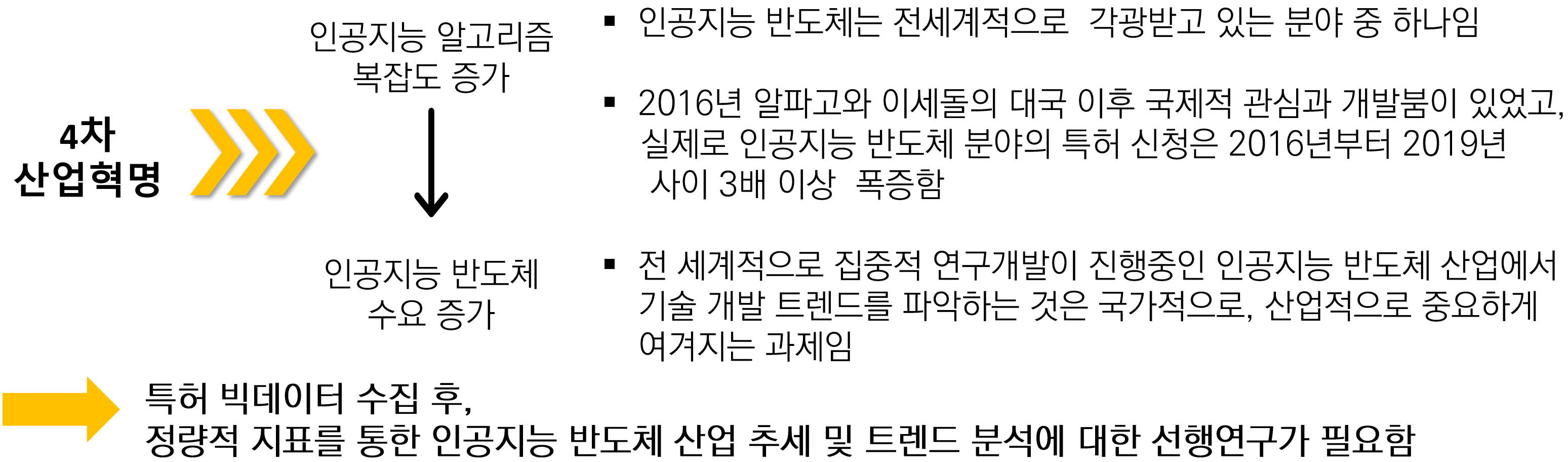


BERTopic을 활용한 인공지능 반도체 트렌드 분석 – 특허 빅데이터를 활용하여

A Study on Artificial intelligence Semiconductor Trend Analysis using BERTopic

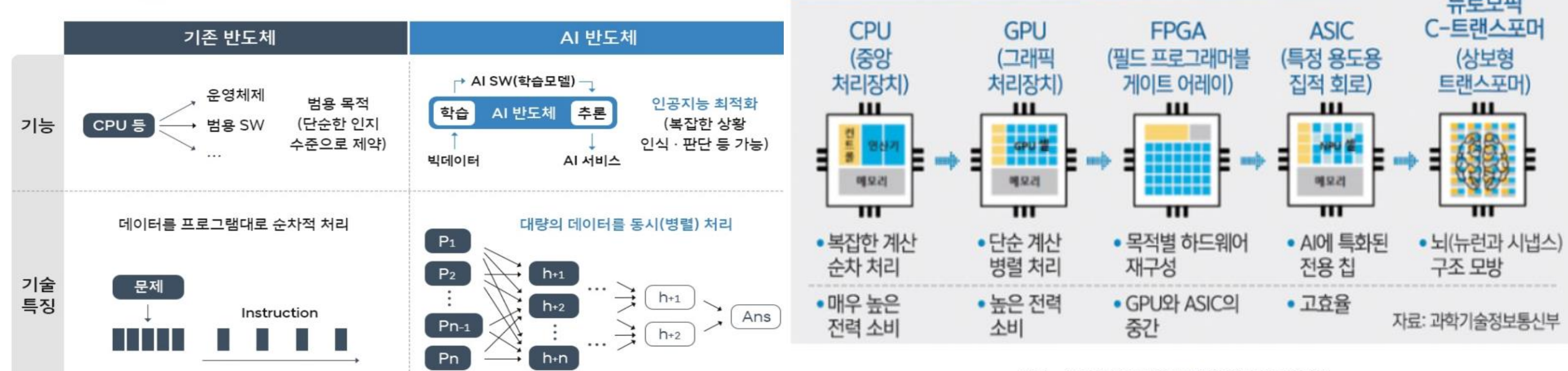
using Patent Big Data

1. 서론



기존 반도체와 AI 반도체의 비교

* 출처: 과학기술정보통신부



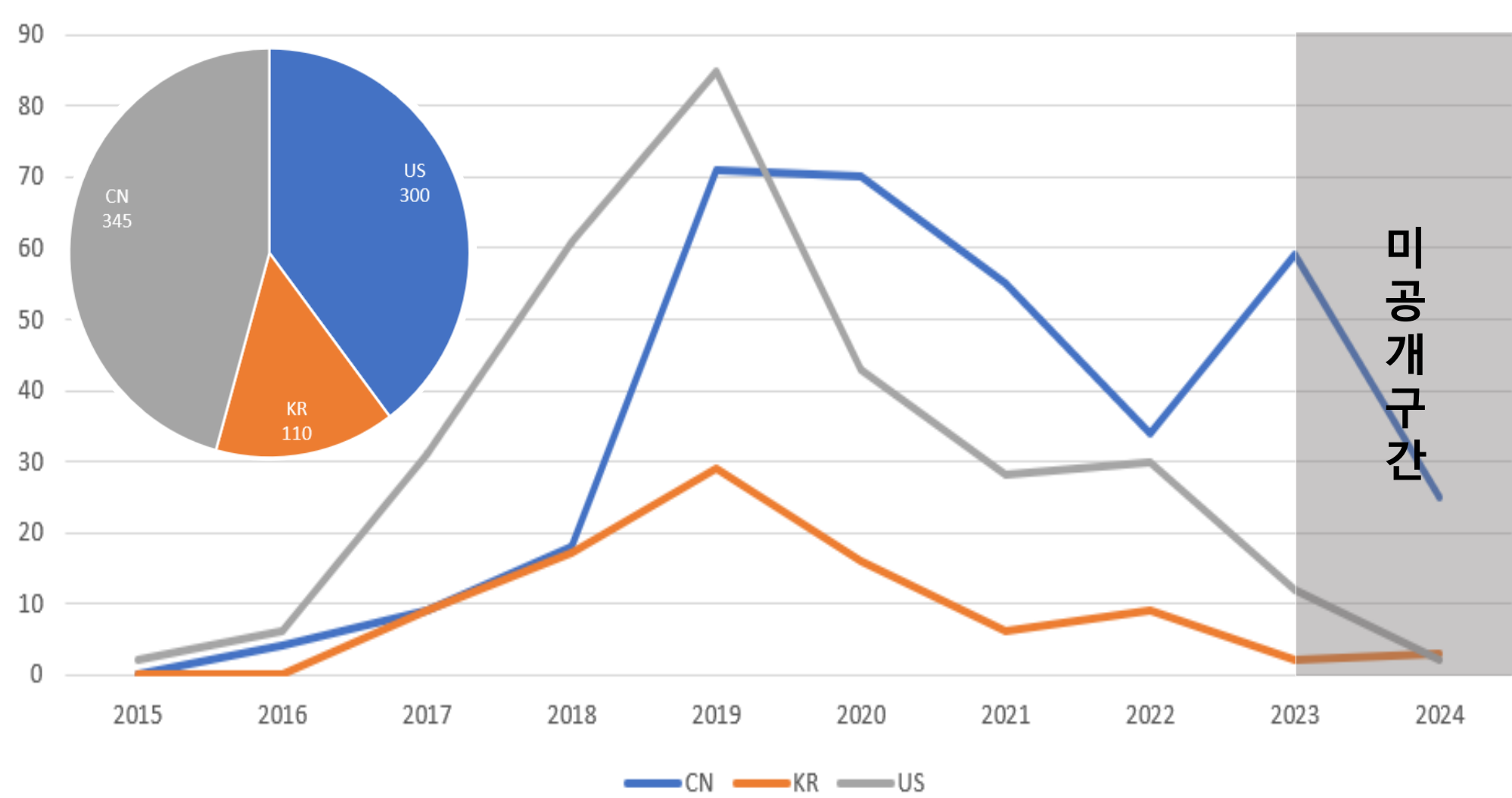
- 인공지능 알고리즘을 효율적으로 처리하기 위해 특화된 하드웨어 반도체를 의미함
- 기존의 범용 CPU나 GPU와 달리, 인공지능 연산에 최적화된 ASIC, FPGA, 뉴로모픽 칩 등이 존재함
- 인공지능 서비스 및 애플리케이션의 확산에 따라 고성능, 저전력 연산을 위해 개발되고 있으며, 연구자와 기업에 의해 다양한 정의와 분류가 이루어지고 있음

2. 관련 연구

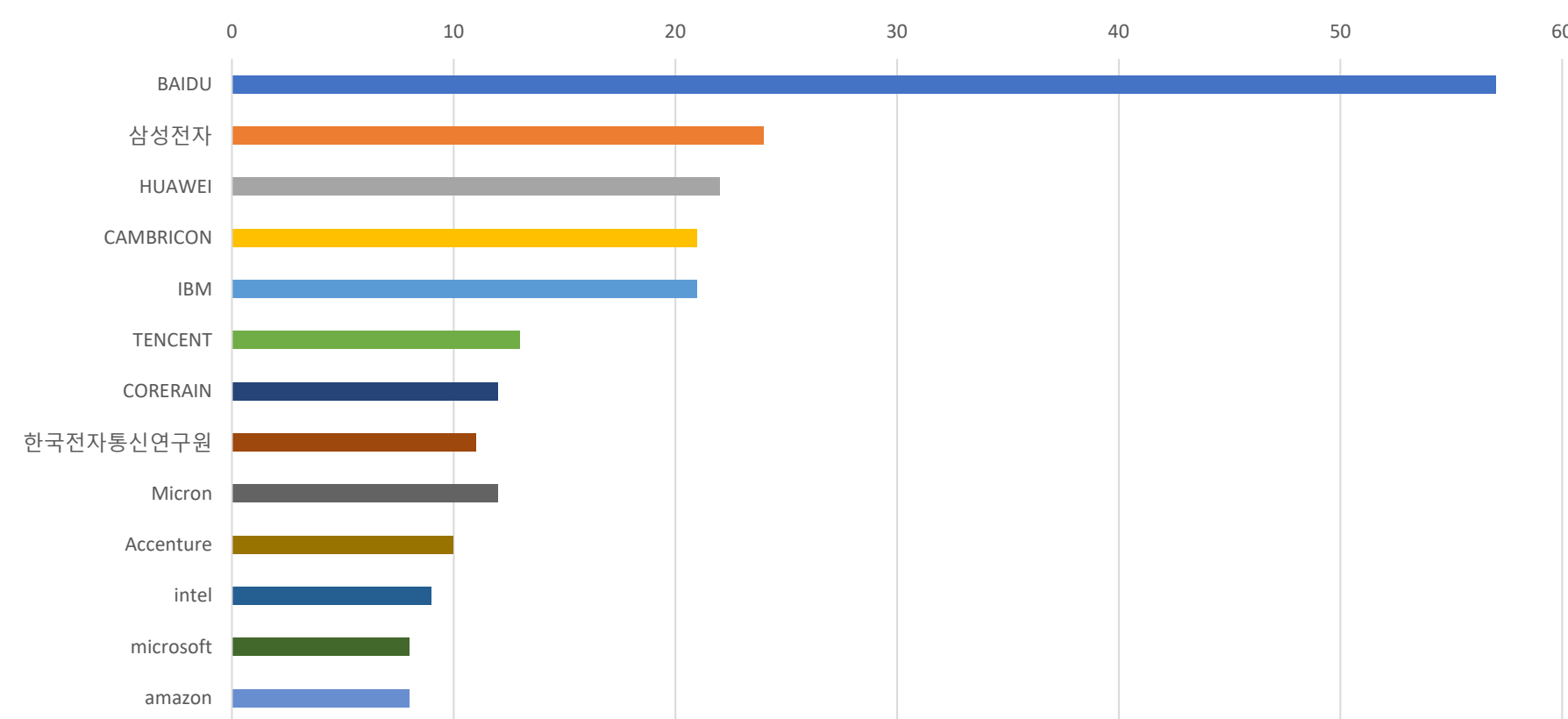
- 특허 도메인에 BERTopic을 사용하는 연구개발은 꾸준히 이루어지고 있다.
- “Iwanta”는 일본 특허 분류코드를 최신 트렌드에 적합하게 수정하기 위해 동적 토픽 모델(DTM)을 활용하였음
- 시간적 변화를 반영할 수 있는 모델을 활용하여 기술 주제의 변화를 분석하였음
- “Risch”는 정보 엔트로피를 기반으로 한 토픽 모델링을 제안하였고, 여러 도메인별에 적용시켰음
- 특허, 논문을 텍스트 데이터로 활용하였으며, 기존 토픽 모델링을 활용하였을 때 분류 정확도를 상승시키고 다양한 도메인 간의 차이점을 효과적으로 학습하였음

3. 실험 및 결과

3-1 연도별 출원 동향



3-2 상위 출원인 동향

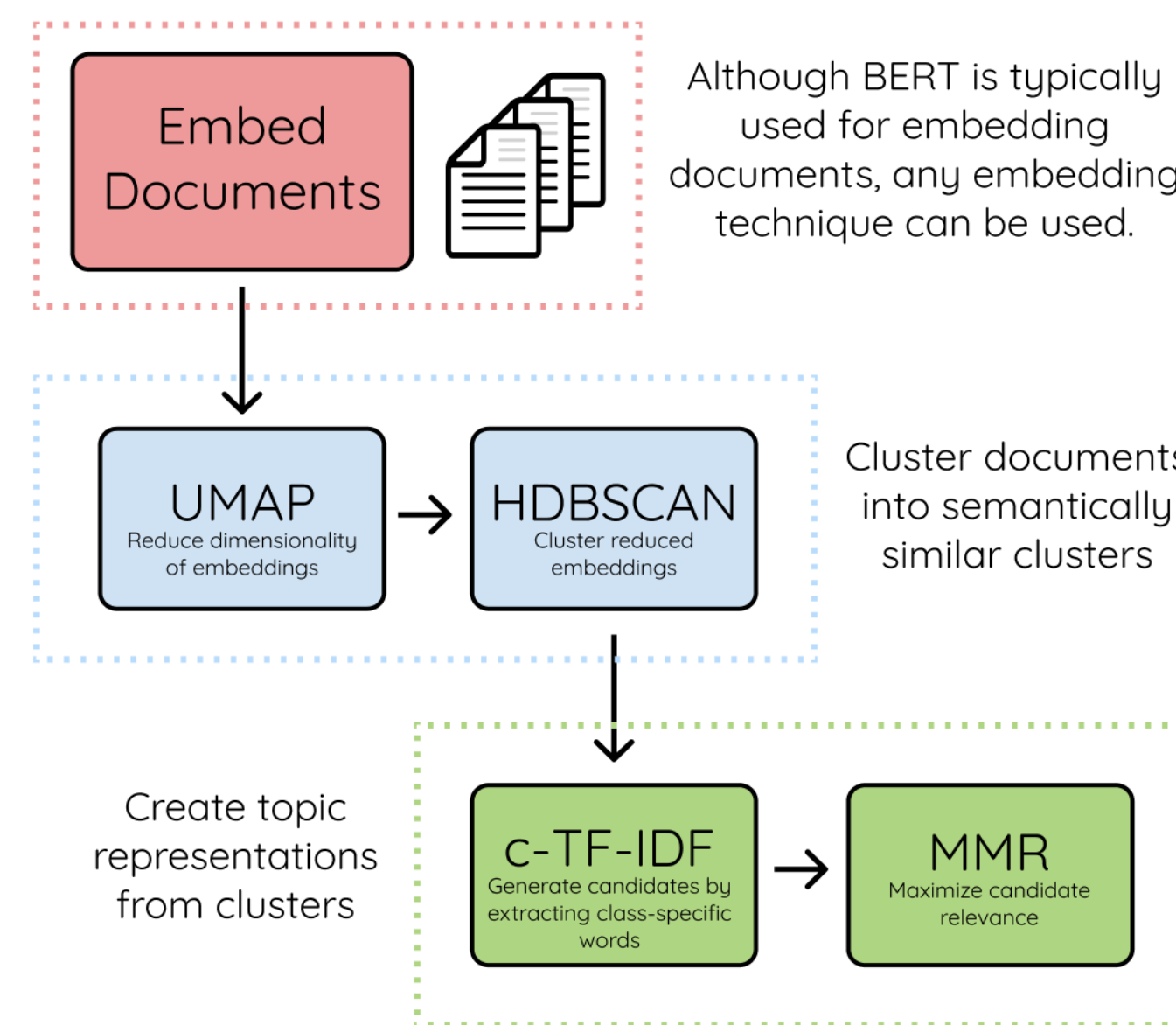


- 중국, 미국 등에서 전체의 약 80% 이상의 특허 출원이 이루어졌음
- 2014년 생성형 AI의 핵심 기술인 ‘생성형 대립 신경망’의 등장으로 본격적인 연구개발이 시작되었으며, 이후 특허 출원은 빠른 속도로 증가하고 있음

- 인공지능 기술 관련 업계에서 많은 출원이 이루어짐
- 컴퓨터, 기계 관련 기업에서도 적지 않은 수의 출원이 이루어짐

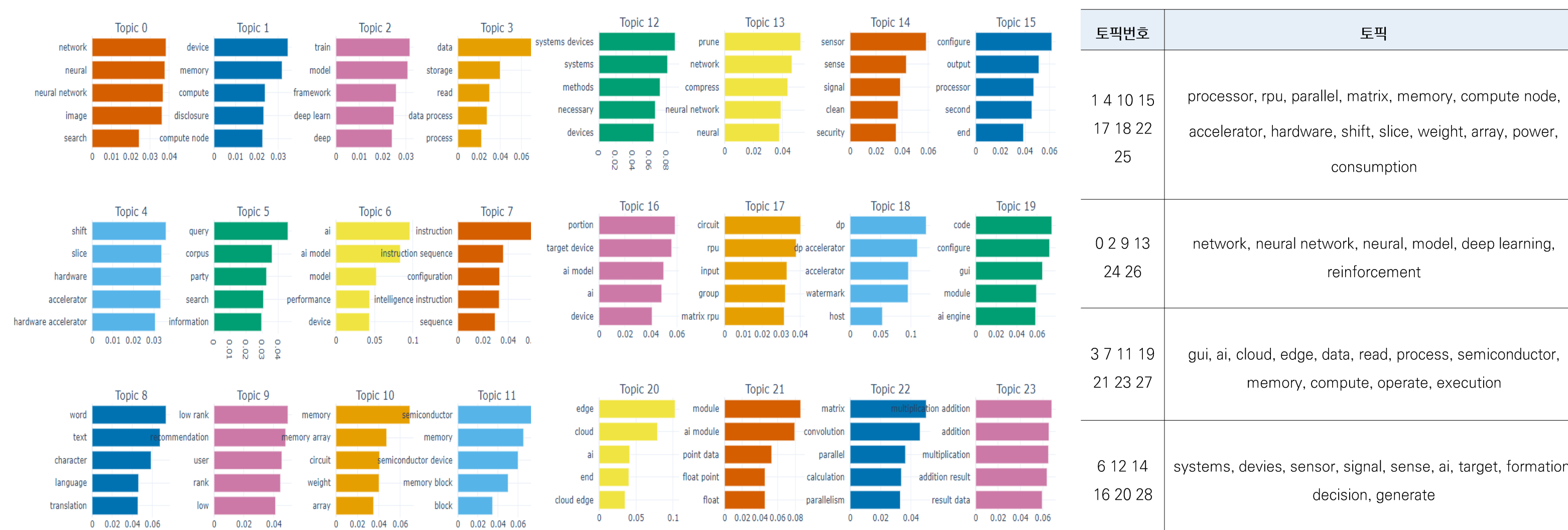
3. 실험 및 결과

3-3 BERTopic 흐름도

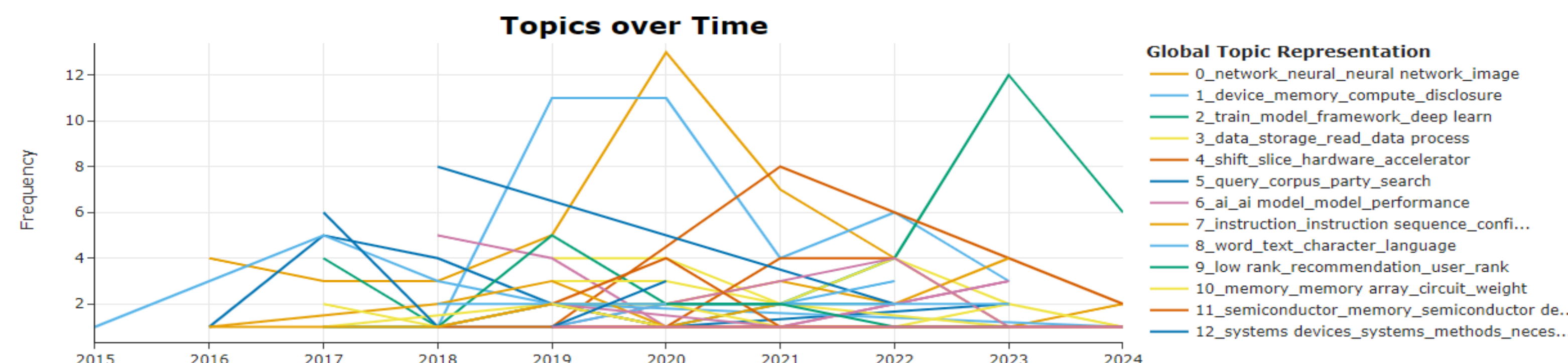


- BERTopic은 텍스트 기반의 문서 데이터 내에서 토픽의 구성을 분석하는 기법임
- 강력한 사전 학습된 언어 모델인 BERT 임베딩을 사용해 특허 문서를 고차원 임베딩으로 벡터화함
- 차원 축소 기법인 UMAP 기법을 사용해 임베딩 벡터를 저차원으로 변환하고, 임베딩 벡터에 대해 밀도 기반 클러스터링 기법인 HDBSCAN 기법을 적용하여 유사한 주제를 가진 텍스트들을 그룹화시킴
- 이후 c-TF-IDF와 같은 기법을 활용하여 토픽 내 대표 단어를 계산하고, 도출함

3-4 BERTopic 분석



- 인공신경망, 병렬처리, 연산, 가속 관련 단어의 출현 빈도가 높은 것으로 도출됨
- 병렬처리, 가속, 비휘발성 관련 단어가 같은 군집으로 도출된 것을 통해 비교적 동시에 연구개발이 진행됨을 알 수 있음
- 인공지능 반도체는 엣지컴퓨팅, 클라우드컴퓨팅 등에서도 활용됨



- 토픽별 시계열 추세를 나타낸 그림을 통해 인공지능 반도체를 활용한 엣지컴퓨팅과 클라우드컴퓨팅에 대한 연구개발이 활발히 이루어지는 것으로 도출됨

- 메모리 반도체와 인공지능 프로세서를 결합한 저전력 고성능 PIM 기술을 위한 메모리, 인공지능 전용 프로세스에 최적화된 컴파일러, 열 관리 및 외부 메모리를 최적화하는 기술이 개발되어야함

4. 결론 및 향후 연구

- 인공지능 반도체 산업은 초기시장인 만큼 해당 산업에서 특허와 같은 정량적 지표를 활용한 기술 트렌드 분석에 대한 선행 연구가 부족한 상태임
- 본 연구에서는 특허 데이터를 활용하여 기술 추세를 분석하였고 중국과 미국에서 높은 점유율을 차지하고 있는 것으로 확인됨
- 또한, GPU, NPU 등 CPU에서 진화된 인공지능 반도체가 주를 이루는 것으로 예상됨
- 본 연구는 특허 데이터만을 이용했다는 점에서 최신 기술 트렌드를 파악하기에는 한계가 있음
- 이에, 향후 연구에서는 관련 논문 및 문헌을 추가한다면 지금보다 더 최신의 기술 트렌드를 파악할 수 있을 것으로 기대됨

