발’ 기술 분야가 가장 많은 1,474건으로 확인되었고, ‘제조 공정 관리 및 제어를 위한

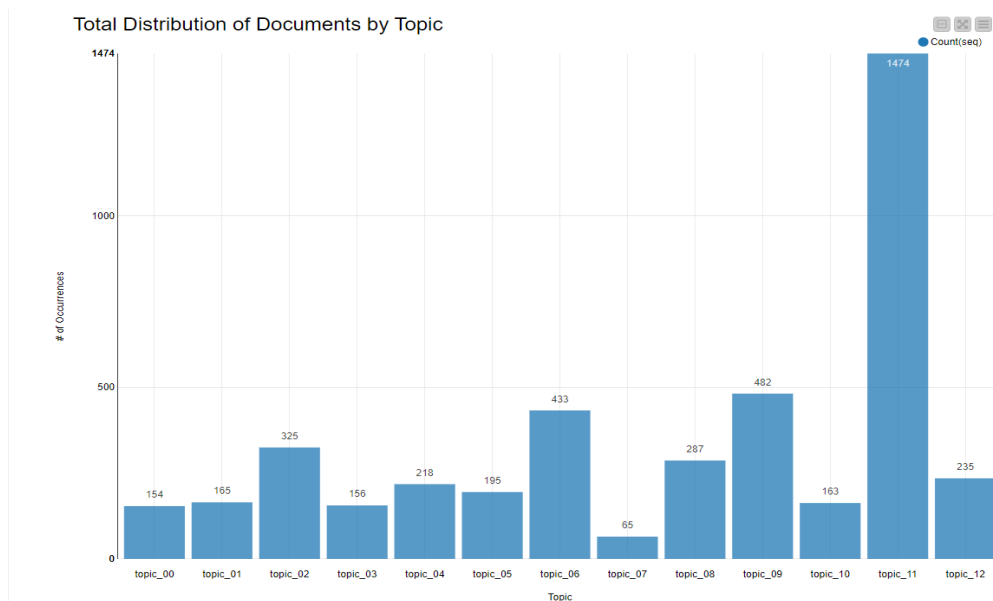
지능형 자동화 시스템 개발’ 482건, ‘설비 관점에서 생산 공정 최적화와 제품 품질 향상을 위한 데이터 기반 해결 방법 및 지능형 제조 시스템 사용’ 433건, ‘IoT 기술 및 데이터 기반 의사 결정’ 325건, ‘지능형 자동화 제조 장비 개발’ 287건 순으로 분석되었다.

4. 스마트팩토리 연관 특허 분석 결과

최종적으로 스마트팩토리 DT 분야의 기술 주제별로 주요 국가 간의 양적 및 질적 수준을 비교 분석하였다. 본

<Table 6> The Final Topic Semantic by Experts' Reviews

Topic id	Topic	Counts
topic_00	Development of Intelligent Automation Manufacturing Systems and Devices for Electronic Component Assembly	154 (3.5%)
topic_01	Development of Automation Manufacturing Systems and Equipment for Material Handling and Packaging in the Mobile Device Industry	165 (3.8%)
topic_02	IoT Technology and Data-Based Decision-Making	325 (7.5%)
topic_03	Development of Advanced Manufacturing Processes and Equipment for Material Processing and Finishing in the Basic and Foundational Sectors	156 (3.6%)
topic_04	Development of Intelligent Automation Manufacturing Systems and Equipment for Production Process Flexibility	218 (5.0%)
topic_05	Development of Intelligent Automation Manufacturing Systems and Processes for Quality Improvement and Cost Reduction from a Product Perspective	195 (4.5%)
topic_06	Data-Based Solutions and the Use of Intelligent Manufacturing Systems for Production Process Optimization and Product Quality Enhancement from a Facility Perspective	433 (9.9%)
topic_07	Development of Advanced Manufacturing Technologies and Systems	65 (1.5%)
topic_08	Development of Intelligent Automation Manufacturing Equipment	287 (6.6%)
topic_09	Development of Intelligent Automation Systems for Manufacturing Process Management and Control	482 (11.1%)
topic_10	Development of Intelligent Automation Systems for Detection and Control of Machining Processes	163 (3.7%)
topic_11	Development of Intelligent Automation Systems for Product Assembly and Manufacturing Utilizing Mechanical Components	1474 (33.9%)
topic_12	Development of Intelligent Automation Systems for Fluid and Gas Management and Control	235 (5.4%)



<Figure 7> The Total Distribution of Documents by Final Topic

연구에서 특허 문헌의 양적 수준을 분석하기 위해, 특허 출원 발생빈도, 출원 특허의 인용 수 및 피인용 수, 패밀리 특허 출원 발생빈도를, 질적 수준을 분석하기 위해서 청구항 수와 독립항 수를 비교 분석하였다.

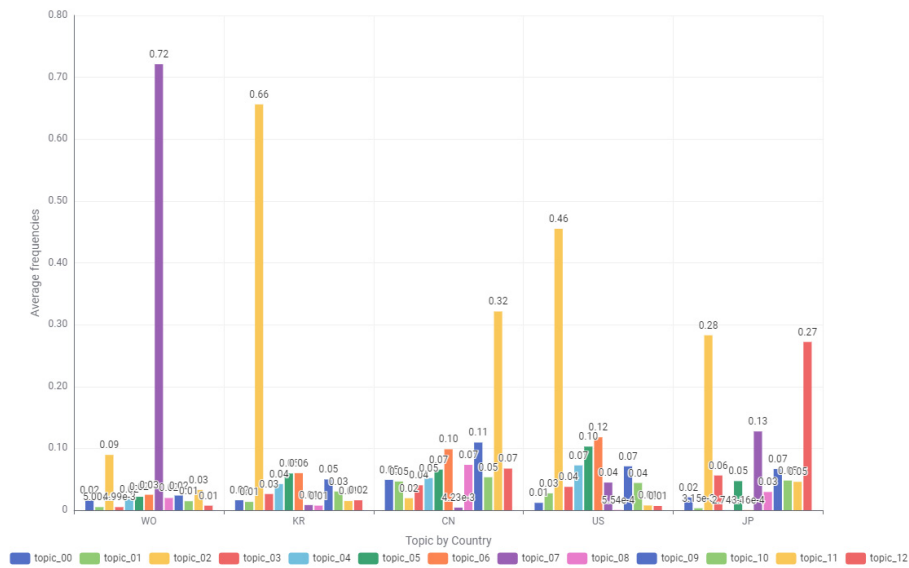
4.1 스마트팩토리 토픽별 평균 발생빈도 분석

PCT 출원 특허인 WO의 경우에 가장 높은 발생빈도를 보이는 주제가 topic_07(첨단 제조 기술 및 시스템 개발)이며, KR은 topic_02(IoT 기술 및 데이터 기반 의사 결정), CN은 topic_11(기계 구성요소를 활용한 제품 조립 및 제조

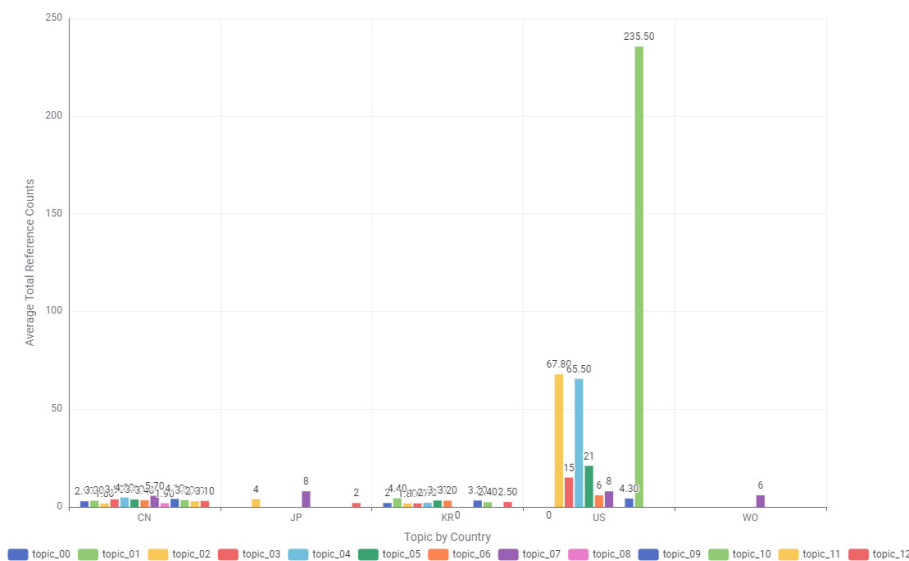
지능형 자동화 시스템 개발), US는 한국과 동일한 topic_02로 나타났고, JP도 근소한 차이로 topic_02로 분석되었다. CN은 타 국가 대비 주제별로 대체로 고른 분포를 보이고 있었으며, KR은 US 분포와 상당히 유사한 것이 가장 큰 특징이다.

4.2 스마트팩토리 토픽별 총피인용수 분석

US가 topic_10(가공 공정 감지 및 제어를 위한 지능형 자동화 시스템 개발) 평균 235.5회, topic_02(IoT 기술 및 데이터 기반 의사 결정) 평균 67.8회, topic_04(생산 공정



<Figure 8> The Avg. Occurrence by Topic



<Figure 9> The Avg. Number of Total Forward Citations by Topic

유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발) 65.5회로 총피인용수가 높았으며, CN은 모든 topic들이 대체로 2회에서 5회 정도로 피인용 되는 것으로 나타났다. KR도 모든 topic들에 대해서 전반적으로 평균 2회에서 4회 정도로 총 피인용 되는 것으로 나타났으며, topic_01(모바일 장치 산업의 재료 처리 및 포장을 위한 자동화 제조 시스템 및 장비 개발) 4.4회, topic_05(제품 관점에서 품질 향상과 비용 절감을 위한 지능형 자동화 제조 시스템 및 프로세스 개발), topic_09(제조 공정 관리 및 제어를 위한 지능형 자동화 시스템 개발)가 평균 3.3회로 나타났다.

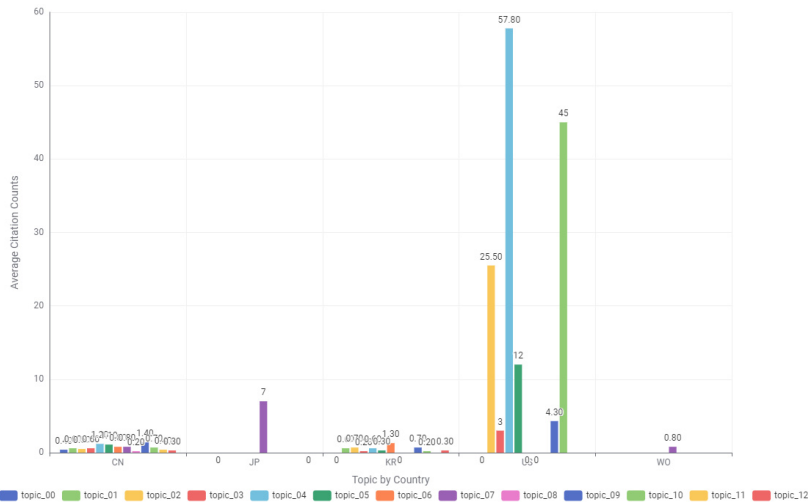
4.3 스마트팩토리 인용 수 분석

US가 topic_04(생산 공정 유연화를 위한 지능형 자동화

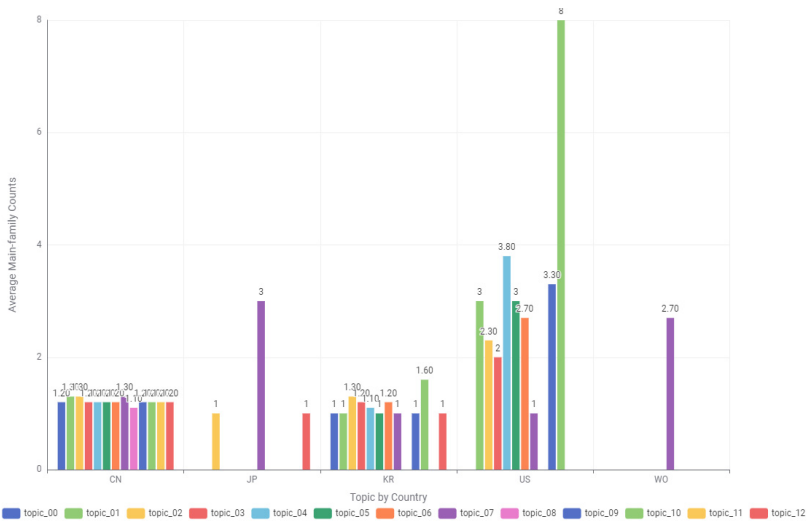
제조 시스템 및 장비 개발) 평균 57.8회, topic_10(가공 공정 감지 및 제어를 위한 지능형 자동화 시스템 개발) 평균 45회, topic_02(IoT 기술 및 데이터 기반 의사 결정) 25.5회로 높은 것으로 분석되었다. CN과 KR은 전반적으로 1회대 이하로 분석되었다.

4.4 스마트팩토리 패밀리특허 건수 분석

CN은 각 주제별로 평균 패밀리특허 건수가 1개를 약간 웃도는 수준이며, KR도 거의 비슷한 양상으로 나타났다. US의 경우, topic_10(가공 공정 감지 및 제어를 위한 지능형 자동화 시스템 개발)이 평균 8건, topic_04(생산 공정 유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발)가 3.8건, topic_09(제조 공정 관리 및 제어를 위한 지능형



<Figure 10> The Avg. Number of Backward Citations by Topic



<Figure 11> The Avg. Number of Family Patent Counts by Topic

자동화 시스템 개발)가 3.3건, topic_01(모바일 장치 산업의 재료 처리 및 포장을 위한 자동화 제조 시스템 및 장비 개발)과 topic_05(제품 관점에서 품질 향상과 비용 절감을 위한 지능형 자동화 제조 시스템 및 프로세스 개발)가 3건으로 분석되었다.

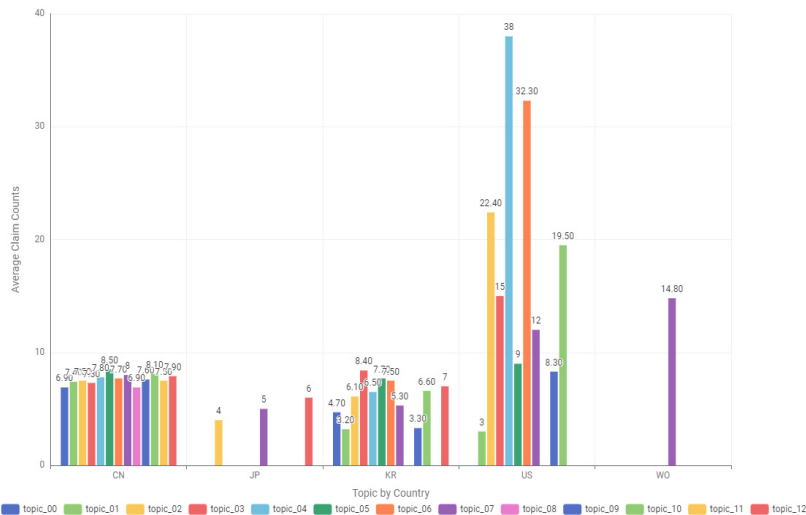
4.5 스마트팩토리 청구항 수 분석

CN은 평균 청구항 수가 7~8개 수준에서 모든 주제에 대해서 비슷한 수준의 경향을 보였으며, KR은 주제별로 약간 다르게 topic_03(기초 및 원천 부문의 재료가공과 마무리 공정을 위한 첨단 제조 공정 및 장비 개발)이 8.4건으로 가장 높았고, topic_05(제품 관점에서 품질 향상과 비용 절감을 위한 지능형 자동화 제조 시스템 및 프로세스 개발)

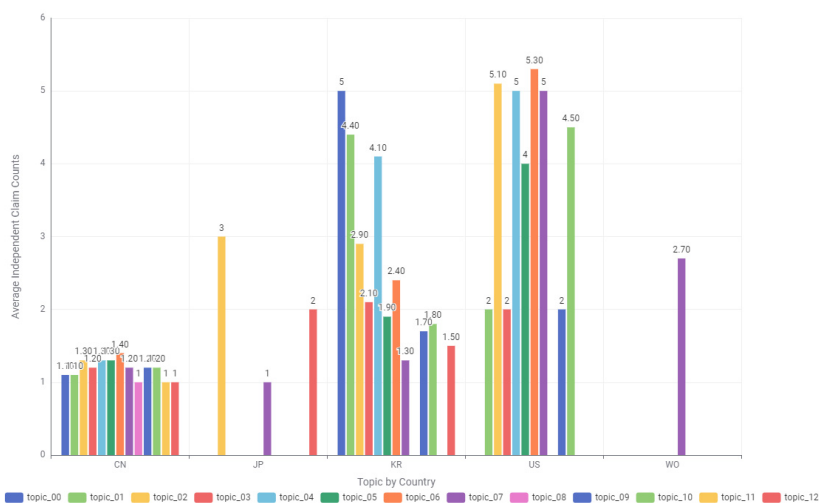
7.7건, topic_06(설비 관점에서 생산 공정 최적화와 제품 품질을 향상하기 위한 데이터 기반 해결 방법 및 지능형 제조 시스템 사용) 7.5건 순으로 높게 나타났다. US는 topic_04(생산 공정 유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발)가 38건으로 가장 높은 평균 청구항 수를 보였으며, topic_06(설비 관점에서 생산 공정 최적화와 제품 품질을 향상하기 위한 데이터 기반 해결 방법 및 지능형 제조 시스템 사용) 38건, topic_02(IoT 기술 및 데이터 기반 의사 결정) 22.4건, topic_10(가공 공정 감지 및 제어를 위한 지능형 자동화 시스템 개발) 19.5건 순으로 높은 것으로 나타났다.

4.6 스마트팩토리 독립항 수 분석

CN은 독립항 수가 모든 주제에 대해서 평균 2개 이하로



<Figure 12> The Avg. Number of Claim Counts by Topic



<Figure 13> The Avg. Number of Independent Claim Counts by Topic

전반적으로 낮은 수준으로 분석되었으며, KR은 topic_00(전자부품 제조 조립을 위한 지능형 자동화 제조 시스템 및 장치 개발) 5개, topic_01(모바일 장치 산업의 재료 처리 및 포장을 위한 자동화 제조 시스템 및 장비 개발) 4.4개, topic_04(생산 공정 유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발) 4.1개 순으로 분석되었다. 나머지 주제들인 topic_02(IoT 기술 및 데이터 기반 의사 결정) 2.9개, topic_06(설비 관점에서 생산 공정 최적화와 제품 품질을 향상하기 위한 데이터 기반 해결 방법 및 지능형 제조 시스템 사용) 2.4개로 나타났다. US는 topic_06(설비 관점에서 생산 공정 최적화와 제품 품질을 향상하기 위한 데이터 기반 해결 방법 및 지능형 제조 시스템 사용) 5.3개, topic_02(IoT 기술 및 데이터 기반 의사 결정) 5.1개, topic_04(생산 공정 유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발) 5개, topic_07(첨단 제조 기술 및 시스템 개발) 5개 순으로 높은 것으로 나타났다.

4.7 연구 결과

4.7.1 토픽별 평균 발생빈도 비교분석

PCT 출원 특허인 WO는 “첨단 제조 기술 및 시스템 개발” 주제가 가장 높은 빈도를 차지하였고, US와 KR은 “IoT 기술 및 데이터 기반 의사 결정” 관련 주제가 가장 높은 빈도를 차지하였다. CN은 “기계 구성요소를 활용한 제품 조립 및 제조 지능형 자동화 시스템 개발” 관련 주제가 상대적으로 높은 빈도를 차지하였으며, CN을 제외한 나머지 국가들의 경우(WO 포함)에, “IoT 기술 및 데이터 기반 의사 결정” 관련 주제가 상대적으로 높은 빈도를 차지하고 있는 것으로 분석되었다. 특히, CN은 13개 주제별로 대체로 고르게 분포되는 양상을 보이고 있으며, KR은 US와 대체로 유사한 패턴을 유지하고 있는 것으로 나타나, 각 주제별 특허출원의 발생빈도 양상이 유사한 것으로 분석되었다.

4.7.2 토픽별 평균 총피인용수 비교분석

US는 “가공 공정 감지 및 제어를 위한 지능형 자동화 시스템 개발”, “IoT 기술 및 데이터 기반 의사 결정”, “생산 공정 유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발”이 높게 나타났고, CN과 KR은 대체로 비슷한 수준인 2회 ~ 4, 5회 수준에서 총피인용수 평균값을 유지하고 있는 것으로 분석되었다.

4.7.3 토픽별 평균 인용 수 비교분석

US는 피인용 수와 유사하게 각 주제별 평균값이 타 국가 대비 두드러지게 높은 것으로 나타나, 선행 유사 기술들에 대한 출원 참조를 타 국가 대비 많이 하는 것으로

분석되었다. “생산 공정 유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발”은 평균 57.8회, “가공 공정 감지 및 제어를 위한 지능형 자동화 시스템 개발”은 평균 45회, “IoT 기술 및 데이터 기반 의사 결정”은 평균 25.5회로 나타난 것에 반해, CN과 KR은 거의 모든 주제에서 전반적으로 평균 1회 대 이하를 유지하고 있는 것으로 나타났다.

4.7.4 토픽별 평균 패밀리특허 수 비교분석

CN과 KR은 각 주제별로 평균 패밀리특허 건수가 평균 1건을 약간 상회 하는 수준으로 나타났다. US는 “가공 공정 감지 및 제어를 위한 지능형 자동화 시스템 개발”이 평균 8건, “생산 공정 유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발”이 3.8건, “제조 공정 관리 및 제어를 위한 지능형 자동화 시스템 개발”이 3.3건, “모바일 장치 산업의 재료 처리 및 포장을 위한 자동화 제조 시스템 및 장비 개발”과 “제품 관점에서 품질 향상과 비용 절감을 위한 지능형 자동화 제조 시스템 및 프로세스 개발”이 각각 3건으로 분석되었다. US 경우에, 기술사업화 가능성이 높은 주제는 상기 5개 주제인 “모바일 장치 산업의 재료 처리 및 포장을 위한 자동화 제조 시스템 및 장비 개발”, “생산 공정 유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발”, “제품 관점에서 품질 향상과 비용 절감을 위한 지능형 자동화 제조 시스템 및 프로세스 개발”, “제조 공정 관리 및 제어를 위한 지능형 자동화 시스템 개발”, “가공 공정 감지 및 제어를 위한 지능형 자동화 시스템 개발”인 것으로 나타났다.

4.7.5 토픽별 평균 청구항 수 비교분석

CN은 각 주제별로 평균 7~8개 이상의 청구항 수를 유지하고 있으며, KR은 각 주제별로 평균 3~8개 이상의 청구항 수를 유지하고 있는 것으로 나타났다. 이에 반해, US는 상당히 높은 수준으로 청구항 수를 유지하고 있으며, “생산 공정 유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발”, “설비 관점에서 생산 공정 최적화와 제품 품질 향상을 위한 데이터 기반 해결 방법 및 지능형 제조 시스템 사용”, “IoT 기술 및 데이터 기반 의사 결정”, “가공 공정 감지 및 제어를 위한 지능형 자동화 시스템 개발” 순으로 각각 38회, 32.3회, 22.4회, 19.5회로 나타났다. 결과적으로, US는 CN과 KR에 비해 상대적으로 많은 청구항 수를 포함하고 있는 것으로 나타났다.

4.7.6 토픽별 평균 독립항 수 비교분석

CN은 각 주제별로 1건~1.4건, KR은 1.3건~5건, US는 각 주제별로 평균 2건~5.3건 정도로 나타났다. KR은 5건이 “전자부품 제조 조립을 위한 지능형 자동화 제조 시스템 및 장치 개발”, “모바일 장치 산업의 재료 처리 및 포

장을 위한 자동화 제조 시스템 및 장비 개발” 4.4건, “생산 공정 유연화를 위한 지능형 자동화 제조 시스템 및 장비 개발” 4.1건, “IoT 기술 및 데이터 기반 의사 결정 ” 2.9건, “설비 관점에서 생산 공정 최적화와 제품 품질을 향상하기 위한 데이터 기반 해결 방법 및 지능형 제조 시스템 사용” 2.4건, “기초 및 원천 부문의 재료가공과 및 마무리 공정을 위한 첨단 제조 공정 및 장비 개발” 2.1건으로 분석되었다. 결과적으로, 독립항 수는 KR과 US가 비슷한 수준으로 나타났다.

5. 결론

본 연구에서는 LDA 토픽모델링 분석 방법과 생성형 인공지능 ChatGPT를 활용하여 스마트팩토리 분야에서 연관 특허 빅데이터 분석을 통해 글로벌 DT 기술 개발 추세와 국가 간의 차이를 분석할 수 있는 체계를 새롭게 제시하였다. 특히, 과학기술 문헌인 특허 문서를 대상으로 LDA를 수행할 때, 각 토픽을 구성하는 전문용어(단어)를 근거로 토픽 의미를 추론하는데 어려움을 해결하기 위하여 ChatGPT 추론 엔진을 도입한 점은 과학기술 정책입안자나 기술 기획 전문가들에게 실무적인 측면에서 중요한 역할을 할 것으로 예상된다. 이외에도 LDA 토픽 모형 분석 결과인 기술 주제들에 대해서 상호배타적이며 상호포괄적으로 기술 주제를 구성하고 있는지에 대한 검증 수단으로도 ChatGPT를 활용하고자 하였다. 마지막으로 스마트팩토리 전문가들의 기술적 검토를 통해 앞서 분석된 기술 주제들의 의미(또는 정의), 내용, 명칭들에 대한 적절성을 확보할 수 있도록 하였다.

이와 같은 절차를 통해 확보된 기술 주제들을 대상으로 글로벌 특허 문헌 비교분석을 수행하였으며, 양적 및 질적 수준을 분석하기 위해 6가지 세부 지표를 활용하였다.

특히 본 연구에서는 스마트팩토리 분야를 대상으로 새롭게 제안한 분석 방법론을 활용하여 사례분석을 수행하였으며, 기술 주제별 상대적인 발생비율이 한국과 미국의 너무나 비슷한 경향을 보인다는 사실을 확인하였다. 이러한 기술 추세는 국내와 미국 간의 스마트팩토리 기술 적용 환경과 시장수요 양상이 두 국가 간에도 유사한 상황으로 볼 수 있을 것인지에 대한 검토가 필요하다는 사실을 알 수 있었다. 특히 중국의 경우에 한국 및 미국과는 전혀 다른 발생비율로 기술 주제들에 대해 특허 문헌들이 등록되고 있어 시장수요와 기술 적용 환경이 어떻게 다른지에 대한 보다 면밀한 검토가 필요할 것으로 예상되었다.

한국과 중국은 총피인용수가 상당히 낮은 거에 반해, 미국은 전반적으로 모든 기술 주제에 대해 높은 피인용 수 수준을 보여, 논문의 질적 수준 측면에서도 차이가 있다는

사실을 확인하였다. 특히 시장개척 및 시장 확대 측면에서 중요한 패밀리 특허의 규모 분석 결과에서는 한국과 중국은 평균 1건을 약간 웃도는 수준인 거에 비해, 미국은 대체로 3~8건 정도로 나타나 해외시장 개척이나 진입을 위한 해외 특허 출원의 필요성을 확인하였다. 다만 중국과 한국 두 국가를 비교할 때, 전체 출원 건수가 중국은 3,928건 한국은 335건인 점을 고려하면, 해외시장개척 및 확대를 위한 사업다각화 효율성은 한국이 중국보다 높다는 점을 확인할 수 있었다.

특허 문헌의 질적 수준을 평가하는데 중요한 지표인 청구항 수와 독립항 수의 규모 측면에서는 중국이 청구항 수가 평균 7~8개, 한국은 3~8개, 미국은 3~38개로 분석되었다. 이처럼 미국은 중국과 한국이 비해 상대적으로 많은 청구항 수를 포함하고 있어 특허의 권리 범위를 나타내는 질적 수준이 상대적으로 높을 가능성이 존재하였다. 독립항 수 측면에서는 중국이 주제별로 1~1.4건, 한국 1.3~5건, 미국은 2~5.3건으로 분석되어, 한국과 미국이 비슷한 수준으로 나타났으며, 중국에 비해 특허 내용의 깊이를 나타내는 질적 수준이 상대적으로 높을 가능성이 존재한다는 사실을 확인하였다.

이처럼 연관 특허 빅데이터를 전처리하고, LDA 토픽 모형과 생성형 인공지능 모형을 적용하여 KNIME 분석 도구를 활용해 시각화함으로써, 글로벌 특허 환경을 더 잘 이해하고 국내 특허 환경과 비교 분석할 수 있도록 하는 체계를 제시한 점과 국가 연구개발 및 기술사업화 투자 방향에 대한 상이한 특징과 투자 방향을 탐색할 수 있다는 점은 본 연구의 가장 큰 기여로 예상된다.

본 연구에서는 최근 기술 개발 추세를 분석하기 검색 기간을 최근 6년으로 한정하였고, 기간 확대 시 분석 결과가 달라질 가능성은 존재한다. 또한 기술 주제 분석 결과에 대한 신뢰도를 확보하기 최종적으로 전문가 검토를 수행하는 절차를 반영하고는 있지만, 최근 지속적인 개선이 진행 중인 ChatGPT에 의한 기술 주제 의미추론 결과의 재현성은 다소 떨어질 가능성이 존재한다는 사실은 본 연구의 한계로 남는다. 향후 추가적인 연구가 필요한 내용은 각 주제별로 구체적인 기술 분류(IPC 또는 CPC 등) 측면에서 발생 양상이 어떻게 나타나고 있는지에 대한 후속 연구는 본 연구 결과를 지원하는 데 도움이 클 것으로 예상된다.

Acknowledgement

This research received partial funding from two programs (K-23-L05-C02-S16, K-23-L03-C04-S01) provided by the Korea Institute of Science and Technology Information (KISTI) in South Korea.

References

- [1] Abdirad, M. and Krishnan, K., Industry 4.0 in Logistics and Supply Chain Management: A Systematic Literature Review, *EMJ - Engineering Management Journal*, 2020, Vol. 33, No. C., pp. 1-15.
- [2] Ampornphan, P. and Tongngam, S., Exploring technology influencers from patent data using association rule mining and social network analysis, *Information (Switzerland)*, 2020, Vol. 11, No. 6.
- [3] Blei, D. M., Ng, A. Y., and Jordan M. I., Latent Dirichlet allocation, *Journal of Machine Learning Research*, 2003 Vol. 3, No. 4-5, pp. 993-1022.
- [4] Cao, L. and Fei-Fei, L., Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes, in *Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, ICCV, 2007*, pp. 1-8.
- [5] Chang, T. and Yang, H., Latent semantic analysis of research papers on smart factory, *ICIC Express Letters*, 2017, Vol. 11, No. 4, pp. 899-904.
- [6] Choi, D. and Song, B., Exploring technological trends in logistics: Topic modeling-based patent analysis, *Sustainability (Switzerland)*, 2018, Vol. 10, No. 8.
- [7] Daim, T., Bukhari, E., Bakry, D., Vanhuis, J., Yalcin, H., and Wang, X., Forecasting technology trends through the gap between science and technology: The Case of software as an e-commerce service, *Foresight and STI Governance*, 2021, Vol. 15, No. 2, pp. 12-24.
- [8] Govindarajan, U., Trappey, A., and Trappey, C., Immersive Technology for Human-Centric Cyberphysical Systems in Complex Manufacturing Processes: A Comprehensive Overview of the Global Patent Profile Using Collective Intelligence, *Complexity*, 2018, Vol. 2018.
- [9] Griffiths, T. L. and Steyvers, M., Finding scientific topics, *Proceedings of the National Academy of Sciences*, April 2004.
- [10] Jung, Y. and Chang, T., Text mining based online news analysis about smart factory, *ICIC Express Letters, Part B: Applications*, 2018, Vol.9, No. 6, p.559-565.
- [11] Kim, M., Park, Y., and Yoon, J., Generating patent development maps for technology monitoring using semantic patent-topic analysis, *Computers and Industrial Engineering*, 2016, Vol. 98, pp. 289-299.
- [12] Kwon, K., Jun, S., Lee, Y., Choi, S., and Lee, C., Logistics Technology Forecasting Framework Using Patent Analysis for Technology Roadmap, *Sustainability (Switzerland)*, 2022, Vol. 14, No. 9.
- [13] Lee, C., Liu, C., Trappey, A., Mo, J., and Desouza, K., Understanding digital transformation in advanced manufacturing and engineering: A bibliometric analysis, topic modeling and research trend discovery, *Advanced Engineering Informatics*, 2021, Vol. 50.
- [14] Lee, J., Park, S., and Lee, J., Study on the Technology Trend Screening Framework Using Unsupervised Learning, *Applied Sciences (Switzerland)*, 2022, Vol.12, No. 17.
- [15] Li, R., Wang, X., Liu, Y., Zhang, S., and Hanif, O., Research status and collaboration analysis based on big data mining: an empirical study of Alzheimer's disease, *Technology Analysis and Strategic Management*, 2021, Vol.33, No. 4, pp. 379-395.
- [16] Nam, D. and Choi, G., The Identification of Emerging Technologies of Automotive Semiconductor, *KSHI Transactions on Internet and Information Systems*, 2023, Vol.17, No. 2, pp. 663-677.
- [17] Park, J., Kim, N., Choi, H., and Han, E., A new forecasting system using the latent dirichlet allocation (LDA) topic modeling technique, *WSEAS Transactions on Environment and Development*, 2018, Vol.14, p.363-373.
- [18] Wang, J. and Hsu, C., A topic-based patent analytics approach for exploring technological trends in smart manufacturing, *Journal of Manufacturing Technology Management*, 2021, Vol.32, No. 1, pp. 110-135.
- [19] Wang, J., Li, K., and Feng, L., Tracing the technological trajectory of coal slurry pipeline transportation technology: An HMM-based topic modeling approach, *Frontiers in Energy Research*, 2022, Vol. 10.
- [20] Wang, Y., Feng, L., Wang, J., Zhao, H., and Liu, P., Technology Trend Forecasting and Technology Opportunity Discovery Based on Text Mining: The Case of Refrigerated Container Technology, *Processes*, 2022, Vol. 10, No. 3.
- [21] Wang, Z., Gao, L., and Suo, H., Designing petrochemical smart plant of the future: State of the art, comparison and prospects; 未来石化智能工厂顶层设计:现状、对比及展望, *Huagong Jinzhan/Chemical Industry and Engineering Progress*, 2022, Vol. 41, No. 7, pp. 3387-3401.
- [22] Yang, H., Chang, T., and Choi, Y., Exploring the research trend of smart factory with topic modeling, *Sustainability (Switzerland)*, 2018, Vol. 10, No. 8.

ORCID

Sang-Gook Kim | <https://orcid.org/0000-0001-7018-9716>
Minyoung Yun | <http://orcid.org/0000-0001-8672-8155>

Taehoon Kwon | <https://orcid.org/0000-0002-6306-1244>
Jung Sun Lim | <https://orcid.org/0000-0002-2737-6118>