**연구노트**

**(수학과)이 재 봉**  2017.07.27 (목)

|  |  |  |
| --- | --- | --- |
|  | | |
| **연 구**  **주 제** | **특허분석을 통한 미래기술 예측** | |
| **관련 논문 조사** | |
| **구 분** | **연구내용** | **진행률** |
|  | 전성해 (2009). 특허정보 분석을 위한 효율적인 텍스트 마이닝  특허정보 분석  현재의 특허분석은 텍스트 레이블의 빈도 등요약통계(summary statistic)와 시각화(visualization)에 기반을 두고 있다. 막대그래프(bar chart), 산점도(scatter plot), 원도표(piechart), 별도표(star chart), 거품도(bubble chart)등을 이용한 특허데이터의 시각화를 통한 의사결정은 이 분야의 전문가에 의한 주관적 해석이 필요하다. 특허데이터의 서지적 정보뿐만 아니라기술적 정보까지 포함하여 기술특허들 간의 권리정보를 효율적으로 도표화한 기술지도(patentmap; PM)는 현재 가장 많이 사용되는 특허분석기법이다. 아직 인공신경망(artificial neural  networks), 통계적 학습이론(statistical learningtheory), 진화연산(evolutionary computing) 등고급의 데이터 마이닝 분석기법을 사용하지 못하는 이유 중 하나는 기본적으로 특허데이터는 번호와 텍스트 등 이들 고급의 마이닝 기법  에 바로 사용되기 어려운 데이터형태(data type)이기 때문이다. 본 논문에서는 텍스트 마이닝과분류, 회귀, 그리고 군집화의 데이터분석기법을유기적으로 연결하여 정량적인 데이터분석을 가능하게 하여 보다 객관적이고 정교한 분석이 가능할 수 있는 방안을 제시한다.  텍스트 마이닝을 이용한 효율적인 특허데이터의 분석  텍스트 마이닝과 분류, 회귀, 그리고 군집화의 데이터분석 기법을 유기적으로 연  결하여 기술예측을 위한 효과적인 방안을 제시한다. 전 세계의 검색사이트로부터 다운로드된특허데이터는 주로 특허에 대한 서지적 정보와기술적 정보를 포함하고 있다. 이 데이터는 기본적으로 엑셀 또는 텍스트 파일로 얻을 수 있다.이 데이터는 바로 분석하기에는 어려움이 있기때문에 텍스트 마이닝을 이용하여 1차적인 전처리(preprocessing) 과정을 거친다. 다음으로 전처리된 데이터(revised data)에 대한 변형(transformation) 과정을 거친다. 이 과정을 통하여 텍스트 위주의 데이터는 연속형(continuous)수치데이터로 바뀐다. 최종적으로 얻게 되는 학습데이터(training data)는 기존의 분류, 회귀, 그리고 군집화의 데이터 마이닝 기법들을 적용할수 있게 된다. 이를 통하여 최종적으로 구축된모형은 연구개발 등을 계획할 수 있는 기술예측(technology forecasting) 모형이 된다. |  |
| **차 주**  **계 획** | ■ 계획  - 관련 논문 조사 | |
| **이 슈** | ■ **해당사항 없음** | |

|  |  |
| --- | --- |
|  |  |