

특허 기반 유사기업 추천 모델

장광선¹, 정현지², 김윤정³

¹한국과학기술정보연구원 NTIS센터 선임연구원

²국립공주대학교 인공지능학부 조교수

³한국과학기술정보연구원 NTIS센터 책임연구원

gsjang@kisti.re.kr, hjeong@kongju.ac.kr, miso@kisti.re.kr

Patent-Based Similar Company Recommendation Model

Gwangseon Jang¹, Hyun Ji Jeong², Yunjeong Kim³

¹NTIS Center, Division of National S&T Data, Korea Institute of Science and Technology Information(KISTI)

²School of Artificial Intelligence, Kongju National University

³NTIS Center, Division of National S&T Data, Korea Institute of Science and Technology Information(KISTI)

요약

본 연구는 기업 간 협력과 경쟁력 강화를 위한 특허 기반 유사 기업 추천 모델을 제안한다. 제안 모델은 특허 데이터와 한국표준산업분류(KSIC) 정보를 활용하여, 특허 정보를 기반으로 기업 간 유사성을 평가하고 유사한 기업을 추천한다. 제안 모델은 특허 초록 정보와 한국표준산업분류를 사용하여 기술 측면에서 기업별 특성을 고려한 기업 대표 맵터를 생성한다. 또한, 기업의 특허 수를 고려하여 정확한 유사기업 추천을 제공합니다. 제안 모델은 기업들이 협력 파트너를 찾고 새로운 비즈니스 기회를 모색하는 데에 도움을 줄 수 있으며, 현재는 NTIS(www.ntis.go.kr)의 분류기반 특허분석 서비스에서 사용 중이다.

1. 서론

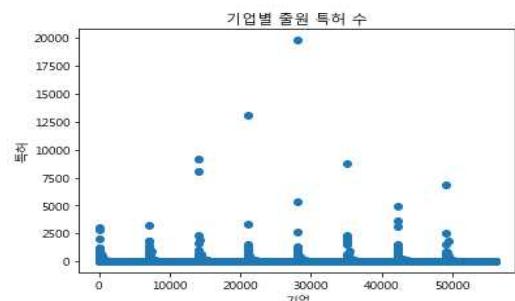
현대 비즈니스 환경에서 기업들은 특허 정보의 가치를 점차적으로 깨닫고 있으며, 이러한 정보를 효과적으로 활용하는 방법을 모색하고 있다. 특허 데이터는 기업의 기술 및 연구 개발 현황을 중요하게 반영하며, 이를 통해 기업의 성공과 지속 가능성에 큰 영향을 미친다. 특히 기술 분야에서는 기업들이 경쟁 상대들의 기술 동향을 파악하고 혁신적인 아이디어를 찾는 데에 있어 특허 정보가 핵심 자원으로 작용한다.

이러한 맥락에서, 특허 기반 유사 기업 추천은 기업들이 경쟁 환경에서 미래를 모색하고 혁신적인 협력 파트너를 찾는 데에 큰 도움이 된다. 기업들은 유사한 기술 영역에 속하는 다른 기업들과의 협력을 통해 경쟁력을 향상시키고 새로운 시장 기회를 발견할 수 있다. 그러나 이러한 유사 기업을 찾는 것은 쉽지 않는 문제이며, 기업들은 종종 이러한 정보를 얻기 위해 상당한 시간과 비용을 투자해야 한다.

2. 사용 데이터

본 연구에서 사용된 데이터는 KPRIS OpenAPI를 통해 수집하였다. 총 432,259개의 국내 특허 정보가 사용되었으며, 이 데이터는 개별 특허의 초록 및 한국표준산업분류(KSIC) 정보를 포함한다.

총 56,107개 기업이 1개 이상의 특허를 소유하고 있으며, 기업당 평균 7.7개의 특허 출원을 보유하고 있다. 그림 1에서 제시한 것과 같이 최소 1개부터 최대 19,877개의 특허를 소유한 기업들이 존재한다. 특허 등록 상태는 거절, 등록, 공개, 소멸, 포기, 취하, 무효 등 7가지로 구분되며, 연구에서는 무효 상태값을 제외한 특허 데이터만을 활용하였다.



(그림 1) 기업별 출원 특허 수

기업 대표 벡터를 효과적으로 생성하기 위하여 기업의 특허 초록 정보와 함께 한국표준산업분류(KSIC)를 활용하였다. 한국표준산업분류는 총 76종의 분류코드로 이루어져 있으며, 특허당 평균 1.5개의 KSIC 분류코드를 가지고 있으며, 최소 1개부터 최대 9개의 한국표준산업분류 코드를 가진 특허도 존재한다. 한국표준산업분류 코드가 부여되지 않은 2,667개의 특허 데이터는 본 연구에서 사용하지 않았다.

3. 제안 모델

국가과학기술표준분류체계 기반으로 논문을 분류한 논문 연구분야 분류 데이터[1]를 사용하여 학습한 인코더 모델을 통하여 각 특허 초록에 대한 임베딩 벡터를 생성한다. 해당 인코더 모델은 Sentence Transformer[2] 기반으로 RoBERTa[3] Base 버전을 KorSTS[4] 데이터와 1,162,984개의 논문 본문 데이터로 만든 논문전용 STS 데이터를 학습하였다. 인코더를 통하여 각 특허별로 임베딩 벡터 V_i 를 생성한다.

특허별 임베딩 벡터, 특허별 한국표준산업분류 코드, 기업별 특허 수 정보를 사용하여 수식 (1)과 같이 기업 대표 벡터를 생성한다. 기업의 특허가 가장 많이 포함된 한국표준분류산업분류 4개의 분류코드를 선정한다. 선정된 분류코드 별로 해당 분류코드를 가지는 특허의 임베딩 벡터를 합하여 분류 벡터를 만든다. 그다음, 기업의 전체 특허 수에서 해당 분류코드에 해당하는 특허 수의 비율을 각 분류 벡터에 곱하여 가중합을 하여 기업 대표 벡터 V_A 를 생성한다.

$$V_A = \frac{n_1}{N} \sum_{i=0}^{n_1} V_i + \frac{n_2}{N} \sum_{i=0}^{n_2} V_i + \frac{n_3}{N} \sum_{i=0}^{n_3} V_i + \frac{n_4}{N} \sum_{i=0}^{n_4} V_i \quad (1)$$

N 은 기업A의 전체 특허수, n_i 는 해당 분류코드 i에 해당하는 기업A의 특허 수, V_i 는 각 특허의 임베딩 벡터를 나타낸다.

유사기업 추천시 기업의 규모를 고려하기 위하여, 기업 대표 벡터가 유사도 산정시 기업의 특허수도 함께 고려한다. 특허수와 기업 규모의 양의 상관관계를 고려하여 특허 수가 많으면 기업 규모도 크다고 가정하였으며, 데이터 분석 결과 특허 수와 기업 규모가 실제 양의 상관관계를 가지는 것을 확인하였다.

기업 대표 벡터가 유사도를 계산할 때 기업의 보유 특허 수를 반영하기 위하여 수식 (2)와 같이 기업 대표 벡터간의 코사인 유사도와 기업의 보유 특허 수에 대한 가중치를 곱한 점수를 기반으로 기업 간 최종 유사도를 산정한다. 기업의 보유 특허수에 대한 가중치를 적용하기 위하여 시그모이드(Sigmoid) 함수를 변형하여 적용하였다.

$$f(V_A, V_B, x_B) = \frac{V_A \cdot V_B}{\|V_A\| \|V_B\|} \cdot \frac{1}{1 + e^{-0.035x_B + 7.7}} \quad (2)$$

$f(\cdot)$ 은 2개의 기업의 유사도를 계산하는 비교함수, V_A 는 질의 기업의 기업 대표 벡터, V_B 는 비교 기업의 기업 대표 벡터, x_B 는 비교 기업의 특허수이다.

3. 실험 결과

현재 유사기업 성능 측정을 위한 정답 데이터셋을 구축 중이므로, 향후 연구에서 유사기업 추천 모델의 정량적 성능을 측정할 예정이다. 이에 따라 본 논문에서는 대기업, 중소기업, 반도체 기업별로 유사기업 추천 결과를 제시한다.

대표적인 대기업에 대한 유사기업 추천결과로 삼성전자 주식회사와 현대자동차 주식회사에 대한 유사결과 추천결과를 표 1과 표2를 통해 확인할 수 있다.

<표 1> 삼성전자 주식회사의 유사기업 추천 결과

구분	기업명	유사도	특허수
Q	삼성전자 주식회사		26,333
1	한양대학교	0.939	2,699
2	한국전자통신연구원	0.935	6,734
3	에스케이하이닉스 주식회사	0.929	3,624
4	한국과학기술원	0.927	4,881
5	서울과학기술대학교	0.924	714

<표 2> 현대자동차주식회사의 유사기업 추천 결과

구분	기업명	유사도	특허수
Q	현대자동차주식회사		14,144
1	기아자동차주식회사	0.999	13,482
2	한국자동차연구원	0.977	456
3	현대모비스 주식회사	0.975	3,138
4	주식회사 만도	0.959	2,211
5	르노삼성자동차 주식회사	0.954	303

중소기업에 대한 유사기업 추천결과로 와이즈넛 주식회사와 주식회사 국순당에 대한 유사결과 추천 결과를 표 3과 표 4를 통해 확인할 수 있다.

<표 3> 주식회사 와이즈넷의 유사기업 추천 결과

구분	기업명	유사도	특허수
Q	주식회사 와이즈넷		30
1	주식회사 솔트룩스	0.905	33
2	(주)플랜아이	0.899	8
3	한국과학기술정보연구원	0.897	173
4	주식회사 리얼타임테크	0.893	29
5	주식회사 피씨엔	0.882	11

<표 4> 주식회사 국순당의 유사기업 추천 결과

구분	기업명	유사도	특허수
Q	주식회사 국순당		10
1	롯데칠성음료주식회사	0.790	19
2	샘표식품 주식회사	0.752	54
3	주식회사 웰빙엘에스	0.752	11
4	(주) 성운파마코파이	0.741	17
5	재단법인 발효미생물산업진흥원	0.741	167

반도체 관련 기업에 대한 유사기업 추천결과로 네메시스 주식회사와 주식회사 넥스트칩에 대한 유사결과 추천결과를 표 5와 표 6을 통해 확인할 수 있다.

<표 5> 네메시스 주식회사의 유사기업 추천 결과

구분	기업명	유사도	특허수
Q	네메시스 주식회사		8
1	어보브반도체 주식회사	0.849	33
2	(주)필콘시스템	0.837	9
3	(주)인피니어	0.821	12
4	(주)센서테크	0.820	6
5	주식회사 유비콤팩트놀로지	0.814	2

<표 4> 주식회사 국순당의 유사기업 추천 결과

구분	기업명	유사도	특허수
Q	주식회사 넥스트칩		20
1	한국전자통신연구원	0.863	6734
2	주식회사 텔레칩스	0.852	26
3	주식회사 에스원	0.849	249
4	삼성에스디에스 주식회사	0.842	493
5	주식회사 한컴지엠디	0.837	7

4. 결론

본 논문은 기업의 특허 정보를 활용하여 유사기업을 추천하는 모델을 제안한다. 특허 초록 정보를 사용함으로써 기술적인 측면에서 유사기업을 추천할 수 있도록 하였다. 또한 기업 대표 벡터 생성시 한국표준산업분류를 적용하여 여러 분야 사업을 진행 중인 기업에 대해서도 정확한 유사기업 추천이 가능하도록 하였으며, 특허 수를 고려한 기업 대표 벡터 생성 및 비교 함수를 통하여 유사기업 추천시 기업 규모를 고려할 수 있도록 하였다. 개발된 유사기업 추천 모델은 NTIS(www.ntis.go.kr)의 분류기반 특허분석 서비스 등에서 활용용 중이다[5].

감사의 글

이 논문은 2023년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다.(과제번호: (KISTI) K-23-L01-C05, (NTIS) 1711196501)

참고문헌

- [1] Roger S. Pressman "Software Engineering A Practitioners' Approach" 3rd Ed. McGraw Hill
- [1] 한국과학기술정보연구원 “논문 연구분야 분류 데이터”. Version 1.0. <https://doi.org/10.23057/50> (2022).
- [2] Reimers, Nils, and Iryna Gurevych. "Sentence-bert: Sentence embeddings using siamese bert-networks." arXiv preprint arXiv:1908.10084 (2019).
- [3] Liu, Yinhan, et al. "Roberta: A robustly optimized bert pretraining approach." arXiv preprint arXiv:1907.11692 (2019).
- [4] Ham, Jiyeon, et al. "KorNLI and KorSTS: New benchmark datasets for Korean natural language understanding." arXiv preprint arXiv:2004.03289 (2020).
- [5] Ministry of Science and ICT, KISTI, National Science and Technology Knowledge Information Service.