

지도학습 기반의 차원축소 모델을 이용한 특허 빅데이터 예측에 관한 연구*

이 주 현** · 이 준 석*** · 강 지 호** · 박 상 성*** · 장 동 식** · 홍 성 욱**** · 김 선 영****

A Study on prediction of patent big data using supervised learning with dimension reduction model

Lee Juhyun · Lee Junseok · Kang Jiho · Park Sangsung · Jang Dongsik · Hong Sungwook · Kim Sunyoung

〈Abstract〉

Patents are system to promote the development of industry by disclosing technology. The importance of recent patent is being emphasized. For this reason, companies apply for many patents. And they analyze the patent. Patent analysis helps to protect and foster their technology. Previously this method has been carried out by experts. Expert-based patent analysis, however, has the disadvantage of being time-consuming and expensive. Consequently, we try to solve this problems by developing prediction model. Therefore, this paper proposes a data-based patent analysis method using quantitative indicator and textual information. We confirmed the practical applicability of the proposed method through 1,831 autonomous vehicle patents. As a result, it was possible to confirmed that safety and lane detection related technologies are important.

Key Words : Patent Big Data, Patent Embedding, Dimension Reduction, PLS, IP-R&D

I. 서론

특허권은 개발된 기술을 공개함으로써 얻는 배타적 지배권이다[1]. 이러한 특허는 2018년 10월 기준

한국 특허청에 누적 건수로 339,294건이 출원되었다. 특히, 2018년 10월에만 특허가 15,017건이 출원되었다. 많은 양의 특허 출원 트렌드는 기업이 자사의 기술력을 보호하기 위한 연구활동 중 하나이다. 기업은 기존의 특허 출원을 통한 기술권리 확보에 머물지 않고 경쟁사의 특허를 분석하여 자사의 기술력을 높이고 경쟁사의 기술에 대응하기 위하여 포트폴리오를 작성한다. 특허 분석에는 특허가 가지고 있는 다양한 형태의 정보를 이용한다. 분석에 사용하는 특허정보로는 텍스트와 이미지 등이 있다. 이러한 특허의 정형 및 비정형 정보들은 빠르게 생성되고 방대한 양으로 누적되기 때문에 빅데이터의 형태를 갖추고 있다.

* 본 논문은 2019년 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.(한국연구재단-NRF-2017R1A2B1010208). 본 연구는 과학기술정보통신부 및 과학기술인력지원사업 2019년 연구산업서비스업역량 강화지원사업의 지원을 받아 수행하였음.[2019K000305].

** 고려대학교 산업경영공학과

*** 청주대학교 빅데이터 통계학과 조교수

(교신저자, hanyul@cju.ac.kr)

**** 엠아이큐브 솔루션

***** 주식회사 에스와이프 / 에스와이프 특허법률사무소

특허 빅데이터 분석은 IP-R&D(Intellectual Property-R&D)에 적극 활용되고 있다. IP-R&D는 특허를 분석하여 연구개발에 활용하는 방법론으로 중복된 연구개발을 사전에 방지하고 기업의 연구개발 전략을 수립하는데 이용된다. 한국지식재산전략원[2]에 따르면 종래의 IP-R&D는 전문가의 집단에 의해서 실시되어 왔고 이러한 이유로 정성적인 분석이 주를 이루어 많은 시간과 비용이 요구되며, 전문가의 주관적인 의견에 편향될 수 있다는 단점이 있다. 또한, 빅데이터인 특허는 전문가가 검토할 수 있는 특허 문서의 양이 한정될 수 있다는 단점도 있어 텍스트 마이닝 기반 방법과 함께 분석될 필요가 있다.

최근, 전문가 기반 IP-R&D의 단점을 개선하기 위해 특허 빅데이터 기반의 IP-R&D 방법론에 대한 연구가 활발히 진행 중이다. 전성해[3]에 의하면, 특허 빅데이터 기반 IP-R&D 방법론은 전문가 기반의 방법에 비해 객관적인 결과를 얻을 수 있다는 장점을 갖는 것으로 알려져 있다. 그리고 방대한 양의 특허도 신속하게 처리할 수 있으며, 새롭게 업데이트되는 특허들도 반영하여 IP-R&D 전략을 수립할 수 있다. 대표적인 방법론으로는 전문가들의 정성적 판단에 의해 결정되는 특허의 가치평가를 개선하기 위하여 자동화된 특허가치평가 시스템이 있다. 자동화된 시스템은 특허가 가지고 있는 권리잔존일 수, 청구항 수, 인용문헌 수, 피인용문헌 수, 패밀리 특허 수와 같은 많은 정량지표를 포함한다. 이러한 정량지표를 활용하면 특허를 질적으로 평가할 수 있다. 특허 평가 시스템은 정량지표를 활용하여 핵심·유망특허를 쉽고 빠르게 발굴할 수 있게 도와준다. 그러나 박상성[4]에 의하면 현재 운영되고 있는 많은 특허 평가 시스템은 특허의 서지 정보를 반영하지 않는 단점이 있다. 따라서 본 연구에서는 기존의 특허 평가 시스템을 활용하면서 특허의 서지 정보를 활용한 개선된 특허 평가 모델을 제안한다.

II. 관련 연구

최근 특허 빅데이터 기반의 IP-R&D 방법론은 다양한 형태로 연구되고 있다[5-10]. 특허 빅데이터 기반 IP-R&D는 주로 특허의 청구항 수, 인용특허 수, 피인용특허 수, 발명자 수, 패밀리 특허 수 등과 같은 정량적 지표를 머신러닝 모델에 활용하여 특허의 군집화, 평가등급 산출, 로열티 산출을 통한 가치 측정 등으로 연구되고 있다. Trappey[5]는 권리범위 잔존일 수, 인용 수, 기술분류코드 등의 정량지표를 산출하고 이를 활용하여 뉴럴네트워크 기반의 특허 가치 산출 모델을 제안하였다. 양동홍[6]은 특허의 발명자 수, 독립항의 수, 인용 수와 같은 정량지표를 활용하여 로열티율을 산출하고 이를 통해 특허의 가치를 산정하는 연구를 진행하였다. Wu[7]는 특허의 패밀리 특허 수, 인용 수와 같은 특허 품질 지표를 산정하고 이를 활용하여 특허의 품질을 평가하는 모델을 제안하였다. 우호영[8]은 한국발명진흥회에서 개발한 특허 평가 지표들을 활용하여 베이지안 기반의 모델을 제안하는 연구를 진행하였다. 해당 연구에서는 베이지안 기반의 모델을 사용함으로써, 소표본의 특허도 평가가 가능한 장점이 존재했다. 최재현[9]은 특허의 정량지표와 기업의 매출액, 정부의 지원 규모를 활용하여 기업을 기술 성숙도에 따라 분류하는 연구를 진행하였다.

특허 빅데이터 기반의 IP-R&D 방법에서 특허의 서지 정보를 활용하는 것은 중요하다. 왜냐하면, 특허는 개발된 기술의 법적 권리범위를 행사하기 위해 청구항에 자세한 내용이 서술되어 있기 때문이다. 앞선 많은 연구들은 주로 정량지표만을 사용한 특허평가 모형을 만들어 특허가 가지고 있는 주요한 기술에 대한 가치를 반영하지 못하는 단점을 가지고 있다. 이러한 단점을 보완하기 위해 최근 서지정보를 반영한 연구가 진행되고 있다. 이준석[10]은 핵심특허 예측을 위해 특허 서지 정보를 정형화된 구조로 변형하는 연구를 진행하였다. 해당 연구에서는 정형화된 특허 서

지 정보를 로지스틱 회귀모델에 적용하여 핵심특허를 예측한 바 있다. 박상성[4]은 한국발명진흥회에서 개발한 특허 평가등급과 서지 정보를 결합하고 이를 활용하여 특허를 군집화하는 연구를 진행한 바 있다. 해당 연구에서는 정량지표를 통해 산출한 평가등급과 서지 정보를 결합한 방법을 제안했다는 점에서 다른 연구들과의 차별성을 가졌다. 최근 연구들의 한계점은 특허의 서지정보만을 주로 사용하여 방법론을 제시하였다는 것이다.

본 연구에서는 특허의 정량지표와 서지 정보를 결합하는 모델을 제안한다. 그리고 제안된 모델을 통해 핵심키워드 추출과 핵심특허의 예측하는 방법을 제안한다.

III. 이론적 배경

3.1 Partial Least Square algorithm

최소제곱법은 관측값과 예측값의 차이가 최소가 되도록 방정식의 해를 구하는 방법이다. 부분최소제곱법(PLS, Partial Least Square)은 반복적인 최소제곱법을 이용하여 반응변수와 예측변수의 공분산이 최대인 축을 찾는 지도학습 기반의 차원축소 방법이다. 즉, PLS는 예측변수 X 와 w 의 선형결합인 PLS 성분 t 와 반응변수의 최소제곱법의 해를 구하고 잔차에 대해 다시 최소제곱법을 통해 해를 구하는 방식으로 학습을 진행한다. 그러면 식(1)을 최소화하도록 예측변수의 차원을 축소할 수 있다[11-15].

$$\begin{aligned} Cov(t, Y) &= Corr(t, Y) \sqrt{Var(t)} \sqrt{Var(Y)} \\ &\propto Corr(t, Y) Var(t) \end{aligned} \quad (1)$$

식(1)에 의해 차원을 축소하면, 반응변수의 정보를 활

용하면서 예측변수의 분산을 최대화하도록 차원을 축소하게 된다. 따라서 식(1)을 만족하는 최적의 w^* 는 $X^T Y$ 이다. 그러므로 w^* 를 표본의 크기가 n 인 데이터에 적용하면 식(2)를 통해 반응변수를 예측하는 것이 가능하다.

$$\hat{Y} = \sum_{i=1}^k t_i \beta_i, i = 1, 2, \dots, n \quad (2)$$

식(2)에서 β_i 는 t_i 와 i 번째 PLS 성분이 설명하지 못하는 부분에 대한 y_{i+1} 의 선형회귀계수이다.

3.2 Patent Evaluation System

특허는 기술에 대한 상세한 내용을 포함한다. 상세한 내용은 기술의 우수성, 시장성 등과 같은 요인들에 의해 평가된다. 종래의 특허 평가는 기술 전문가에 의해 진행되어왔다. 그러나 전문가 기반의 평가는 많은 시간과 비용이 소모되며 전문가의 의견에 편향될 수 있다는 단점을 갖는다. 최근에는 전문가 기반 특허 평가 시스템을 개선하기 위해 특허 빅데이터 분석 기반의 특허 평가 시스템들이 개발되고 있다[16].

한국발명진흥회의 SMART3는 특허의 청구항 수, 피인용 수, 청구항의 텍스트 길이와 같은 정량지표를 활용하여 특허를 9개의 등급으로 평가한다. 특허정보진흥센터에서 개발된 K-PEG 특허 평가 시스템도 특허의 정량지표를 활용하여 특허의 생존기간을 예측하며 텍스트 정보를 활용하여 유사한 특허를 탐색하는 two-track 서비스를 제공한다. 기술보증기금의 KPAS는 딥뉴럴네트워크를 이용하여 출원인 수, 청구항 수, 청구항의 단어 수와 같은 정량지표를 학습하여 특허의 권리 이전 여부를 예측하는 시스템이다. 위트인텔리전스는 한국발명진흥회의 SMART3를 실제 비즈니스 환경에 부합하도록 변형한 특허 평가 서비스 KEV를 운영 중에 있다. KEV는 출원인의 주요 기업 여부, 패밀

리국가별 조정계수, 보유특허 조정계수 등을 요인으로 분석하여 특허를 평가하는 시스템이다. 위즈도메인은 발명자의 등록건수, 패밀리 특허 건수, 청구항 수와 같은 정량지표를 활용하여 특허를 등급으로 평가한다.

현재 운영되고 있는 대표적인 특허 평가 시스템은 모두 특허의 서지 정보를 활용하지 않는 공통점을 가지고 있다. K-PEG은 특허의 텍스트 정보를 활용하여 유사특허를 탐색해주는 서비스를 제공하고 있지만, 정량지표와 함께 결합하여 운영되고 있지 않다. 따라서 특허의 다양한 정량지표와 함께 특허의 서지 정보를 결합하여 특허를 평가하는 시스템의 개발이 필요하다.

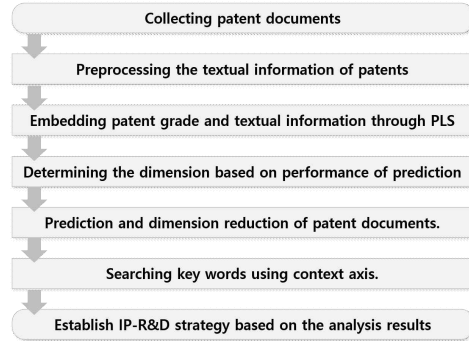
IV. 제안 방법론

본 논문에서는 PLS 방법을 이용하여 특허 평가점수와 텍스트 정보를 함께 학습하고 특허 평가점수를 예측하는 방법을 제안한다. 아래 그림1은 본 연구에서 제안하는 방법의 흐름을 나타낸 것이다. 먼저, 특허 DB로부터 특허 문서를 수집한다. 그리고 특허의 서지 정보를 전처리하여 문서×단어 행렬을 구축한다. 다음으로 문서×단어 행렬을 PLS 알고리즘으로 학습한다. 이때, 특허 등급을 예측하는 방향으로 학습하며 예측력이 가장 우수한 경우를 최적의 차원 수 k^* 로 정하고 예측변수를 임베딩한다.

이러한 과정을 거쳐 얻은 PLS 모델은 특허의 정량 지표와 서지 정보를 결합한 핵심특허 예측 모델이 된다. 그리고 PLS 모델로부터 k^* 개의 PLS 성분과 식(1)을 만족하는 T 를 얻을 수 있다.

$$V \leftarrow \arg \text{Min} \left(\cos^{-1} \frac{T \cdot V}{|T||V|} \right) \quad (3)$$

식(3)은 핵심특허와 상관이 높은 단어 벡터를 찾는



<그림 1> 제안된 방법론의 흐름도

수식이다. k^* 개의 PLS 성분은 특허 평가등급과 상관이 높은 벡터이다. 따라서 PLS 성분 벡터와 T 벡터는 내적을 통해 두 벡터끼리 이루는 각 θ 의 코사인 값을 구할 수 있다. V 는 PLS 성분 벡터와 T 벡터가 이루는 각이 최소인 벡터이다.

특허 평가등급은 PLS를 통해 특허의 서지 정보와 함께 k^* 차원으로 임베딩된다. k^* 차원의 공간에서 PLS 축은 서지 정보를 반영하여 특허 평가등급을 잘 설명하는 축이다. 이를 이용하여 PLS 축과 단어 벡터의 내적을 구하고 이를 역변환하면 두 축의 각을 구할 수 있다. 제안된 모델을 활용하면 핵심특허를 예측하고, 핵심특허와 상관이 높은 단어 벡터를 탐색하는 것이 가능하다.

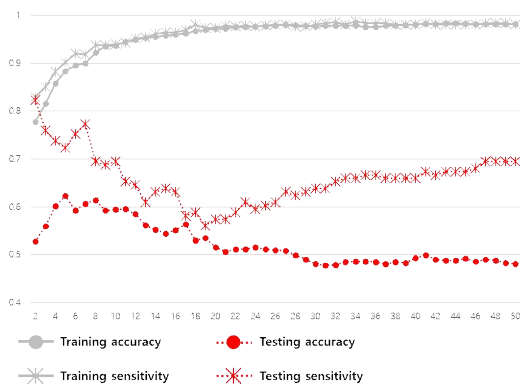
V. 실험 및 결과

본 연구는 특허 평가등급과 텍스트 정보를 함께 학습한 모델을 통해 핵심특허를 예측하고, 핵심특허와 상관이 높은 단어를 탐색하는 방법을 제안한다. 제안된 방법의 실제 적용 가능성을 확인하기 위해 실험을 실시하였다. 실험 데이터로는 자율주행자동차 관련 특허를 선정하였다. 자율주행자동차 관련 특허는 위즈도메인 특허 DB를 통해 수집하였다. 수집된 특허

는 1993년부터 2017년까지 미국에 출원된 1,831건이다. 수집된 데이터는 제안된 모델에 적용하기 위하여 워즈도메인에서 제공하는 특허 평가등급을 우수등급과 비우수등급으로 나누었다. 나누어진 데이터 중 우수등급을 받은 특허는 556건이다.

수집된 특허는 제안된 모델에 적용하기 위하여 문서×단어 행렬로 전처리하였다. 전처리된 단어 수는 모두 9,244개이다. 이때, 문서×단어 행렬의 원소는 Tf-Idf 값을 사용하였다[17]. 제안된 모델의 성능을 측정하기 위하여 전체 데이터의 70%에 해당하는 2012년 이전 데이터는 학습 데이터로 사용하였고 2013년 이후 데이터는 검정 데이터로 사용하였다.

아래 그림2는 학습 데이터와 검정 데이터를 이용하여 최적 PLS 성분 수를 찾는 과정을 보여준다. 실험을 위해, 부분최소제곱법을 나이브베이시안 모델에 적용하였다. 따라서 $m-1$ 번째 나이브베이시안 모델이 반응변수를 설명하지 못하는 부분을 m 번째 나이브베이시안의 반응변수에 적용한다. m 은 2부터 50까지 변화시키면서 진행하였고, 모델의 성능은 정확도(Accuracy)와 민감도(Sensitivity)를 사용하여 측정하였다.



〈그림 2〉 PLS 성분 수에 따른 모델의 성능 측정 결과

정확도와 민감도를 이용하여 최적 PLS 성분 수를

탐색한 결과, PLS 성분 수가 2일 때, 검정 데이터에 대한 민감도가 0.83으로 가장 우수하였다. 따라서 본 실험에서의 최적 PLS 성분 수는 2로, 9,244개의 단어 벡터를 2차원에 임베딩하였다.

제안된 모델의 예측 성능을 주성분 분석(PCA, Principal Component Analysis)을 사용한 경우와 비교한다. 주성분 분석은 대표적인 비지도 학습 기반의 차원축소 방법이다. 수집된 데이터에 주성분 분석을 적용한 결과, 622차원으로 축소했을 때, 97.5%의 정보량을 유지하는 것이 가능하였다. 훈련 데이터를 622차원으로 차원축소한 뒤, 로지스틱 회귀모델(PCA_LR, PCA based Logistic Regression)과 나이브 베이시안 모델(PCA_NB, PCA based Naïve Bayes)로 특허 등급을 학습하였다. 모델 학습이 이루어진 후, 검정 데이터에 대해 예측 성능을 측정한 결과, 로지스틱 회귀모델은 정확도가 0.59, 민감도 0.71이었고 나이브 베이시안 모델은 정확도가 0.57, 민감도가 0.59이었다. 아래 표1은 제안된 모델과 PCA 기반의 모델의 성능 비교 결과이다. 성능을 비교한 결과, 제안된 모델이 PCA 기반의 로지스틱 회귀모델과 나이브 베이시안 모델보다 정확도와 민감도가 모두 높았다.

〈표 1〉 제안된 모델과 PCA 기반 모델의 비교 결과

Model	Test Accuracy	Test Sensitivity
Proposed Model	0.62	0.83
PCA LR	0.59	0.71
PCA NB	0.57	0.59

자율주행자동차 관련 특허는 제안된 모델에 의해 2차원의 PLS 성분에 임베딩되었다. 2차원의 PLS 성분에 대해 식(3)을 만족하는 단어 벡터 리스트는 아래 표2와 같다. 첫 번째 PLS 성분은 특허등급을 예측하기 위한 가장 많은 정보를 포함하는 축이다. 첫 번째 PLS 성분과 상관이 높은 단어 벡터 리스트는 1st_PLS에 기재되어 있으며, 두 번째 PLS 성분과 상관이 높

은 단어 벡터 리스트는 2nd_PLS에 기재되어 있다. 그리고 첫 번째 PLS 성분과 두 번째 PLS 성분의 평균 벡터와 상관이 높은 단어 벡터 리스트는 Avg_PLS에 기재하였다.

〈표 2〉 PLS 성분과 상관이 높은 단어 벡터 리스트

Components	Word vector List
1 st _PLS	belt, seat, guide, brake, engage, end, gear, transmission, web, pressure
2 nd _PLS	unit, lane, line, mark, detect, image, extract, inform, road, candidate
Avg_PLS	unit, gear, automate, transmission, line, brake, shift, park, pedal, mark

첫 번째 PLS 성분과는 belt, seat, guide 순으로 가까웠다. 이 단어들은 주로 안전과 관련된 단어들이었다. 두 번째 PLS 성분과는 unit, lane, line 순으로 가까웠으며 해당 단어 벡터는 차선 감지와 관련된 단어들이었다. 마지막으로 두 개의 PLS 성분의 평균 벡터와는 unit, gear, automate 순으로 가까웠고, 해당 단어 벡터들은 구동과 통제에 관련된 단어들이었다. 즉, 안전, 차선 감지 순으로 관련된 단어가 핵심특허에 가장 많은 요인을 설명하는 것을 알 수 있다[18-21].

VI. 결론 및 향후 연구

최근 특허 빅데이터 기반 IP-R&D 방법론에 대한 연구가 활발히 진행 중에 있다. 특허 빅데이터 기반 IP-R&D 방법은 종래의 방법에 비해 시간과 비용을 절약할 수 있고, 객관적인 분석 결과를 얻을 수 있는 장점을 지닌다. 선행연구들에서는 특허의 정량지표를 머신러닝 모델에 적용하여 핵심특허의 예측, 특허의 가치산정 등에 관한 연구를 주로 수행하였다. 그러나 특허는 개발된 기술의 내용이 청구항에 상세히 기재되어 있어 이를 활용한 모델의 개발이 필요하다.

본 논문에서는 특허의 정량지표와 서지정보를 결합하고 특허문서 중 핵심키워드 산출, 핵심특허 예측이 가능한 모델을 제안하였다. 제안된 모델의 실제 적용 가능성을 확인하기 위해 자율주행자동차 관련 특허 1,831건을 활용하였다. 자율주행자동차 관련 특허를 제안된 모델에 적용한 결과, 검정 데이터에서 민감도 0.83의 성능을 보이는 것으로 확인되었다. 그리고 제안된 모델에 의해 탐색된 PLS 성분과 높은 상관성을 갖는 단어 벡터를 탐색하였다. 그 결과, 안전, 차선 감지, 구동과 통제에 관한 단어들이 핵심특허에 가장 많은 요인을 설명하는 것을 알 수 있다. 그러므로 안전과 차선 감지, 이를 위한 구동과 통제에 관한 자율주행자동차 기술이 관련 분야에서 핵심 기술이라고 할 수 있다.

향후 연구에서는 제안된 모델보다 예측 성능이 우수한 모델 구조에 대한 연구가 진행될 필요가 있다. 본 연구에서는 제안된 모델의 최적 성분 수를 민감도로 결정하였다. 그러나 보다 우수한 예측모델링을 위해서는 민감도와 특이도(specificity) 성능 측도가 모두 높은 모델이 요구된다. 따라서 향후 연구에서는 본 연구의 한계점인 예측성능을 개선할 필요가 있다. 그리고 제안된 모델에 의해 도출된 핵심키워드를 반영할 수 있는 특허 빅데이터 기반의 IP-R&D 방법에 대한 연구도 필요하다.

참고문헌

- [1] 특허청 · 한국발명진흥회, 특허와 정보분석, 경성문화사, 서울, 2007.
- [2] 특허청 · 한국지식재산전략원, 특허관점의 R&D 혁신전략, 서울, 2013.
- [3] 전성해 · 최준혁, “특허와 빅데이터, 그 연결은?,” 한국지능시스템학회 논문지, 제24권, 제2호, 2014, pp.43-44.

- [4] 박상성, "통계분석을 이용한 기업의 기술 지원 전략에 관한 연구," 한국지능시스템학회 논문지, 제29권, 제4호, 2019, pp.316-320.
- [5] Trappey, A., Wu, C. and Lin, C., "A patent quality analysis for innovative technology and product development," *Advanced Engineering information*, Vol. 26, No. 1, 2012, pp.26-34.
- [6] 양동홍 · 김성철 · 장근석, "특허가치평가를 위한 합리적 로열티율 산정 방안-손해액산정모형을 중심으로," *기술혁신학회지*, 제15권, 제3호, 2012, pp.700-721.
- [7] Wu, J., Chang, C., Tsao, C. and Fan, Y., "A patent quality analysis and classification system using self-organizing maps with support vector machine," *Applied soft computing*, Vol. 41, 2016, pp.305-316.
- [8] 우호영 · 광정애 · 임창원, "구조방정식 모형의 베이시안 접근법 기반의 특허평가 모델링에 대한 연구," *응용통계연구*, 제30권, 제6호, 2017, pp.901-916.
- [9] 최재현, "지속성장 가능한 기업 탐색을 위한 방법," *한국지능시스템학회 논문지*, 제28권, 제4호, 2018, pp.417-421.
- [10] 이준석 · 이준혁 · 김갑조 · 박상성 · 장동식, "데이터 마이닝을 통한 기술경영 전략 수립에 관한 연구," *한국지능시스템학회 논문지*, 제25권, 제2호, 2015, pp.126-132.
- [11] Geladi, P. and Kowaski, B., "Partial least-square regression: A tutorial," *Analytica Chimica Acta*, Vol. 185, 1986, pp.1-17.
- [12] Barker, M. and Rayens, W., "Partial least squares for discrimination," *Journal of Chemometrics*, Vol. 17, 2003, pp.166-173.
- [13] Tseng, Y., Lin, C. and Lin, Y., "Text mining techniques for patent analysis," *Information Processing & Management*, Vol. 43, No. 5, 2007, pp.1216-1247.
- [14] Fordellone, M., Bellincontro, A. and Mencarelli, F., "Partial least squares discriminant analysis: A dimensionality reduction method to classify hyperspectral data," eprint arXiv:1806.09347, 2018.
- [15] Tibshirani, R., Hastie, T., Narasimhan, B. and Chu, G., "Diagnosis of multiple cancer types by shrunken centroids of gene expression," In *Proceedings of the National Academy of Sciences*, Vol. 99, No. 10, 2002, pp.6567-6572.
- [16] Wold, S., Trygg, J. and Eriksson, L., U.S. Patent No. 8244498. MKS Instruments : United States Patent and Trademark Office.
- [17] Ramos, J., "Using tf-idf to determine word relevance in document queries," In *Proceedings of the first instructional conference on machine learning*, Vol. 242, 2003, pp.133-142.
- [18] Park, S., Jeong, H., Kim, L. and Yun, I., "Development of Safety Evaluation Scenario for Autonomous Vehicle Take-over at Expressways," *The Journal of The Korea Institute of Intelligent Transport Systems*, Vol. 17, No. 2, 2018, pp.142-151.
- [19] Hancock, P., Nourbakhsh, J. and Stewart, J., "On the future of transportation in an era of automated and autonomous vehicles," In *Proceedings of the National Academy of Sciences*, Vol. 116, No. 16, 2019, pp.7684-7691.
- [20] 진병욱 · 차시호, "차량인터넷을 포함한 V2X 환경에서 안전한 차량 통신 서비스 제공을 위한 해시 트리 기반 통신 프로토콜," *한국디지털산업정보학회 논문지*, 제14권, 제1호, 2018, pp.22-34.
- [21] 최병관, "자율 주행자동차 오픈플랫폼 온톨로지

구축을 위한 스마트디바이스 연구,” 디지털산업 정보학회 논문지, 제15권, 제3호, 2019, pp.1-14.

■ 저자소개 ■



이 주 현
Lee, Juhyun

2018년 3월~현재
고려대학교 산업경영공학과
석·박사통합과정
2018년 2월
청주대학교 통계학과(이학사)
관심분야 : Patent analysis, Statistical learning, NLP
E-mail : leejuu@korea.ac.kr



이 준 석
Lee, Junseok

2019년 9월~현재
엠아이큐브 솔루션
2019년 8월
고려대학교 산업경영공학과
(공학박사)
2013년 2월
Zhejiang University, Mechanical Engineering and Automation
(공학사)
관심분야 : Patent analysis, Data Mining, Management of technology
E-mail : jxli12@korea.ac.kr



강 지 호
Kang, Jiho

2015년 3월~현재
고려대학교 산업경영공학과
박사과정
2014년 2월
아주대학교 산업경영공학
(공학석사)
2012년 2월
아주대학교 산업정보시스템공학부
(공학사)
관심분야 : Management of technology, Innovation strategy, Data science, Text mining
E-mail : kangmae@korea.ac.kr



박 상 성
Park, Sangsung

2019년 3월~현재
청주대학교 소프트웨어융합학부
조교수
2015년~2018년
고려대학교 기술경영전문대학원
조교수
2006~2014년
고려대학교 산업경영공학부
연구교수
2006년
고려대학교 산업시스템정보공학과
(공학박사)
관심분야 : Patent analysis, Data mining, Management of technology
E-mail : hanyul@cju.ac.kr



장 동 식
Jang, Dongsik

1989년~현재
고려대학교 산업경영공학부 교수
1988년
Texas A&M 산업공학과(공학박사)
1985년
Texas 주립대학교 산업공학과
(공학석사)
1979년
고려대학교 산업공학과(공학사)
관심분야 : Project management, Pattern recognition, Data mining
E-mail : jang@korea.ac.kr



홍 성 욱
Hong, Sungwook

2013년 1월~현재
주식회사 에스와이피 / 에스와이피
특허법률사무소 대표 변리사
2010년 2월
성균관대학교 기술경영학과
(공학석사)
2006년 2월
성균관대학교 정보통신공학부
전기 및 전자공학과(공학사)
관심분야 : 기술권리화, IP컨설팅, 기술거래 및 사업화, 창업지원 및 멘토링
E-mail : hong@syqip.com



김 선 영
Kim, Sunyoung

2017년 12월~현재
주식회사 에스와이피 / 에스와이피
특허법률사무소 변리사
2011년 8월
연세대학교 공학대학원 전파통신
공학과 (공학석사)
2005년 2월
고려대학교 전기·전자·전파
공학부 (공학사)
관심분야 : 기술권리화, IP컨설팅, 연구소기업
설립, 기술사업화 및 투자연계
E-mail : sykim@syqip.com

논문접수일 : 2019년 11월 26일

수정일 : 2019년 12월 4일

게재확정일 : 2019년 12월 6일