

## 2020-05-04 & 2020-05-11 멧쟁이터틀

### [논문 정리] 추론모델 기반 성과지표 네트워크를 이용한 특허 빅데이터 시각화에 관한 연구

-> 특허의 텍스트 정보와 질적 성과지표 간의 관계를 추론하고 이를 시각적으로 표현

청구항에 기술 내용이 기재된 특허는 출원된 뒤에 IPC 코드 할당되며, 신규성, 권리성, 기술성 등의 관점에서 평가. 평가 뒤 등록된 특허는 다른 특허에 인용되기도 하고, 다른 국가에서든 법적 권리 범위를 행사하기 위해 패밀리 특허를 출원

특허는 주로 청구항 수, 피인용 수, 패밀리 특허 수(국가 수)등에 의해 평가

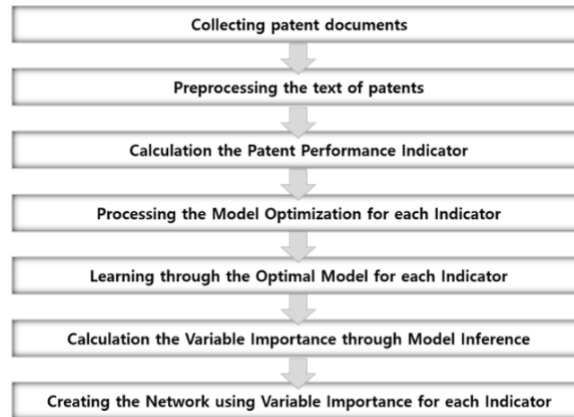
청구항 수는 특허의 권리성과, 피인용 수는 특허의 기술성과, 패밀리 특허는 특허의 시장성과 관련

특허청의 가이드라인에 제시된 대표적인 질적 우수성 지표는 권리성, 기술성, 시장성, 기술 다양성

#### \* 추론 모델

복잡도가 낮은 선형회귀모델의 경우 회귀계수를 통해 예측변수의 중요도 파악 및 비교 가능 (랜덤포레스트 모델 사용)

- 1) 특허 문서를 분석 목적에 맞게 기술 중심 또는 출원인 중심으로 수집
- 2) 특허 문서의 서지정보를 전처리 (비정형 데이터인 문자를 활용하기 위해서는 공백, 숫자, 구두점, 불용어 제거 등의 과정 요구)
- 3) 빈도 기반의 DTM(Document-Term Matrix)을 구성
- 4) 각 문서 별로 특허 성과지표를 산출



DTM과 성과지표를 통해 추론이 가능한 모델에 적용 (모델의 최적화가 필요한 경우에는 선형하여 진행)

추론모델을 통해 성과지표의 우수성에 영향을 주는 단어를 탐색하여 네트워크로 표현

n 개의 특허가 갖는 성과지표를 p개의 단어와 시각화하는 것이 가능

또한, 제안된 방법은 수집된 특허의 성과지표가 연관된 단어를 파악할 수 있는 장점

#### \* 실험 결과

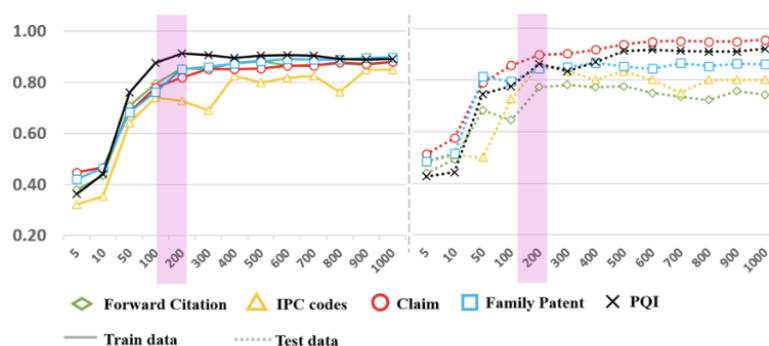


그림 3. 양상블 수에 따른 훈련데이터와 검정데이터의 정밀도

정밀도는 우수-클래스로 분류된 것 중 실제 우수-클래스의 비율

정밀도가 높은 추론모델은 우수-클래스에 영향을 주는 변수들의 중요도가 높아지도록 학습

성과지표별 변수 중요도를 네트워크로 표현하기 위해 Min-Max 스케일링

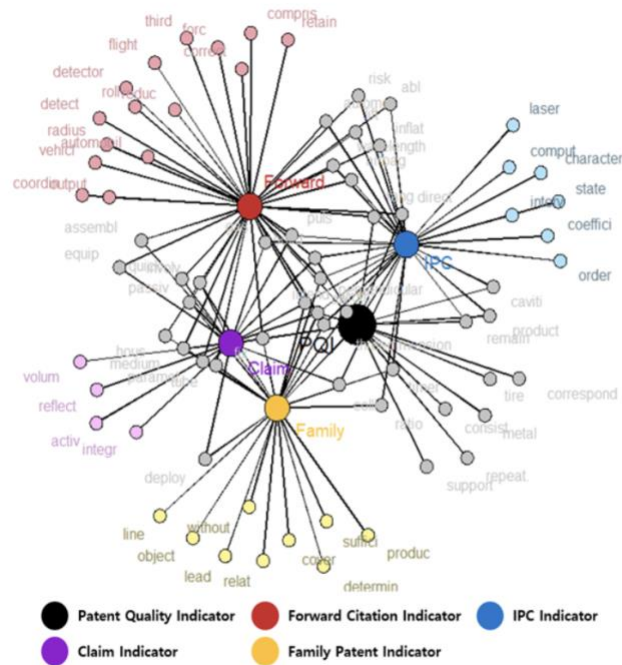


그림 5. 특허 성과지표 기반 네트워크 분석 결과

그림5는 성과지표 별 변수 중요도를 네트워크로 표현한 결과

PQI 를 중심으로 네 가지 성과지표는 모두 연결되어 있으며 각 성과지표에 높은 영향을 주는 단어는 서로 연결

표 1. 성과지표별 중요 변수 목록

Table 1. Important variable list of each indicator

PI	Words
Forward Citation	solid, wavelength, quick, width, involve, airbag, passive, lift, automotor, automobile
IPC	direct, perpendicular, wavelength, airbag, interval, width, axis, timer, intend, lift
Claim	perpendicular, passive, collision, medium, integration, involve, volum, gas, deploy, reflect
Family Patent	tube, parameter, deploy, threedimension, ratio, timer, fix, perpendicular, axis, intend
PQI	solid, consist, axis, remain, product, tire, metal, direct, support, cavity
bi-PI	wavelength, airbag, tube, timer, quick, parameter, lift, deploy, inflate, remain
ter-PI	width, passive, direct, threedimension, involve, medium, fix, pulse, range, gas

## 2) 시각화 방안 논의

### - 기존의 문단 방식

1. A subframe for a motor vehicle comprising:					
2. A subframe according to claim 1, wherein the side plates, together with the conti					
3. A subframe according to claim 1, wherein the bearings of the subframe on the si					
4. A subframe according to claim 3, wherein the bearings of the subfran					
5. A subframe according to claim 1, wherein the side plates are castings and each c					
6. A subframe according to claim 5, wherein between the struts and the					
7. A subframe according to claim 5, wherein the chords and the struts o					
8. A subframe according to claim 7, wherein the ribs on the s					
9. A subframe according to claim 1, wherein the bearings in the side plates as well					
10. A subframe according to claim 9, wherein the bearings for a lower c					
11. A subframe according to claim 1, wherein the side plates have respective locally					
12. A subframe according to claim 1, wherein in the lower horizontal plane in a coi					
13. A subframe according to claim 12, wherein the third cross member c					
14. A subframe according to claim 12, wherein the additional cross mer					
15. A subframe according to claim 14, wherein the legs of th					
16. A subframe according to claim 14, wherein the bridge me					
17. A subframe according to claim 16, wherein th					
18. A subframe according to claim 14, wherein a drive unit is					
19. A subframe according to claim 12, characterized in that the cross me					
20. A subframe according to claim 1, wherein one of the cross members is fastened					
21. A subframe according to claim 1, wherein the cross member which relative to th					
22. A subframe for a motor vehicle with rear wheel steering system and two side plates extend					
23. A subframe according to claim 22, wherein the console is made of a non-rigid r					
24. A subframe according to claim 22, wherein a rear wheel control element compri					

### - 청구항간의 관계 시각화 다른방안 : 테크트리(오른쪽)



### 인용분석

인용 depth별 문헌 간 인용관계 분석부터  
심사과정 중 활용 단계까지 파악할 수 있습니다.



### 청구항분석

시각화된 청구항 계층구조 분석과  
등록 과정 중 청구항 보정내용을 알 수 있습니다.

- 관련 연구: 유사특허의 청구항 기반 특허 분석법
- 기술 분야의 전문성, 서술형 문장 등으로 인한 테크트리 구축의 시간, 자원의 소모가 큼
- 기본적으로 청구항간의 구조는 계층식
- 테크트리의 유사 특허 분석을 도와줄 수 있는 keyword 방식의 matrix 제안
- 즉 청구항에서 정보를 keyword형식으로 추출하여, 다른 청구항들과 비교하여 테크트리의 유사한 특허들이 모여있는 brach를 쉽게 파악할 수 있도록 시간을 단축
- 결과: 하나의 주제에 따른 특허들에서 3개의 브랜치를 발견

- 그러나 맥락을 이해하지 못한다는 단점, 따라서 keyword의 맥락을 파악하여 동의어, 및 반복어를 처리해 브랜치를 구분한다면 향후 더 나은 결과를 기대할 수 있음