



데이터 마이닝을 통한 기술경영 전략 수립에 관한 연구

Establishment of Strategy for Management of Technology Using Data Mining Technique

이준석* · 이준혁* · 김갑조* · 박상성** · 장동식*†

Junseok Lee*, Joonhyuck Lee*, Gabjo Kim*, Sangsung Park**, and Dongsik Jang*†

*고려대학교 산업경영공학과

†Department of Industrial Management Engineering, Korea University

**고려대학교 기술경영전문대학원

‡Department of Intellectual Property, Korea University

요약

기술예측은 현재까지 관측된 특정기술에 대한 데이터를 바탕으로 미래에 그 기술이 어떠한 상태가 될 지를 알아보는 것으로서 기술경영 전략 수립 시 유용하게 사용된다. 현재는 전문가 의견을 바탕으로 한 분석법을 이용하여 기술예측을 실시하고, 국가, 기업 그리고 연구자는 이를 근거로 연구개발의 방향 및 전략을 수립한다. 전문가의 의견을 바탕으로 하는 정성적 기술예측은 전문가마다 다른 결과를 예상할 수 있고, 여러 전문가의 의견을 수집하여야 하므로 많은 시간과 비용을 필요로 한다. 이러한 문제점을 극복하고 예측에 대한 객관성을 확보하여 기업의 연구개발 의사결정을 돕기 위해 정량적 예측법을 바탕으로 한 기술예측 방법이 연구되고 있다. 본 논문에서는 정량적 분석법에 기반 한 기술예측 방법론에 대한 연구를 제안한다. 제안된 방법은 데이터 수집, 주성분 분석, 그리고 데이터마이닝 기법 중 하나인 로지스틱 회귀분석을 이용한 예측 단계로 구성되어 있다. 본 연구에서는 무인자동차에 관련된 특허 문서를 이용하여 데이터를 수집 및 추출하고, 특허문서의 텍스트를 마이닝하여 분석이 가능한 형태로 구축한다. 주성분분석 후 추출된 주성분 점수를 이용하여 로지스틱 회귀분석을 실시하며 이를 바탕으로 개발현황 분석 및 기술예측을 시행한다.

키워드 : 기술예측, 특허, 로지스틱 회귀분석, 무인자동차, 주성분 분석

Abstract

Technology forecasting is about understanding a status of a specific technology in the future, based on the current data of the technology. It is useful when planning technology management strategies. These days, it is common for countries, companies, and researchers to establish R&D directions and strategies by utilizing experts' opinions. However, this qualitative method of technology forecasting is costly and time consuming since it requires to collect a variety of opinions and analysis from many experts. In order to deal with these limitations, quantitative method of technology forecasting is being studied to secure objective forecast result and help R&D decision making process. This paper suggests a methodology of technology forecasting based on quantitative analysis. The methodology consists of data collection, principal component analysis, and technology forecasting by logistic regression, which is one of the data mining techniques. In this research, patent documents related to autonomous vehicle are collected. Then, the texts from patent documents are extracted by text mining technique to construct an appropriate form for analysis. After principal component analysis, logistic regression is performed by using principal component score. On the basis of this result, it is possible to analyze R&D development situation and technology forecasting.

Key Words : Technology Forecasting, Patent, Logistic Regression, Autonomous vehicle, Principal component analysis

Received: Sep. 14, 2014
Revised : Sep. 28, 2014
Accepted: Jan. 20, 2015
†Corresponding author(jang@korea.ac.kr)

본 논문은 BK21 플러스사업(고려대학교 제조·물류 분야에서의 빅데이터 운용 사업팀)으로 지원된 연구임. 본 논문은 2012년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (한국연구재단-NRF-2010-0024163)

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

경제가 고도화되고 기업들의 경쟁이 가속화됨에 따라 기존의 경제 환경에서는 경쟁력 확보가 어렵게 되었다. 따라서 기업들은 경쟁 기업과의 차별화와 경쟁력우위를 확보하기 위해 보다 창의적이고 혁신적인 제품을 개발하고자 노력하고 있다. 이와 동시에 많은 기업들은 개발된 제품의 아이디어와 기술을 보호받고자 지식재산권의 중요성을 인식하고 특허를 출원 및 등록함으로써 특허권을 확보하고 있다.

특허는 기술에 대한 권리를 보장하는 배타적 지배권을 말한다[1]. 발명을 장려하고 권리를 보호하기 위해 만든 특허권은 기업의 경쟁력이자 자산으로 평가받고 있다. 현재 기업의 자산으로써 특허를 활용하는 대표적인 기업으로는 쉘컴이 있으며, 최근에는 삼성전자, 애플, 구글 등도 특허권을

자산과 권리보호를 위해 적극적으로 활용하고 있다. 특허를 출원하기 위해서는 특허명세서가 제출해야하며, 출원인은 특허명세서의 내용을 바탕으로 권리를 보호받게 된다. 기업의 입장에서 특허는 자산과 기술에 대한 권리를 보호하는 역할로서 활용하고 있지만, 연구개발자의 입장에서 특허는 기술의 의미를 모두 담고 있는 기술 문서(Technology Description: TD)로 활용되고 있다.

현재까지는 기술예측을 전문가의 의견 수렴을 통하여 기술예측을 실시하였고, 이러한 정성적 방법에는 전문가 의견조사, 델파이(Delphi)기법, 시나리오 분석 등이 있었다. 이러한 방법들은 기술예측방법으로써 주류를 이루고 있지만 전문가의 개인적 의견을 바탕으로 한 예측법이기에 때문에 다분히 주관적이고 실험설계를 위해서 많은 시간과 노력이 필요하다. 하지만 많은 시간과 노력이 소요됨에도 불구하고 객관성을 확보하기 어렵기 때문에 정량적 분석기법의 필요성이 대두되었다.

최근에는 정성적 기술예측 방법의 한계점을 극복하기 위해 정량적 기술예측 방법에 대한 많은 연구가 시도되고 있다. 정량적 기술예측 방법은 현재의 기술동향 뿐만 아니라 예측하기 어려운 미래의 상황에 대한 예측도 가능하게 할 수 있다. 이러한 기술동향 파악과 기술예측은 경쟁사에 대한 정보획득의 신속성, 기술 분야의 핵심기술 도출, 타사와의 특허분쟁을 미연에 방지할 수 있는 회피설계, 혁신제품 개발 등으로 연결시킬 수 있을 것으로 예상된다. 많은 정량적 기술예측 방법이 연구되고 있으며 그 중 특허데이터를 활용한 여러 기술예측 방법이 시도되고 있다. 따라서 특허명세서를 충분히 분석한다면 현재의 연구동향을 파악하고 기술예측을 하는데 많은 근거를 얻을 수 있을 것이다.

본 논문에서는 정량적 기술예측 방법을 제안하기 위해 특허문서를 이용하여 분석을 실시한다. 분석방법으로는 대표적인 데이터 마이닝 방법 중 하나인 로지스틱 회귀분석을 이용하고 그 결과를 바탕으로 기술예측과 연구개발 방향을 제안한다.

2. 배경이론 및 선행연구

2.1 기존 연구

기술예측이란 어떤 유용한 기능에 이바지하는 예견 가능한 기술 혁신, 특정과학의 발전, 그리고 과학적 발견에 대한 가장 실현 가능한 시간을 지적하는 것을 말한다[2]. 현재까지의 기술예측은 전문가 의견조사, 델파이(Delphi) 기법, 시나리오 분석 등의 주관적인 방법을 사용하였다[2]. 이러한 방법들은 많은 인력을 필요로 하며 또한 많은 시간과 비용이 투입되어야 하는 한계점을 지니고 있다. 현재의 문제점을 극복하기 위해 여러 연구가 진행되고 있으며 그 중 특허분석을 이용한 기술예측 연구가 활발히 진행되고 있다.

Tseng, Lin(2007)은 특허분석을 위한 텍스트 마이닝 기법에 관하여 설명하였고, 이를 진행하기 위한 방법으로는 문서분할(Text segmentation), 요약추출(Summary extraction), 특징추출(Feature selection), 단어 연관성(Term association), 군집화(Cluster generation), 주제 식별(Topic identification) 마지막으로는 정보지도작성(Information mapping)을 설명하였다[3].

전성해, 박상성, 신영근, 장동식, 정호석(2009)은 특허분석 시스템 기술에 필요한 공백기술을 찾기 위해 특허분석 시스템 특허를 Raw-data로 이용하고 매트릭스 분석과 데이터 마이닝 기법인 자기조직화 지도를 사용하여 공백기술을 예측하였다[4].

전성해(2011)는 지능형 시스템의 정량적 기술예측을 위해서 지능

형 시스템 기술관련 특허를 바탕으로 SVD-PCA를 사용하여 차원 축소 특징추출을 하였고, 대표적인 군집화 방법 중 하나인 K-means 알고리즘을 사용하여 각 특허의 특징을 군집화 하였다. 군집된 결과는 특허분석을 위한 데이터로 이용하여 공백기술 예측을 시도하였다[5].

전성해(2013)에서는 사례분석으로써 효율적인 빅 데이터 학습 절차를 제안하고자 특허문서를 다중선형회귀분석을 이용하여 특허분석을 실시하였고, 그 결과를 바탕으로 기술예측을 실시하였다[6].

2.2 특허데이터

지식재산권은 인간이 만든 창작물 중 보호할 가치가 있는 것에 대하여 그 창작자에서 배타적 권리를 부여하는 것이다[1]. 이러한 지식재산권에는 특허, 실용신안, 의장, 상표 등과 같은 산업재산권, 문학, 예술과 같은 사상과 감정을 표현한 저작권, 컴퓨터프로그램보호권, 반도체 집적회로 배치 설계권과 같이 급변하는 사회 환경에서 새롭게 나타난 창작물인 신지식재산권으로 나뉘어진다. 이 중, 특허는 특허명세서라는 기술문서(TD)를 가지고 있다. 특허명세서에는 기본적인 서지사항과, 기술에 대한 요약서, 권리범위, 도면 등으로 구성되어 있다. 또, 특허는 출원 후 일정기간 이후 자동 공개가 되므로 관련 기술내용을 파악할 수 있고 R&D를 위해 자료를 비교적 손쉽게 수집하여 분석할 수 있는 장점을 가지고 있다. 이러한 이유로 많은 연구자들은 IP(Intellectual Property) R&D에서 특허를 적극적으로 활용하여 핵심기술, 부상기술, 공백기술 등을 추출할 수 있다. 추출된 기술들을 통하여 R&D 진행 시 제품설계 및 빠른 출시가 가능하도록 할 수 있다. 더욱이 많은 경쟁기업들과의 특허분쟁을 대비하여 회피설계를 가능하게 하고 확보된 권리를 통하여 경쟁기업들에게 라이선스를 주장함으로써 기업 경쟁력 확보 및 수익창출을 가능하다.

특허 데이터의 접근은 무료와 유료의 방법이 있다. 무료 데이터베이스의 경우는 한국특허정보원에서 운영하고 있는 KIPRIS가 있으며 유料的 경우 WIPSON, ACCLAIMIP 등이 있다.

2.3 무인자동차

미래자동차는 친환경자동차와 무인자동차가 개발될 것으로 예측되고 있다. 이러한 미래자동차를 실현시키기 위해 많은 자동차업체와 관련 업체들은 부단히 연구와 개발에 힘쓰고 있다. 전기자동차, 수소자동차 등은 대표적인 상용화된 친환경자동차이다.

무인자동차는 기존의 자동차에 운전자 대신 컴퓨터를 탑재하여 목적지까지 운행하는 자동차이다. 운전자가 담당하는 인지, 판단의 역할을 해야 하므로 자동차는 주변정보를 수집할 수 있는 하드웨어와 그것을 운용할 소프트웨어의 개발이 필수적이다. 국내외 여러 기업, 대학 및 연구소에서 무인자동차 상용화를 위한 연구개발을 진행하고 있으며, 특히 구글에서 많은 연구를 진행하고 있고 실제 시험주행도 진행하고 있는 상태이다[8].

무인자동차 역시 친환경자동차와 같이 상용화를 위해 관련 기술들을 지속적으로 연구개발하고 있다. 무인자동차 분야의 대표적인 상용화 기술로는 자동 주차시스템, 차선이탈 경보장치 등이 있으며 이러한 기술들은 운전자의 편의성을 극대화하고 있다. 구글을 비롯한 많은 언론보도 그리고 연구 보고서를 통하여 무인자동차의 완전 상용화가 임박함을 알 수 있다[9, 10].

2.4 로지스틱 회귀분석

로지스틱 회귀분석(Logistic Regression)은 대표적인 지도학습도

형 중의 하나로써, 선형회귀분석과 다르게 독립변수를 통해 범주형인 종속변수를 분류하는 기법이다. 일반적으로 범주형의 결과를 설명 또는 예측하기 위해 사용 한다[11, 12].

로지스틱 회귀분석 식은 아래 식(1)과 같이 나타낼 수 있으며 로지스틱 반응함수(Logistic Response Function)라 불리고 이 때, 종속 변수(Target variable) p 는 [0, 1]사이의 확률 값으로 표현된다.

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n)}} \quad (1)$$

또한 특정 클래스에 속할 측도는 Odds를 이용할 수 있으며, 특정 클래스($Y=0$)에 속할 Odds는 클래스 1에 속하는 확률에 대한 클래스 1에 속하는 확률의 비이며, 식(2)와 같이 나타낼 수 있다.

$$Odds = \frac{p}{1-p} \quad (2)$$

본 연구에서는 종속변수를 기술을 대표할 수 있는 단어로 설정하고, 독립변수에 따라서 종속변수가 출현할 수 있는지를 확인하여 유의한 단어 선정을 통해 기술예측을 실시한다.

3. 제안방법

본 연구에서는 로지스틱 회귀분석을 이용하여 유의한 키워드를 발견하고, 발견한 키워드를 바탕으로 기술예측 및 연구개발 전략을 제안한다.

특허명세서에는 유·무료의 특허검색 데이터베이스에서 수집할 수 있다. 다음은 본 연구에서 사용한 특허문서를 수집하기 위해 사용된 특허검색 식이다.

표 1. 작성된 검색 식
Table 1. Search formulae

Groups	Data Base	Search formulae
Google	U.S (WIPS)	((autonomous adj (vehicle or car or motor))) and google
Others		((autonomous adj (vehicle or car or motor))) and not google

수집된 특허명세서는 제목, 출원인, 발명인, 명세서에 대한 요약, 청구항 등 다수의 서지적 정보와 출원하고자 하는 기술에 대한 설명이 포함되어 있다. 그림1에서는 본 연구에서의 제안방법에 대한 Flow chart를 나타내었다.

특허명세서에는 문자, 그림, 숫자 등의 형태로 표현되어 있기 때문에 분석이 가능할 수 있도록 전처리과정이 필요하다. 본 연구에서는 텍스트 마이닝을 이용하여 요약에 포함되어 있는 텍스트들을 전처리하여 분석하였다. 텍스트 마이닝 과정을 통하여 각각의 특허 문서는 특허번호와 단어들로 이뤄진 매트릭스 형태로 저장된다. 하지만 이러한 형태의 매트릭스는 단어의 수(Column)가 문서의 수(Line)보다 월등히 많아 차원의 저주(Curse of dimensionality)가 발생하게 됨에 따라 분석이 불가능하게 된다. 이러한 문제점을 해결하고자 본 연구에서는 특징선택(Feature selection)의 방법으로써 키워드의 빈도수를 기준으로 분석에 사용할 키워드를 선정한다. 로지스틱 회귀분석은 특징추출을 통해 만들어진 새로운 매트릭스를 이용한다. 이 때 분석에 사용할 매트릭스는 여전히 크기 때문에 로

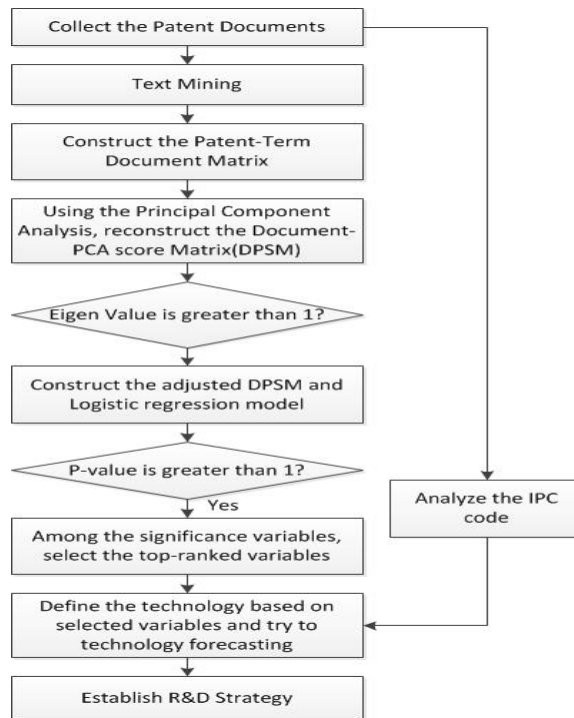


그림 1. 제안된 방법의 순서도
Fig. 1. Flow chart of Suggested Methodology

지스틱 회귀분석을 진행하는데 있어 어려움이 존재한다. 또한, 결과에 영향을 미칠 수 있는 회귀분석의 다중공선성 문제 역시 존재한다. 차원의 크기와 다중공선성 문제를 모두 해결하기 위하여 주성분분석(Principal Component Analysis)을 실시하여 매트릭스의 크기를 축소한다. 이 때 주성분의 개수는 고유 값이 1이상인 주성분만을 선택하여 차원축소를 실시한다[13].

주성분분석 결과를 통하여 차원 축소된 매트릭스는 로지스틱 회귀분석에 사용된다. 이 때 종속변수(Target variable)는 기술을 대표하는 단어의 출현확률로 하여 모델을 생성한다. 생성된 모델에서 각각의 매개변수(parameter) 중 유의한 변수들만을 추출하여 그 변수에 포함되어 있는 키워드를 핵심키워드로 선정한다. 선정된 핵심 키워드의 결과를 통하여 기술정의와 기술예측을 실시하며, 이를 바탕으로 향후 연구개발 전략을 수립한다.

- Step 1 : 예측하고자 하는 기술 분야를 선정하고 관련된 특허문서를 수집한다.
- Step 2 : 수집된 특허데이터를 텍스트 마이닝하고 분석에 불필요한 불용어, 기호 등을 제거한다.
- Step 3 : 모델 생성이 가능하도록 차원을 축소하고 다중공선성의 문제 등을 제거하기 위해 주성분 분석을 실시한다.
- Step 4 : 주성분분석 결과를 이용하여 로지스틱 회귀모델을 생성하고 이 중 유의한 변수를 확인한다.
- Step 5 : 유의한 변수들에 속한 키워드를 핵심키워드로 선정하며, 선정된 키워드와 수집된 특허데이터의 IPC코드를 비교하여 기술정의를 실시한다.
- Step 6 : 기술정의 결과를 바탕으로 필요한 연구개발 전략을 제안한다.

4. 실험 및 결과

앞서 3장에서 언급한 바와 같이, 본 연구를 진행하기 위해 무인 자동차에 관련된 특허를 수집하였으며, 구글(Google)과 다른 자동차 관련 업체(Others)로 구분하여 특허검색을 진행하였다. 검색결과 는 아래 표2와 같다.

표 2. 무인자동차 관련 특허 검색 결과
Table 2. Patent search result: Autonomous vehicle

Groups	Search results	Available patents
Google	35	31
Others	564	495

특허검색 결과, Google과 Others군에서 각각 35건과 564건의 특허가 검색되었다. 이 중에서 중복특허, 거절특허 등 노이즈 제거 후 유효특허로써 각각 31건, 495건이 선정되었다.

특허명세서는 선 공개를 별도로 신청하지 않을 경우, 출원 후 1년 6개월이 지나면 자동으로 공개가 된다. 따라서 논문작성일 기준으로 하였을 때, 확인 가능한 특허는 2013년 이전에 출원된 특허이다. 그러므로 추세분석에서는 2013년 이후에 출원된 특허는 제외하도록 한다.

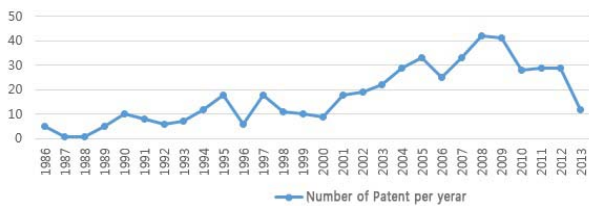


그림 2. 연도별 무인자동차 관련 특허출원 건수
Fig. 2. Illustration of applied patents by year

검색에 대한 연도별 특허출원 추세를 살펴보면, 1986년 이후로 관련 특허가 꾸준히 출원되는 것을 알 수 있다. 그 후 2008년과 2009년을 기점으로 2010년까지 특허 출원건수가 감소하였다. 2010년부터 2012년까지는 별다른 변동이 없는 것을 확인할 수 있다.

3장의 제안방법에 따라 특허분석을 실시하기 위해 전처리과정으로 텍스트 마이닝을 실시하여 키워드를 추출하였다. 이때, R-project의 "tm" 패키지를 이용하였다[14]. 추출된 키워드 중 "are", "is", "can" 등과 같은 기술예측 하는데 있어 불필요한 단어는 제거하고 findFreqTerms 함수를 이용하여 일정빈도 수 이상의 키워드를 추출하였다. 추출된 키워드는 표3에 나타내었다.

표 3. 최종 선정된 키워드
Table 3. Selected core keywords

Groups	Google	Others
Common Keywords	Driving, Data, Control, Autonomous, Sensor, Information, Location	
Different Keywords	Aspect, Computer, Environment, Map	Detection, Line, Obstacle, Processing, Signal, Unit, Apparatus, System, Position, Image, Navigation

선정된 키워드는 비구조화의 형태로 되어 있으므로 분석이 가능하도록 구조화된 매트릭스의 형태로 생성해야한다. 특허 문서와 단어로 구성된 구조화된 매트릭스를 Patent-Term Document Matrix(PTDM)라 하며 구조화된 매트릭스를 구축함으로써 분석이 가능하게 한다[6].

PTDM을 구성한 후 로지스틱 회귀모델을 생성하기 위해서는 다중공선성의 문제와 매트릭스의 희소(sparse)문제를 우선 해결해야한다. 이러한 문제점은 주성분분석(Principal component analysis)을 통해 해결할 수 있다. 본 연구에서는 고유 값이 1이상인 주성분만을 선택하여 모델생성에 사용한다[13]. Google과 Others의 고유 값을 구한 결과, 그림3과 그림4에 나타내었다.

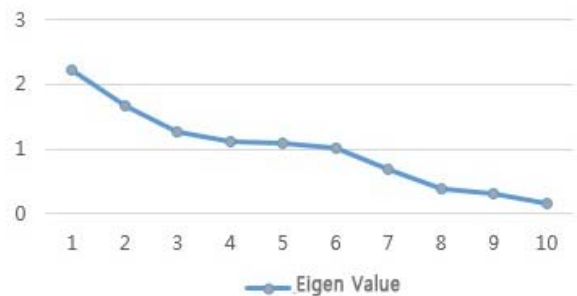


그림 3. Google 매트릭스의 고유 값
Fig. 3. Eigen Value of Google Matrix

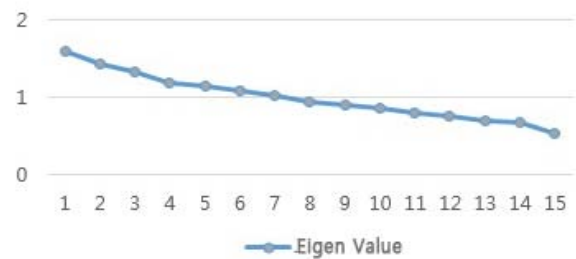


그림 4. Others 매트릭스의 고유 값
Fig. 4. Eigen Value of Others Matrix

Google과 Others 매트릭스의 고유 값 계산결과, Google매트릭스에서는 제7주성분부터 고유 값이 1미만이었고, Others매트릭스에서는 제8주성분부터 고유 값이 1미만이였다. 따라서 본 실험에서는 모델생성을 위해 각각 6개와 7개의 주성분을 선택하였다.

주성분분석 결과를 바탕으로 추출된 단어들과 "autonomous" 사이의 통계적으로 유의미한 키워드를 추출하고자 "autonomous"의 출현여부를 종속변수로 하는 로지스틱 회귀모델 구축을 시도하였고 모델구축 결과는 표4와 표5에 각각 나타내었다.

표 4. 로지스틱 회귀모델 생성결과(Google)
Table 4. Logistic regression model(Google)

Variable	Coefficient	p-value
Intercept	5.9776	0.0196308
Comp.2	3.02438	0.0303575
Comp.3	-0.94654	0.252180
Comp.4	2.17669	0.344816
Comp.6	4.5936	0.0595

표 5. 로지스틱 회귀모델 생성결과(Others)

Table 5. Logistic regression model(Others)

Variable	Coefficient	p-value
Intercept	-0.46218	0.0000308
Comp.1	0.52908	0.0000875
Comp.2	-0.21644	0.026680
Comp.3	0.17099	0.078016
Comp.4	-0.19368	0.091975
Comp.5	0.36975	0.001311
Comp.6	0.43901	0.000452
Comp.7	0.08968	0.441232

다중선형회귀모델 또는 로지스틱 회귀모델에서 매개변수/의 p-value가 0.05이하 일 때 유의하다고 할 수 있다[5]. 이를 바탕으로 Google 매트릭스를 이용한 로지스틱 회귀모델 구축결과, 모델에서는 Comp.2만 유의한 변수로 나타났다.

같은 방법으로 Others 매트릭스를 이용한 로지스틱 회귀모델에서는 Comp.1, Comp.2, Comp.5, Comp.6이 유의한 매개변수/로 나타났다.

핵심키워드를 추출하기 위해 유의한 변수의 주성분 계수(PC loadings)를 추출하였고, 그 중 상위 3개에 속하는 키워드만을 선정하여 다음 표6과 표7에 나타내었다.

표 6. 유의한 변수들의 주성분 계수(Google)

Table 6. PC Loadings(Google)

Variable	Loadings
aspect	0.603
sensor	0.475
information	0.293

표 7. 유의한 변수들의 주성분 계수(Others)

Table 7. PC Loadings(Others)

Comp.1 Variables	Loadings	Comp.2 Variables	Loadings
system	0.213	position	0.430
navigation	0.165	data	0.345
position	0.151	image	0.322

Comp.5 Variables	Loadings	Comp.6 Variables	Loadings
location	0.514	drive	0.670
detect	0.499	apparatus	0.260
sensor	0.374	sensor	0.145

종속변수인 “autonomous”에 유의한 영향을 미치는 변수들로 Google에서 제2주성분이 있었으며, Others에서는 제1, 2, 5, 6주성분이 존재하였다. 다시 유의한 변수들 중에서 영향을 미치는 단어를 알아보기 위해 주성분 계수를 나타내었고, Google에서는 “aspect”, “sensor”, “information” 순으로 영향을 미치고 있는 것을 알 수 있었다. Others에서는 “system”, “navigation”, “position”, “data”, “image”, “location”, “detect”, “drive”, “apparatus”의 단어가 “autonomous”에 영향을 미치고 있는 것을 확인할 수 있었다.

기술정의는 추출된 키워드와 IPC코드에 따른 분류를 종합하여 실시하였고 그림5와 그림6에 수집한 Google과 Others 특허들의

IPC코드 분포를 나타내었다.

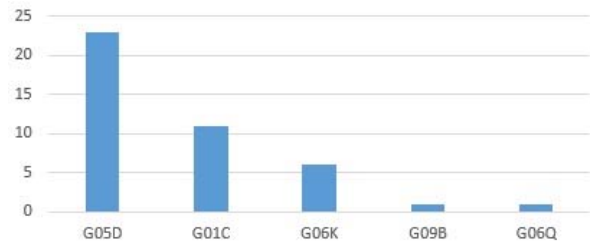


그림 5. 수집한 Google 특허의 상위 IPC코드

Fig. 5. Top 5 ranked IPC code: Google

표 8. 상위 5개 IPC코드의 내용: Google[15]

Table 8. Definition of top 5 ranked IPC codes: Google[15]

IPC code	Definition
G05D	System for controlling or regulating non-electric variables
G01C	Gyroscopic instruments
G06K	Recognition of Data; Presentation of Data; Record carriers; Handling record carriers
G09B	Educating; Cryptography; display;
G06Q	Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes

수집된 Google 특허의 IPC코드는 G05D, G01C, G06K, G09B, G06Q순으로 나타났다. IPC코드는 의미는 다음 표8에 나타내었으며 IPC코드와 추출된 핵심키워드를 바탕으로 Google에 대한 기술예측을 실시하였다.

검색된 Google의 특허문서를 이용한 실험에서 추출된 키워드 “aspect”, “sensor”, “information”와 상위 IPC코드의 내용을 고려하여 다음 표9과 같이 기술정의를 실시하였다.

표 9. Google의 핵심기술 정의

Table 9. Core technology definition: Google

Google	Techno-logy group	Extracted Keyword	Technology Definition
	Techno-logy 1	aspect, sensor, information	Technology or method for processing information obtained by sensor

수집된 Others의 특허문서를 이용하여 동일한 방법으로 상위 5개 IPC코드를 그림6에 나타내고, IPC코드에 관련된 설명을 표10에 나타내었다.

수집된 Others의 특허문서들의 상위 IPC코드를 확인한 결과, G06F, G05D, G01C, G06K, H04N 순으로 나타났고, 또한 앞서 로지스틱 회귀분석을 통해 얻은 유의한 키워드들은 “system”, “navigation”, “position”, “data”, “image”, “location”, “detect”, “sensor”, “drive”, “apparatus”로 나타났다. 두 개의 결과를 종합적으로 고려하여 Others의 핵심기술을 다음 표11에 나타내었다.

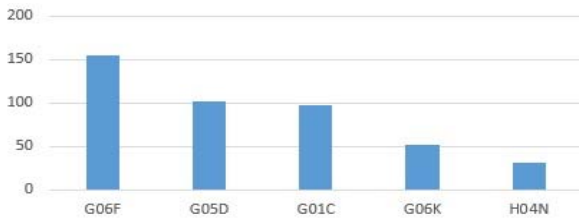


그림 6. 수집한 Others 특허의 상위 IPC코드
Fig. 6. Top 5 ranked IPC code: Others

표 10. 상위 5개 IPC코드의 내용 : Others[15]

Table 10. Definition of top 5 ranked IPC codes: Others[15]

IPC code	Definition
G06F	Electric digital data processing
G05D	System for controlling or regulating non-electric variables
G01C	Gyroscopic instruments
G06K	Recognition of Data; Presentation of Data; Record carriers; Handling record carriers
H04N	Pictorial communication

표 11. Others의 핵심기술 정의

Table 11. Core technology definition: Others

Techno-logy Group	Extracted Keywords	Technology Definition
Others	No. 1 system, navigation, position	Navigation system for operating autonomous vehicle
	No. 2 position, data, image	Technology for processing image or data
	No. 3 location, detect, sensor	Technology for external condition by sensor
	No. 4 drive, apparatus, sensor	Self-operating apparatus or device by using sensor

기술정의 결과, Google과 Others 두 집단에서 모두 센서를 이용한 기술에 집중하고 있는 것을 확인할 수 있었다. 하지만 Others에서는 무인자동차 운행을 위한 경로 시스템에 대한 기술을 연구하고 있다는 것을 알 수 있었다.

본 연구를 통해 센서를 활용하여 수집, 처리, 활용하는 방법에 대한 기술에 대하여 더욱 깊이 있는 조사를 통해 연구개발 진행이 필요할 것으로 예상되었다.

5. 결론 및 향후 연구

본 논문에서는 기술예측 및 연구개발 전략을 제안하기 위해 로지스틱 회귀모델을 이용하여 유의한 키워드를 추출하고 키워드와 IPC 코드를 바탕으로하여 기술정의를 실시하였다.

Google에서는 센서를 통해 획득한 정보를 처리하는 방법론에 관한 기술을 핵심기술로 판단하였다. Others는 운행, 경로 시스템에

관한 기술과 이미지 또는 데이터의 처리 기술, 센서를 이용한 외부 요인 탐지 기술, 마지막으로 센서를 이용한 운행 가능한 기계 혹은 장비로 판단하였다.

이러한 결과를 바탕으로 할 때, 현재 두 집단에서 모두 센서를 이용하는 기술에 관한 특허가 다수 출원되고 있는 것을 확인할 수 있었다. 또한 본 연구에서는 고려하지 않았지만 거절된 특허의 상당수는 선행기술 조사의 미흡으로 인한 중복투자로 인한 것이라는 보고서를 참고할 때 상당 수 부분의 연구는 이미 중복되었을 가능성이 높을 것으로 생각할 수 있다[16].

따라서 본 논문에서는 현재의 중복되는 부분으로 발견된 부분을 Others에서는 계속해서 연구를 진행하기 보다는 Google과의 전략적 제휴, Cross-licencing과 같은 협력할 수 있는 방안을 모색하여 중복투자를 막고, 현재 연구가 미흡한 부분에 투자를 진행한다면 무인자동차의 상용화시기를 현재 예측시점보다 더욱 앞당길 수 있을 것으로 예상된다.

향후 연구에서는 기술정의를 보다 정량적으로 할 수 있는 연구가 필요할 것으로 보인다. 또한 제안방법에서는 주성분분석을 실시하기 전 키워드 추출 시 빈도수만을 고려하여 선정하였지만 앞으로는 키워드의 중요도를 반영하여 분석에 반영한다면 더욱 정확한 결과를 도출하는데 큰 기여를 할 수 있을 것으로 생각된다.

References

- [1] Korea Intellectual Property Office, Korea Invention Promotion Association, "Patent and Information Analysis(for researchers)", Seoul: Kyungseong Moonhwa, 2007.
- [2] C. Kim, J. Lee, C. Lee, Management of Technology, Vision Project Management, Seoul, 2010.
- [3] Yuen-Hsien Tseng, Chi-Jen Lin, Yu-I Lin, "Text mining techniques for patent analysis," *Information Processing and Management*, no.43, pp. 1216 - 1247, 2007.
- [4] S. Jun, S. Park, Y. Shin, D. Jang, H. Chung, "Forecasting Vacant Technology of Patent Analysis System using Self Organizing Map and Matrix Analysis", *The Journal of Korea Contents Association*, 10th, no.2, pp. 462 - 480, 2009.
- [5] S. Jun, "Technology Forecasting of Intelligent Systems using Patent Analysis", *Journal of Korean Institute of Intelligent Systems*, vol. 21, no. 1, pp. 100 - 105, 2011.
- [6] S. Jun, "A Big Data Learning for Patent Analysis", *Journal of Korean Institute of Intelligent Systems*, vol. 23, no. 5, pp. 465 - 411, 2013.
- [7] J. Yoo, "Patent Information Investigation Theory and Actuality", Seoul: Hyungseul Publishing Networkds, 2004.
- [8] Google Autonomous Vehicle 'openly' Driving in City, Available: <http://news.donga.com/3/01/20140429/63140319/1>, [Accessed: Nov, 15, 2014]
- [9] The latest chapter for the self-driving car: mastering city street driving, Available: <http://googleblog.blogspot.kr/2014/04/the-latest-chapter-for-self-driving-car.html>, [Accessed: Nov, 15, 2014]
- [10] H. Jun, "R&D Trend of Autonomous Vehicle Technology",

NIPA, 2012.

[11] Galit Shmueli, Nitin R. Patel, Peter C. Bruce, "Data Mining for Business Intelligence", WILEY, 2010.

[12] M.Heo, Y.Lee, "Data Mining Modeling and Case", Seoul: Hannarae Publishing, 2008.

[13] Kaiser, H.F, "The application of electronic computers to factor analysis", *Educational and Psychological Measurement*, vol.20, pp. 141 - 151, 1960.

[14] tm package, <http://cran.r-project.org/web/packages/tm/tm.pdf>

[15] WIPS ON, "<http://www.wipson.co.kr>"

[16] S. Kim, "Policy Research on Intellectual Property Rights in Japan", KIIP, 2003.

Phone : +82-2-3290-3900
 E-mail : iguana751@korea.ac.kr



김갑조(Gabjo Kim)

2011년 : 숭실대학교 산업정보시스템공학과 공학사
 2011년~2015년 : 고려대학교 산업경영공학부 박사
 2015년~현재 : 한국지식재산전략원

관심분야 : Technology Forecasting, Data mining
 Phone : +82-2-3290-3900
 E-mail : kkjjo@korea.ac.kr

저 자 소 개



이준석(Junseok Lee)

2013년 : Zhejiang University, Mechanical Engineering and Automation, 공학사
 2014년~현재 : 고려대학교 대학원 산업경영공학부 석사과정

관심분야 : Patent analysis, Data Mining, Management of Technology
 Phone : +82-2-3290-3900
 E-mail : jxli12@korea.ac.kr



박상성(Sangsung Park)

2006년 : 고려대학교 산업시스템정보공학과 공학 박사
 2006년~2014년 : 고려대학교 산업경영공학부 연구교수
 2014년~현재 : 한국지식재산전략원 평가위원
 2014년~현재 : 지식재산창조기업협의회 위원
 2015년~현재 : 고려대학교 기술경영전문대학원 조교수

관심분야 : Patent Analysis, Data Mining, Management of Technology, Technology evolution
 Phone : +82-2-3290-4618
 E-mail : hanyul@korea.ac.kr



이준혁(Joonhyuck Lee)

2012년 : 한국항공대학교 정보통신공학과 공학사
 2014년 : 고려대학교 산업경영공학부 공학석사
 2014년~현재 : 고려대학교 대학원 산업경영공학부 박사과정

관심분야 : Predicting Firm Performance, Management of Technology



장동식(Dongsik Jang)

1979년 : 고려대학교 산업공학과 공학사
 1985년 : 텍사스 주립대학 산업공학과 공학석사
 1988년 : 텍사스 A&M 산업공학과 공학박사
 1989년~현재 : 고려대학교 산업경영공학부 교수

관심분야 : PM, Pattern Recognition, Data Mining
 Phone : +82-2-3290-3387
 E-mail : jang@korea.ac.kr