

BERTopic을 활용한 텍스트마이닝 기반 인공지능 반도체 기술 및 연구동향 분석

김현경* · 이정훈** · 강선구***

Topic Modeling on Patent and Article Big Data Using BERTopic and Analyzing Technological Trends of AI Semiconductor Industry

Hyeonkyeong Kim* · Junghoon Lee** · Sunku Kang***

Abstract

The Fourth Industrial Revolution has spurred widespread adoption of AI-based services, driving global interest in AI semiconductors for efficient large-scale computation. Text mining research, historically using LDA, has evolved with machine learning integration, exemplified by the 2021 BERTopic technology. This study employs BERTopic to analyze AI semiconductor-related patents and research data, generating 48 topics from 2,256 patents and 40 topics from 1,112 publications. While providing valuable insights into technology trends, the study acknowledges limitations in taking a macro approach to the entire AI semiconductor industry. Future research may explore specific technologies for more nuanced insights as the industry matures.

Keywords : Artificial intelligence Semiconductor, Topic Modeling, BERTopic

Received : 2024. 01. 08. Revised : 2024. 02. 28. Final Acceptance : 2024. 02. 28.

* First Author, Manager, KT, e-mail : karenkim26@gmail.com

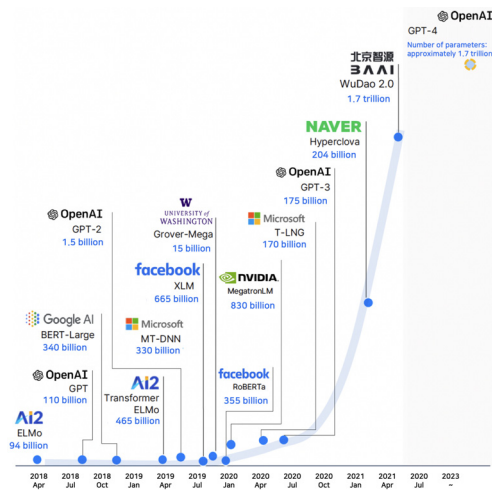
** Corresponding Author, Professor, Graduate School of Information, Yonsei University, New Millennium Hall, 50, Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea, Tel : +82-2-2123-4529, e-mail : jhoonlee@yonsei.ac.kr

*** Co-Author, Master student, IoT Service Convergence Graduate School of Information, Yonsei University, e-mail : kty8623@yonsei.ac.kr

1. Introduction

1.1 연구의 배경

4차 산업 혁명으로 인해 촉발된 산업 전반의 Digital Transformation으로 인공지능 기반 서비스가 확장되고, 공공과 민간을 막론하고 사회 전반에 인공지능이 도입되었다. 2018년 AI2에서 만든 ELMo라는 알고리즘은 파라미터 수가 9,400만에 불과하였으나, 2023년 출시 예정인 Open AI 사의 자연어 처리모델인 GPT-4는 100조가 넘는 파라미터 수를 가질 것으로 예상되고 있다. 이처럼 인공지능발전에 따라 데이터 처리수요 또한 폭증하고 있으며[Choi and Yeon, 2020], 이에 따라 증가하는 인공지능 알고리즘의 복잡도로 연산을 효율적으로 처리할 수 있는 프로세서에 대한 수요도 증가하고 있다[Lee et al., 2020].



<Figure 1> Increasing Number of Configuration Parameters by Artificial Intelligence Model

이에 인공지능 반도체 시장을 독점하고 있던 GPGPU (General-Purpose Graphic Processing Unit) 대비 저비용 고효율로 인공지능 연산을 처리할 수 있는 인공지능 반도체, NPU(Neural Processing Unit)가 각광받고 있다[Tripathy and Rajput, 2022]. 이에 한 시장조사 기관의 조사에 따르면 실제로 인공지능 반도체 시장은 2030년까지 연평균 37.41%의 성장률을 보이며 1,940억 달러의 시장으로 성장 할 것으로 전망된

다[Allied Market Research, 2021]. 이러한 산업 흐름에 따라 정부는 차세대 반도체 성장 동력으로 시스템 반도체, 그 중 인공지능 반도체를 선정하고, 5년간 1조 규모의 인공지능 반도체 산업 성장 지원 대책을 발표한 바 있다[Ministry of Science and ICT, 2022]. 1조 200억 원은 인공 지능 반도체 첨단기술 연구개발에 투입되며, 인공지능 반도체산업의 초격차 기술력을 확보하려는 목적으로 차세대 지능형 반도체 개발(20년~29년), PiM 반도체 개발(22년~28년) 등에 사용된다[Ministry of Science and ICT, 2022].

현재까지는 인공지능 서비스를 제공하는데 있어서 GPU가 가장 범용적으로 사용되고 있으나, 인공지능연산에 특화된 아키텍처를 기반으로 한 빅테크 기업의 자체 인공지능 반도체 개발(Google-TPU, AWS-Inferencia)과 더불어 혁신적 반도체 기술 기반의 스타트업들이 본격적으로 출현하여 성장 중에 있다[Korea Institute for Advancement of Technology, 2022]. 한편, 국내 인공지능 반도체 업체는 대부분의 기업들이 스타트업으로, 한정된 투자금액 대비 효율적인 결과를 산출하기 위한 R&D 기획 및 기술 개발 방향성이 중요하다. 금융감독원의 전자공시시스템에 따르면 국내 팹리스 상위 18개의 매출 대비 연구개발(R&D) 비용이 차지하는 비중의 평균은 약 17.7%로 집계되었다[Sisajournal, 2021]. 많은 연구를 통해 R&D는 기업의 성공을 결정하는 중요한 요소로 작용한다고 입증되어 왔다[Pessoa, 2007; Shefer and Frenkel, 2005]. 특히, 자금조달이 쉽지 않은 스타트업 특성 상 초기 R&D 투자 비용이 매우 높은 반도체 산업에서 유망한 인공지능 반도체 기술 트렌드를 파악하는 것이 매우 중요한 실정이다. 이처럼, 전 세계적으로 집중적으로 육성중인 인공지능 반도체 산업에서 기술 개발 트렌드를 파악하는 것은 국가적으로, 산업적으로 매우 중요하게 당면한 과제이다. 하지만 인공지능 반도체 산업이 초기시장 인 만큼 해당 산업에서 특허와 같은 정량적 지표를 활용한 기술 트렌드 분석에 대한 선행 연구가 부족한 실정이다. 이에, 본 연구는 객관적인 방법론으로 사용하여 인공지능반도체 산업의 기술동향 및 연구 동향을 분석하고자 한다.

1.2 연구의 목적

최근 인공지능은 다양한 제품과 형태로 제공되고 있다.

인공지능 기술은 산업 전반에 적용되어 효율성을 개선하기도 하고, 사회 전반의 혁신적인 변화를 유도한다. 이에 인공지능 기술은 국가를 막론하고 사회가 나아가야 할 방향으로 인식되고 있다. 이러한 인공지능의 부상에 따라 인공지능을 활용한 다양한 서비스들이 출시되고 있고, 인공지능 서비스의 사용자가 증가할수록 인공지능 인프라도 크게 확대된다. 하지만 GPU의 높은 비용과 높은 전력 소모량으로 인해 기업의 TCO(총 소유비용)가 증가하는데, 이러한 부담을 경감시켜줄 수 있는 것이 인공지능 반도체이다. 이에 연구의 목적은 아래와 같다. 본 논문의 연구 목적은 인공지능 반도체와 관련된 기술 특허와 논문데이터를 객관적인 방법론으로 분석하여 실증적인 기술 토픽을 제시하고, 이를 바탕으로 인공지능 반도체 기업, 및 국가 연구소등 인공지능 반도체 산업의 다양한 연구 주체에 기초적인 시사점을 제공하는 것이다. 인공지능 반도체는 전 세계에서 국가적, 산업적으로 중요성이 부각되고 있지만, 정량적인 데이터를 바탕으로 기술 트렌드를 도출하는 연구는 전무한 실정이다. 한편, 기술 트렌드 도출을 위해 가장 많이 사용되는 지표인 특허는, 등록과 게재에 시일이 소요되어 최신 기술 트렌드를 반영하는 데는 한계가 있다. 이러한 특허의 후행성을 보완하기 위해 논문[Lee et al. 2017], Github[Lee, 2021] 등의 데이터를 추가하기도 한다. 또한 특허 내에서도 여러 가지 지표를 추가하기도 하는데, 예를 들어 Han의 연구는 유효 특허 및 비 유효 특허 간의 관계를 분석하여 기술 수명 주기를 도출, 중국의 3D 프린팅 시장에 대한 기술에 실증 한 바 있다[Batra et al., 2019].

한편, 정보통신기술의 발달로 텍스트 중심의 데이터가 급격하게 증가하면서, 이러한 문서에서 의미 있는 주제를 찾기 위한 노력도 지속되고 있다. 토픽모델링은 다양한 문서 집합 안에서 내재된 토픽을 찾기 위한 방법론으로, 1980년대 생성적 확률 모델링에서 분리되어 지속적으로 발전되어온 개념[Liu et al., 2016]이다. 최근까지 LDA(잠재적 디레클레 할당, Latent Dirichlet Allocation)이라는 토픽 모델링 방법론이 가장 일반적으로 활용되고 있는데[Vayansky and Kumar, 2020], 머신 러닝이 토픽모델링에 응용되기 시작하면서, 기존 방법의 단점을 보완하여 정확도를 높이려는 여러 시도들이 이루어지고 있다. 대표적으로 2022년, 클러스터링 기술과 클래스 기반 TF-IDF를 활용하여

일관된 토픽을 생성하는 BERTopic 기술이 제안되었는데, 다른 토픽모델링 대비하여 높은 주제 일관성과 다양성을 보이는 것으로 확인되었다[Grootendorst, 2022]. BERTopic은 Transformer모델의 우월한 성능과 UMAP, HDBSCAN등과 같은 머신러닝 기술을 활용하여 토픽모델링에서 더 나은 결과를 보여준다는 평가가 있다[Atzeni et al., 2022]. 이에, 본 논문에서는 아직 초기시장에 머물러 있는 인공지능 반도체 산업의 특성을 반영하여 논문데이터를 추가하여, 한국 및 미국의 특허와 논문 데이터에 대한 BERTopic 기반의 토픽모델링을 통해 인공지능 반도체 기술 구조를 도출하고 연구동향을 분석한다.

1.3 연구의 방법 및 범위

본 연구에서는 인공지능 반도체의 특허와 논문 데이터를 바탕으로 하여, 2022년 새롭게 제시된 BERTopic이라는 토픽모델링 방법론을 활용하여 인공지능 반도체 기술의 동향을 분석한다. 본 논문은 총 5장으로 구성되어 있으며, 제1장은 연구의 배경과 목적, 방법 및 목적으로 이루어져 있다. 제1장에서는 전체적인 연구의 방향과 개요에 대하여 서술한다. 제2장은 본 연구와 분석 대상, 분석 방법, 분석 도메인에 대한 이론적 배경과 선행연구를 분석하였다. 제3장은 연구 방법에 대한 내용으로, 연구 방법과 연구문제, 연구 프레임워크와 사용되는 환경에 대해 기술한다. 세부 내용으로는 연구 프레임워크에 따른 데이터 수집 사이트, 데이터 처리 과정, 및 토픽모델링 절차에 대하여 기술하였다. 제4장은 연구방법에 의해 도출된 특허와 논문의 주요 토픽 및 연구주체를 도출하고, 도출 결과를 해석한다. 제5장은 연구 결과 요약 및 학술적, 실무적 시사점을 서술하였으며, 연구의 한계 및 향후 연구 방향을 제시하였다.

2. 이론적 배경

전장에서 제시한 연구 목적은 인공지능 반도체 산업에서, 객관적인 방법론을 통해 토픽을 도출하여 기초적인 시사점을 제시하는 것이다. 이러한 연구목적을 이루기 위해 설정한 분석대상(인공지능 반도체), 분석 방법(토픽 모델링), 분석 도메인(특허 및 논문)에 대한 선행 연구를 통해 이론적 배경을 도출한다.

2.1 인공지능 반도체의 개념 및 기술에 대한 선행연구

2.1.1 인공지능 반도체 개념에 관한 선행연구

인공지능 반도체는 인공지능을 위한 연산을 처리하는데 최적화 된 반도체로 연구자 및 연구기관마다 다양한 정의를 가지고 있다. 인공지능 반도체에 대한 정량적 연구를 수행하기 이전에, 관련 데이터를 추출하기 위한 인공지능 반도체에 대한 범위를 설정하는 것이 선행되어야 한다. 지금까지 인공지능 반도체에 대하여 다양한 선행 연구가 진행되었으나, 인공지능 반도체의 정의는 연구자마다 다르게 표현되었다. Kim[2018]은 인공지능 반도체를 “데이터센터 서버 또는 엣지 디바이스에서 인공지능 알고리즘을 보다 효율적으로 계산하는 데 최적화 된 반도체”로 정의하고, 하드웨어 타입에 따라 GPU, FPGA, ASIC, ASSP, 뉴로모픽을 협의의 인공지능 반도체, 협의의 인공지능 반도체의 정의에 CPU를 추가하여 광의의 인공지능 반도체로 정의한 바 있다 [Kim, 2018]. 또한 Mckinsey[2019]에서는 인공지

능 하드웨어를 ASIC, GPU, FPGA로 한정하여 정의한 바 있다. 또한 Lee et al.[2022]는 인공지능 반도체 산업의 기술 발전 방향에 대한 연구에서 인공지능 반도체를 세대별로 나누어 CPU와 GPU를 1세대 인공지능 반도체, FPGA와 ASIC을 2세대, Neuromorphic을 3세대 인공지능 반도체로 정의하였다[Batra et al., 2019]. 반면 인공지능 반도체와 관련된 해외 연구에서는 CPU, GPU와 별도의 개념으로 NPU(Neural network Processing Unit)를 정의하였으며, 비디오 및 이미지와 같은 방대한 멀티미디어 데이터를 처리하는데 뛰어난 “데이터 기반 병렬 컴퓨팅” 아키텍처를 채택하고 있는 반도체로 정의하였다(Wang, 2020). 인공지능 연산을 수행하기 위해서는, CPU, GPU 등 다양한 하드웨어가 모두 사용될 수 있다. 하지만 최근에는 인공지능 서비스가 많아지며, 고가의 GPU, 범용적인 CPU를 대체하여 인공지능 연산에 특화된 반도체를 설계하고 있고, 이를 제품화 하여 판매하고 있는 기업들도 다수 생겨나고 있다.

<Table 1> Synthesis of Prior Research on Artificial Intelligence Semiconductor Technology

Research Literature	Research Contents
Park Ji Min [2020], Analyzing AI semiconductor trends and global business activities using patent data	Leveraging Data Envelopment Analysis (DEA), we assess the synergy between patent applications, patent quality efficiency, and financial performance in AI semiconductor technologies (GPU, FPGA, ASIC, Neuromorphic) for business groups. Limitations: exclusive focus on financial performance, no quantitative technology trend analysis using big data.
Lee Jong-ho, Oh Chul [2022], “A study on the technological development path of the AI semiconductor industry and the catch-up chance for latecomers: Focusing on technical patent analysis as the view of the technological life cycle”, Journal of the Korean Society of Innovation	AI Semiconductor Technology Life Cycle Analysis by 1st, 2nd, and 3rd generation chips based on the number of patent re-appropriations U.S. patents/generation technology life cycle analysis. Limitations: Simple numerical comparison, not big data analysis, limited to U.S. patents
Kwon Young-soo [2018], “Artificial intelligence processor technology trends”, Electronic communication trend analysis, 33(5), 121-134.	Introduction of artificial intelligence semiconductor technology (creating) features and major corporate-based technology trends. Limit: Simple introduction of artificial intelligence semiconductor
Batra, G., Jacobson, Z., Madhav, S., Queirolo, A., and Santhanam, N.[2019], Artificial-intelligence hardware: New opportunities for semiconductor companies. McKinsey and Company, January, 2.	Comparison of Market Trends and Appearance Background of AI Semiconductor Companies by Semiconductor Type. Limitations: Market outlook focus, not technical approach, simple explanation by semiconductor hw type
Wang, Z. [2020], NPU Development Overview. Scientific Journal of Economics and Management Research Volume, 2	Summary of Natural Processing Unit (NPU) birth background and singularity. Limitations: General NPU technology related content

2.1.2 인공지능 반도체 기술 트렌드에 관한 선행연구

2018년도부터 인공지능 반도체 기술 및 기술 트렌드에 대한 연구는 다양하게 수행되어 왔으나, 대부분 관련 연구기관의 연구자의 지식을 바탕으로 한 주관적인 트렌드 분석이라는 한계가 있었다. Chang은 CPC코드로 특허 데이터를 취득하여 네트워크 분석을 진행하고, 뉴로모픽 시스템의 기술 트렌드에 관한 연구를 수행한 바 있다[Chang, 2018]. 또한, Park[2018]은 GPU, FPGA, ASIC 및 Neuromorphic으로 정의되는 인공지능 반도체 기술과 관련된 기업집단들을 대상으로 기업의 재무성과에 따른 특허출원 및 특허품질 효율성, 특허출원에 따른 특허품질 효율성 및 특허출원 대비 재무성과 효율성분석을 진행하였다. 그러나 빅데이터 기반 정량적 기술트렌드 분석이 아닌, 기업대상 재무성과 분석에 그친다는 한계점이 있다. 또한 Lee et al. [2022]는 특허 재인용수를 기반으로 1세대/2세대/3세대 칩 별 인공지능 반도체 기술 수명 주기 분석 미국 특허/세대별 기술 수명 주기 분석을 진행하였으나, 특허 재인용수를 기반으로 한 분석이라는 차이점이 있고, 분석 범위가 미국 특허에 한정되어 있다[Lee et al., 2022]. Kwon[2018]은 인공지능 반도체 기술(병렬 처리)의 특징을 설명하였으며, 인공지능 반도체 주요 기업 기반 기술 동향을 소개하였으나, 인공지능 반도체에 대한 단순 소개에 그친다는 한계가 있다. Batra G는 인공지능 반도체 기업별 시장 동향 및 등장 배경, 그리고 반도체 하드웨어 타입별 장단점을 비교하였으나, 인공지능 반도체의 설계 기술이 아닌 하드웨어의 특성을 분석하였다. 또한 Shibata et al.[2018]은 IPC 코드를 활용하여 FPGA기술관련 기업의 특허출원을 분석하여 트렌드를 도출하였으나, FPGA로 범위가 한정되어 있다.

2.1.3 인공지능 반도체 개념 및 기술에 관한 선행연구 결과

인공지능 반도체는 인공지능을 위한 연산을 처리하는 데 최적화된 반도체로 연구자 및 연구기관마다 다른 정의를 내리고 있다. 선행연구를 정리하면, 광의의 개념으로서의 인공지능 반도체는, 인공지능을 처리하는데 사용될 수 있는 모든 하드웨어 타입을 포함하지만, 최근에는 NPU, Neuromorphic등 인공지능 데

이터를 처리할 수 있는 형태로 인공지능 컴퓨팅 전용 반도체들이 나오고 있다. 따라서 기존의 범용성을 가지고 있는 GPU, FPGA, ASIC, ASSP등은 인공지능처리와 관련된 키워드를 추가하여 데이터를 추출하고, 그 자체로 인공지능 반도체라고 할 수 있는 Neuromorphic 및 NPU를 추가하여 대상 데이터를 확장 할 수 있다. 이에, 선행 연구를 통해 도출한 특허 검색식은 다음과 같다.

Patent Search Formula

((“AI” or “Artificial Intelligence” or “Machine Learning” or “Neural Network” or “Deep Learning”) and (“GPU” or “Graphic process” or “FPGA” or “Field programmable” or “ASIC” or “ASSP” or “semiconductor” or “accelerator” or “CPU”) or (“cerebras” or “Groq” or “graphcore” or “sapeon” or “sambanova” or “hailo” or “tenstorrent” or “esperanto” or “mythic” or “deepx” or “furiosaAI” or “rebellions”).ap.

기존에 인공지능반도체 관련 연구에서 가장 많이 채택된 GPU, FPGA, ASIC등 하드웨어 타입 별 세대별 분류를 적용하고, 인공지능의 연산에 필요한 하드웨어로 한정하기 위해 AI, Deep Learning, ML, Neural Network등의 키워드와 And 조건으로 동시 검색한다. 또한 유효한 검색데이터를 늘리기 위해, 현재 인공지능반도체 산업을 영위하고 있는 회사를 검색식에 추가하여 인공지능반도체 회사 별 인공지능 반도체 관련 특허를 추가로 추출하였다. 이를 통해 수집된 데이터들을 바탕으로 토픽 모델링을 진행하여 인공지능 반도체의 기술 구조를 파악하는 것을 목표로 한다.

또한 학술적인 용도로 발행되는 논문의 경우에는, 위의 특허에 사용된 기술중심의 검색식을 적용할 경우 너무 광범위한 결과 나오게 되어 오히려 토픽 모델링의 성능을 떨어트리게 된다. 또한 세부기술 중심으로 작성되어 있는 특허 청구항과 다르게, 논문은 인공지능 반도체 그 자체를 주제로 하는 학술 연구들이 진행이 되기 때문에 논문 검색식은 AI Accelerator, NPU등으로 한정하여 인공지능 반도체 관련 연구로 범위를 좁혀 신뢰성을 높였다. 논문을 통해 연구 동향을

알아보기 위해 아래 검색식으로 도출된 토픽에 동적 토픽 모델링을 적용하여, 연도 별 연구 주제의 흐름을 알아본다.

Thesis Search Formula

"AI accelerator" | "AI semiconductor" | "Neural Processing Unit" | "Neuromorphic Hardware"

또한 인공지능 반도체 기술 트렌드에 관한 빅데이터 분석 선행연구 분석 결과 정량 수치 기반 통계비교는 있었으나, 빅데이터 분석 방법론을 적용하여 구체적인 기술 토픽을 추출해 내는 연구는 없었다. 또한 특허데이터를 통한 분석은 있었으나, 논문 데이터를 기반으로 한 빅데이터 분석은 연구된 바 없다. 따라서, 본 논문에서는 현재 한국과 미국의 특허 및 논문 등을 분석하여 현재 연구가 어느 기술들을 중심으로 진행되고 있는지를 도출한다.

2.2 정량 데이터를 통해 기술 동향을 파악한 선행연구

2.2.1 논문을 통해 연구 동향을 파악한 선행 연구

논문은 그 자체로 연구 결과의 산출물이기 때문에, 다양한 산업분야에서 연구동향을 파악하기 위해 논문 데이터를 사용하는 연구가 지속해서 있어 왔다. 논문데이터는 연구되는 기술의 트렌드를 파악하고, 앞으로의 연구동향을 파악할 수 있는 중요한 지표가 될 수 있다 [Lee, 2017]. 논문을 통해 산업의 연구 동향을 파악한 선행연구로는 컴퓨터 공학 분야 논문 데이터에서 키워드 네트워크 분석을 진행하여 구조적 변화와 특징을 파악한 연구 [Kwon, 2013], 태양광에너지 기술에 대한 논문데이터와 특허데이터 기반으로 토픽 모델링을 진행하여 기술 발전 흐름을 도출한 연구, Lee et al. (2017), SCI저널의 인공지능 관련 논문을 키워드 네트워크 분석과 빈도분석을 수행하여 인공지능 분야 주요 연구 동향 분석을 진행한 연구 [Jung et al., 2017] 등이 있다. 논문과 저널의 영향도를 분석하는 연구들도 존재한다. 최근 2년간 수록된 논문 수 대비 최근 2년간 논문이 인용된 수를 산출하여 상대적 중요성을 결정하는 피인용 지수 [Garfield, 1994], 특정한 저널이 다른 저널 대비 가지는 상대적 영향력을 반영하여

인용 데이터에 기반을 두어 분석하는 Eigen factor Score [Bergstrom et al., 2008] 등이 대표적이다.

2.2.2 특허를 통해 기술 동향을 파악한 선행연구

기업의 지식재산은 기업의 시장의 경쟁 속에서 상대적 우위와 성과를 유지하는데 중요한 자원으로 여겨진다 [Drucker, 1995]. 기업의 지적 자산을 파악할 수 있는 매체인 특허는 개발된 기술의 정보를 파악할 수 있는 중요한 원천이다 [Lee, 2021]. 또한 특허는 기업의 R&D 성과와 기업의 매출에 영향을 주고 있으며 기업의 기술 혁신 전략에 깊이 관련되어 있는 자산이다 [Grzeszczyk and Grzeszczyk, 2021; Tekic et al., 2013]. 특허 데이터에 포함된 정보들은 제목, 초록, 설명, 청구항 등의 텍스트 형태로 이루어져 있는데, 약 80%의 기술정보가 특허 정보에서만 찾을 수 있기 때문에, 이러한 데이터들은 기술정보로서 큰 가치를 가지고 있다 [Lee, 2021; Tekic et al., 2013]. 이에 따라 특허에 기록되어 있는 핵심 기술과 관련된 정보를 기반으로 빅데이터 분석을 실행, 기술 동향을 파악하려는 시도는 지속적으로 있어왔다. 인공지능기술 특허 초록 데이터를 활용하여 LDA 방법론을 적용, 인공지능 토픽을 추출한 연구 [Park et al., 2017], 특허데이터의 토픽모델링을 활용하여 핀테크 산업에 대한 기술동향을 분석한 연구 [Kim et al., 2016], LDA를 활용하여 스마트 제조 기술 특허의 동향을 분석한 연구 [Wang and Hsu, 2020] 등이 있다. 특히, 인공지능을 통한 자연어처리 (Natural Language Processing) 기술이 발전하면서, 특허 데이터 베이스에 설명된 핵심 기술과 관련하여 자연어 처리 알고리즘을 사용하여 분류 정확도를 향상시키고 기술적인 기회를 발견하는데 기여할 수 있다 [Grzeszczyk and Grzeszczyk, 2021].

2.2.3 정량 데이터를 통해 기술 동향을 파악한 선행연구 결과

특허와 논문은 그 문서가 가지고 있는 정보들로 인해 기술과 연구 동향을 파악할 수 있는 대표적인 데이터로 여겨진다. 이에 산업의 기술 동향을 파악하기 위한 객관적인 지표로서 논문과 특허가 대표적으로 사용되어 왔다 [Gu et al., 2017]. 이에 본 연구에서는 특

〈Table 2〉 Synthesis of Prior Research on Artificial Intelligence Semiconductor Technology

Research Literature	Research Contents
Park, C. [2018]. Early detection of valuable patent using a deep learning model.	Using CNN and RNN to propose a patent rating framework based on context. Identify USPTO semiconductor patents, building a model with more than 75% accuracy
Ko, Namuk [2022]. A patent trademark data linking approach to business analytics for technology-based firms	Patent-Trademark Data Connection Business Competition Analysis/Business Diversification Analysis through Word Embedding
Oh, Seung-hyun [2020]. A link prediction approach to patent analytics for discovering product development directions Naver.	Predicting the direction of technology convergence by analyzing the patent CPC code as link prediction
Lim, Seonyong [2020]. An analysis of patent trends in AIoT (AIoT) core technology using big data.	A study on the direction of technology development through the analysis of the patent data network/centrality to derive core technologies and the analysis of networks such as Korea, the U.S., and Japan
Cho, Kyung-cheol [2014]. A sectoral comparison of the influence of the intellectual property rights system on technological innovation and financial performance: Korean pharmaceutical, semiconductor and shipbuilding industries	The path analysis of the semiconductor industry knowledge property rights system that affects the technology innovation and development/san Industrial Knowledge Property rights system that affects corporate technology innovation and management and management performance Application
Triulzi, G. and Magee, C. L. [2020]. "Functional performance improvement data and patent sets for 30 technology domains with measurements of patent centrality and estimations of the improvement rate".	30 technology areas of functionality, patent sets, patent centrality measurements and estimates of annual technology performance improvement rates

허 데이터와 논문 데이터가 기술동향을 파악할 수 있는 가장 적절한 자원이라고 보고, 분석 대상을 인공지능반도체의 특허와 논문 데이터로 선정하였다. 특히 논문과 특허데이터를 기반으로 진행된 선행 연구를 분석해보면 특허와 논문 데이터의 통계를 기반으로 각 특허와 논문의 중요성을 평가하거나, 내용을 분석하여 기술 동향을 파악하기도 한다. 대량의 텍스트에서 패턴을 찾아내는 텍스트 마이닝 기법 중, 기술동향을 파악하는데 있어서 가장 많이 사용 되는 분석 방법론은 토픽 모델링이다. 토픽 모델링은 데이터를 평가하는데 범용적으로 사용되는 분석 도구로써, 오랜 기간 연구되어 다양한 방법이 제시되어 왔다[Vayansky and Kumar, 2020]. 이에 다음 절에서는 토픽 모델링 관련 선행 연구 분석을 진행하여, 가장 적절한 분석 방법론을 도출한다.

2.3 텍스트 마이닝 관련 기존 연구

2.3.1 텍스트 마이닝 관련 기존연구

텍스트 마이닝은 데이터 마이닝의 한 분야로, 텍스트

에 숨어있는 패턴을 찾아내는 것 이라고 할 수 있다 [Hearst, 2003]. 텍스트 마이닝은 Feldman and Dagan이 텍스트 데이터 기반의 지식의 발견이라는 개념으로 1995년 처음 제시한 개념이다. 특히 최근 온라인에서 많은 정보들이 생산되고 축적되며 계량화 된 정형 데이터 뿐만 아니라 다양한 형태의 비정형 데이터들이 생성되고 있으며, 텍스트 데이터는 그중 가장 대표적인 비정형 데이터로 여겨진다. 방대한 텍스트 데이터들이 쌓이며 텍스트 기반 데이터를 분석하고 활용하는 방법론과 연구들이 등장하기 시작했는데, 뉴스 텍스트 데이터의 분석을 통해 투자 의사결정 모형을 개발하는 연구 [Hahm and Lee, 2016], 사용자 리뷰 텍스트 마이닝을 통해 자산관리 어플의 서비스 개발에 있어서 사용자 니즈에 대해 분석한 연구(Park, 2022), 음식점 리뷰 감성 분석을 통해 평점을 예측 한 연구(So and Shin, 2020) 등 텍스트 마이닝은 다양한 산업에서 다양한 용도로 활용되고 있다.

주로 사용되는 텍스트 마이닝 분석 기법으로는 형태소 및 키워드 분석, 오피니언 마이닝, 토픽모델링, Word2Vec 등이 있다(Kim et al., 2016). 먼저, 텍스

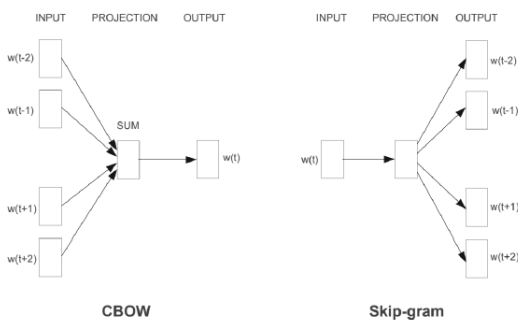
트 마이닝을 위해서는 문장으로 이루어진 데이터를 형태소 분석을 통해 개별 단어로 분리하는 과정이 필요하다. 형태소 및 키워드 분석 기법은 이러한 형태소 분석을 통해 명사, 용언 등의 형태소를 추출하여 발생 빈도를 파악하고, 문서의 주제 또는 문서에 대한 태도나 감성을 추론할 수 있으며, 키워드 간의 동시 출현 빈도를 파악하여 단어들 간 네트워크 분석을 통해 관련성을 파악할 수 있는 기법이다(Kim et al., 2016). 두 번째로, 오피니언 마이닝은 감성분석, 감성분류, 오피니언 추출 등을 포함하는 개념으로, 주로 소셜 미디어 콘텐츠로부터 이용자들의 의견을 분석하는 데 있어 사용되는데 최근 여러 연구에 의해 AI등 다양한 분석 기법들이 적용되며 성숙되어오고 있다(Kang et al., 2013; Kim et al., 2016). 세 번째로, 토픽 모델링은 가장 범용적으로 사용되는 텍스트 마이닝 기법으로, 다음 절에서 자세히 살펴본다. 마지막으로 Word2vec은 일정 규칙에 따라 단어를 벡터로 변화시켜 공간에 투사하는 워드 임베딩 방식 중 하나로, 토픽모델링에서도 자주 사용되는 기법이다. 먼저 한 단어에 대해 인접하게 출현하는 관련단어들을 인공 신경망을 사용하여 학습시킨다. 단어의 순서가 근접하게 자주 출현할 수록 유사한 값을 가지게 되며 특정 단어의 앞뒤에 위치한 단어들의 분포를 고려하여 벡터 값으로 변환하여 신경망에 학습시킨다(Mikolov et al., 2013). 학습 알고리즘으로는 CBOW(Continuous Bag of Words)와 Skip Gram이 있으며, 각각의 구조는 아래와 같다. 앞선 텍스트 마이닝 기법들은 불용어들을 삭제하는 데이터 전 처리 과정이 필요하지만, Word2vec기법은 불용어를 삭제하지 않은 텍스트 그 자체를 학습 데이터로 사용한다는 장점이 있다. 그러나 Word2vec은 문장 내 동시 출현 정보에 기반하여 벡터 값으로 단어를 학습하

로, 어떤 단어가 다른 의미로 사용되었다라든가 그 차이를 구별하지 못한다는 단점이 있다(Mikolov et al., 2013).

정리하면, 텍스트 마이닝은 텍스트 기반 빅데이터에서 구조화된 패턴을 추출하는 것을 의미하며 비정형 데이터 안에서 구조화된 패턴을 찾아 데이터를 정제한 주요 토픽을 추출할 수 있다(Nam et al., 2013). 본 연구의 목적인 인공지능 반도체 산업의 기술 동향을 파악하기 위해 가장 적절한 방법론은 토픽 모델링으로, 토픽 모델링에 관한 선행연구를 통해 어떤 토픽 모델 방법론을 채택하는 것이 가장 적절한지 살펴본다.

2.3.2 토픽 모델링 방법론 관련 선행연구

토픽 모델링은 텍스트 마이닝 기법 중에서 가장 많이 활용되는 기법 중 하나로, 다양한 문서 집합에 내재한 토픽, 즉 주제를 파악할 때 쓰는 방법이다. 따라서 토픽 모델링은 예측보다는 내용의 분석 자체를 목적으로 하는 기법이라 할 수 있다. 토픽 모델링은 단어의 동시 출현 빈도에 따라 계산된 유사한 단어를 하나의 주제로 군집하는 것인데, 단순 유사도를 기준으로 군집시키는 것이 아니라, 그 단어가 관찰된 문서와 그 단어에 대한 통계적 확률이나 우도와 같은 수치에 기반하고 있다. 토픽 모델링은 다양한 연구에서 변화 및 동향 분석을 위해 사용된다. 토픽 모델링은 1980년대의 생성적 확률 모델링(Generative Probabilistic Modeling)에서 분리된 개념으로(Liu et al., 2016), 통계적 관계를 고려하여 대규모 데이터의 분류 및 요약 수행할 필요에 의해 제시되었다. 이런 토픽 모델링을 수행하기 위해 개발된 첫 번째 방법은 TF-IDF 감소 체계로, Salton과 Macgil에 의해 고안되었다. 문서마다 단어의 중요도를 고려하여 가중치를 주는 통계적인 단어 표현 방법으로 TF(단어 출현빈도, Term Frequency) * IDF(역문서 빈도, Inverse Document Frequency)로 계산된다. TF는 특정 문서 d에서 특정 단어 t의 출현빈도를 계산한다. IDF는 모든 단어에서 공통적으로 발생하는 단어들(~합니다, ~에서, ~가 등의 서술어나 조사와 같은 불용어)은 오히려 중요성이 떨어지기 때문에, 전체문서 D에서 특정단어 t가 출현한 문서의 역수로 계산한다. 단어 출현 빈도와 역 문서빈도를 모두 고려하기 때문에, 전체 문서 집합에서 문서를 구별하는 단어 집합을 식별하는



(Figure 2) Word2VEC(mikolove et al., 2013)

〈Table 3〉 Summary of Prior Research Related to Topic Modeling Methodology

Research Literature	Research Contents	Application Methodology
A patent trademark data linking approach to business analytics for technology-based firms [Ko, 2021]	Analyzing Business Through Patent-Trademark Data Connection Framework	
An analysis of user needs by user's review, based on text-mining : focusing on Bank-Salad and Mint [Park, 2022]	Analysis of needs through user review text mining	LDA
Research Trend Analysis on Living Lab Using Text Mining [Kim, 2021]	Identify research trends by analyzing living lab research literature	LDA + Keyword Network Analysis
A Study of foresight method based on textmining and complexity network analysis [Jeong, 2010]	Trend prediction method through text mining and network analysis	LDA+Link Forecasting Approach
Changes in Online Shopping Mall Trends Before and After COVID-19 Using Topic Modeling Analysis Techniques Analysis [Cho, 2021]	An Analysis of Shopping Mall Trends Based on News Data	Dynamic Topic Modeling (LDA)

데는 효과적이나, 단어의 빈도 수를 이용하였기 때문에 단어의 의미를 반영하지 못한다는 단점이 있다(Kherwa and Bansal, 2020). 이를 보완하기 위해 1990년 잠재된(Latent)의미를 분석하는 LSA라는 차원 감소 방법이 고안되었다(Deerwester et al., 1990). LSI에서 TF-IDF행렬은 특이 값 분해(Singular Value Decomposition, SVD)에 의해 분해되어 세 개의 개별 행렬을 생성하고, 일부 벡터들을 삭제 시킨 절단된 SVD를 사용하여 차원을 축소시킨다. 이렇게 일부 벡터들을 삭제하여 데이터의 차원을 줄이게 되면 데이터 계산비용이 줄어들고 설명력이 낮은 정보를 삭제하여 기준에 행렬에서는 드러나지 않았던 심층 정보를 알아낼 수 있다는 장점이 있다(Deerwester et al., 1990). 1999년 이러한 LSA에 확률데이터를 기반으로 문헌 내 특정 용어가 들어갈 횟수가 아닌 확률을 기반으로 행렬을 구축하여 차원 축소의 현저한 발전을 보인 pLSA가 제안되었으나(Hofmann, 1999), 새로운 정보에 대한 업데이트가 어렵고, 대상 말뭉치 크기에 따라 과적합이 발생할 수 있다는 단점이 있다. 이후 2003년 문서의 집합으로부터 어떤 토픽이 존재하는지 알아내는 LDA(잠재적 디렉클레 할당, Latent Dirichlet Allocation)이 제안되었다. LDA는 각 문서에 담긴 단어의 분포 현상이 문서별 토픽 분포의 사전 확률과 단어분포를 결정짓는 요소가 된다는 가정에 기반한 알고리즘이다. LDA는 DTM 또는 TF-IDF행렬을 입력으로 하기 때문에 단어의 순서는 고려하지 않는다(Blei et al., 2003). LDA는 정교하게

텍스트에 내재된 의미를 도출할 수 있다고 평가되며(Nam and Lee, 2019). 비지도 생성 모델이기 때문에, 토픽과 관련된 모든 단어를 쉽게 찾을 수 있다는 장점(Blei et al., 2003)이 있다. 또한 Perplexity Score, Coherence Score등을 통해 토픽모델링의 성능을 평가할 수 있으며, 이에 따라 최적 Topic 개수를 조정하여 최적의 결과를 도출할 수 있어 텍스트 기반 토픽 모델링에 가장 범용적으로 활용되고 있다. 〈Table 3〉을 통해 알 수 있듯이, 근래 진행되는 대부분의 토픽모델링 연구는 LDA에 기반하여 네트워크 분석이나 링크예측 기법 등을 적용하거나, 혹은 시계열적으로 분석하여 동적 토픽모델링을 수행하기도 한다. 그러나 너무 짧은 문서에 대해서는 성능이 떨어지며(Lee et al., 2010). 의미가 다른 동일한 단어는 다른 주제에 할당 될 수 없고, Mimno et al.[2007] 문장에서 단어의 맥락을 설명하지 않기 때문에 토픽 사이의 관계를 모델링 할 수 없다(Rani and Kumar, 2021)는 한계를 가진다. 또한 데이터 전처리가 필요하고 사용법이 어려우며, 하이퍼파라미터를 어떻게 설정하느냐에 따라 전혀 다른 결과를 야기할 수 있다(Kim, 2022).

2.3.3 BERTopic

앞선 토픽모델링 방법들은 문서를 BoW(Bag of Word)형태로 설명하고, 각 문서를 잠재 주제의 혼합물로 구성한다. 하지만 단어들을 BoW형태로 표현함으로써 단어간 의미적 관계가 무시된다는 한계가 있다

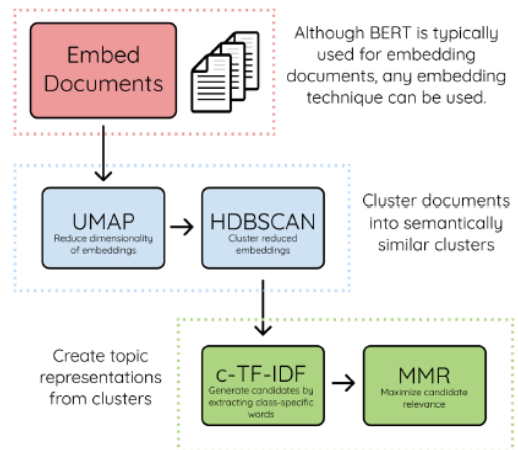
[Grootendorst, 2022]. 이러한 문제의 해결책으로써 BERT등과 같은 양방향의 트랜스포머 인코더 표현 Devlin et al.[2018]에서 파생된 연구들은 문맥적 해석이 필요한 단어를 생성하거나 문장단위의 벡터 표현을 생성하는데 큰 성과를 보였다. Doc2Vec의 단어와 문서 표현을 활용하여 공통 주제 및 단어 벡터를 학습하는 Top2Vec[Angelov, 2020] 등의 방법이 제시되었으며, Sia et al.[2020]은 밀도 기반 HDBSCAN을 활용하여 클러스터를 생성, 단어 임베딩 및 클러스터링과 동시에 가중치 클러스터링과 문서 정보를 통합하는 연구도 진행하였다. 하지만 이러한 토픽모델링 기술은 클러스터의 중심에 가까운 단어를 토픽인 것으로 가정하지만, 실제로는 클러스터가 클러스터 중심에 항상 위치하는 것은 아니므로 모든 문서 클러스터와 해당 클러스터의 표현에 대한 가정이 성립될 수 있고, 성능이 떨어질 수밖에 없다 [Grootendorst, 2022]. 이에 Grootendost[2022]는 클러스터링 기술과 c-TF-IDF의 클래스 기반 변형을 활용하여 BERTopic 기법을 제안하였다.

BERT(Bidirectional Encoder Representations from Transformers)는 언어 모델 기반의 학습을 이용하여 언어에 대한 이해를 높이는 데 목적이 있다. 기존 토픽모델링을 위해 광범위하게 사용되던 LDA나 LSA의 경우 주제의 개수에 따라 성능의 차이가 나거나 사전에 불용어처리, 형태소 분석 등의 데이터 전처리가 필요하고, 단어 간의 의미관계를 포함하지 않는다는 한계가 있다(Ko et al., 2022). BERTopic은 사전 훈련된 Transformer기반 언어 모델로 문서 임베딩을 생성하고, 이러한 임베딩을 클러스터링한 다음, 클래스 기반 c-TFIDF 절차로 토픽을 생성한다. 이러한 차원축소와 클러스터링을 바탕으로 BERTopic은 더 많은 토픽을 도출할 수 있고, 조금 더 과감한 결과가 나온다는 특징이 있다(Kim, 2022). 또한 문서 내부에 노이즈를 일으키는 HTML Tag등이 포함되지 않는 한, LDA등과 같은 BOW기반 토픽 모델링과는 달리 문서 전체의 General Topic을 파악하기 위해 전처리를 요하지 않는다는 장점이 있다[Grootendorst, 2022].

BERTopic의 토픽 추출 과정을 정리하면 아래와 같다. 단계별로 각각 SBERT, UMAP, HDBSCAN, c-TF-IDF 등이 Default로 설정되어 있으나, 제안자인 Grootendorst의 GitHub에는 처리되는 데이

터 및 이용자의 따라 다양하게 변형 가능하다고 설명하고 있다.

- 1) Embed Document: 문서를 자어 처리 모델을 활용하여 벡터 형태로 표현한다. 기본적으로 SBERT, Transformer, spaCy가 사용되며 이외 다른 임베딩 기술을 적용할 수 있다.
- 2) Cluster Embeddings: 클러스터링이 용이할 수 있도록 UMAP, PCA, Truncated SVD 등을 사용하여 임베딩 된 벡터 표현들의 차원을 축소한다. 이후 HDBSCAN, K-Means, BIRCH 등을 사용하여 축소된 벡터를 클러스터링, 의미론적으로 유사한 문서를 도출해 낸다.
- 3) Topic Representation: 도출된 토픽을 BoW (Bag of Words)형태로 토큰화하고, 클래스 기반 TF-IDF 로 문서를 대표하는 토픽을 추출한다.



〈Figure 3〉 BERTopic Topic Extraction Process [Grootendorst, 2020]

BERTopic을 사용하여 토픽 모델링을 진행한 연구로는 Twitter 데이터를 바탕으로 BERTopic을 활용하여 사이버 테러리스트들을 식별하고 테러 커뮤니티를 분류하는 프레임워크를 제시한 연구[Saidi et al., 2022], 법률문서에서 토픽모델링을 위해 Legal-BERT로 긴 텍스트의 법률문서를 임베딩하고 품질을 평가한 연구 [Silveira et al., 2021] LDA와 BERTopic을 사용하여 전 세계 민화(Folktales)들의 공통된 경향을 분석한

연구[Werzinsky et al., 2022] 등 학계 전반에 걸쳐 다양한 도메인에서 BERTopic을 사용하여 토픽모델링을 수행해 왔다. 하지만 아직 연구 초기 단계로, 실제로 Google Scholar에 LDA를 검색하면 1,140,000개의 검색 결과가 도출되는 것과 대비하여, BERTopic은 408개의 검색 결과밖에 나오지 않는 등 대중적으로 널리 채택된 토픽모델링 방법론이라고 할 수 없다. 이에, 본 연구에서는 기존 통계적 방법론을 적용한 토픽모델링 방법 대비 단점들을 개선하여 2021년 제시된 BERTopic으로 토픽 모델링을 실행하여 인공지능 반도체 관련 연구의 토픽들을 탐색하고자 한다.

3. 연구문제 및 방법

3.1 연구문제

본 연구는 최근 제안된 토픽모델링 방법론인 BERTopic을 통해 인공지능 반도체 산업의 주요 기술들을 도출하는 것을 목표로 한다. 연구 목표를 달성하기 위해 설정된 연구문제는 다음과 같다.

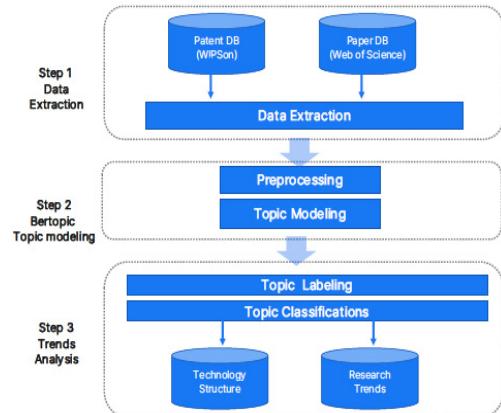
첫째, 인공지능 반도체를 분류 할 수 있는 핵심 키워드는 무엇인가? 선행 연구를 통해, 인공지능 반도체를 분류 할 수 있는 핵심 키워드를 도출한다. 둘째, 도출된 핵심 키워드로 특허, 논문 빅데이터를 추출한다. WIPSon, Web of Science을 통해 특허와 논문 빅데이터를 추출한다. 선행 연구를 통해 도출된 핵심 키워드를 바탕으로 인공지능 반도체 관련 데이터를 수집할 수 있는 검색식을 도출한다. 셋째, BERTopic을 활용하여 추출된 빅데이터에서의 핵심 키워드를 도출하고, 국가별, 연구도메인(특허, 논문) 별 기술개발, 연구 동향 토픽을 도출한다. WIPSon, Web of Science을 통해 추출된 데이터를 바탕으로 BERTopic기반 토픽모델링을 수행한다. 이를 통해 특허와 논문 별 연구 동향 토픽들을 도출한다.

3.2 연구 방법

3.2.1 연구 프레임 워크

본 연구는 <Figure 4>와 같이 구성된다. 1단계는 데이터 추출 단계로 각국의 특허 데이터를 보유하고 있는 WIPSon을 통해 특허 데이터를 수집하고, Web

of Science를 통해 논문 데이터를 수집한다. 2단계는 BERTopic 기반 토픽모델링 단계로 사전에 널리 학습되어 있는 BERT를 활용하여 Topic Modeling을 진행한다. 3단계는 학술논문 데이터 베이스와 특허 데이터 베이스를 통해 토픽 모델링을 수행한 결과를 바탕으로 기술 토픽과 연구 동향을 도출한다.



<Figure 4> Research Framework

3.2.2 데이터 수집

2절 선행연구를 통해 기관 및 학계에서 내리는 인공지능 반도체에 대한 개념을 도출하였다. 광의의 개념으로써의 인공지능 반도체는 인공지능을 처리하는데 사용될 수 있는 모든 하드웨어 타입을 포함하지만 [Kwon, 2018], 최근에는 인공지능 연산에 최적화된 형태의 인공지능 컴퓨팅 전용 반도체가 개발되고 있다. 인공지능 반도체의 특허를 분석한 선행연구 [Park, 2020; Lee and Oh, 2022]를 기반으로 기존에 인공지능 반도체 특허 관련 연구에서 채택된 검색식에, 유효한 데이터를 추가하기 위해 인공지능 반도체 설계 전문 업체들의 특허를 추가하여 정확도를 높이고자 하였다. 검색을 위해 특허 전문 검색 도메인인 WIPSon을 통해 데이터를 추출하였고, 인공지능 반도체 관련 연구가 가장 활발하게 이루어지고 있는 한국, 미국, 유럽의 등록, 심사 중, 출원, 등록 예정특허를 분석하였다. 검색된 3,542건의 검색결과 중 등록 우선 조건으로 중복된 데이터를 제거하여, 총 2,256건의 데이터를 추출할 수 있었다. 출원인은 INTEL CORPORATION(160건), GRAPHCORE LIMITED

(156건), SAMBANOVA SYSTEMS(928건), SAMSUNG ELECTRONICS(77건) 순으로 많았고, 미국 1,588건, 유럽 425건, 한국 243건 순으로 많았다. 해당 방식으로 추출된 특허 데이터의 청구항을 대상으로 토픽 모델링을 수행하였다.

3.2.3 토픽 모델링 방법

BERTopic은 Bertopic을 처음 제안한 Grootendrost의 Github에 이용 방법이 자세하게 나와 있다.¹ 기존에 학습되어 널리 사용되고 있는 자연어 처리 모델인 BERT를 사용하기 때문에 LDA와 같이 번거로운 전처리 과정이 필요 없고, 사용 방법도 매우 용이하다. 실험 환경은 Google Colab에서 GPU로 가속기를 추가한 환경에서 사용하였다. 임베딩, 차원축소, 클러스터링, 토픽 추출 단계에서는 각각 Default로 설정된 SBERT, UMAP, HDBSCAN, C-TF-IDF를 사용하였다. 또한, LDA와는 다르게 BERTopic은 토픽의 개수를 스스로 정하는데, 1,000개 이하의 문서에서는 토픽이 너무 적게 도출되는 현상이 있다. 분석 데이터가 상대적으로 적은 논문 데이터의 경우, min_topic_size api를 통해 별도로 최소 토픽의 개수를 6개로 설정하였다. 데이터가 너무 많은 특허 청구항과 같은 경우, 기본 설정 값 대로 토픽모델링을 수행하는 경우 a, the, this 등과 같은 불용어(Stop words)가 결과로 도출되는 문제가 있었다. 이에 임베딩 및 클러스터링 이후 Count Vectorizer의 Stop_Words 파라미터를 추가하여 불용어를 제거한 토픽을 도출하였으며, 이외에는 별도 설정 없이 토픽 모델링을 수행하였다. 또한 Bertopic은 차원 축소 과정에서

사용되는 UMAP의 통계적 특성 상 BERTopic은 같은 코드와 데이터를 사용하더라도 다른 결과가 도출될 수 있다. 임베딩을 초기에 생성한 이후 다양한 파라미터들을 추가하여 가장 적합한 결과가 도출될 수 있도록 유도할 수 있다. 따라서 상기된 불용어 제거 등 파라미터 등을 추가하며 최적의 결과가 도출될 수 있도록 작업을 수행하였다.

4. 연구결과

4.1 인공지능 반도체 기술 토픽

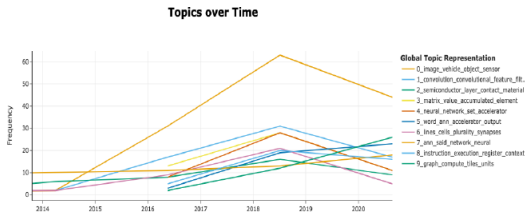
위의 검색식을 적용하여 도출된 한국, 미국, 유럽의 특허 2,256건을 모델링하여 총 48개의 토픽이 생성되었고, 상위 10개의 토픽과 각 토픽 별 주요 단어는 <Table 4>와 같다.

주제 0은 이상치 문서(Outliers)들로, 토픽이 할당되지 않은 문서들을 나타낸다. 따라서 주제 0에서 도출된 단어들은 특별한 의미를 갖고 있다고 보기 힘들다. [주제 1]은 Image, Vehicle관련 내용으로 자율주행을 위해 차량으로 수집되는 영상 정보를 처리하는 FPGA (VEHICLE-MOUNTED CONTROL UNIT, AND FPGABASED VEHICLE AUTOMATIC DRIVING METHOD AND DEVICE, BAIDU ONLINE NETWORK TECHNOLOGY(BEIJING) CO., LTD.), GPU 기반 자율주행차량 컨트롤러 (SYSTEMS AND METHODS FOR SAFE AND RELIABLE AUTONOMOUS VEHICLES, NVIDIA Corporation) 관련 특허 등이 있었다. 특허

<Table 4> Top 10 Patent Topics for Artificial Intelligence Semiconductors

topic	No. of Doc	Word 1	Word2	Word 3	Word 4
0	821	data	network	second	processing
1	151	image	vehicle	object	sensor
2	67	convolution	convolutional	feature	filter
3	56	semiconductor	layer	contact	material
4	52	matrix	value	accumulated	element
5	47	neural	network	set	accelerator
6	45	word	ann	accelerator	output
7	44	lines	cells	plurality	synapses
8	43	ann	said	network	neural
9	41	instruction	execution	register	context
10	40	graph	compute	tiles	units

데이터를 기반으로 동적 토픽 모델링을 수행한 결과, 자율주행 관련 연구가 활발해진 2010년대부터 해당 토픽이 늘어나는 것을 알 수 있었다.



<Figure 5> Patent Data Dynamic Topic Modeling Results

[주제 2]와 [주제 5], [주제 8]은 뉴럴 네트워크 처리 관련 토픽으로, 사람의 신경망 원리와 구조를 모방하여 만든 ANN(Artificial Neural Network)과, 은닉층을 깊게 만든 DNN을 응용하여 정보추출, 문장분류, 얼굴 인식 등의 분야에서 가장 범용적으로 사용되는 네트워크인 합성곱 신경망 CNN(Convolutional Neural Network)을 효율적으로 처리(Accelerating)할 수 있는 기술 관련 특허이다. CNN을 가속하는 기술에 관한 특허는 157건으로 가장 많았고 이는 CNN이 이미지처리에서 가장 범용적으로 사용되는 시장의 수요를 반영하였다고 할 수 있다. [주제 6]과 같이 ANN을 가속하는 Word line과 Bit line 관련한 특허가 많았다 (SYSTEM AND MEMORY FOR ARTIFICIAL NEURAL NETWORK, DEEP X), (Memory for an Artificial Neural Network Accelerator, ARM Limited) [주제 3]은 Pre-programmed resistive cross-point array for neural network

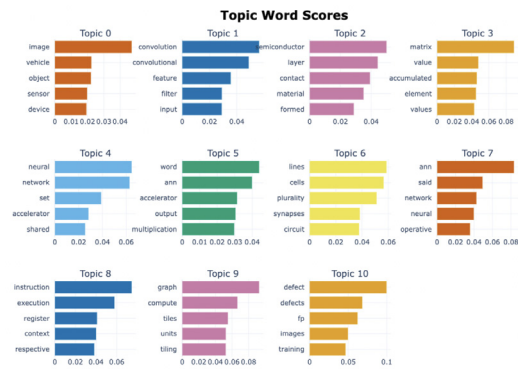
(INTERNATIONAL BUSINESS MACHINES CORPORATION)등 인공지능 반도체와 연관된 특허도 있었으나, Artificial Intelligence가 아닌 알루미늄(원소기호 AI)이 함께 검색되어 추출된 반도체 생산 관련 데이터들이 추가되어 전체 데이터를 살펴보면 실질 인공지능반도체 관련 기술 토픽이 줄어들게 된다. [주제 4]는 신경망출력을 연산하는 것과 관련된 매트릭스 연산 관련 설계 기술로 Machine Learning 연산의 효율적인 처리 방법 및 이를 하드웨어로 구현할 때의 구조에 관한 특허라고 할 수 있다(Mobile speech recognition hardware accelerator, Google), [주제 7]은 Memory Cell Array를 활용한 Matrix 연산 및 인공지능 연산 방법 및 구조와 여러 종류의 메모리에서의 In-Memory Computing 방법에 관한 특허(DEEP LEARNING NEURAL NETWORK CLASSIFIER USING NON-VOLATILE MEMORY ARRAY, SILICON STORAGE TECHNOLOGY, INC. | The Regents of the University of California) 등 메모리 관련 된 기술 토픽이 추출되었다. [주제 9]는 다수 core와 thread를 가지는 컴퓨터 구조에서 workload를 스케줄링하는 방법으로, 주로 Graphcore사의 Architecture관련 특허들이 해당 토픽으로 묶였기 때문에, 다소 지엽적인 결과가 도출되었다. [주제 10]은 GPU(및 NPU)를 활용한 Machine Learning 알고리즘 최적 동작 방식으로써, 다수의 CPU+GPU 시스템을 활용한 Machine Learning 동작 방식에 관한 기술 특허였다(System and method for accelerating and optimizing the processing of machine learning techniques using a graphics processing unit,

<Table 5> Top 10 AI Semiconductor Topics

topic	No. of Doc	Word 1	Word2	Word 3	Word 4
0	388	neural	neuromorphic	hardware	learning
1	65	snns	spiking	networks	snn
2	47	ai	convolutional	accelerator	performance
3	42	robot	control	robotic	neuromorphic
4	40	memtistor	memristors	memristive	crossbar
5	38	neurons	neuron	neuromorphic	analog
6	27	neuromorphic	systems	hardware	nengo
7	27	cnn	quantization	convolution	dnn
8	24	hardware	mapping	partitioning	snnbased
9	24	test	faults	fault	criticality
10	23	vision	eventbased	stereo	sensor

Microsoft Corporation) [주제 4], [주제 7], [주제 9], [주제 10]은 설계 기술 및 하드웨어 구현 방법에 대한 특허로, 설계 기술은 회사마다 다른 특허를 가지고 있기 때문에 특허 9와 같은 경우 Graph Core 사의 설계 기술에 대한 토픽으로 다소 지엽적으로 도출되었다.

설계기술에 대한 특허 토픽의 경우 회사마다 설계 기술에 따른 특징적인 기술들이 도출되기 때문에 산업 전반의 기술 동향을 파악하기 어렵다는 한계가 있었다.



(Figure 6) Patent Data Dynamic Topic Modeling Results

4.2 인공지능 반도체 연구 토픽

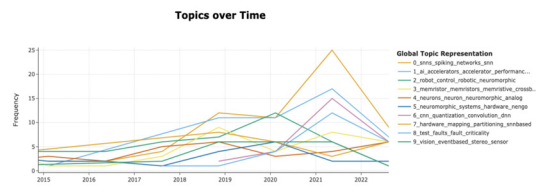
선행 연구를 통해 도출된 검색어를 통해 Web of Science에서 검색한 결과, 총 1,112개의 결과가 도출되었다. 이 중, 동적 토픽모델링을 수행하기 위해 연도 결과가 없는 데이터를 삭제하고, 토픽 모델링을 수행하기 위한 초록 정보가 없는 데이터를 삭제하여 총 1,084건의 논문을 분석 데이터로 하였다. Min Topic 수를 6으로 별도로 설정하였고, 이를 통해 총 40건의 토픽이 생성되었고, 주요 10개 토픽은 <Table 5>와 같다. 이상치(Outlier) 데이터인 [주제 0]과 가장 많은 빈도수로 등장하는 [주제 1]은 SNN관련 토픽으로 인공지능 반도체 관련 선행연구에서 4세대로 분류되는 Neuromorphic[주제 6] 관련 토픽이다. SNN(Spiking Neural Network)은 Spiking이 정보의 단위가 되는 인공신경망의 한 종류로써, 특정 시간에 특정 뉴런에서 Spike가 발생했는지에 대한 이산정보를 뉴런과 시냅스로 이루어진 인공지능 네트워크를 통해 전달한다. 기존 CNN과 같은 딥러닝 네트워크가 Tensor나 Float를 주고받는 것과 대조적이다. 아날로그[주제 5]방식으로 구현이 가능하여 낮은

전력소모와 낮은 지연속도가 강점이기 때문에 Robotics [주제 3] 산업에서의 응용이나 Edge Computing 시장에서 효율적으로 사용될 것으로 예상된다.

[주제 4]는 Memristor 기반의 Neuromorphic Processor(System) 혹은 이러한 시스템을 활용할 수 있도록 하는 학습, mapping, scheduling 등의 방법 및 프레임워크(PHAX: Physical Characteristics Aware Ex-Situ Training Framework for Inverter-Based Memristive Neuromorphic Circuits, Ansari M) 등 Memristor (Memory + Resistor) 관련 토픽이다.

[주제 7]은 Neural Network, 특히CNN에서의 Quantization 방법 혹은 이를 활용하는 Accelerator에 관한 연구 토픽으로, 가속기 연산 효율 향상을 위한 DL algorithm 경량화(quantization 및 pruning 등) 경량화된 딥러닝 알고리즘을 효율적으로 처리하기 위한 HW 구조에 관한 연구 등이 포함되어 있다(Optimized Quantization for Convolutional Deep Neural Networks in Federated Learning, Kim and Hong, 2020).

[주제 10]은 Event-driven(dynamic) vision sensor를 이용한 SNN(spiking neural network)기반 저전력(stereo) vision system에 관한 연구로, SNN의 특성이 저전력임을 감안하였을 때, 저전력 Vision Application을 위한 효율적인 딥러닝 알고리즘 관련 연구라고 할 수 있다. 위와 같이, [주제 1], [주제 3], [주제 4], [주제 5], [주제 6], [주제 7], [주제 10]은 Neuromorphic Hardware 관련 연구로, 아래 동적 토픽 모델링 연구결과를 보면 최근에서야 관련 연구가 급격하게 증가했음을 알 수 있다.



(Figure 7) Thesis Data Dynamic Topic Modeling Results

[주제 2]는 인공지능 반도체의 성능을 향상시키기 위한 모델 개선 방안에 관한 연구토픽이다. CNN실행

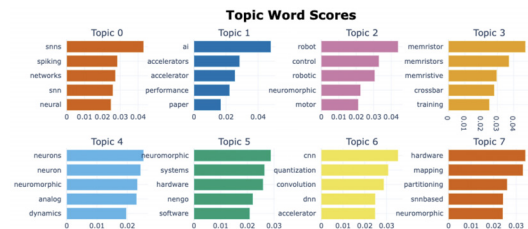
시간을 추정하기 위한 분석 모델, 내결함성(Fault Tolerance)을 향상시킬 수 있는 뉴럴넷 개선방안, 신경망 알고리즘 라이브러리, Testing 방안 등 인공지능 반도체를 효율적으로 사용할 수 있는 DNN 개선 방안에 관한 연구들이 있었다. [주제 8]은 DNN이나 SNN에서 하드웨어를 효율적으로 이용하기 위한 최적화 및 매핑 방법으로 DNN(deep neural network)에서 SNN(spiking neural network)의 전환이나, SNN and neuromorphic HW co-design 등과 같은 Hardware compiling/synthesis(mapping/scheduling) 방법에 대한 연구 토픽이다. [주제 9]는 AI반도체의 성능 향상을 위한 하드웨어 개선에 관한 연구 토픽으로 Systolic Array 방식의 인공지능 반도체의 Processing Elements의 구조적 결함을 분류하는 연구(Special Session: Fault Criticality Assessment in AI Accelerators, Chaudhuri et al., 2024), synaptic weight storage Elements의 결함을 도출해 내는 연구(Post-Silicon Validation Methodology for Resource-Constrained Neuromorphic Hardware, Lee et al., 2022), 자율 주행 보조시스템에 필요한 인증을 취득하기 위한 인공지능 반도체 Architecture에 관한 연구(Highly Efficient Test Architecture for Low-Power AI Accelerators, Ibtesam et al., 2019) 등이 있었다.

[주제 1], [주제 3], [주제 4], [주제 5], [주제 6], [주제 10]은 SNN 관련 토픽으로, 인공지능기술에서의 이슈중 하나인 “전력을 최소화 할 수 있는 컴퓨팅 기술” 측면에서, 기존 신경망 대비 저전력으로 인공지능 연산을 수행할 수 있을 것으로 예상되어 각광받고 있는 SNN 관련 연구가 가장 많았다. 이러한 시장 동향에 맞추어 대한민국 정부는 2022년 12월 13일 초고속 저전력 국산 인공지능 반도체 개발을 위해 8,000여억 원 규모의 K-Cloud 추진전략을 발표하였고, 1단계는 국산 NPU 고도화, 2단계는 저전력 PiM(Processing in Memory) 개발, 3단계는 극저전력 PiM 구현이라는 계획을 발표하였다. 저전력 고효율이라는 인공지능 반도체의 목표는 인공지능과 인공지능을 위한 하드웨어 발전방향과 일치하는 것을 알 수 있다. 또한 [주제 2]는 인공지능 반도체 성능 향상을 위한 모델 개선 방안, [주제 9]는 하드웨어 개선 방안으로, 인공지능 반도체 성능 향상을 위해 Hardware와 AI의 가속화가 동시에 연구됨을 알 수

있었다.

<Table 6> Top 10 Artificial Intelligence Papers Topics

No.	Topic
0	Deleted
1	Self-Driving Image Information Processing Semiconductor
2	CNN Acceleration Method Topics
3	SNN Hardware in Robotics
4	The design technology of matrix operation at the time of neural network output
5	Neuralet-specific acceleration technology
6	Neuromorphic related topics
7	CNN Quantization/ Accelerator
8	Optimization and mapping methods for efficient use of hardware in DNN or SNN
9	Hardware Improvements for Improving the Performance of Artificial Intelligence Semiconductors
10	SNN (spiking neural network)based low-power (stereo) using event driven (dynamic) vision sensor vision system



<Figure 8> Topic by Word Score

5. 결 론

5.1 연구의 시사점

토픽 모델링은 대용량의 문헌이나 텍스트 데이터를 문서 내에 존재하는 키워드를 바탕으로 문서의 주제를 통계 기반으로 찾아내는 방법론으로, 텍스트로 이루어진 빅데이터에 숨겨진 패턴을 발굴하여 기존엔 파악하기 어려웠던 동향 및 주제를 파악해 낼 수 있다는 장점이 있다. 이에 토픽모델링은 대량의 문헌 분석을 통해 학술적인 연구주제(Cho, 2018)를 파악하기도 하며, 후기 등의 데이터를 수집하여 사용자 의도 분석을 통한 제품이나 서비스의 니즈 분석을 진행하거나(Park, 2022) 혹은 SNS데이터를 수집하여 특정 주제에 대

한 여론을 파악하는 등 다양한 분야에 사용되고 있다. 기존에는 확률 기반으로 통계학적 방법론을 사용하여 기존 방법론의 단점을 보완하며 토픽모델링이 발전되어 왔으나, 인공지능을 활용한 자연어처리의 발전과 함께 새로운 토픽 모델링 방법론이 제안되고 있다. 본 연구는 비교적 최근 제안된 토픽모델링 방법론인 BERTopic을 활용하여 유망 산업인 인공지능 반도체 관련 데이터를 분석하였다.

연구 결과는 아래와 같다.

첫째, 인공지능 반도체 기술 토픽 분석 결과, 시장 수요에 맞춘 기술 개발 흐름을 볼 수 있었다. 자율주행의 발전과 함께 자율 주행을 위해 차량으로 수집되는 영상정보를 처리하는 인공지능 반도체 하드웨어에 대한 기술이 많았다. 이는 자율주행을 처리하는데 있어서 다양한 인공지능 기술이 활용되고, 각각의 기술마다 가장 최적의 형태로 인공지능 반도체를 도입하는 시장의 수요가 크다는 뜻으로, 인공지능 반도체의 주요한 수요처가 Automotive 산업이 될 수 있음을 시사한다. 또한 인공지능 신경망 처리를 가속하는 것과 관련한 토픽이 가장 많았는데, 이 중 CNN을 가속하는 기술에 관한 특허가 가장 많았다. 이는 CNN이 이미지 처리에서 가장 범용적으로 사용되는 시장의 수요를 반영하였다고 할 수 있다. 마지막으로 설계기술에 대한 토픽들이 있었는데, 설계기술에 대한 특허 토픽과 같은 경우, 회사마다 설계 기술에 따른 특징적인 기술들이 토픽으로 분류되었기 때문에 설계기술의 경우 산업 전반의 기술 동향을 파악하기 어려운 데이터라는 토픽모델링의 한계가 있었다.

둘째, 인공지능 반도체 연구 동향 분석 결과 인공지능 반도체 기술에서의 이슈 중 하나인 저전력 고효율의 하드웨어라는 측면에서, 기존 신경망 대비 저전력으로 인공지능 연산을 수행할 수 있을 것으로 예상되어 각광받고 있는 SNN 관련 연구가 가장 많았다. 저전력 고효율이라는 인공지능 반도체의 목표는 인공지능과 인공지능을 위한 하드웨어 발전방향과 일치하는 것을 알 수 있다. 또한 [주제 2]는 인공지능 반도체 성능 향상을 위한 모델 개선 방안, [주제 9]는 하드웨어 개선 방안으로, 인공지능 반도체 성능 개선을 위해 Hardware와 AI가 동시에 연구됨을 알 수 있었다.

이 연구를 통한 시사점은 다음과 같다.

첫째, 본 연구는 그동안 별도로 연구되어 오지는 않

았으나 그 중요성이 커지고 있는 인공지능 반도체 산업에 대해, 새로운 토픽 모델링 방법론인 BERTopic을 사용하여 기술 및 연구동향을 도출해 내었다는 점에 있어 학문적 의의를 가진다. BERTopic은 국내 관련 연구 문헌이 100건 이하의 인공지능 기반 신규 토픽 방법론이다. 이를 인공지능 반도체 산업의 특허와 문헌에 대하여 실증적으로 검증함으로써 국내 토픽모델링 연구자들에게 학문적인 시사점을 제공하였다. 또한, 기존에 CPU, GPU로도 처리가 가능했으나, 최근 인공지능 산업이 성숙해 지며 그 규모가 커지게 되자, 인공지능 연산만을 위한 별도의 반도체가 개발되기 시작하였다. 본 연구는 상대적으로 초기 단계의 인공지능 반도체 산업에서 주로 연구 및 개발되고 있는 기술적 토픽에 대하여 제시하였다는 점에서 실무적 의의를 가진다.

5.2 연구 한계 및 제언

본 연구를 진행한 결과, 본 연구의 한계와 후속 연구를 위한 제언은 다음과 같다. 첫째, 본 연구는 인공지능반도체라는 다소 넓은 범위의 연구 주제를 설정하였기 때문에, 검색 키워드 및 검색 도메인에 따라 토픽 모델링의 결과가 상이할 수 있다. 이러한 한계를 보완하고 정확한 raw data를 도출하기 위해, 선행 연구 분석을 통해 범용적으로 사용되는 키워드들을 사용하고, 인공지능반도체 전문 업체들의 특허를 추가하여 검색식을 고도화 하였으나, 필요에 따라 인공지능 반도체 설계 트렌드, 인공지능 반도체 응용 산업 등 주제를 세분화하여 분석 대상을 좁힌다면 조금 더 세분화된 주제를 도출할 수 있을 것이다. 또한 데이터 모수의 한계로 인해 본 연구에서는 한국, 미국, 유럽에 대한 특허 전체를 대상으로 토픽모델링을 수행하였다. 미국과 유럽, 중국은 각국의 반도체 산업을 위해 법안을 발의하여 국가 정책의 하나로 삼고 있는 만큼, 인공지능 반도체 관련 산업이 어느 정도 성숙해진 시점에 추가된 국가별 특허데이터를 대상으로 토픽모델링을 진행하여 국가별로 어떤 연구들이 진행되는지 비교해보는 연구도 유의미할 것이다.

둘째, 본 연구는 별도의 모델 학습을 거치지 않고 기존에 범용적으로 사용되는 자연어 처리 모델을 그대로 채택하여 결과를 도출하였다. 따라서 토픽 모델링

을 하는 산업 도메인에 맞추어 모델을 구축한 후 토픽 모델링을 수행한다면 더 나은 결과를 도출할 수 있을 것으로 기대된다. 또한 1,000건 이하의 문서에 대해서는 너무 작은 토픽이 도출될 수 있기 때문에, 적절한 수준의 양질의 데이터를 바탕으로 토픽모델링을 수행한다면 좀 더 정확한 결과를 도출할 수 있을 것이다.

마지막으로, BERTopic은 임베딩 기반 토픽 모델링 방법이기 때문에, 토픽 개수를 별도로 설정하지 않는 한, 클러스터 개수를 지정하지 않고 클러스터링 시 자동으로 토픽 개수가 결정된다[Wang et al., 2018]. 따라서 기존의 토픽모델링 방법론과 다르게, 임베딩 기반 방법론들이 과도하게 클러스터를 생성하는 경향이 있다는 연구 결과[Thielmann et al., 2021]도 존재하며, BERTopic을 별도로 평가할 수 있는 방법이 없다고 알려져 있다. 이에, BERTopic 토픽 개수 설정을 위해 LDA와 결합하여 새로운 토픽 모델링 방법론을 제시하거나[Kim, 2022], 도출된 토픽을 Visualization을 통해 2차원 영역으로 시각화 한 후, 가까운 토픽들을 병합하여 토픽을 병합하는 등[Ebeling et al., 2021] 다양한 방법으로 임베딩 기반 토픽모델링의 한계를 개선하려는 시도들이 이어지고 있다. 또한 FGI(Focus Group Interview)등의 탐색적 연구와 결합하여 도출된 키워드들을 바탕으로 심층적인 분석을 진행한다면, 더욱 조금 더 정밀한 결과를 낼 수 있을 것이다.

References

- [1] Angelov, D., "Top2vec: Distributed representations of topics", arXiv preprint arXiv:2008.09470, Aug 19, 2020.
- [2] Atzeni, D., Bacciu, D., Mazzei, D., and Principe, G., "A Systematic Review of Wi-Fi and Machine Learning Integration with Topic Modeling Techniques", *Sensors*, Vol. 22, No. 13, 2022, p. 4925.
- [3] Batra, G., Jacobson, Z., Madhav, S., Queirolo, A., and Santhanam, N., "Artificial-intelligence hardware: New opportunities for semiconductor companies", *McKinsey and Company*, January 2, 2019.
- [4] Bergstrom, C. T., West, J. D., and Wiseman, M. A., "The eigenfactor™ metrics", *Journal of Neuroscience*, 2019, Vol. 28, No. 45, pp. 11433-11434.
- [5] Blei, D. M., Ng, A. Y., and Jordan, M. I., "Latent dirichlet allocation", *Journal of Machine Learning Research*, Vol. 3, Issue, Mar, 2003, pp. 993-1022.
- [6] Chaudhuri, A., Talukdar, J. and Chakrabarty, K., "Special Session: Fault Criticality Assessment in AI Accelerators", 2022 IEEE 40th VLSI Test Symposium (VTS), San Diego, CA, USA, Apr, 2022, pp. 1-4.
- [7] Choi, S. and Yeon, S., "ETRI AI Strategy# 2: Strengthening Competencies in AI Semiconductor & Computing Technologies", *Electronics and Telecommunications Trends*, Vol. 35, No. 7, 2020, pp. 13-22.
- [8] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R., "Indexing by latent semantic analysis", *Journal of the American Society for Information Science*, Vol. 41, No. 6, 1990, pp. 391-407.
- [9] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding", arXiv preprint arXiv:1810.04805, 2018.
- [10] Drucker, P. F., "Managing in a time of great change", Oxford: Butterworth Heinemann, 1995.
- [11] Ebeling, R., Sáenz, C. A. C., Nobre, J., and Becker, K., "The effect of political polarization on social distance stances in the brazilian covid-19 scenario", *Journal of Information and Data Management*, 2021, Vol. 12, No. 1.
- [12] Garfield, E., "The impact factor", *Current Contents*, Vol. 25, No. 20, 1994, pp. 3-7.

- [13] Grzeszczyk, T. A. and Grzeszczyk, M. K., "Improving the discovery of technological opportunities using patent classification based on explainable neural networks", *European Research Studies Journal*, Vol. 24, No. 3, 2021, pp. 402-409.
- [14] Gu, J., Lee, J., Chung, M., and Lee, J., "Electric Vehicle Technology Trends Forecast Research Using the Paper and Patent Data", *Journal of Digital Convergence*, Vol. 15, No. 2, 2017, pp. 165-172.
- [15] Hahm, Y. and Lee, S., "The distinctiveness of big data business model in its components: A comparative analysis of Korea-US cosmetic big data business cases", *Journal of Information Technology and Architecture*, 2016, Vol. 13, No. 1, pp. 63-75.
- [16] Hearst, M., "What is text mining?" SIMS, UC Berkeley, Oct, 17, 2003.
- [17] Hofmann, T., "Probabilistic latent semantic indexing", Paper presented at the Proceedings of the 22nd annual international ACM SIGIR Conference on Research and Development in Information Retrieval, 1995.
- [18] Hung, M., Park, S., Chae, B., and Lee, J., "Analysis of Major Research Trends in Artificial Intelligence through Analysis of Thesis Data", *Journal of Digital Convergence*, Vol. 15, No. 5, May 28, 2017, pp. 225-233.
- [19] Ibtisam, M., Solangi, U., Kim, J., Ansari M., and Park, S., "Highly Efficient Test Architecture for Low-Power AI Accelerators", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 41, No. 8, Aug, 2022, pp. 2728-2738
- [20] Jung, M., Park, S., Chae, B., Lee, J., "Analysis of major research trends in artificial intelligence through analysis of thesis data", *Journal of Digital Convergence*, Vol. 15, No. 5, 2017, pp. 255-233.
- [21] Kang, B., Song, M., and Cho, H., "A Study on Opinion Mining of Newspaper Texts based on Topic Modeling", *JAMS*, Vol. 47, No. 4, Nov, 2013, pp. 315-334.
- [22] Kherwa, P. and Bansal, P., "Topic Modeling: A Comprehensive Review", *EAI Endorsed Transactions on Scalable Information Systems*, Vol. 7, No. 24, 2020, pp. 1-16.
- [23] Kim, S., "Topic Model Augmentation and Extension Method using LDA and BERTopic", *KOSIM*, Vol. 39, No. 3, 2022, pp. 99-132.
- [24] Kim, S., Cho, H., and Kang, J., "The Status of Using Text Mining in Academic Research and Analysis Methods", *The Journal of Information Technology and Architecture*, Vol.13, No. 2, 2016, pp. 317-329.
- [25] Kim, T., Choi, H., and Lee, H., "A Study on the Research Trends in Fintech using Topic Modeling", *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 17, No. 11, Nov. 30, 2016, pp. 670-681.
- [26] Kim, Y., "A Study on the Analysis of AI Semiconductor Industry Trends and Implications", *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, Jun, 2018, pp. 45-1104.
- [27] Ko, Y., Lee, S., Cha, M., Kim, S., Lee, J., Han, J., and Song, M., "Topic Modeling Insomnia Social Media Corpus using BERTopic and Building Automatic Deep Learning Classification Model", *JKOSIM*, Vol. 39, No. 2, 2022, pp. 111-129.
- [28] Kwon, Y., "AI Processor Technology

- Trends”, *Electronics and telecommunications trends*, Vol. 33, No. 5, Oct 1, 2018, pp. 121-134.
- [29] Kwon, Y., “Understanding of structural changes of keyword networks in the computer engineering field”, *KIPS Transactions on Software and Data Engineering*, Vol. 2, No. 3, 2013, pp. 187-194.
- [30] Lee, J. and Oh, C., “A study on the technological development path of the AI semiconductor industry and the catch-up chance for latecomers: Focusing on technical patent analysis as the view of the technological life cycle”, *Innovation studies*, Vol. 17, No. 3, 2022, pp. 113-133.
- [31] Lee, J., “Research on Trend of Solar-Photovoltaic(PV) Technology Using Papers and Patents Data: Using LDA Algorithm”, Graduate School of Information, Yonsei University, 2017, Seoul.
- [32] Lee, J.-H., Lee, I.-S., Jung, K.-S., Chae, B.-H., and Lee, J.-Y., “Patents and papers trends of solar-photovoltaic (PV) Technology using LDA algorithm”, *Journal of Digital Convergence*, Vol. 15, No. 9, 2017, pp. 231-239.
- [33] Lee, M., Chung, J., Lee, J., Han, J., and Kwon, Y., “Trends in AI processor technology”, *Electronics and Telecommunications Trends*, Vol. 35, No. 3, 2020, pp. 66-75.
- [34] Lee, S., Song, J., and Kim, Y., “An empirical comparison of four text mining methods”, *Journal of Computer Information Systems*, Vol. 51, No. 1, 2010, pp. 1-10.
- [35] Lee, W., “Analyzing technological structure and trends of artificial intelligence: Using patent and open source project data”, 2021.
- [36] Lee, Y., Nambiar, V., Goh, K. and Tuan Do, A., “Post-Silicon Validation Methodology for Resource-Constrained Neuro-morphic Hardware”, *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, Singapore, Oct, 2020, pp. 3836-3840.
- [37] Liu, L., Tang, L., Dong, W., Yao, S., and Zhou, W., “An overview of topic modeling and its current applications in bio-informatics”, *SpringerPlus*, Vol. 5, No. 1, 2016, pp. 1-22.
- [38] Mikolov, T., Chen, K., Corrado, G., and Dean, J., “Efficient estimation of word representations in vector space”, *Proceedings of Workshop at ICLR*, 2013.
- [39] Mimno, D., Li, W., and McCallum, A., “Mixtures of hierarchical topics with pachinko allocation”, Paper presented at the *Proceedings of the 24th International Conference on Machine Learning*, 2017.
- [40] Nam, S. and Lee, H., “Airline Passenger Characterizations Using LDA Topic Modeling”, *Korean management science review*, Vol 36, No. 3, Sep, 2019, pp. 67-85.
- [41] Noh, S., “Analysis of Issues Related to Artificial Intelligence Based on Topic Modeling”, *Journal of Digital Convergence*, Vol. 18, No. 5, May 28, 2020, pp. 75-87.
- [42] Park, D., “An analysis of user needs by user’s review, based on text-mining : focusing on Bank-Salad and Mint”, Graduate School of Information, Yonsei University, 2022, Seoul.
- [43] Park, J., “Analyzing AI semiconductor trends and global business activities using patent data”, Graduate School of Information, Yonsei University, 2020, Seoul.
- [44] Park, J., Hong, S., and Kim, J., “A Study

- on Science Technology Trend and Prediction Using Topic Modeling”, *Journal of the Korea Industrial Information Systems Research*, Vol. 22, No. 4, August 31, 2017, pp. 19-28.
- [45] Pessoa, A., “Innovation and Economic Growth: What is the actual importance of R&D?”, *Universidade do Porto, Faculdade de Economia do Porto, FEP Working Papers*, No. 254, Nov, 2007.
- [46] Rani, S. and Kumar, M., “Topic modeling and its applications in materials science and engineering”, *Materials Today: Proceedings*, Vol 45, 2021, pp. 5591-5596.
- [47] Saidi, F., Trabelsi, Z., and Thangaraj, E., “A novel framework for semantic classification of cyber terrorist communities on Twitter”, *Engineering Applications of Artificial Intelligence*, Vol. 115, 2022, pp. 105271.
- [48] Shefer, D. and Frenkel, A., “R&D, firm size and innovation: An empirical analysis”, *Technovation*, Vol. 25, No. 1, 2022, pp. 25-32.
- [49] Silveira, R., Fernandes, C., Neto, J. A. M., Furtado, V., and Pimentel Filho, J. E., “Topic modelling of legal documents via legal-bert”, *SSRN Electronic Journal*, 2021.
- [50] So, J. and Shin, P., “Rating Prediction by Evaluation Item through Sentiment Analysis of Restaurant Review”, *Journal of the Korea Society of Computer and Information*, Vol. 25, No. 6, June 30, 2020, pp. 81-89.
- [51] Tekic, Z., Kukolj, D., Drazic, M., and Vitas, M., “Towards Understanding The Role and Value of Patents in a Knowledge-Based Economy”, *DAAAM International Scientific Book*, 2003.
- [52] Thielmann, A. F., Weisser, C., Kneib, T., and Saefken, B., “Coherence-Based Document Clustering”, *2023 IEEE 17th International Conference on Semantic Computing (ICSC)*, Laguna Hills, CA, USA, 2023, pp. 9-16.
- [53] Vayansky, I. and Kumar, S. A. P., “A review of topic modeling methods”, *Information systems*, Oxford, Vol. 94, 2021, p. 101582.
- [54] Wang, J. and Hsu, C., “A topic-based patent analytics approach for exploring technological trends in smart manufacturing”, *Journal of Manufacturing Technology Management*, Vol. 193, No. 1, 2020, pp. 110-135
- [55] Wang, Y., Shi, Z., Guo, X., Liu, X., Zhu, E., and Yin, J., “Deep embedding for determining the number of clusters”, Paper presented at the *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.
- [56] Werzinsky, J., Zhong, Z., and Zou, X., “Analyzing Folktales of Different Regions Using Topic Modeling and Clustering”, *arXiv.2206.04221*, 2022.

〈부록 1〉 특허 데이터 토픽 모델링결과

번호	주제	문장개수		단어1	단어2	단어3	단어4
0	-1	821	36.39%	data	network	second	processing
1	0	151	6.69%	image	vehicle	object	sensor
2	1	67	2.97%	convolution	convolutional	feature	filter
3	2	56	2.48%	semiconductor	layer	contact	material
4	3	52	2.30%	matrix	value	accumulated	element
5	4	47	2.08%	neural	network	set	accelerator
6	5	45	1.99%	word	ann	accelerator	output
7	6	44	1.95%	lines	cells	plurality	synapses
8	7	43	1.91%	ann	said	network	neural
9	8	41	1.82%	instruction	execution	register	context
10	9	40	1.77%	graph	compute	tiles	units
11	10	39	1.73%	defect	defects	fp	images
12	11	38	1.68%	ladder	10	11	rails
13	12	38	1.68%	switching	program	fabric	processing
14	13	37	1.64%	model	semiconductor	training	process
15	14	34	1.51%	instructions	machine	learning	deterministic
16	15	33	1.46%	exponent	bits	bit	number
17	16	31	1.37%	gpu	graphics	shaders	instruction
18	17	31	1.37%	deep	gpu	job	learning
19	18	31	1.37%	tensor	tensors	mask	scalar
20	19	30	1.33%	reconfigurable	configurable	configuration	block
21	20	26	1.15%	transformed	inference	layers	deep
22	21	26	1.15%	frame	drawer	10	receptacle
23	22	26	1.15%	read	memory	data	buffer
24	23	26	1.15%	tile	tiles	synchronization	phase
25	24	25	1.11%	circuitry	operations	fpga	gate
26	25	24	1.06%	gpu	application	instance	learning
27	26	24	1.06%	memory	device	number	intelligence
28	27	23	1.02%	workpiece	manufacturing	semiconductor	workpieces
29	28	23	1.02%	gateway	subsystem	synchronisation	exchange
30	29	22	0.98%	said	means	charge	address
31	30	21	0.93%	nodes	processing	links	link
32	31	20	0.89%	group	alkyl	aryl	membered
33	32	18	0.80%	wafer	metrology	wafers	tool
34	33	16	0.71%	dp	watermark	ai	desultory
35	34	16	0.71%	resource	ml	dcs	cloud
36	35	16	0.71%	layout	pattern	design	opc
37	36	15	0.66%	multiplier	operand	operands	circuits
38	37	14	0.62%	virtual	vms	tier	hypervisor
39	38	14	0.62%	current	node	summation	sources
40	39	14	0.62%	ph	antigenbinding	dissociation	domain
41	40	14	0.62%	nvm	cim	nonvolatile	synaptic
42	41	14	0.62%	clock	delay	frequency	generator
43	42	12	0.53%	dataset	systems	transformations	transformed
44	43	12	0.53%	ai	accelerator	request	hardware
45	44	12	0.53%	fpga	architecture	characters	neural
46	45	12	0.53%	quantization	quantized	conv	fc
47	46	12	0.53%	transistor	terminal	electrically	voltage
48	47	10	0.44%	filter	chip	latency	block

〈부록 2〉 논문 데이터 토픽 모델링 결과

번호	주제	문서 개수	단어1	단어2	단어3	단어4
0	-1	388	neural	neuromorphic	hardware	learning
1	0	65	snn	spiking	networks	snn
2	1	47	ai	accelerators	accelerator	performance
3	2	42	robot	control	robotic	neuromorphic
4	3	40	memristor	memristors	memristive	crossbar
5	4	38	neurons	neuron	neuromorphic	analog
6	5	27	neuromorphic	systems	hardware	nengo
7	6	27	cnn	quantization	convolution	dnn
8	7	24	hardware	mapping	partitioning	snnbased
9	8	24	test	faults	fault	criticality
10	9	23	vision	eventbased	stereo	sensor
11	10	23	neuromorphic	computing	hardware	brain
12	11	22	photonic	optical	dl	laser
13	12	22	classification	embedded	power	image
14	13	21	ai	edge	mobile	cloud
15	14	17	switching	oxygen	resistive	rs
16	15	15	ferroelectric	thickness	film	films
17	16	15	nan	nand	npu	drambased
18	17	14	spinnaker	simulation	model	time
19	18	14	optoelectronic	synaptic	optical	plasticity
20	19	14	reservoir	rc	deep	prediction
21	20	13	learning	recurrent	truenorth	rnns
22	21	12	rram	reram	accuracy	memory
23	22	12	atomic	plasticity	synaptic	switch
24	23	11	sampling	stochastic	random	markov
25	24	10	routing	cell	mia	violations
26	25	10	stdp	rule	plasticity	learning
27	26	9	organic	posi	transistor	synaptic
28	27	9	neuron	circuit	circuitry	circuits
29	28	9	superresolution	srnpu	sr	cams
30	29	9	stdp	learning	snn	unsupervised
31	30	8	device	stp	rram	synapse
32	31	8	memristor	magnetic	memristors	memristive
33	32	8	learning	onchip	capability	speed
34	33	8	federated	mobile	learning	transfer
35	34	7	radar	gesture	processing	waveform
36	35	7	ferroelectric	antiferroelectric	polarizatio...	
37	36	7	olfactory	odor	classification	multivariate
38	37	7	ly	alpha	galaxies	galaxy
39	38	6	temperature	cooling	npu	thermal
40	39	6	ssa	patterns	vlsi	neurons

■ Author Profile



김 현 경

연세대학교 정보대학원에서 정보시스템학(IoT서비스융합) 석사학위를 취득하고, 현재 KT에서 매니저로 재직 중이다. 주요 관심 분야는 IoT 비즈니스 모델, 스마트시티, 디지털전환 등이다.



강 선 구

연세대학교 정보대학원 IoT서비스융합트랙 석사과정에 재학 중이다. 주요 관심분야는 IoT 플랫폼 서비스, 스마트시티, 디지털전환 등이다.



이 정 훈

영국 Manchester 대학교 전기 전자공학(학사), 영국 Manchester 대학교 정보시스템공학(석사) 및 영국 London School of Economics 경영정보학(석사), University of Cambridge, Institute of Manufacturing 산업공학경영(박사), 현재 연세대학교 정보대학원 정교수로 재직 중이다. 스마트시티, IT 성과관리 및 거버넌스 관련 다수 프로젝트에 참여하였으며 주요 연구 연구분야는 스마트시티, IT 성과관리 및 거버넌스, IoT, 빅데이터 서비스, 시민참여 등을 포함한다.