특허기술 고객충성도에 대한 기계학습 예측모델링 연구*

A Modeling Study of Machine Learning Prediction on
Patent Technology Customer Loyalty

박진희 (Park, Jin - Hee)** · 신건권 (Shin, Gun - Kwon)***

〈요 약〉

본 연구에서는 기계학습 알고리즘을 적용하여 특허기술의 고객충성도의 결정요인이 무엇인지를 탐색하고자 하였다. 기계학습 알고리즘인 선형모델, 선형회귀모델 및 인공신경망모델을 적용하여 예측모델링을 수행한 결과 인공신경망모델이 가장 예측력의 정도가 높은 것으로 분석되었다. 인공신경망모델을 통해 분석한 결과, 기술충성도의 영향요인으로 매출액 변화율과 연구개발비 변화율이 다른 예측변인들에 비해 중요한 것으로 나타났다.

본 연구는 기술충성도의 예측을 위해서 기계학습 기법을 통한 예측모델링을 수행하고 실무적인 시사점을 도출했다는 점에서 의의가 있다. 특히 본 연구에서는 기술충성도 예측모델의 성능평가를 통해 선형회귀모델에 의한 것보다 인공신경망이 예측오차가 가장 적어서 예측성능이 우수하다는 것을 검증하였다. 이와 함께 기술충성도의 영향요인으로 매출액 변화율과 연구개발비 변화율 등이 다른 예측변인들에 비해 중요하다는 점을 알 수 있었다.

본 연구는 기술충성도의 예측변수로 특허기술의 특성변수인 질적 우수성과 활용가능성만을 채택했다는 점에서 한계가 있을 수 있다. 향후 연구에서는 다양한 예측변수(특허경영활동, 자원역량, 조직역량, 성과확산 역량 등)를 도입하여 연구할 필요가 있다. 또한 연구대상을 A연구원만으로 제한했다는 점에서 연구 결과를 일반화하는 데 한계가 있다고 볼 수 있다. 따라서 향후 연구에서는 연구대상 기관 수를 증가시키는 것이 필요할 것이다.

한글검색어 : 특허기술, 고객충성도, 기계학습, 기술이전성과, 기업성과

• 접수일자 : 2019.5.25 • 1차수정일 : 2019.6.15 • 최종수정일 : 2019.6.26 • 게재확정일 : 2019.6.28

* 본 논문은 2018학년도 기술경영경제학회 하계학술대회에서 발표한 논문과 박진희의 박사학위논문을 일부 발췌하여 수정 보완한 것입니다.

** 호서대학교 기술경영전문대학원 기술경영공학박사, 제1저자, e-mail : pjh@kitech.re.kr

*** 호서대학교 기술경영전문대학원 교수, 교신저자, e-mail : shingok@hoseo.edu

I. 서 론

정부에서는 정부출연연구소와 같은 공공연구기관들이 보유한 특허기술들을 기업으로 이전할 수 있도록 지원하고 있다. 그러나 공공연구기관들의 특허기술 이전과 확산은 매우 미흡하며 개선이 요구되는 실정이다. 특히 연구기관의 특허 수입을 증대시키기 위한 모든 노력은 고객의 관점으로 전환되어 수행되어야 한다. 이러한 노력 중의 하나가 바로 어떻게 하면 연구기관들이 보유한 특허들이 기술 수요자들에게 지속적으로 이전되도록 할 것인가에 초점을 맞추는 것이다. 이것이 바로 고객충성도(customer loyalty)를 확보하는 전략의 수행이다.

고객충성도는 전통적으로 “고객이 제품 구매 상황에 대해 깊이 관여되어 향후에도 계속해서 선호하는 상품이나 서비스, 아이디어를 다시 구매 또는 이용하는 것”으로 정의되어 왔다(Oliver, 1999). 이후의 선행연구들에서도 고객충성도는 구전의도, 추천의도, 재이용의도 등과 같은 유사한 개념으로 이용되고 있다. 고객충성도는 급속히 변화하는 경영환경 속에서 기업이 생존할 수 있는 핵심요소 중의 하나라고 볼 수 있다.

특허와 관련된 선행연구들은 주로 기술이전 당사자인 기술 수요자의 관점이나 기술 사업화를 주제로 한 연구가 많이 수행되어 왔다. 그러나 기술공급자의 관점에서 특허기술의 우수성이나 활용 가능성 등이 기술이전성과나 기업성과 그리고 기술충성도에 미치는 영향에 대한 연구는 매우 미흡한 실정이다(양동우와 김수정, 2008 ; 황경연과 성을현, 2014 ; 구본철, 2014 등).

기계학습을 통한 예측모델링의 선행연구들은 특허분석이 기업이나 개인 그리고 국가의 기술경쟁력을 제고시킬 수 있는 중요한 분야로 인식하고 다양한 연구들을 수행해 오고 있다. 주요한 연구 분야로는 키워드 특허분석 연구, 특허인용네트워크 기반 소셜네트워크분석 연구, 특허문서 분류 연구, 특허문서 마이닝 연구 및 특허관련 예측모델 연구 등이 있다(박진희, 2018). 기업들은 자체적으로 기술개발을 수행하기도 하지만 경우에 따라서 기술이전을 통해서도 기술을 획득할 수 있다. 결국 기술이전을 적극적으로 추진하기 위해서는 기계학습 기반의 예측모델을 개발하여 기업의 연구개발 의사결정을 효율적으로 지원할 필요성이 있다. 왜냐하면 많은 기업은 대부분 기술사업화를 위한 특허 선정 시에 주로 전문가 의견(expert opinion)에 의존하는 경우가 많은데, 이는 전문가가 해당 분야에 대한 충분한 지

식이 없는 경우에 기술이전이나 사업화 가능성이 높은 특허를 잘못 판별할 수 있기 때문이다. 따라서 다수의 특허를 보유하고 있는 A연구원의 측면에서 볼 때 기술마케팅을 효과적으로 수행하기 위해서는 기술충성도에 미치는 영향요인이 무엇인가를 규명하고, 더욱 객관적인 기계학습 기반의 예측모델(predictive model)을 구축하는 것이 필요하다. 그러나 국내에서는 특허데이터를 이용한 기계학습 기반의 예측모델링 연구가 많이 활성화 되어 있지 않은 실정이다. 따라서 향후 연구에서는 공공연구기관들이 특허기술에 대한 경쟁력을 확보할 수 있도록 지원하기 위해서 기술공급자의 관점에서 기술충성도의 결정요인이 무엇인가를 탐색하는 것이 필요하다.

위와 같은 연구의 필요성에 따라서 본 연구에서는 기계학습 알고리즘을 적용하여 기술충성도를 추정하고 예측하는 연구모델을 생성하고 평가하였다.

본 연구의 자료수집방법은 A연구원이 2017년 말 보유하고 있는 3,942건의 특허 중에서 기술평가지표가 존재하고 이전된 기업의 특성변수를 확인할 수 있는 746건으로 하였다. 기계학습을 통한 예측모델링을 위한 통계처리프로그램으로는 SPSS Modeler 17을 사용하였다.

II. 기계학습을 이용한 예측모델링의 선행연구

2.1 기계학습과 분류 체계

예측은 인간 생활 속의 다양한 분야에서 그 필요성과 중요성이 높다. 기후, 주가, 질병 등의 예측에서부터 경제학, 의학 등과 같은 다양한 분야에서 예측을 위한 연구들이 수행되고 있다. 최근에 이러한 예측을 용이하게 수행하기 위한 기법으로 기계학습(machine learning)이 많이 사용되고 있다. 기계학습은 데이터를 이용해 학습을 진행하고 이를 통해 새로운 데이터에 대해 학습된 알려진 속성을 기반으로 예측하는 방법이다.

일반적으로 기계학습 알고리즘을 적용하면 목표변수(종속변수)에 대한 예측모델(predictive model)을 생성할 수 있다. 이를 보통 예측모델링이라 한다. 기계학습은 크게 지도학습(supervised learning)과 비지도학습(unsupervised learning)으로 구분된다(데이터솔루션 빅데이터 러닝센터, 2018). 전자는 지도를 받아서 학습하는 것을 말하며 이에는 분류모델링

(classification modeling)과 추정모델링(estimation modeling)이 있다. 후자는 스스로 알아서 학습하는 것을 말하며 이에선 군집화모델링(clustering modeling)과 연관성모델링(association modeling)이 있다. 이외에도 지도학습에서 강화학습(reinforcement learning)을 분리하여 세 가지 유형으로 분류하기도 한다.

2.2 특허관련 기계학습 연구

기계학습을 통한 연구 분야는 공학과 의학 등을 중심으로 활발하게 수행되고 있다. 공학 분야의 적용 사례로는 로봇 기술의 관점에서 시각과 청각에 활용될 수 있는 제어, 영상, 음향 등에서 분류, 판단 등을 예측하는 분야이다. 의학 분야에서는 환자나 의료기관의 편의를 증진하기 위한 목적의 연구가 수행되어 왔다. 대다수 기계학습기법을 적용한 연구들은 이론적 측면의 연구보다는 실용적으로 응용할 수 있는 연구나 실험 등이 주류를 이루고 있다(Choi et al., 2015 ; 이호현, 정승현 및 최은정, 2016). 의학과 공학 분야에서 모두 SVM(Support Vector Machine), 인공신경망(Artificial Neural Network), 의사결정나무(Decision Trees) 등의 기계학습기법이 많이 적용되고 있다.

특허와 관련된 선행연구들을 보면 크게 키워드 기반 특허분석 연구(최진호, 김희수 및 임남규, 2011), 특허인용네트워크 기반 소셜네트워크분석 연구(Social Network Analysis : SNA)(박준형과 광기영, 2013), 특허문서 분류 연구(박찬정, 성동수, 이건배, 2012 ; 이주현 외, 2018), 특허문서 마이닝 연구(김현우 외, 2015), 특허관련 예측모델 연구(유재복과 정영미, 2010 ; 구정민 외, 2010 ; 전성해, 2011, 2013 ; Jun and Park, 2013 ; Choi et al., 2015 ; 정병기, 고남욱 및 윤장혁, 2016) 등으로 구분할 수 있다. 연구의 분석기법으로는 SNA와 기계학습기법들이 사용되었다.

특허관련 예측모델을 구축한 선행연구들을 살펴보면 <표 1>과 같다. 먼저, 특허문서에 대한 텍스트 마이닝 기술을 이용해 공백기술의 예측모델을 구축한 연구가 있다. 전성해(2011)는 특허가 가지고 있는 서지사항, 제목, 초록, IPC(국제특허분류 코드) 등을 이용해 지능형 시스템에 관한 공백기술의 예측모델을 구축하였으며 2개의 공백기술 군집(지능형 입출력에 관한 세부기술, 지능형 통신에 관한 세부기술)을 발견하였다.

〈표 1〉 특허관련 기계학습 선행연구 요약

연구자	방법	주요 연구 결과
유재복과 정영미 (2010)	다중회귀분석	특허의 인용에 영향을 미치는 주요 변수들을 이용해 특허의 피인용 횟수를 예측하기 위한 모델을 제시
구정민 외 (2010)	텍스트 마이닝	텍스트 마이닝 기법을 이용한 특허문서 분석을 통하여 특허 등록 및 거절 여부를 예측하는 방법을 제안
전성해 (2011)	텍스트 마이닝 군집분석	특허를 가지고 서지사항, 제목, 초록, IPC(국제특허분류 코드) 등을 이용해 지능형 시스템에 관한 공백기술의 예측모델을 구축
전성해 (2013)	텍스트 마이닝 다중회귀분석	특허문서에 대해 회귀모델을 통한 빅 데이터 학습을 적용하여 특허분석을 수행하고 이 결과를 기술예측에 적용하는 방법을 제시
Choi et al. (2015)	SNA, 선형회귀분석 의사결정나무	특허분석을 통해 기술이전에 대한 예측모델을 구축하기 위해서 SNA, 선형회귀분석 및 의사결정나무 모델링을 적용
정병기, 고남욱 및 윤장혁 (2016)	심층신경망	대량의 미국 특허로부터 추출된 정보와 특허유지비용의 납부 데이터 그리고 심층신경망 알고리즘을 적용해 특허수명 예측모델을 구축

또한 전성해(2013)는 특허문서에 대해 회귀모델을 통한 빅 데이터 학습을 적용하여 특허 분석을 수행하고 이 결과를 기술예측에 적용하는 방법을 연구하였다. 연구 결과 기술예측의 실제 적용에 대한 지식으로 빅 데이터 관련 기술의 개발을 위해서 통계적으로 유의한 관련성이 있는 단어들을 발견하였고 이 단어들로 이루어진 기술들에 대한 선행 개발이 필요하다는 결론을 내었다. 다음으로, 특허수명과 특허 인용횟수의 예측모델 연구가 있다. 유재복과 정영미(2010)는 특허의 인용에 영향을 미치는 주요 변수들을 이용해 특허의 피인용 횟수를 예측하기 위한 모델을 제시하였다. 분석 결과 예측변수로 사용된 독립변수 중에서 특허 피인용횟수에 가장 영향력이 높은 변수는 ‘문헌 간 유사도’로 나타났다. 구정민 외(2010)는 텍스트마이닝 기법(text mining)을 이용한 특허문서 분석을 통하여 특허 등록 및 거절 여부를 예측하는 방법을 제안하였다. 특허 거절 기준값을 선정하기 위해서 군집화 알고리즘을 적용하여 실험하고 분석한 결과, 거절된 특허문서와 유사한 특허문서는 거절될 가능성이 높다는 결론을 얻었다. 정병기, 고남욱 및 윤장혁(2016)은 대량의 미국특허로부터 추출된 정보와 특허유지비용의 납부 데이터 그리고 심층신경망 알고리즘을 적용해 특허수명 예측모델을 구축하였다. 이들의 연구는 특허의 기대수명이라는 예측하기 어려운 지표를 모델을 통해

객관적으로 정량화시켰다는 점에서 기여점이 있다고 하겠다. Jun and Park(2013)은 SNA를 적용해 애플사의 미래기술 예측을 위해서 모든 특허를 대상으로 분석하였다. 이를 통해 애플사의 기술추세와 애플사의 기술들 간의 관계를 파악하였다. 분석 결과 애플사의 공백기술(vacant technologies)과 기술혁신을 위한 핵심기술을 발견하였다. 마지막으로, Choi et al.(2015)는 특허분석을 통해 기술이전에 대한 예측모델을 구축하기 위해서 SNA, 선형회귀분석 및 의사결정나무 모델링을 적용하였다. 이 연구의 기본 목적은 기술이전 여부에 가장 영향을 크게 미치는 예측변수가 무엇인지를 규명하는 것이었다. 이를 위해서 기술이전 결과를 포함한 특허데이터를 수집하여 데이터 전처리를 하였다. 분석을 시행한 결과 첫째, 국가연구기관 특허에 대한 기술이전 영향요인에는 권리성, 권리만료예상일, 패밀리수, 자국 인용횟수와 신규성의 순이었다. 둘째, 국립대 특허의 영향요인에는 신규성, 권리성, 자국과 인용횟수의 순으로 나타났다. 셋째, 사립대 특허의 영향요인에는 패밀리국가수, 권리성과 발명자 수, 권리만료예상일, 자국인용횟수이었다. 공동연구 특허의 영향요인에는 신규성, 권리성, 발명자 수와 권리만료예상일이었다. 마지막으로 국가 특허의 영향요인에는 신규성, 권리만료예상일, 청구항수의 순으로 나타났다. 본 연구는 특허정보와 통계 및 기계학습 알고리즘을 결합해 기술의 잠재적 가치를 반영할 수 있는 방법론을 제안하였고, 특허지표를 활용함으로써 보다 객관적인 기술정보를 반영했다는 점에서 의의가 있다고 하겠다. Kang et al.(2018)은 기계학습의 앙상블기법 중의 하나인 AdaBoost 알고리즘에 기반한 기술이전예측모델을 제안하는 것을 목적으로 하고 있다. 이를 위해 특허 데이터를 수집하였고, 수집된 특허의 기술 분야를 식별하기 위해서 토픽모델링을 적용하였다. 목표변수는 기술이전 여부(더미변수)이며 예측변수는 특허 후행 인용 수, 적용에서 등록까지의 기간, 등록된 청구항수, 패밀리 특허의 수, 패밀리국가의 수, 특허 참조의 수, 특허이외의 논문, 책 등의 참조수, 토픽모델링에 의한 토픽 수(더미변수)이었다. 제안된 방법의 성과 비교를 위해서 대표적인 분류 알고리즘인 KNN, 서포트벡터머신, 신경망을 사용하였다. 분석 결과 본 연구에서 제안한 모델은 충분한 분류 성과를 가지는 것으로 확인하였다.

III. 예측모델링을 위한 연구방법론

3.1 변수의 조작적 정의와 측정

본 연구의 예측모델링에 적용된 독립변수(특징변수)와 종속변수(레이블변수)에 대한 조작적 정의와 측정방법은 <표 2>와 같다.

<표 2> 연구변수의 조작적 정의와 측정변수의 측정방법

연구변수	조작적 정의	측정변수	출처	자료 특성	비고
특허기술의 질적 우수성	특허기술의 특허집중도, 질적 우수성 및 시장확보력의 정도	특허집중도	기관자체 진단결과	연속형 (1-5점)	
		질적 우수성	SMART K-PEG		
		시장확보력	기관자체 진단결과		
특허기술의 활용가능성	특허기술 시장성의 정도	특허활동력	기관자체 진단결과	연속형 (1-5점)	
		사업화 성공률	기관자체 진단결과		
		특허영향력	기관자체 진단결과		
기업성과	특허기술의 활용 결과로서의 기업성과	매출액 변화율	신용평가 회사의 기업 재무자료	$\frac{[(\text{이전 후 매출액} - \text{이전 전 매출액}) / \text{이전 전 매출액}] \times 100}{}$	정규성 충족을 위해 log변환
		연구개발비 변화율		$\frac{[(\text{이전 후 R\&D비용} - \text{이전 전 R\&D비용}) / \text{이전 전 R\&D비용}] \times 100}{}$	
		종사자 수 변화율		$\frac{[(\text{이전 후 종사자 수} - \text{이전 전 종사자 수}) / \text{이전 전 종사자 수}] \times 100}{}$	

〈표 2〉 연구변수의 조작적 정의와 측정변수의 측정방법 (계속)

연구변수	조작적 정의	측정변수	출처	자료 특성	비고
기술이전 성과	특허기술을 이용한 기술료 수입 정도	계약금액	기관 보유자료	계약금액	정규성 충족을 위해 log변환
		계약형태		0=무상매각, 1=통상, 2=전용, 3=유상매각	
		실시범위		0=국내 1=해외포함	
기술충성도	특정 특허기술과 유사 특허기술을 지속적으로 이전한 건수	해당 특허의 이전 수	기관 보유자료	연속형 (1-7점)	

3.2 특허기술 고객충성도의 예측모델

본 연구에서는 특허기술에 대한 고객충성도의 결정요인에 대한 분석을 위해서 [그림 1]과 같이 기계학습에 입각한 예측모델(predictive model)의 적용을 통해서 추가적으로 A연구원이 효율적인 특허관리(patent management)를 지원할 수 있는 방안을 모색하고자 한다. 이러한 분석은 A연구원이 어떠한 예측변수에 초점을 두고 특허관리를 해야 하는지에 대한 대안을 제공해 줄 수 있을 것이다.

예측모델에서 특허기술의 질적 우수성의 세 가지 측정변수(시장확보력, 질적 우수성, 특허집중도), 특허기술의 활용가능성의 세 가지 측정변수(사업화성공율, 특허영향력, 특허활동력), 기술이전성과의 세 가지 측정변수(계약금액, 계약형태, 실시범위), 기업성과의 세 가지 측정변수(매출액 변화율, 연구개발비 변화율, 종사자 수 변화율) 등의 12개 측정변수들은 기계학습 모델링에서 입력변수(독립변수, 혹은 예측변수)로 투입된다¹⁾. 이에 반해 기술충성도의 한 가지 측정변수(기술이전 수)는 목표변수(종속변수)로 투입된다.

1) 측정변수 중에서 실시범위와 종사자 수 변화율은 본 연구의 최종적인 변수의 신뢰도와 타당도검증 결과 임계치에 미치지 못하여 삭제하는 것으로 결정했으며, 이후 분석에서는 이 변수들을 연구모델에서 제외하였다.

[그림 1] 기술충성도의 예측모델

예측모델의 추정을 위해서 SPSS Modeler 기계학습 알고리즘 중에서 선형알고리즘(linear algorithm), 선형회귀알고리즘(linear regression algorithm), 인공신경망알고리즘(neural network algorithm)을 적용한다.

IV. 특허기술 고객충성도의 예측모델링과 평가

4.1 특허기술 고객충성도 예측모델의 추정

4.1.1 선형모델에 의한 고객충성도의 예측모델 추정

선형모델 알고리즘에 의해 추정된 기술충성도의 예측모델에서 목표변수인 이전 수를 예측하는데 있어서 예측변수의 중요도(예측자 중요도)는 매출액 변화율(0.84), 계약형태(0.12),

연구개발비 변화율(0.02), 질적 우수성(0.01), 시장확보력(0.00), 계약금액(0.00), 사업화 성공률(0.00), 특허영향력(0.00), 특허집중도(0.00), 특허활동력(0.00)의 순이었다.

선형모델에 의한 예측모델을 구축한 결과는 <표 3>과 같다.

<표 3> 선형모델 알고리즘에 의한 예측모델 구축 결과 요약

예측변수(입력변수)	계수	유의성	중요도
절편	-19.722	0.000	
매출액 변화율	4.310	0.000	0.842
계약형태(0)	1.517	0.000	0.118
계약형태(1,2)	-0.421	0.000	0.118
연구개발비 변화율	0.335	0.000	0.023
질적 우수성	-0.262	0.039	0.007
시장확보력	0.082	0.075	0.005
계약금액	-0.022	0.174	0.003
사업화성공률	-0.044	0.362	0.001
특허영향력	0.032	0.476	0.001
특허집중도	-0.009	0.843	0.000
특허활동력	0.000	0.997	0.000

기술충성도인 이전 수를 예측하는데 있어서 가장 중요한 변수는 매출액 변화율이었고 가장 중요하지 않은 변수는 특허활동력인 것으로 나타났다. 이외에도 이전 수에 유의한 영향을 미치는 변수는 계약형태(0), 계약형태(1,2), 연구개발비 변화율 및 질적 우수성이었다. 그러나 계약형태(1,2)는 목표변수인 이전 수에 부의 영향을 미친 것으로 나타났다.

4.1.2 선형회귀모델에 의한 고객충성도의 예측모델 추정

선형회귀모델 알고리즘에 의해 추정된 기술충성도의 예측모델에서 목표변수인 이전 수를 예측하는데 있어서 예측변수의 중요도(예측자 중요도)는 매출액 변화율(0.45), 연구개발비 변화율(0.41), 계약형태(0.05), 특허활동력(0.03), 특허집중도(0.02), 시장확보력(0.02), 특허영향력(0.02), 질적 우수성(0.01), 계약금액(0.00), 사업화 성공률(0.00)의 순이었다.

전체 회귀식에 대한 분산분석표에 의하면 $F=121.009$ 이고, 유의확률은 0.000이어서 선형

회귀모델 알고리즘을 통해 도출된 회귀식은 의미가 있는 것으로 나타났다. 구체적인 선형회귀모델 추정 결과를 요약하면 <표 4>와 같다.

<표 4> 선형회귀모델 알고리즘에 의한 예측모델의 추정 결과

모델	비표준화 계수		표준화 계수	t	유의확률
	B	표준오차	베타		
(상수)	- 8.366	.349		- 23.953	.000
특허집중도	.051	.054	.037	.930	.353
질적우수성	- .157	.152	- .042	- 1.030	.304
시장확보력	- .034	.053	- .022	- .640	.523
특허영향력	- .005	.054	- .003	- .084	.933
사업화성공률	- .094	.058	- .063	- 1.623	.105
특허활동력	.072	.064	.049	1.123	.262
매출액 변화율	1.769	.116	.668	15.251	.000
연구개발비 변화율	.545	.069	.312	7.939	.000
계약금액	- .058	.010	- .237	- 5.835	.000
계약형태	.093	.060	.055	1.544	.123

분석 결과에 의하면 매출액 변화율, 연구개발비 변화율, 계약금액은 목표변수인 이전 수에 유의한 영향을 미쳤다. 매출액 변화율과 연구개발비 변화율은 이전 수와 정의 관계를 보였으나 계약형태는 이전 수와 부의 관계를 나타냈다. 이는 매출액 변화율과 연구개발비 변화율이 높아질수록 목표변수인 이전 수가 많아질 수 있음을 시사해 주는 것이다. 반면에 계약형태의 결과는 기술특허의 제공을 유상보다는 무상으로 했을 때 기술충성도가 높아질 수 있음을 의미한다.

4.1.3 인공신경망모델에 의한 고객충성도의 예측모델 추정

인공신경망모델 알고리즘에 의해 추정된 기술충성도의 예측모델에서 목표변수인 이전 수를 예측하는데 있어서 예측변수의 중요도(예측자 중요도)는 매출액 변화율(0.65), 연구개발비 변화율(0.10), 질적 우수성(0.04), 사업화성공률(0.04), 계약금액(0.04), 특허영향력(0.03), 특허집중도(0.03), 특허활동력(0.03), 시장확보력(0.02), 계약형태(0.02)의 순이었다. 결과적

으로 인공신경망모델 알고리즘을 적용한 경우에는 목표변수 예측에 있어서 매출액 변화율이 가장 중요한 영향을 미쳤고 계약형태가 가장 영향을 적게 미쳤음을 알 수 있다. 매출액 변화율에 이어서 목표변수에 중요한 영향을 미치는 변수는 연구개발비 변화율로 분석되었다.

4.2 특허기술 고객충성도 예측모델의 비교 평가

기계학습 알고리즘을 통해 기술충성도의 예측모델을 추정했으며, 각 모델의 설명력을 요약하면 <표 5>와 같다. 선형회귀모델의 설명력은 68.5%, 선형모델은 79.3%, 인공신경망모델은 97.7%로 본 연구에서 적용한 모델 중에서 인공신경망모델이 가장 설명력이 높은 것으로 나타났다.

<표 5> 기술충성도 예측모델의 설명력 비교

기계학습 모델	설명력(수정된 R^2)
선형모델	79.3%
선형회귀모델	69.5%
인공신경망모델(MLP)	97.7%

기술충성도 예측모델의 성능평가 결과는 <표 6>과 같다. 선형모델 알고리즘은 평균절차 오차(Mean Absolute Error : MAE)가 훈련과 테스트모델 모두에서 0.5 정도를 나타내고 있어 기술충성도인 이전 수 예측에 있어서 0.5건 정도의 오차가 있음을 의미한다. 선형회귀모델 알고리즘을 적용한 경우에도 평균절대오차는 훈련모델이 0.5건, 테스트모델이 0.6건 정도의 오차를 나타냈다. 반면에 인공신경망모델의 망 성능을 평가한 결과 훈련과 테스트모델에서 모두 0.3건 정도의 오차를 보이고 있다. 결국 기술충성도 예측모델의 성능평가 결과에 의하면 인공신경망이 예측 성능이 가장 높고, 다음으로 선형모델과 선형회귀모델의 순인 것으로 나타났음을 알 수 있다.

〈표 6〉 기술충성도 예측모델의 성능평가 결과

파티션	선형모델		선형회귀모델		인공신경망	
	훈련	테스트	훈련	테스트	훈련	테스트
최소 오류	- 1.416	- 1.416	- 2.570	- 1.949	- 0.803	- 0.920
최대 오류	1.847	1.856	2.871	2.739	1.660	1.700
평균 오류	- 0.000	0.045	- 0.000	0.074	- 0.003	0.035
평균절대오차	0.500	0.506	0.573	0.601	0.345	0.336
표준편차	0.633	0.631	0.768	0.832	0.451	0.459
선형 상관관계	0.893	0.903	0.837	0.832	0.950	0.956
발생	528	218	528	218	528	218

[그림 2]는 기술충성도 예측모델의 이익도표를 나타낸 것이다. 이는 모델 추정 결과를 바탕으로 추정 정확도를 평가하는 방법이다. 가로축에는 0-100%까지의 백분위수를, 세로축에는 [그 비율에 들어가는 Hit 사례 수]/[전체 Hit 사례 수]×100%를 넣어 구성한 도표이다. 선형모델, 선형회귀모델 및 인공신경망모델의 종합적인 누적이익도표를 보면 훈련모델의 경우는 백분위수 9에서 이익이 100%로 나타났고, 테스트모델의 경우는 백분위수 11에서 이익이 100%로 나타났다. 기술충성도 예측에 사용된 세 가지 알고리즘은 모두 가파르게 상승하는 곡선을 그리고 있어 좋은 적합도를 보이며, 모두 예측력이 우수한 것으로 평가된다.

[그림 2] 기술충성도 예측모델의 이익도표

4.3 특허기술 고객충성도 예측모델의 분석 결과 요약과 논의

본 연구에서 수행된 예측모델 추정 결과를 제시하고 논의하면 다음과 같다.

첫째, 예측모델의 설명력은 선형모델이 79.3%, 선형회귀모델이 69.5% 그리고 인공신경망 모델이 97.7%로 나타났다. 분석 결과에 의하면 예측모델의 설명력은 인공신경망모델이 가장 높고 다음으로 선형모델, 선형회귀모델의 순으로 나타났다.

둘째, 훈련데이터를 사용한 예측모델의 평균절대오차는 선형모델이 0.5, 선형회귀모델이 0.5, 인공신경망모델이 0.3으로 나타났다. 예측오차가 가장 작게 나타난 예측모델은 인공신경망모델로 이전 수를 예측하는데 있어서 0.3건의 오차가 있었다. 따라서 예측모델의 성능 평가 결과는 인공신경망모델이 가장 높고, 다음으로 선형모델과 선형회귀모델의 순임을 알 수 있다.

셋째, 선형모델에 의해 기술충성도를 추정한 결과 매출액 변화율의 변수 중요도가 가장 높게 나타났고 특허활동력이 가장 중요도가 낮은 변수로 나타났다.

넷째, 선형회귀모델에 의해 기술충성도를 추정한 결과 매출액 변화율의 변수 중요도가 가장 높게 나타났고 사업화성공률이 가장 낮은 중요도를 갖는 것으로 나타났다.

다섯째, 인공신경망모델에 의해 기술충성도를 추정한 결과 매출액 변화율의 변수 중요도가 가장 높았고 계약형태가 가장 중요하지 않은 변수로 나타났다. 두 번째 중요도가 있는 변수는 연구개발비 변화율이었다.

이상에서 살펴 본 것처럼 기계학습 기법을 사용하여 기술충성도의 예측모델을 추정한 결과 본 연구에서는 인공신경망모델이 가장 설명력과 예측력이 높은 것으로 나타났다. 또한 세 가지 예측모델 모두에서 공통적으로 가장 중요한 예측변수는 매출액 변화율인 것으로 분석되었다. 따라서 기술공급자인 A연구원의 입장에서 볼 때 기술충성도를 지속적으로 확보하기 위해서는 매출액 규모가 높은 기업들을 대상으로 기술마케팅을 시행하는 것이 중요하다고 판단된다.

V. 결 론

5.1 연구 결과의 요약

공공연구기관들이 기술마케팅을 효과적으로 수행하기 위해서는 기술충성도에 미치는 영향요인이 무엇인가를 규명해야 한다. 이를 위해서는 객관적인 기계학습 기반의 예측모델(predictive model)을 구축과 활용이 요구된다. 그러나 국내에서는 특허데이터를 이용한 기계학습 기반의 예측모델링 연구가 많이 활성화 되어 있지 않은 실정이었다. 이러한 연구의 필요성에 따라 본 연구에서는 기계학습 알고리즘을 적용하여 특허기술의 고객충성도의 결정요인이 무엇인지를 탐색하고자 하였다.

본 연구에서 기술충성도에 대한 예측모델을 구축하고 추정한 결과를 요약하면 다음과 같다. 먼저, 기계학습 알고리즘인 선형모델, 선형회귀모델 및 인공신경망모델을 분석한 결과 인공신경망모델이 가장 예측력의 정도가 높은 것으로 분석되었다. 다음으로, 인공신경망모델을 통해 분석한 결과, 기술충성도의 영향요인으로 매출액 변화율과 연구개발비 변화율이 다른 예측변인들에 비해 중요한 것으로 나타났다.

5.2 연구 결과의 시사점과 한계 그리고 향후 연구방향

본 연구는 기술충성도의 예측을 위해서 기계학습 기법을 통한 예측모델링을 수행하고 실무적인 시사점을 도출했다는 점에서 의의가 있다. 특히 본 연구에서는 기술충성도 예측모델의 성능평가를 통해 선형회귀모델에 의한 것보다 인공신경망이 예측오차가 가장 적어서 예측성능이 우수하다는 것을 검증하였다. 이와 함께 기술충성도의 영향요인으로 매출액 변화율과 연구개발비 변화율 등이 다른 예측변인들에 비해 중요하다는 점을 알 수 있었다.

본 연구는 기술충성도의 예측변수로 특허기술의 특성변수인 질적 우수성과 활용가능성만을 채택했다는 점에서 한계가 있을 수 있다. 향후 연구에서는 다양한 예측변수(특허경영활동, 자원역량, 조직역량, 성과확산 역량 등)를 도입하여 연구할 필요가 있다. 또한 연구대상을 A연구원만으로 제한했다는 점에서 연구 결과를 일반화하는 데 한계가 있다고 볼 수 있다. 따라서 향후 연구에서는 연구대상 기관 수를 증가시키는 것이 필요할 것이다.

본 연구의 결과를 토대로 A연구원이 참조할 수 있는 몇 가지 정책적 제언을 하면 다음과 같다. 먼저, 기계학습 기반의 타겟기업을 도출하는 등 우수기술의 선택과 잠재수요기업을 확장하는 전략의 추구가 요구된다. 다음으로, 심층신경망 기반의 정보검색시스템에 의한 정보검색 엔진을 개발해 기업의 기술수요 대응과 기술이전 마케팅을 활성화하는 것이 필요하다. 이를 위해서는 다양한 분야에서 심층신경망 기법을 보유했던 기술 조희에 적용하여 기술이전 수요자의 니즈를 충족시키는 것도 대안이 될 수 있을 것이다.

참고문헌

〈국내문헌〉

- 구본철. (2014). 연구성과의 기술이전 및 사업화 촉진요인 도출 및 실증분석. *벤처창업연구*, 9(5), 69 - 81.
- 구정민, 박상성, 신영근, 정원교, 장동식. (2010). 특허 등록 예측을 위한 특허 문서 분석 방법. *한국과학기술학회 논문지*, 11(4), 1458 - 1467.
- 김현우, 김종찬, 이준혁, 박상성, 장동식. (2015). 핵심 기술 및 특허 추출을 위한 IP 마이닝에 관한 연구. *한국지능시스템학회 논문지*, 25(4), 392 - 397.
- 데이터솔루션 빅 데이터 러닝센터. (2018). SPSS Modeler와 예측분석 교육자료.
- 박준형, 광기영. (2013). 특허 인용 관계가 기업 성과에 미치는 영향 : 소셜네트워크분석 관점. *지능정보연구*, 19(3), 127 - 139.
- 박진희. (2018). 특허기술의 질적 우수성과 활용가능성이 기술이전성과, 기업성과와 기술충성도에 미치는 영향. 호서대학교 *기술경영전문대학원 박사학위논문*.
- 박찬정, 성동수, 이진배. (2012). 기계학습을 이용한 특허 문서의 자동 IPC 분류. *한국정보기술학회 논문지*, 10(4), 119 - 128.
- 양동우, 김수정. (2008). 기술공급자 (R&D 기관)의 기술이전애로요인에 관한 기초연구. *대한경영학회지*, 21(1), 205 - 227.
- 유재복, 정영미. (2010). 특허인용 예측모형 구축에 관한 연구. *정보관리학회지*, 27(4), 239 - 258.
- 이주현, 강지호, 박아람, 박상성, 장동식. (2018). 기계학습을 이용한 특허 분류의 성능 비교에 대한 연구. *한국컴퓨터정보학회 하계학술대회 논문집*, 26(2), 539 - 540.
- 이호현, 정승현, 최은정. (2016). 기계학습 응용 및 학습 알고리즘 성능 개선방안 사례연구. *Journal of Digital Convergence*, 14(2), 245 - 258.
- 전성해. (2011). 특허분석을 이용한 지능형시스템의 기술예측. *한국지능시스템학회 논문지*, 21(1), 100 - 105.
- 전성해. (2013). 특허분석을 위한 빅 데이터학습. *한국지능시스템학회 논문지*, 23(5), 406 - 411.
- 정병기, 고남옥, 윤장혁. (2016). 딥 뉴럴 네트워크 기반의 특허수명 예측모형 개발. *대한산업공학회 추계학술대회 논문집*, 981 - 985.
- 최진호, 김희수, 임남규. (2011). 기술예측을 위한 특허 키워드 네트워크 분석. *지능정보연구*, 17(4), 227 - 240.
- 황경연, 성을현. (2014). 기술의 특성, 기술사업화성과 및 경영성과 간의 관계에 관한 실증연구 : 대덕 연구개발특구 정부출연 연구기관에서 이전된 기술을 중심으로. *생산성논집*, 28(2), 25 - 49.

〈외국문헌〉

- Choi, J., Jang, D., Jun, S. & Park, S. (2015). A predictive model of technology transfer using patent analysis. *Sustainability*, 7, 16175 – 16195.
- Jun, S., & Sung Park, S. (2013). Examining technological innovation of Apple using patent analysis. *Industrial Management & Data Systems*, 113(6), 890 – 907.
- Kang, L.J., Jun, J.H., Lim, H., Jang, D., & Park, S. (2018). Ensemble modeling for sustainable technology transfer. *Sustainability*, 10(7), 2278.
- Oliver, R L.(1999). Whence customer loyalty?. *Journal of Marketing*, 63, 33 – 44.

“본 논문은 학회윤리규정을 준수하였으며 논문에 대한 판권을 학회에 위임합니다.”

<Abstract>

A Modeling Study of Machine Learning Prediction on Patent Technology Customer Loyalty

Park, Jin - Hee* · Shin, Gun - Kwon**

In this study, we tried to find out the determinants of customer loyalty of patent technology by applying machine learning algorithm. As a result of predictive modeling using linear learning model, linear regression model and artificial neural network model, artificial neural network model has the highest predictive power. As a result of analyzing through the artificial neural network model, the rate of change of sales and the rate of change of R&D expenditure are more important than other predictors.

This study is meaningful in that it predicts model loyalty through machine learning technique and draws practical implications for prediction of technology loyalty. Especially, in this study, the performance evaluation of the technology loyalty prediction model proved that the artificial neural network has the least prediction error than the linear regression model, and the prediction performance is excellent. In addition, we found that the rate of change in sales and the rate of change in R&D expenditure are more important than other predictors.

This study has limitations in that it adopts only qualitative excellence and application possibility which is a characteristic variable of patent technology as a predictive variable of technology loyalty. In future research, it is necessary to study various predictive variables (such as patent management activities, resource capabilities, organizational capabilities, and performance diffusion capabilities). In addition, there is a limit to the generalization of research results because it limits research subjects to only A researchers. Therefore, it will be necessary to increase the number of institutions to be studied in future research.

Key words : Patent loyalty, Customer loyalty, Machine learning, Prediction

* Ph.D. of Management of Technology, Hoseo Graduate School of MOT, First Author,
e-mail : pjh@kitech.re.kr

** Professor, Hoseo Graduate School of MOT, Corresponding Author, e-mail : shingk@hoseo.edu