

오피니언 마이닝 및 특허분석을 통한 사용자 니즈기반 이종영역 기술기회 탐색

장혜진 · 노태연 · 윤병운[†]

동국대학교 산업시스템공학과

User Needs-Based Technology Opportunities in Heterogeneous Fields Using Opinion Mining and Patent Analysis

Hyejin Jang · Taeyeoun Roh · Byungun Yoon

Department of Industrial and System Engineering, Dongguk University

In a digital economy, users actively express their needs in many ways. Thus, many researchers analyze what users need and whether they are satisfied or not through opinion mining. In addition, they begin to find technology opportunities in heterogeneous technology fields. But they did not connect users' opinion to technology development process, only focused on natural language processing or marketing or manufacturing area. Also, heterogeneous technology fields are focused on fusion technology. Thus, this study suggests a novel approach that is based on sentimental value and can be applied to exploring technology opportunities in heterogeneous fields. Sentimental value is calculated from users' opinion through sLDA. The heterogeneous technology opportunity is explored by patent analysis. This research contributes to suggesting a hybrid methodology through patent and users' opinion. In addition, it can provide managerial efficiency by suggesting base data onto decision making.

Keywords: Opinion Mining, Supervised LDA, Technology Opportunity, Heterogeneous Technology, Patent Analysis

1. 서론

사용자의 참여로 인한 정보 공유 및 개방이 활발히 일어나는 Web 2.0 시대가 도래하고, 이를 활용하기 위한 오피니언 마이닝 연구가 활발하게 일어나고 있다. 오피니언 마이닝은 설문조사, 여론조사 등 고객 의견정보를 수집하는데 필요한 시간과 경제적 비용을 획기적으로 줄일 수 있는 방법으로, 수요 견인형 기술개발에 활발하게 사용되고 있다. 수요 견인형 기술개발은 기술개발 관점 중 하나로서,

고객이 선호하는 가치 혹은 고객이 바라는 기술의 특성을 충족시키기 위하여 신기술을 개발하는 것(Paradigms and Trajectories, 1982)을 의미한다.

오피니언 마이닝은 고객들이 자유롭게 서술한 의견을 객관적이고 정량적인 정보로 변환시킬 수 있는 방법으로, 다양한 유형의 연구가 진행되고 있다. Ghazizadeh *et al.*(2014)은 특정 단어의 출현여부를 통해 문장의 극성을 판단하였으며, Lee (2009), Lee *et al.*(2008), Wu *et al.*(2014)은 문장의 구성형식과 연관관계 분석을 통해 고객의 요구사항을 도출 하였다. 또한, 오피니언 마이닝을 통하여 의미 있는 리뷰를 평가하고(Liu *et al.*, 2008; Liu *et al.*, 2013), 기존고객의 니즈를 분석하여 잠재적인 고객을 예측하는 모형(Thoreleuchter *et al.*, 2012)을 개발 하였다.

최근 기술개발을 위한 새로운 기회로서 기존에 주력하고 있던 동종 분야가 아닌 다른 분야에서 기회를 탐색하는 연구가

본 논문은 한국연구재단의 인문공동연구지원사업의 지원을 받아 진행되었음(NRF-2014S1A5A2A03065010).

[†] 연락저자 : 윤병운 교수, 04620 서울특별시 중구 필동로 1길 30 동국대학교 공과대학 산업시스템공학과, Tel : 02-2260-8743, Fax : 02-2269-2212,

E-mail : postman3@dongguk.edu

2016년 9월 17일 접수; 2016년 12월 23일 수정본 접수; 2017년 1월 20일 게재 확정.

증가하고 있다. 이종 산업과 동종 산업과의 특징 및 연관 정도를 파악하여 융합을 통한 기술 기회, 신사업 기회 등을 탐색하는 연구가 진행되고 있다.

오피니언 마이닝 연구를 통해 고객의 의견을 분석하여 니즈를 도출한 연구가 많이 진행되었지만, 이를 산업과 연결시켜 기술개발에 직접적으로 활용한 연구는 매우 드물다. 또한 오피니언 마이닝 연구는 자연어 처리를 강조한 오피니언 마이닝 모형 자체에 대한 연구가 대부분이었으며, 활용분야가 제품 및 마케팅 분야에 국한되어 있다. 이종영역 연구의 경우 융합 기술을 중심으로 수행되고 있으며, 특히 데이터만을 분석하여 기술기회 방법론을 제시한 연구가 주를 이루고 있다.

따라서 본 연구에서는 사용자의 니즈를 대변할 수 있는 리뷰 데이터에서 오피니언 마이닝을 통해 핵심적인 정보를 추출하고 이를 이용하여 실질적인 산업에 활용될 수 있는 이종영역 기술기회 탐색 방법론을 제안하고자 한다. 이를 위하여, 텍스트 마이닝의 일종인 supervised Latent Dirichlet Allocation(sLDA)을 활용한 오피니언 마이닝을 통해 리뷰들의 주제와 각 주제의 감성 극성 값(sentimental value), 대상이 되는 부분(part), 속성(attribute)을 라벨링한다. 또한 분석대상 특허를 수집하고 Latent Dirichlet Allocation(LDA)을 활용하여 이종·동종 기술 분야를 정의한다. 이후 사전에 정의된 리뷰의 주제와 기술 분야를 코사인 유사도를 통해 연결시키고, 기술 분야의 특징, 리뷰들의 감성 극성 값을 이용하여 외향형(Outbound) 기술기회와 내향형(Inbound) 기술기회를 탐색한다.

2. 배경이론

2.1 오피니언 마이닝

오피니언 마이닝은 자연어 처리과정, 언어학, 텍스트 마이닝을 바탕으로 작성자가 자유롭게 작성한 의견 데이터에서 객관적인 정보를 추출하고 정의하는 것을 말한다. 텍스트 형태의 데이터에서 긍정, 부정, 중립 의견으로 분류하므로 감성 분석이라고도 한다. 이때, 분석의 결과는 긍정적인 의견은 양수로 부정적인 의견은 음수로 표현되며, 표현되는 값을 감성 극성 값이라고 한다.

오피니언 마이닝은 감성분석과 내용분석의 두 가지 관점에서 수행된다. 감성분석은 문서의 감성 극성 값을 구하는 것을 목적으로 하며, 내용기반 분석은 문장의 품사, 구조 등을 분석함으로써 핵심이 되는 내용을 추출하는 연구이다. 감성 분석 기반의 연구의 경우, 긍정과 부정과 관련 있는 단어를 사전에 정의한 감성 사전을 기반으로 하여 감성 극성 값을 분석하거나(Penalver et al., 2014; Singh et al., 2013), 의견과 함께 제시된 평점을 이용하여 의견에 대한 감성 극성 분류 및 예측 모형을 개발하는 연구가 진행되었다(Ghiassi et al., 2013; Min and Park, 2012).

내용 기반 분석은 목적에 따라 보다 구체적인 정보를 획득하기 위하여 사용되었다. Thoreleuchter(2012)는 의견을 작성한

고객정보를 활용하여 잠재적인 미래의 고객을 예측하는 모델을 제안하였고, Lee(2009), Lee et al.(2008)은 사용자 의견에서 단어 간 연관관계, 문장구조 분석, 특정 단어 출현 유무 등을 활용하여 고객의 니즈를 추출하기 위한 방법론을 개발하였다.

성공적인 사용자의 의견을 기술개발 과정에 활용하기 위해서는 의견의 긍정 및 부정의 정도뿐 아니라 사용자의 구체적인 요구사항을 도출하는 것이 필수적이다. 따라서 본 연구에서는 오피니언 마이닝을 기술 개발 과정에 직접적으로 활용하기 위해 감성 극성 값을 제품에 대한 사용자의 만족도로 해석하고 감성분석의 관점과 내용분석의 두 가지 관점을 모두 고려한 오피니언 마이닝을 수행하고자 한다.

2.2 특허분석을 통한 이종기술 기회 탐색

동종 산업 내에서 경쟁사보다 경쟁우위를 확보하기 위해 보다 복잡하고 고도화된 기술개발의 필요성이 강조됨에 따라 기술기회 탐색의 전략적 중요성이 증가하고 있다(Lee et al., 2015). 기술기회 탐색은 ‘다양한 환경요인을 고려하여 미래에 기회를 지닌 기술을 찾고, 이와 관련한 사회적 이슈와 유망성 등을 측정하는 활동’으로 정의된다(KISTI, 2012). 동종 산업 내에서 기술기회를 탐색하는 것에서 나아가 다양한 분야와의 융합을 통한 기술기회 탐색이 진행되고 있으며, 이를 위해 이종영역에서의 기술기회에 주목하고 있다.

이종영역이란 특정산업에서의 기술개발에 있어 개발하고자 하는 기술이 속한 산업 외의 해당산업 이외의 영역이라 정의된다. 본 연구에서는 이종영역에 해당하는 기술을 이종기술이라 정의한다. 예를 들어, 자동차산업에 포함되는 문에 관한 기술을 개발하고자 한다면, 개발 목표가 되는 산업은 자동차 산업이며, 자동차 이외의 산업에 해당하는 문과 관련되는 기술은 이종기술이라 할 수 있다.

특허는 다양한 기술정보를 포함하고 있으므로, 특허를 분석하여 기술 분야의 특징을 도출할 수 있다. 특허의 정보는 IPC(International Patent Classification)코드, 인용관계정보, 출원인, 등록인 등 구조화 된 정보와 제목, 요약, 청구항 등 비 구조화 된 정보로 나눌 수 있다. 또한 특허는 구조화 된 정보를 활용하여 양적 지표로써 기술력을 평가할 수 있다. 인용수를 활용한 영향력 평가, 여러 국가에서 권리를 파악할 수 있는 패밀리 특허수를 활용한 시장력 평가 등 다양한 지표가 활용되고 있다.

구조화된 정보를 다양한 방식으로 분석하고 해석하는 연구가 다수 진행되었다. Jeong et al.(2014), Lee(2005), Choi et al.(2007)은 구조화된 정보 중 가장 대표적으로 활용되고 있는 등록 특허수를 이용하여 유망분야 및 융합된 기술 분야를 선정하였으며, An et al.(2016)은 IPC코드를 이용하여 ICT분야의 융합기술을 도출하였다. Yun(2011)과 Lim et al.(2015)은 특허의 인용/피인용 횟수를 이용하여 인용관계가 많을수록 특허, 기술의 중요도가 뛰어나다고 해석하였다. Jo et al.(2005)은 독

립 청구항의 수가 많고, 청구항의 내용이 적을수록 뛰어난 특허임을 제시하였다. 또한 Kwon(2012)은 패밀리 특허수가 많은 특허가 보다 중요한 특허임을 보였다.

구조화된 정보에 기반한 특허지표를 이용하여 이중기술을 분석한 연구가 활발히 진행되고 있다. No and Park(2010), Bin et al.(2015)은 IPC코드와 인용 정보를 활용하여 융합 패턴을 파악 하였고, Ko et al.(2014)은 특허에서 얻을 수 있는 정보와 기술 분류와 산업 분류 간 일치정도를 분석하여 산업별 기술 융합 정도를 평가하였다. 또한, 최근에는 텍스트 마이닝을 통해 구조화된 정보와 비 구조화된 정보를 복합적으로 활용한 지식 흐름을 탐색한 연구(No et al., 2015), IPC코드와 텍스트 유사도를 사용한 융합 대상을 선정하는 연구(Huang et al., 2015), 키워드 유사도를 기반으로 IPC별 인문·기술 융합 현황을 파악하는 연구(Kim and Lee, 2016) 등이 진행되었다.

본 연구에서는 특허의 구조화된 정보를 이용하여 각 기술 분야의 기술력을 판단할 수 있는 지표를 제시한다. 이를 바탕으로 각 기술 분야의 기술력을 판단하고, 기술력 대비 니즈유형을 구분하여 두 가지 이중영역 기술 탐색 방향을 모색한다. 이후, 탐색 방향에 따라 지표를 활용하여 각 기술이 사용자의 니즈를 만족시킬 수 있는지를 예측한다.

2.3 토픽 모델링

(1) LDA모형

LDA는 주제를 도출해내는 모형 중 하나로(Blei et al., 2003), 제안된 텍스트 마이닝 방법론 중 가장 뛰어난 효율을 보이고 있어, 텍스트 마이닝 분야에서 활발하게 활용되고 있다. LDA는 Dirichle 확률분포를 기반으로 주제들의 확률 분포와 각 주제에 대하여 특정 단어가 주제에 포함 될 확률을 통해 단어들로부터 특정 문서가 만들어질 확률을 분석하는 모형이다(Blei et al., 2003).

LDA 모형은 주제별 단어분포를 확인 할 수 있다는 점을 활용하여 클러스터링의 일종으로 사용되고 있다. 각 텍스트의 단어 분포를 기준으로 군집화 되므로, 다른 클러스터링 방법론에 비하여 포함하고 있는 의미가 반영된다는 장점이 있다. Kim et al.(2016)은 이를 이용하여 특허 로드맵 작성을 위해 특허의 키워드를 활용한 기술 분야를 도출하였고, Lee and Sohn(2015), Jeong and Lee(2016)와 Park and Song(2013)은 연구 동향을 파악하기 위해 논문 제목 및 초록을 바탕으로 주제를 도출하였다. Jin et al.(2013)은 트위터 데이터를 대상으로 네트워크 분석을 통해 이슈 변화를 탐색하였다. 본 연구에서는 LDA를 통해 도출된 주제별 키워드를 포함하고 있는 특허문서를 해당 주제에 할당하는 방법으로 특허문서를 군집화 하고, 키워드 목록을 이용하여 니즈와 기술 분야 간 연관성을 분석한다.

(2) LDA 확장 모형

LDA 모형은 시계열 분석, 감성 극성 분석 등 다른 분석 모형

을 통해 보다 많은 의미를 부여 할 수 있는 모형으로 확장이 가능하다. 이를 이용하여 Griffiths and Steyvers(2004)는 시계열 분석으로 확장된 모형을 통해 논문의 최신 연구주제와 연구가 줄어들고 있는 주제를 도출하였다. Mcauliffe and Blei(2008)은 예측 값을 포함한 데이터를 기반으로 지도학습 모델로 확장시켜 sLDA 모형을 제안하였다. 이는 기존 LDA 모형에서 사용되는 단어와 문서영역을 그대로 사용하지만, 예측 값을 포함하는 정보를 기반으로 하여 모형에 적용된다. 본 연구에서는 문서별로 오피니언 마이닝을 통해 감성 극성 값이 계산된 문서를 sLDA 모형에 적용하여, 사용자 리뷰로부터 주제와 관련된 단어 목록을 도출하고, 각 주제의 감성 극성 값을 도출한다. 이를 이용하여 사용자의 리뷰에서 핵심적인 주제를 도출하고, 각 주제에 대한 사용자가 느끼는 긍정 또는 부정의 정도를 구할 수 있다.

3. 프레임워크

3.1 연구 아이디어

본 연구의 목표는 특정 제품 및 서비스의 리뷰데이터를 분석하여 사용자의 요구사항을 도출하고, 이와 연관된 동종영역과 이중영역의 기술을 분석하여 이중영역의 기술기회를 탐색하는 것이다. 이를 위해 <Figure 1>과 같이 각 사용자 니즈별로 동종·이중 기술 영역에서 니즈와 관련된 기술 분야를 찾고, 기술 역량과 감성 극성 값을 기준으로 하여 외향형, 내향형 두 가지 유형의 기술기회를 탐색한다. 이때, 동종기술 영역의 기술역량이 낮음에도 불구하고 요구사항을 만족하는 경우와, 동종기술의 영역이 기술역량이 높음에도 불구하고 요구사항을 만족하지 못하는 경우는 이중영역의 기술이 사용자의 요구사항을 만족 할 수 있다고 보장할 수 없으므로 기술기회 탐색분야에서 제외 한다.

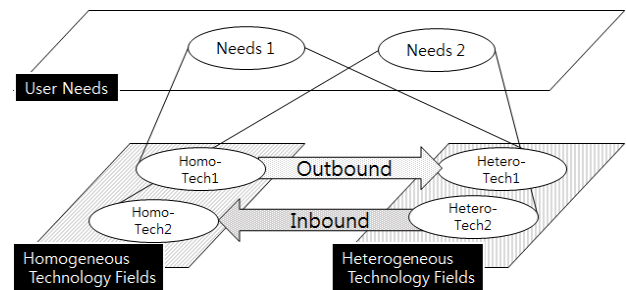


Figure 1. Research Concept

3.2 연구 프로세스

본 연구는 두 가지 유형의 기술기회를 탐색하기 위해 <Figure 2>와 같이 세 가지 단계로 구성된다. 먼저, 사용자의 의견을 오피니언 마이닝과 sLDA를 사용하여 사용자의 니즈를 대표할

수 있는 주제 및 주제에 대한 만족도를 분석한다. 주제와 관련된 기술 분야를 선정하기 위해 동종기술 특허를 LDA로 클러스터링하고 코사인 유사도를 이용하여 관련된 니즈와 연결시킨다. 이후 기술력에 따른 만족도를 구분하기 위해 특허 평가 지표를 통해 각 기술 분야의 기술력을 평가한다. 이후 니즈별 만족도와 관련 기술의 기술력을 바탕으로 기술기회 탐색 방향을 선정하고, 동종 기술영역의 기술역량과 만족도를 이용하여 도출된 다중 선형 회귀식에 이종기술의 기술력을 도입하여 만족도 변화량을 예측한다.

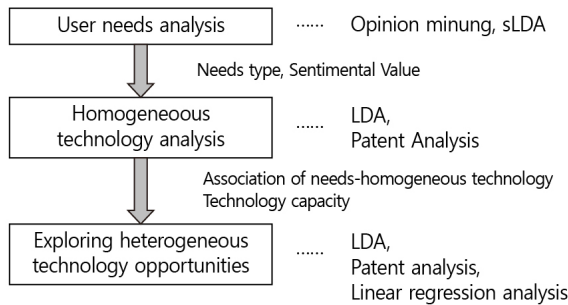


Figure 2. Research Framework

(1) 사용자 니즈 분석

사용자의 니즈를 포함하고 있는 리뷰 데이터를 수집하기 위해 통계분석 프로그램 R의 웹 크롤링 패키지를 통해 수집한다. 수집된 사용자의 리뷰 데이터와 sLDA 모델을 통해 핵심적인 주제와 각 주제의 키워드 리스트를 도출한다. sLDA를 통해 주제의 감성 극성 값을 계산하기 위해서는 각 문서별 감성 극성 값이 필요하다. 이를 위해 Sementria 소프트웨어로 각 문장의 감성 극성 값을 계산하여 모형에 도입한다. 이후 기술과 관련된 키워드를 대상으로 사용자가 의견을 표출하는 구성요소 (part), 사용자가 각 구성요소에 대해서 어떠한 감정을 느꼈는지를 속성(attribute)로 정의하고 분석 대상의 part-attribute 사전을 작성한다.

이후 보다 구체적으로 사용자 니즈와 기술을 연결하기 위하여 각 주제별로 Liu(2012)가 제안한 오피니언의 5개 요소인 목표(object), 대상(entity), 세부 속성(attribute), 감성 극성 값(sentiment), 날짜(date) 중 목표와 대상을 part로 해석하여 각 요소를 라벨링한다. 이때, 사전에 정의된 part-attribute 사전을 이용하여 주제별 키워드를 part, attribute로 구조화하고, 감성 극성 값, 날짜가 라벨링된 데이터를 얻을 수 있다.

사용자 니즈의 유형을 정의하기 위하여 주제별 만족도와 시간적 변화량을 활용한다. 주제별 만족도는 각 기간별로 더 많이 인지된 주제에 가중치를 곱하여 만족도를 보정한다. 가중치는 기간별 전체 리뷰 대비 해당주제의 리뷰 값으로, 동일 기간 내에 많은 리뷰가 작성된 주제에는 높은 가중치를, 적은 주제에는 낮은 가중치를 할당하여 특정 기간 내에 이슈가 된 주제를 보정할 수 있다. 시간적 변화량은 해당 주제에 대한 시간적 변동성으로, 3년 간격으로 의견 수를 계산하고 선형회귀 분

석을 통해 회귀식을 작성한다. 이후 회귀식의 기울기를 시간적 변화량으로 선정하였다. 이 때, 시간적 변화량이 매우 작은 값을 가질 경우, 최근 해당 니즈에 대한 고객 의견 수가 급격히 줄어든 것으로 판단하여, 기술기회 탐색의 대상에서 제외하였다.

시간적 변화량이 유의미한 주제들의 시간적 변화량의 평균 값과 만족도의 0을 기준으로 사용자의 니즈를 <Figure 3>과 같이 4가지 형태로 구분한다. 각 유형별 니즈의 의미는 다음과 같다.

- Long-run : 사용자가 꾸준히 긍정적으로 인지하고 있는 니즈
- Innovative : 사용자들이 긍정적으로 체감하며 의견이 급속도로 증가하고 있는 니즈
- Troublesome : 지속적인 불만이 표출되고 있는 니즈
- Inexperienced : 최근 불만이 급속도로 증가되고 있는 니즈

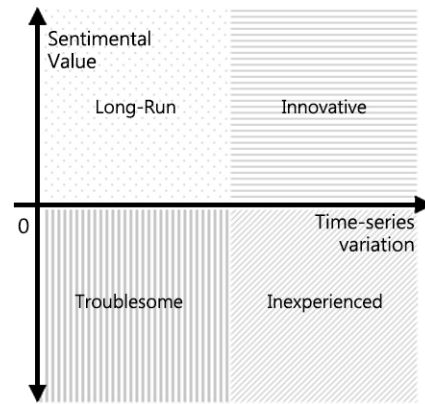


Figure 3. User Needs Type

(2) 동종영역 기술 분석

동종영역의 기술을 분석하기 위하여 특허를 활용한다. 먼저 관심대상인 기술 분야에 해당하는 키워드를 이용하여 검색식을 작성하고, 동종영역 산업과 관련된 키워드를 포함하고 있는 특허를 동종영역 특허로, 포함하고 있지 않은 특허를 이종영역 특허로 정의한다.

먼저, 동종영역의 기술 분야를 정의하기 위하여 특허의 서지정보 중 기술의 핵심내용이 서술되어 있는 전체 청구항의 텍스트를 이용한다. LDA 모델을 이용하여 주제와 키워드 리스트를 얻을 수 있다. 도출된 키워드 리스트를 바탕으로 사전에 정의된 part-attribute 사전을 활용하여 주제를 기술 분야로 정의한다. 기술 분야에 특허를 할당하기 위하여 개별 특허 청구항에 서술된 키워드가 해당 기술 분야의 키워드 리스트에 출현하는 비율을 바탕으로 기술 분야에 특허를 할당한다. 이 때, 특허가 복수의 주제에 할당 될 수 있으므로, 출현비율이 가장 높은 분야를 선택한다.

사용자의 니즈와 기술 분야를 연결하기 위하여 코사인 유사도를 이용한다. sLDA를 통해 얻은 니즈 별 키워드 리스트와 LDA를 통해 얻은 기술 분야의 키워드 리스트 간 코사인유사도

를 분석한다. 이 때, 기술 분야 키워드 리스트는 기술적으로 전문적인 용어가 사용되어 일반적으로 사용되는 단어로 이루어진 니즈 별 리스트와의 유사도를 정확하게 분석하지 못한다. 이를 해결하기 위해 청구항에 서술된 키워드가 아닌, 발명의 필요성과 배경에 대하여 서술한 배경 항목을 이용한다. 또한 키워드 리스트에 일반적으로 사용되는 단어가 포함되어 있을 수 있으므로, TF-IDF(Term Frequency-Inverse Document Frequency) 값을 이용하여 단어의 중요도를 평가하여 가중치로 사용한다.

이후 니즈와 연결된 기술이 사용자의 요구사항을 만족시킬 수 있는 지에 대한 평가를 위해 특정한 의미를 가지는 특허의 구조화된 정보를 활용한다. 등록된 특허의 수는 특정 분야의 특허활동 정도를 의미한다. 피인용 수는 특허를 다른 특허가 인용한 수로, 각 특허의 영향력을 측정하는 데 활용될 수 있다. 패밀리국가 수는 특허가 법적 권리를 인정받을 수 있는 국가의 수로, 기술시장에서의 영향력을 의미한다. 독립 청구항 수는 각 특허가 보호받아야 하는 권리의 수를 의미하며, 후방인용 수는 특허가 해당 특허를 출원하는 과정에서 인용한 특허의 수로, 적을수록 다른 특허의 영향을 적게 받은 특허라고 할 수 있다.

위와 같은 정보의 의미를 활용하여, 각 기술 분야의 기술력을 활동성, 영향력, 시장성, 권리성, 신규성의 다섯 가지 지표로 계산한다. 하나의 기술 분야에는 많은 특허가 포함되어 있으므로, 피인용 수, 패밀리 국가 수, 독립 청구항 수, 후방인용 수는 각 기술 분야의 평균값을 이용하고 등록된 특허의 수는 각 기술 분야에 포함된 특허의 수를 이용한다. 이때, 단위나 크기가 다른 각 지표를 동일한 수준에서 계산하기 위해 max-min 값을 이용한 정규화를 진행하며, 최종적인 기술 분야의 기술력은 각 지표의 평균값으로 계산한다. 각 지표를 구하는 수식은 다음과 같다.

$$\text{활동성} = \frac{\text{등록건수} - \min(\text{등록건수})}{\max(\text{등록건수}) - \min(\text{등록건수})} \quad (1)$$

$$\text{영향력} = \frac{\text{평균 피인용수} - \min(\text{평균 피인용수})}{\max(\text{평균 피인용수}) - \min(\text{평균 피인용수})} \quad (2)$$

$$\text{시장성} = \frac{\text{평균 패밀리국가수} - \min(\text{평균 패밀리국가수})}{\max(\text{평균 패밀리국가수}) - \min(\text{평균 패밀리국가수})} \quad (3)$$

$$\text{권리성} = \frac{\text{평균 독립청구항 수} - \min(\text{평균 독립청구항 수})}{\max(\text{평균 독립청구항 수}) - \min(\text{평균 독립청구항 수})} \quad (4)$$

$$\text{신규성} = 1 - \frac{\text{평균 후방인용 수} - \min(\text{평균 후방인용 수})}{\max(\text{평균 후방인용 수}) - \min(\text{평균 후방인용 수})} \quad (5)$$

(3) 이종영역 기술기회 탐색

이종기술 영역에서의 기술기회를 탐색하기 위해 기술기회의 방향성을 정의하고 각 기술기회의 우선순위를 도출한다. 먼저, 동종 산업 기술의 기술력이 높고 만족도가 높을 경우, 이를 이종 산업에서 활용하여 높은 고객 만족도를 기대 할 수 있으므로, 외향형 기술기회 탐색이 가능하다. 반대로 동종 산업 기술의 기술력과 만족도가 낮은 경우, 이종 산업의 높은 기술

력을 갖는 기술을 동종 산업에 도입하여 고객 만족도를 높일 수 있으므로, 내향형 기술기회 탐색이 가능하다. 방향별로 다양한 기술기회가 존재하므로, 각 기술기회의 우선순위를 도출하기 위하여 만족도와 기술력을 바탕으로 한 판별식을 활용한다. 각 기술기회별 우선순위 판별식은 다음과 같다.

$$\text{외향형 기술기회 우선순위 점수} = \text{사용자 만족도} \times \text{기술 역량} \quad (6)$$

$$\text{내향형 기술기회 우선순위 점수} = |\text{사용자 만족도}| \times \frac{1}{\text{기술 역량}} \quad (7)$$

니즈별로 매칭이 되는 이종영역의 기술을 파악하기 위하여 동종 산업 키워드를 포함하고 있지 않은 특허를 대상으로 동종 산업 특허 분석 프로세스와 동일한 과정을 거쳐 기술 분야 키워드 리스트를 얻은 후 기술 분야 별로 특허를 할당한다. 이후 니즈별 키워드와 각 기술 분야에서 특허들의 배경 서지정보와 각 단어의 TF-IDF 가중치를 이용해서 관련된 니즈-이종 기술 분야를 연결한다.

선정된 이종영역의 기술기회가 채택되었을 때의 사용자 만족도 변화량을 확인하기 위해 선형회귀분석을 사용한다. 본 연구에서 사용된 기술지표만으로 만족도의 변화량을 예측하기 위하여, 변수의 영향력을 직접적으로 판단할 수 있는 선형회귀 분석을 사용하였다. 앞선 분석결과를 통해 유효한 니즈-동종 기술 간 연결 관계 및 니즈 만족도-기술력의 조합을 얻을 수 있다. 이를 활용하여 선형 회귀식을 도출하고, 이종영역과 동종영역의 기술력의 차이를 회귀식에 대입하여 만족도의 변화량을 예측한다. 마지막으로 각 유형별로 가장 높은 만족도 변화량을 갖는 기술기회를 선정하여 각 니즈, 동종-이종 기술의 키워드를 기반으로 기술기회를 해석한다.

4. 실증분석

4.1 분석대상 선정

본 연구에서 분석하고자 하는 자동차 산업은 2014년 KISTEP에서 ‘무인 자동차 및 수송 장치’로 10대 미래 유망기술에 선정되었다. 또한 세계경제포럼에서 10대 유망기술에서 KAIST의 ‘OLEV(On-line Electric Vehicle)가 선정되는 등 기술 유망성이 입증되었다. 자동차 산업은 사용자가 활발하게 리뷰를 작성하여 오피니언 마이닝의 분석 대상으로 활발하게 활용되고 있으며(Bodenorf and Kaiser, 2010; Kessler et al., 2010), 다양한 유망기술이 활용되고 있는 분야이다. 즉, 자동차 산업은 고객의 피드백정보가 풍부하고 기술의 유망성이 높아, 고객 니즈와 기술의 통합적 사고를 적용하기에 적합한 분야로 판단된다.

자동차를 구성하고 있는 많은 기술 중 사용자가 가장 많이 접하는 도어는 엔진, 기어 등 기계적 요소가 다수 포함된 다른 부위에 비하여 전문가가 아닌 일반적인 사용자들이 즉각적이고 다양한 의견을 표출 할 수 있으며, 도어에 관련된 요구사항

을 해결하기 위한 기술간 모순(trade off)이 존재하는 기술간 모순이 적게 발생하는 부위이다. 또한 도어를 구성하는 힌지, 핸들, 자동 잠금장치 등 다양한 기술이 자동차 산업에 국한되어 사용되지 않고 다양한 산업의 기술에 함께 사용되거나 다른 산업의 기술을 활용하고 있어 이종산업 기술기회 탐색에 적합한 대상으로 판단된다. 따라서 본 연구에서는 자동차 산업의 도어 관련기술을 분석대상으로 선정하여 이종기술기회를 탐색하였다.

4.2 사용자 니즈 분석

사용자의 니즈를 분석하기 위하여 온라인 리뷰기반으로 오피니언 마이닝을 실시한다. 이 때, 다양한 리뷰 사이트 중에 자동차 시장조사 기관인 JD Power의 평가 지표와 url 주소 및 클래스의 구조 등 웹 크롤링에 적합한 사이트를 선정하였다. 선정된 사이트는 Cars.com와 Autobytel.com이며, 각각 JD Power 평가 2위와 3위에 랭크 되어있어, 리뷰의 신뢰도가 높아 사용자의 니즈를 도출하기에 적합한 사이트로 판단되었다. 두 사이트에서 2002년부터 2012년 사이에 작성된 총 19,800건의 리뷰를 수집하였고, 도어관련 키워드를 포함하고 있는 리뷰는 2,605건이었다. 리뷰에 세부적인 기술적 요소를 포함하지 않고 단순히 좋고 나쁨을 표현한 리뷰들을 제외하기 위해 5개미만의 단어로 표현된 리뷰를 제외하여 2,594건의 분석대상 리

뷰를 선정하였다.

도어 관련 기술의 세부적 내용을 파악하기 위하여 리뷰에서 출현하는 숫자, 관사, 전치사 등 오피니언 마이닝에 불필요한 단어를 제외한 1,000개의 단어를 사용하여 part-attribute 사전을 정의하였다. 이후 다양한 사용자의 니즈를 주제별로 군집화하기 위하여 각 리뷰에서 불필요한 단어를 제외한 뒤 sLDA 분석을 통해 20개의 주제에 따른 키워드 리스트를 도출하였다. 이후 사전에 정의된 part-attribute 사전을 이용하여 <Table 1>과 같이 사용자 니즈를 기술적 관점에서 정리하였다. 니즈 1은 도어의 핸들(handle)에 관련된 것으로, 도어를 닫을 때(close, shut)의 단단함, 튼튼함(solid)에 대한 니즈, 니즈 3은 창틀(window, panel)의 소음(noise, rattle, wind)에 대한 니즈로 해석할 수 있다. part-attribute 사전으로 니즈를 특정할 수 없는 니즈 2, 4, 11은 분석대상에서 제외하였다.

구조화된 니즈를 바탕으로 각 니즈에 대한 시간별 변화량과 만족도를 계산하였다. 3년 단위의 구간으로 니즈 수를 계산하였으며, 선형회귀 분석을 통한 회귀 계수를 시간적 변화량으로 도출하였다. 이후, 시간적 변화량의 평균값에서 음의 방향으로 1.5×표준편차 이상 차이가 나는 니즈 3, 7은 고객의 니즈가 급격하게 감소하고 있는 니즈로 판단하여, 분석대상에서 제외하였다.

시간적 변화량이 유의미한 니즈들의 시간적 변화량 평균값과 만족도 0을 기준으로 니즈의 유형을 도출하였다. 각 유형의 니즈들의 세부적인 내용은 <Table 2>와 같다. 이후 만족도를 기준으로 내향형과 외향형으로 기술기회 탐색의 방향을 구분하였다.

Table 1. Part -Attribute Structure

Needs	Part 1	Part 2	Part 3	Attribute 1	Attribute 2	Attribute 3
1	Handle			Close	Sound	Solid
2	Handle	Package		Drive	Sport	Open
3	Window	Panel		Noise	Rattle	Wind
4				Drive	Road	Wife
5	Lock	Handle		Drive	Bad	Low
6	Power	Window		Open	Great	Kid
7	Holder	Rest	Storage	Put	Control	small
8				Open	Park	Dinge
9	Power	Hatch		Park	Paint	Road
10	Lock	Power		Time	Work	Time
11	Room			Dinge	Ride	Space
12	Handle			Protect	Open	Small
13	Panel	Dash	Leather	Plastic	Cheap	Clothe
14	Handle	Room		Great	Outdoor	Style
15	Panel	Open	Trim	Open	rust	Stuck
16	Power			Great	Drive	Size
17	Cargo	Switch	Handle	Problem	Light	Open
18	Handle	Switch	Seal	Light	Open	Style
19	Power	Lock	Window	Control	Automate	Auto
20	lock	key	Button	Unlock	Automate	Auto

Table 2. Needs type and contents

Needs type	Needs No.	Contents	
Outbound*	Innova-tive	6	Convenience of power window system, Child protection
		16	Convenience of power door closing
		17	Lightness of opening cargo
		19	Excellent controlling about power lock auto closing
		20	Usability of auto key/button lock system
Inbound**	Long-run	1	Solid door/handle while closing
		14	Great outdoor style about door/handle
	Trouble-some	8	Crashing door on parking
		9	Error about power hatch in driving/parking
		10	Solving malfunction error about power lock
	Inexper-enced	5	Solving malfunction error about handle lock
		12	Protecting while handle opening/shutting
		15	Rusting problem about panel/trim
		18	Handle sill style

*Outbound technology opportunity, **Inbound technology opportunity.

4.3 동종영역 기술 분석

기술기회를 탐색하고자 하는 동종 영역의 기술을 분석하기 위해 특허 데이터베이스를 제공하는 사이트인 위즈도메인 (Wisdomain.Inc)에서 미국특허청(USPTO; United States Patents and Trademark Office)에 등록된 특허의 서지정보를 수집하여 분석하였다. 등록된 특허들 중 문에 관련된 특허임을 알 수 있는 키워드인 door, doors와 자동차에 관련된 특허로 특정 지을 수 있는 키워드인 automotive, automotives, vehicles, car, cars, automobile, automobiles를 포함하고 10년 이내에 등록된 특허를 사용하기 위하여 검색식 (8)을 이용하여 778개의 특허를 수집하였다.

$$TI = (\text{door or doors}) \text{ and } TAC = (\text{automotive or automotives or vehicles or car or cars or automobile or automobiles}) \text{ and } AD > 2005 \quad (8)$$

동종영역 특허로 수집된 778개의 청구항과 LDA 모형을 이용하여 군집화하였다. 군집화를 반복하여 최적의 군집을 형성할 수 있는 20개의 군집으로 기술 분야를 도출하였다. 하나의 특허가 복수의 기술 분야에 속해 있을 경우, 키워드의 분포가 가장 높은 기술 분야에 특허를 할당하였다.

각 기술 분야는 도출된 키워드 리스트를 통해 정의하였으며, 1번 클러스터의 경우 힌지의 마운트 블라켓에 관련된 기술, 2번의 경우 모터 내 회전 지지를 위한 케이블 기술, 3번의 경우 패널 및 트림의 표면관련 기술 분야로 정의하는 등 20개의 분야 중 키워드로 특정 기술 분야를 정의 할 수 없는 5, 11, 12, 15, 18번 기술 분야는 분석대상에서 제외하고 각 기술 분야를 정의하였다.

15개의 기술 분야와 제 4.2절에서 도출된 사용자의 니즈의 관계를 파악하기 위하여 사용자 니즈의 키워드 리스트와 기술 분야에 포함되는 특허들의 배경에서 나타나는 키워드간의 유사도를 TF-IDF 가중치를 적용한 코사인 유사도를 통해 계산하였다. 0.275의 기준치를 이용하여 코사인 유사도가 기준치 이상 일 경우 기술 분야가 니즈와 관련되어 있다고 판단하였다. 니즈 17과 니즈 10의 경우 연관된 기술 분야가 없는 것으로 분석되었다.

4.4 이중영역 기술기회 탐색

각 니즈와 관련된 기술력을 바탕으로 기술기회 탐색 방향을 설정하였다. 외향형 기술기회의 경우 사용자의 만족도가 높고 기술역량이 높을수록, 내향형 기술 기회의 경우 사용자의 만족도와 기술역량이 낮을수록 우선순위가 높다고 할 수 있다. 이를 바탕으로 식 (6), 식 (7)을 활용하여 <Table 3>의 우선순위를 도출하였다.

외향형 기술기회의 경우 Innovative 유형의 니즈가 가장 우선순위가 높은 것으로 나타났으며, 내향형 기술기회의 경우 Inexperienced 유형의 니즈가 우선순위가 가장 높은 것으로 나

Table 3. Priority of Technology Opportunities

Needs Type	Needs No.	Homo-tech* fields	Senti-mental Value	Techno-logy capability	Priority score	Rank	
Outbound	Innova-tive	6	17	0.5801	0.4044	0.2346	1
		16	20	0.5935	0.3873	0.2299	3
		17	-				
		19	6	0.2071	0.3539	0.0733	5
		20	13	0.0446	0.4372	0.0195	6
	Long-run	1	8	0.2464	0.4072	0.1003	4
Inbound	Trouble-some	8	4	-0.0254	0.1289	0.1969	5
		9	17	-0.0261	0.3189	0.8177	6
		10	-				
	Inexpe-rienced	5	20	-0.5087	0.3262	1.5593	2
		12	6	-0.2070	0.3890	0.5322	4
		15	19	-0.5304	0.1286	4.1228	1
		18	8	-0.3670	0.2495	1.4708	3

*Homogeneous technology.

타났다. 니즈와 관련되는 기술이 없는 경우, 우선순위에서 제외 하였으나 이중기술 중 관련된 기술이 존재 할 경우, 새로운 기술기회가 될 수 있으므로 지속적으로 분석에 활용하였다.

이후 이중영역에서의 기술기회를 탐색하기 위해, 자동차 산업의 도어가 아닌 다른 산업의 도어 특허를 수집하였다. 검색식 (9)를 이용하여 전체 도어에 관련된 특허를 수집하였으며, 여기에서 제 4.2절에서 수집된 778개를 제외하여 1,507개의 특허를 수집하였다.

$$TI = (\text{door or doors}) \text{ and } AD > 2005 \quad (9)$$

1,507개의 이중영역 특허의 청구항과 LDA 모형을 이용하여 27개의 기술 분야를 도출 하였으며, 각 기술 분야의 키워드를 이용하여 기술 1의 경우 신호 감지 및 통째에 관련한 기술, 기술2는 톨박스에 관련한 기술 등 키워드로 특정 기술 분야를 정의 할 수 없는 4, 6, 8, 9, 14, 18, 21, 22, 27번 기술 분야는 분석대상에서 제외하고 각 기술 분야를 정의하였다. 이후 제 4.3절의 프로세스를 이용하여 이중기술 분야와 사용자 니즈를 연결하였다.

이중기술 도입에 따른 사용자의 만족도 변화량을 예측하기 위하여 니즈의 만족도와 관련된 동종 기술의 기술력을 이용하여 다중선형 회귀식 (10)을 도출하였다. 회귀식을 도출하기 위하여 통계분석 프로그램 SPSS21을 사용하였으며, 신뢰구간은 90%로 stepwise 변수선택 방법을 채택하였다.

특허의 개수, 평균 패밀리국가 수, 피인용 수, 후방인용 수, 독립 청구항 수 다섯 가지의 모든 지표를 사용하여 선형 회귀식을 도출하고자 하였으나, 계수의 유의확률이 낮은 3개 변수를 제외하고, 평균패밀리 국가수와 tan(특허 수) 만이 회귀식에 사용되었다. 회귀식의 세부지표는 <Table 4>와 같다.

Table 4. Linear Regression Index

Model summary					
R	R square	Adjusted R Square	Standard Estimation Error		
.615	.379	.329	.2928690404		
Variance Analysis					
	Sum of suares	Degree of freedom	Mean square	F	level of significance
Regression Analysis	1.307	2	.654	7.619	.003
Residual	2.144	25	.086		
Sum	3.451	27			
Coefficient Analysis					
	Nonstandard coefficient		Standard coefficient	t	level of significance
	B	Standard error	Beta		
constant	-.185	.103		-1.794	.085
x_1^*	.084	.026	.516	3.209	.004
x_2^{**}	.068	.024	.453	2.819	.009

*Number of family countries, **Tangent(Number of patents).

Table 5. Predicted Sentimental Value Variation from Technology Opportunities

Needs type		Needs No.	Homo-tech fields	Homo-tech capability	Hetero-tech* fields	Hetero-tech capability	Predicted Sentimental Value Variation
Outbound	Innovative	6	17	0.4044	-		0.8807
		16	20	0.3873	19	0.3098	-0.1240
		19	6	0.3539	-		-0.0142
		20	13	0.4372	-		-0.0913
	Long-run	1	8	0.4072	12	0.2458	0.2188
14		6	0.3864	16	0.2484	0.3769	
Inbound	Troublesome	10	-		1	0.5164	0.1641
	Inexperienced	15	19	0.1286	16	0.1933	0.1320
		18	8	0.2495	19	0.3512	0.3218

*Heterogeneous technology.

기술기회의 방향성에 따라 니즈와 관련된 이종·동종 기술이 없을 경우, 만족도를 예측할 수 없으므로, 도출된 회귀식 (10) 과 동종-이종간 기술력차이를 이용하여 만족도의 변화량을 예측하였다. 회귀식에 각 이종-동종간 기술력의 차이를 대입하여 니즈 만족도의 변화량을 계산하였다. 이때 변화량이 음의 값을 가질 경우, 현재의 사용자 만족도보다 낮아지는 경우로, 기술기회에서 제외하였다. 외향형 기술기회의 경우, 동종영역의 기술이 이종기술에 적용되었을 때 동일한 니즈의 만족도 변화를 계산할 수 있으며, 내향형 기술기회의 경우, 이종영역의 기술이 동종영역에 적용되었을 때 동일한 니즈의 만족도 변화를 계산할 수 있다. 만약 이종영역(혹은 동종영역)에 니즈를 만족하는 기존기술이 없을 경우에도 기존기술의 만족도를 0으로 계산하여 만족도 변화량을 예측할 수 있다. 각 기술유형별로 니즈변화량을 예측하여 <Table 5>와 같이 예측하였다.

외향형 기술기회의 경우 기술기회 우선점수가 가장 높은 6번 니즈의 만족도 증가량이 가장 큰 것으로 나타났으며, 내향형 기술기회의 18번 니즈가 가장 높은 만족도 증가량을 나타내고 있다.

다중선형 회귀식 (10)

$$y = 0.084x_1 + 0.068x_2 - 0.185$$

y : 사용자의 만족도,

x_1 = 평균 패밀리국가 수, x_2 = tan(특허 수)

4.5 기술기회 해석

다중선형 회귀식과 기술기회별 우선순위 판별식을 이용하여 각 니즈 유형별 기술기회를 탐색할 니즈와 기술 분야를 도출하였다. 이후 키워드를 통한 기술 분야 선정 및 기술기회 탐색을 위하여 국내 자동차 무빙 시스템 부분에서 5년 이상 종사한 두 명의 전문가에게 자문을 구하여 키워드를 해석하였다. 니즈, 동종 영역 기술, 이종 영역 기술의 LDA 결과로 선정된 키워드 중 가장 니즈 혹은 기술 분야로 특정할 수 있는 키워드를 선택하여 <Table 6>과 같이 해석하였다. 이후 <Table 6>을 바탕으로 선정된 각 니즈, 이종·동종 기술의 주제를 통합하여 해석하였다.

Table 6. Explored Heterogeneous Technology Opportunities

Needs Type		Contents	Homogeneous technology	Heterogeneous technology
Outbound	Innovative	Convenience of power window system, Child protection	Connecting module, Movement through magnetic power	
	Long-run	Outdoor type Vehicle door handle	Recognize wireless/automatic lock signal	Recognize wireless signal
Inbound	Troublesome	Solving malfunction error about power lock		Signal perception and controlling through sensor, receiving switch and alarming, auto selection
	Inexperienced	Handle silling style	Attaching frame on braket mount and protecting	Surface on panel, hinge designing

먼저 외향형 기술기회 중 Innovative형 6번 니즈는 power, window에 해당하는 part에 open, great kid의 attribute를 갖고 있다. 이는 파워 윈도우 시스템 사용 시 편리함과 어린이 보호에 대한 만족감을 나타내고 있다고 해석할 수 있다. 이에 대응하는 기술 분야는, 동종의 17번 기술이며 모듈의 연결, 자기력에 관한 기술 분야이다. 니즈와 관련된 이종분야는 존재하지 않으므로, 이는 모듈 및 자기력을 이용한 도어기술이 자동차 이외의 영역에서도 활용 될 수 있다고 해석 할 수 있다.

Long-run형 14번 니즈는 handle, room에 해당하는 part에 great, outdoor, style의 속성을 갖고 있다. 이는 아웃도어 차량의 도어 손잡이에 만족감을 나타내고 있다고 해석 할 수 있다. 이에 대응하는 동종의 6번 니즈는 도어의 무선잠금장치의 인식률에 대한 기술 분야이며, 이종의 16번 니즈는 신호인식에 관련된 기술이다. 이를 해석하면, 자동차 손잡이에 사용되는 무선 신호 인식 장치를 이종산업에 활용하여 만족도 증가를 기대할 수 있다.

내향형 기술기회 중 troublesome 10번 니즈는 lock, power에 해당하는 part와 time, work의 attribute를 갖고 있다. 이는 자동차의 전원 혹은 잠금장치에 작동, 시간에 대한 불만족을 나타낸다고 해석 할 수 있다. 이와 관련된 동종 기술이 존재하지 않고, 이와 관련된 이종의 1번 기술은 신호를 감지하여 스위치를 통제 하거나 알람을 발생시키는 기술로, 이를 활용하여 전원 혹은 잠금장치의 불만을 해결할 수 있다고 해석 할 수 있다.

Inexperienced형 18번 니즈는 handle, switch, seal에 해당하는 part와 light, open, style의 attribute를 갖고 있다. 이는 핸들과 스위치, 마감 처리 부분에 대한 스타일과 작동에 대한 불만으로 해석 할 수 있다. 관련된 동종의 8번 기술은 브라켓 마운트 프레임 부착 및 보호하는 기술로, Long-run형 1번 니즈를 만족시키고 있지만, 18번 니즈는 만족 시키지 못하고 있다. 따라서 관련된 이종의 패널 표면과 힌지에 관련된 19번 기술을 활용하여 만족도 상승을 기대할 수 있다.

5. 결론

본 연구는 온라인상에 표출된 고객들의 요구사항을 바탕으로 LDA 모형을 통해 사용자의 니즈를 군집화하고 주제별 감성

극성 값을 도출하였다. 이후 니즈와 관련된 특허를 특허 검색식을 기준으로 동종·이종영역을 구분하고, LDA 모형을 통해 각 기술 분야를 구분하였다. 사용자 니즈 유형별 내향·외향형 기술 기회를 탐색하고 기술력과 특허지표를 이용하여 만족도 변화량이 가장 높은 기술을 기술기회로 제시하였다.

본 연구는 특허문서만을 활용한 기술 기회 탐색에서 벗어나 고객들의 피드백 정보와 특허분석을 통한 시장과 기술을 모두 고려한 복합적인 기술기획 방법론을 제안했다는 점에서 학문적 성과를 기대 할 수 있다. 또한 텍스트 형태의 특허문서를 이용하여 기술의 내용적 관점에서 기술기회를 탐색하여 기존 IPC 코드, 인용정보만을 이용한 접근 방식에서 벗어나 이미 융합된 기술 분야가 아닌 잠재적으로 융합 가능한 이종영역과 동종영역의 기술기회를 탐색하였다. 또한 설문조사를 통한 기존 사용자 의견 조사 방법에 비해 시간, 비용적 효율성을 높일 수 있으며, 사용자의 의견에 기반 한 이종기술 기회를 탐색하여 기술개발 및 전략을 수립하는 과정에 활용될 수 있다.

본 연구는 몇 가지 한계점이 존재한다. 먼저 니즈와 기술 분야를 도출하기 위한 클러스터링 과정에서 LDA만을 활용하여 특색이 명확하지 않은 클러스터와 의미가 유사한 클러스터가 도출 되어 기술기회 탐색대상으로 적절치 않을 가능성이 있다. 또한 기술기회로 도출된 분야가 산업현장의 트렌드 혹은 진행 중인 기술개발 과제인지 등 산업현장에서의 피드백을 반영하지 못하였다. 또한 만족도 변화량을 예측하기 위해 구성된 다중선형회귀 식의 설명력이 0.3으로 낮게 나타났다. 이는 사용자 만족도를 예측하기 위한 나이, 성별 등 개인 및 사회적 특성을 배제하고 오직 기술력만을 고려했기 때문이라고 할 수 있다. 위와 같은 한계점을 극복하기 위하여 사용자 니즈, 기술 분야 도출 단계에서 전문가의 피드백과 리뷰 데이터 수집과정에서 내용뿐만이 아닌 다양한 요소를 함께 수집하여 분석에 활용할 필요가 있다.

참고문헌

An, J., Kim, K., Noh, H., and Lee, S. (2016), Identifying Converging Technologies in the ICT Industry : Analysis of Patents Published by Incumbents and Entrants, *Journal of Korean Institute of Industrial*

- Engineers*, **42**(3), 209-221.
- Bin, H., Lucheng, H., Feifei, W., and Hong, M. (2015), Assessment of the Connectivity between Technologies Based on Patent Co-classification, Beijing, China.
- Blei, D. M., Ng, A. Y., and Jordan M. I. (2003), Latent dirichlet allocation, *Journal of machine Learning research*, **3**, 993-1022.
- Choi, C., Kim, S., and Park, Y. (2007), A patent-based cross impact analysis for quantitative estimation of technological impact : The case of information and communication technology, *Technological Forecasting and Social Change*, **74**(8), 1296-1314.
- Ghazizadeh, M., McDonald, A. D., and Lee, J. D. (2014), Text mining to decipher free-response consumer complaints insights from the NHTSA vehicle owner's complaint database, *Human Factors : The Journal of the Human Factors and Ergonomics Society*, **56**(6), 1189-1203.
- Ghiassi, M., Skinner, J., and Zimbra, D. (2013), Twitter brand sentiment analysis : A hybrid system using n-gram analysis and dynamic artificial neural network, *Expert Systems with Applications*, **40**(16), 6266-6282.
- Griffiths, T. L. and Steyvers, M. (2004), Finding scientific topics, *Proceedings of the National academy of Sciences*, **101**(suppl 1), 5228-5235.
- Huang, L., Shang, L., Wang, K., Porter, A. L., and Zhang, Y. (2015), Identifying target for technology mergers and acquisitions using patent information and semantic analysis, *2015 Portland International Conference on Management of Engineering and Technology (PICMET)*.
- Jeong, E. S., Kim, Y. G., Lee, S. C., Kim, Y. T., and Cxhang, Y. B. (2014), Identifying Emerging Free Technologies by PCT Patent Analysis, *The Journal of the Korea Institute of Electronic Communication Sciences*, **9**(1), 111-122.
- Jin, S. A., Heo, G. E., Jeong, Y. K., and Song, M. (2013), Topic-Network based Topic Shift Detection on Twitter, *Journal of the Korean Society for Information Management*, **30**(1), 285-302.
- Jo, C. Y., Lee, Y. M., Sin, H. J., and Park, H. J. (2005), Validation of claim as patent analysis indicator, *Patent 21*, **64**, 14-21.
- Jeong, B. and Lee, H. (2016), Research Topics in Industrial Engineering 2001~2015, *Journal of Korean Institute of Industrial Engineers*, **42**(6), 421-431.
- Kim, M., Park, Y., and Yoon, J. (2016), Generating patent development maps for technology monitoring using semantic patent-topic analysis, *Computers and Industrial Engineering*, **98**, 289-299.
- Kim, J. and Lee, S. (2016), Issues for Technology-Humanities Convergence : Patent Keyword Analysis, *Journal of Korean Institute of Industrial Engineers*, **42**(3), 231-240.
- KISTI (2012), Design and construction of a knowledge base for technology opportunity discovery, Daejun.
- Ko, N., Yoon, J., and Seo, W. (2014), Analyzing interdisciplinarity of technology fusion using knowledge flows of patents, *Expert Systems with Applications*, **41**(4), 1955-1963.
- Kwon, J.-D. (2012), An actual proof study on selecting key patents based on patent index, Hansung University, Seoul, Korea.
- Lee, C., Kang, B., and Shin, J. (2015), Novelty-focused patent mapping for technology opportunity analysis, *Technological Forecasting and Social Change*, **90**(B), 355-365.
- Lee, H.-J. (2005), Analysis of science and technology information necessary for research and development (R&D), *Patent 21*, **60**, 33-39.
- Lee, T. Y. (2009), Automatically learning user needs from online reviews for new product design, *AMCIS 2009 Proceedings*, 22.
- Lee, T. Y., Li, S., and Wei, R. (2008), Needs-centric searching and ranking based on customer reviews, *2008 10th IEEE Conference on E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing*.
- Lee, W. S. and Sohn, S. Y. (2015), Topic model analysis of research trend on spatial big data, *Journal of Korean Institute of Industrial Engineers*, **41**(1), 64-73.
- Lim, C., Yun, D., Park, I., Park, G., Koh, S., and Yoon, B. (2015), Exploring Prospective Research Areas in UI/UX through the Analysis of Patents, *Korean Management Science Review*, **32**(4), 1-18.
- Liu, B. (2012), Sentiment analysis and opinion mining, *Synthesis Lectures on Human Language Technologies*, **5**(1), 1-167.
- Liu, Y., Huang, X., An, A., and Yu, X. (2008), Modeling and predicting the helpfulness of online reviews, *2008 Eighth IEEE International Conference on Data Mining*.
- Liu, Y., Jin, J., Ji, P., Harding, J. A., and Fung, R. Y. (2013), Identifying helpful online reviews : a product designer's perspective, *Computer-Aided Design*, **45**(2), 180-194.
- Mcauliffe, J. D. and Blei, D. M. (2008), Supervised topic models, *Advances in Neural Information Processing Systems*, 121-128.
- Min, H.-J. and Park, J. C. (2012), Identifying helpful reviews based on customer's mentions about experiences, *Expert Systems with Applications*, **39**(15), 11830-11838.
- No, H. J., An, Y., and Park, Y. (2015), A structured approach to explore knowledge flows through technology-based business methods by integrating patent citation analysis and text mining, *Technological Forecasting and Social Change*, **97**, 181-192.
- No, H. J. and Park, Y. (2010), Trajectory patterns of technology fusion: Trend analysis and taxonomical grouping in nanobiotechnology, *Technological Forecasting and Social Change*, **77**(1), 63-75.
- Paradigms, T. and Trajectories, T. (1982), A Suggested Interpretation of the Determinants and Directions of Technical Change, *Research Policy*, **11**, 147-162.
- Park, J. and Song, M. (2013), A study on the Research Trends in Library and Information Science in Korea using Topic Modeling, *Journal of the Korean Society for Information Management*, **30**(1), 7-32.
- Penalver-Martinez, I., Garcia-Sanchez, F., Rodriguez-Garcia, M. A., Moreano, V., Fraga, A., and Sanchez-Cervantes, J. L. (2014), Feature-based opinion mining through ontologies, *Expert Systems with Applications*, **41**(13), 5995-6008.
- Singh, V. K., Piryani, R., Udding, A., and Waila, P. (2013), Sentiment analysis of movie reviews : A new feature-based heuristic for aspect-level sentiment classification, *Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*.
- Thorleuchter, D., Poel, D. V., and Prinzie, A. (2012), Analyzing existing customers' websites to improve the customer acquisition process as well as the profitability prediction in B-to-B marketing, *Expert systems with applications*, **39**(3), 2597-2605.
- Wu, M., Wang, L., Li, M., and Long, H. (2014), An approach of product usability evaluation based on Web mining in feature fatigue analysis, *Computers and Industrial Engineering*, **75**, 230-238.
- Yun, J.-H. (2011), Patent Information Analysis : Tools for Systematic R&D Planning, *Industrial Engineering Magazine*, **18**(3), 23-28.