



Contrastive Loss를 이용한 특허기술 검색 시스템의 효과적인 탐색 모델

Siheon Joe^{1*}, Mi Jin Lee^{2†}, Seung-Woo Son^{1, 2††}

¹Department of Applied Artificial Intelligence, Hanyang University, Ansan 15588, Korea

²Department of Applied Physics, Hanyang University, Ansan 15588, Korea

Contact us

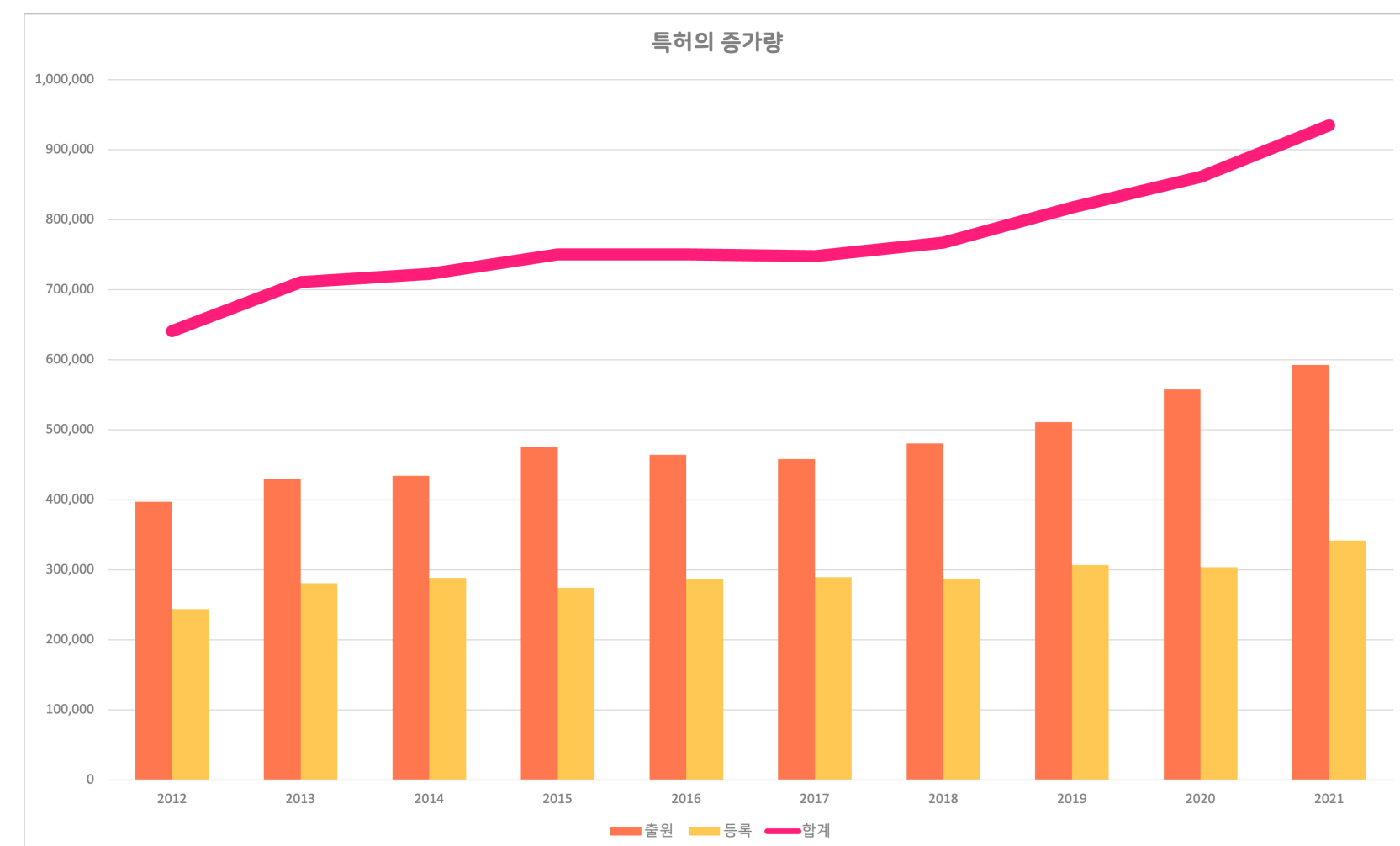
*) josehan12@hanyang.ac.kr

†) mijinlee@hanyang.ac.kr

††) sonswoo@hanyang.ac.kr

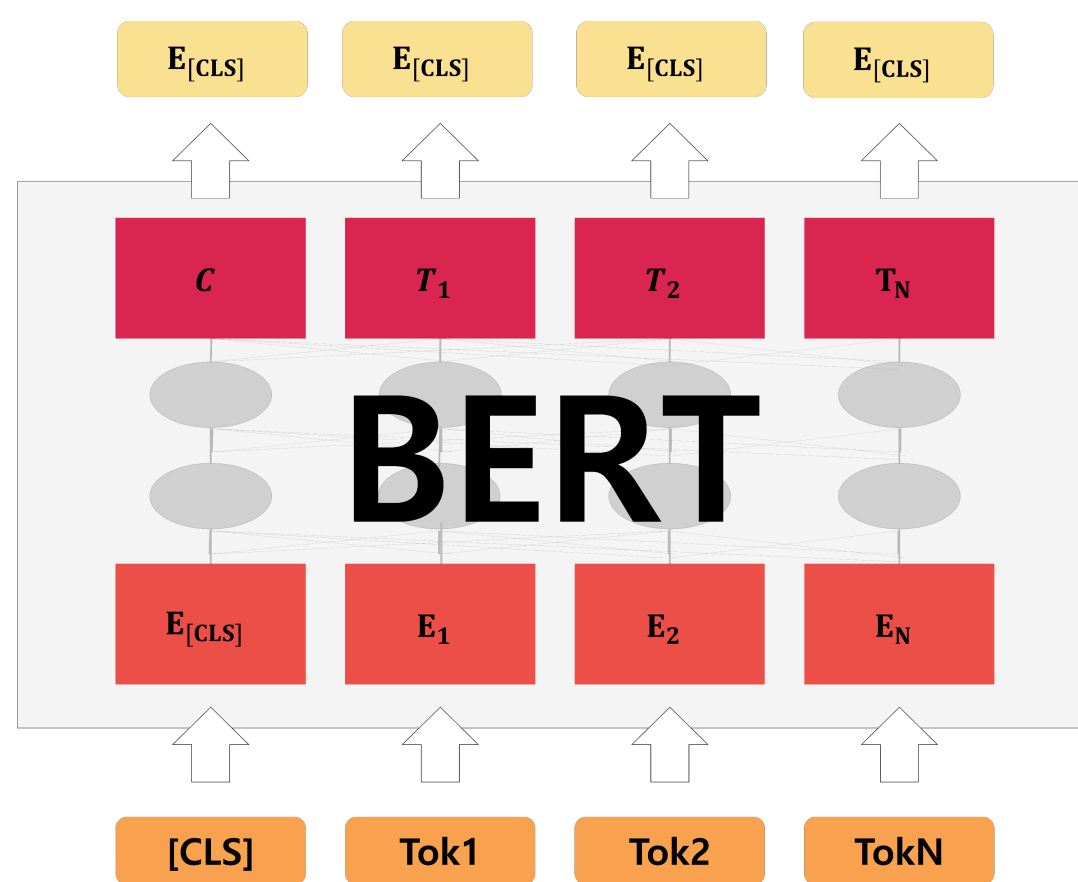
Motivation

- 특허를 출원하기 위해서는 특허의 신규성과 진보성 등을 판단하고자 기존에 출원된 발명과 동일하거나 유사한 기술이 존재하는지 등을 파악하는 특허기술 선행조사를 실시
- 해당 과정은 전문가가 탐색 해야 하는 특허의 범위가 넓어 시간적, 비용적 부담이 크기 때문에 조사의 효율을 높이고자 특허의 탐색 범위를 축소시키는 방법이 필요
- 기존의 방법은 같은 문장을 사용한 빈도수를 유사도로 보아 검색 시스템을 구축했지만 글의 문맥 특징까지 고려하지 못한 한계점이 존재



Model

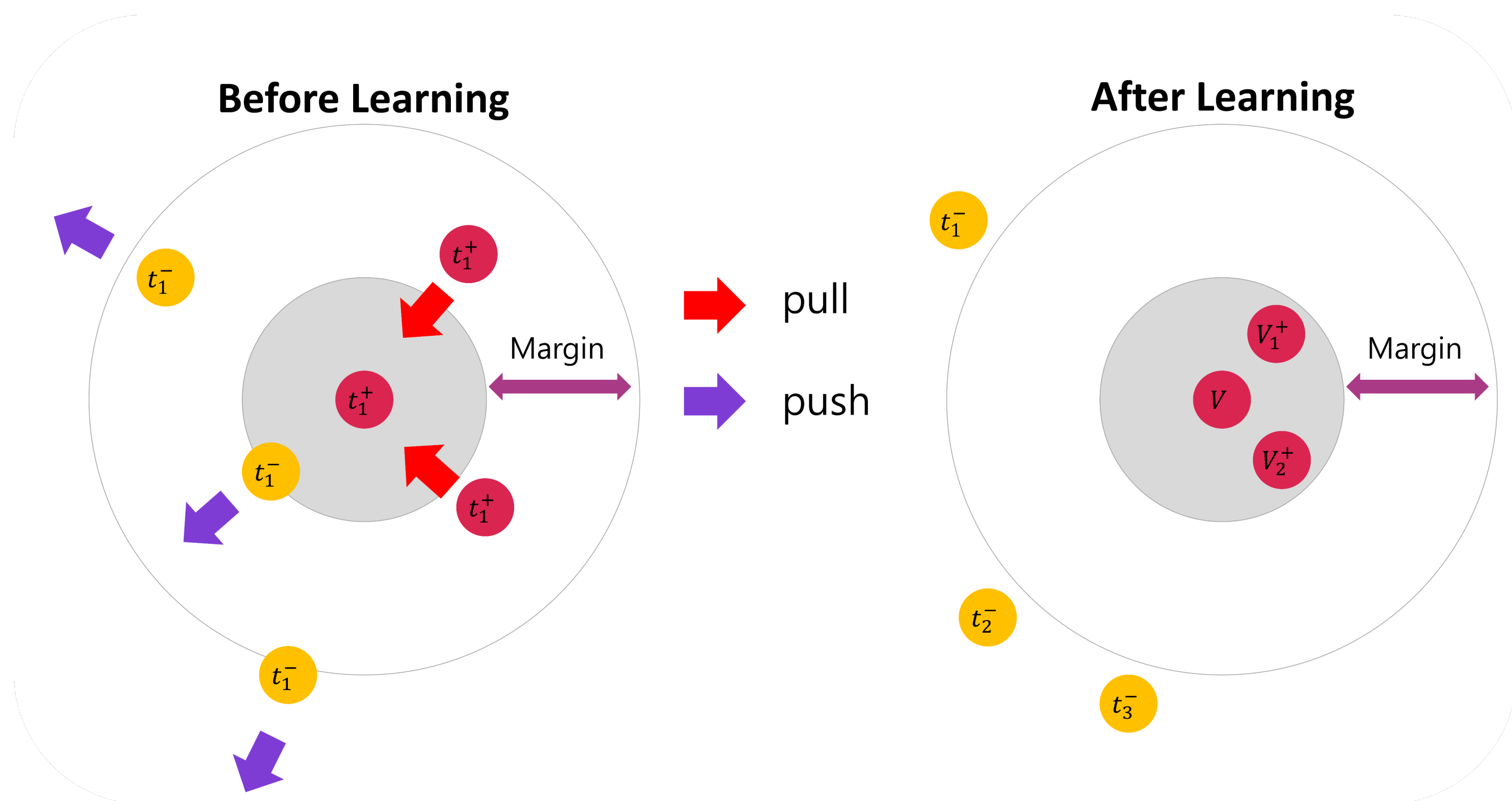
- 글의 문맥 특징을 파악하여 단어를 벡터화 시켜주는 BERT를 사용
- 사전 학습된 임베딩을 통해 특정 과제의 성능 향상
- 다양한 NLP Task에서 우수한 성능



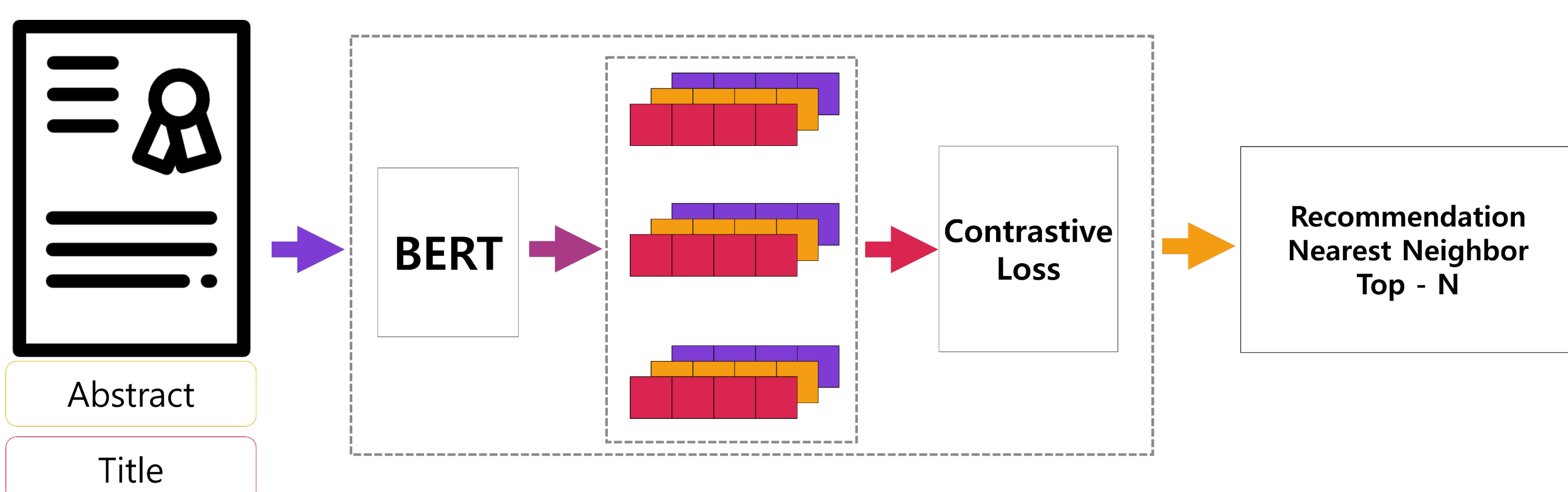
Patent Abstract	Embedding
A data link for an automatic aircraft carrier landing system is described which an ...	[0.5431 -0.13675 -0.10359 ... 0.21088 -0.58176 -0.67708]
A storage type phase coherent pulse doppler radar system wherein return signals containing doppler...	[0.6121705 -0.2982698 -0.0127 3062 ... 0.031723 -0.5511159 -0.38540003]
New chemical compounds bis quaternary carbamates having the generic fora...	[0.13732 0.09446 -0.14751 ... -1.04338 -0.5042 -0.648]

Contrastive Loss

- 대상들의 차이를 좀 더 명확하게 보여줄 수 있도록 학습
- 각 문서들에 대한 자카드 유사도를 기준으로 학습 실행



Framework



- Embedding된 데이터를 Contrastive Loss로 학습 후 주변 Top-N개의 특허 추천

Dataset

Harvard University USPTO Dataset

- 1980-2020 데이터셋
- 자카드 유사도 기반의 데이터

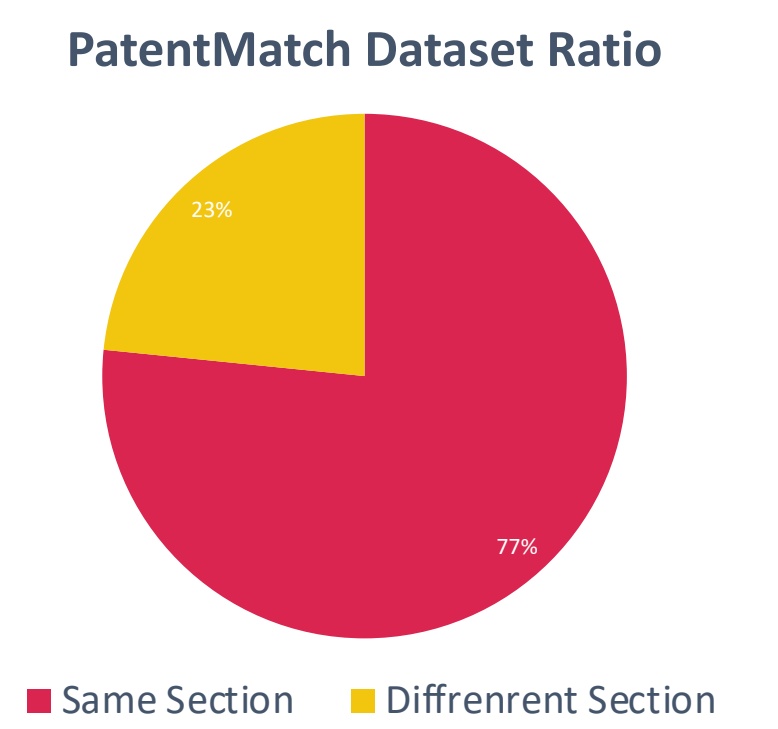
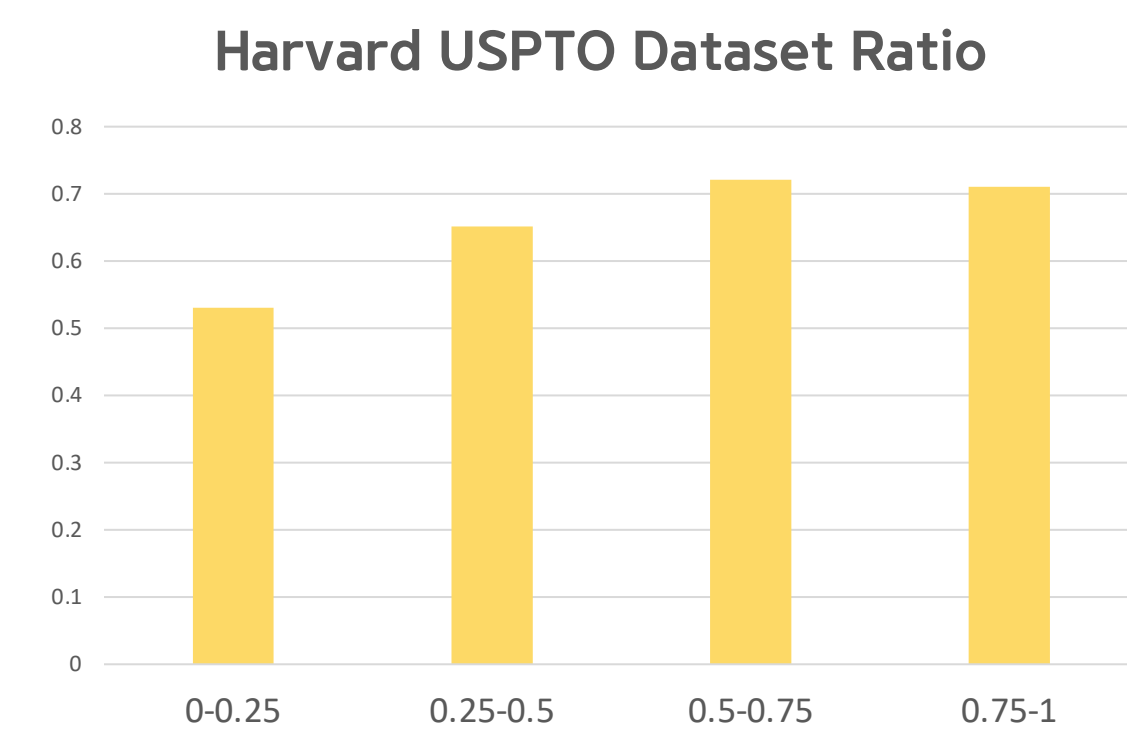
Patent A Abstract	Patent B Abstract	Similarity
A data link for an automatic aircraft carrier landing system is described which an aircraft is guided to the carrier by pitch ...	A data link for an automatic aircraft carrier landing system is described which an aircraft is guided to ...	0.91420001
A submarine detecting and tracking system which utilizes sonouoys to track ...	A probe acquisition and display system is disclosed featuring one ...	0.081600003
A single-ended laser device for locating reflective targets within a ...	This invention relates to an interferometer particularly adapted to ...	0.076899998

PatentMatch Dataset

- European Patent 추출 (1978-2012)
- 신규성 및 진보성을 판단하는데 관련이 있는 경우 해당 단락을 인용

Patent A Abstract	Patent B Abstract	Label
A bottom hole stabilizer, or similar drill string member, is provided with three s...	The device for aerating sewage water by means of minute bubbles ...	1
A television signal conversion apparatus for converting one ...	A pick-up device composed of three, 'red,' 'green' and 'blue'...	0
A process for the preparation of an anhydride of trivalent phosphorus ...	A process for preparing an anhydride of trivalent phosphorus ...	1

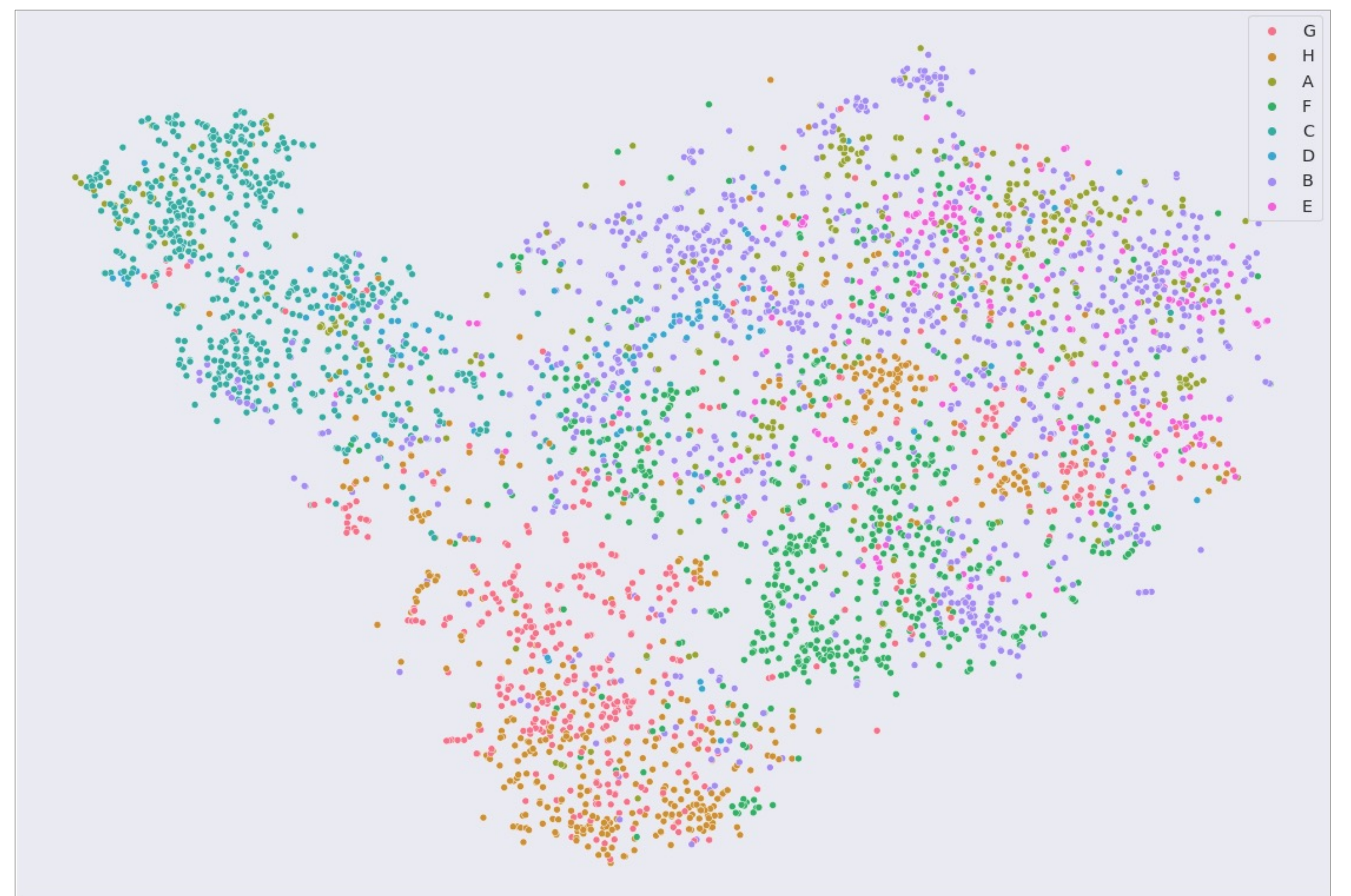
- 자카드 유사도가 높을수록 같은 종류의 특허일 확률이 높음
- 신규성 진보성을 판단하는데 관련이 있는 경우 같은 특허



Result

- 특정 노드 주변의 Top-N개의 노드를 선택해 신규성에 위배되는 특허가 들어 있는지 확인
- 같은 종류의 특허끼리 잘 묶여 있는지 확인

	Top-1	Top-5	Top-10	Top-20	Center Section
BERT	56.72	65.67	70.27	73.33	58.73
Contrastive BERT	57.49	67.29	70.69	76.06	68.32



Conclusion

- Contrastive Loss를 사용시 Center Point로 모아주는 효과를 관찰하였다.
- 특정 도메인으로 학습된 BERT를 사용시 정확도가 좀더 높다.
- 특허 선행조사 처리 속도에 도움이 될 것이다.
- 유사한 내용의 논문 선별이나 표절검사 등 다양한 영역에 사용할 수 있다.