한국통계학회논문집
2010, 17권, 2호, 205-222

특허와 통계학, 그 연결은?

전성해[1,a], 엄대호[b]

[a]청주대학교 바이오정보통계학과, [b]오클라호마주립대학교 통계학과

요약

특허제도는 발명자의 기술을 공개시키고 동시에 일정기간 동안 발명자에게 해당기술의 독점적 사용권을 법으로 보장하는 지식재산권제도이다. 이 제도를 바탕으로 전 세계는 치열한 기술경쟁을 하고 있다. 특허는 물건의 발명뿐만 아니라 방법의 발명도 인정한다. 방법의 발명에는 기업경영을 위한 비즈니스모델뿐만 아니라 다양한 통계적 분석기법도 포함된다. 본 논문에서는 지금까지 출원, 등록된 통계적 분석기법과 관련된 전체특허를 조사하여 기술유형별로 분석한다. 또한 특허데이터 자체에 대한 통계분석의 적용방법을 제안한다. 이를 위하여 국내와 미국의 특허청으로부터 통계분석 기술관련 전체 등록특허를 검색하여 조사, 분석하였다.

주요용어: 특허, 통계분석, 특허데이터분석.

1. 서론

고대 그리스의 시바리스(Sybaris)라는 도시에서는 특별한 요리법을 창안한 자에게 1년간 독점권을 주어 그 발명에 대한 보상을 하였다 (나까야마 노부히로, 2001). 기원전에 이미 개인의 창의적인 발명에 대한 보호 및 보상제도가 있었음을 알 수 있다. 1883년에 지식재산권 전체에 대한 최초의 국제조약인 파리협약(Paris convention for protection of industrial property) 이후 전 세계의 모든 지식재산권은 국제연합(UN)의 전문기관인 세계지식재산권기구(World Intellectual Property Organization; WIPO) 에 의해 보호되고 있다. 현재의 지식기반사회에서 창의적 지식재산은 국가의 기술경쟁력을 좌우할 수 있기 때문에 전 세계는 자국의 지식재산권을 보호하고 강화하는 정책을 펴고 있다. 대표적인 지식재산권인 특허(patent)는 각국이 자국의 산업발전을 위하여 총력을 기울이는 분야이다. 특허의 어원은 라틴어로 'Patere'이며 '공개되어진 것(be opened)'이라는 의미이다 (황종환, 2001). 즉 창의적이고 유용한 기술을 공개하여 모든 사람들이 알 수 있도록 하여 국가의 경제발전에 기여하는 목적을 가지고 있는 동시에 특허권은 새로운 기술에 대한 일정기간 독점적 권리를 부여하여 발명자에게 이익을 주는 배타적인 지배권이다 (특허청 정보기획팀 등, 2007). 최근 기술에 대한 특허권을 바탕으로 전 세계의 국가들은 치열한 기술경쟁을 하고 있다 (제대식 등, 2000). 특허를 위한 기술분야는 전자, 정보, 바이오기술 뿐만 아니라 최근에는 기업의 경영모델(business model; BM)에까지 확장되고 있다. 특히 바이오, 정보기술 분야의 방대한 데이터베이스와 더불어 고객 마케팅과 기업경영에서 발생하는 온라인, 오프라인 데이터의 통계적 분석방법에 대한 기술특허가 출원, 등록되고 있다. 본 논문은 이 부분에 대한 통계학의 필요성과 특허와의 연관성에 대하여 연구한다. 현재 국내의 많은 대학들은 특허교육에 대한 관심을 보이고 있다. 경영공학, 기술경영학 분야에서는 정규교과과정으로 '특허와 지식재산권', '특허와 정보분석' 등의 과목이 학부와 대학원 과정에 개설되어 있다. 특허청에서도 대학과 공동으로 특허교육을 진

[1] 교신저자: (360-764) 충북 청주시 상당구 내덕동 36 청주대학교 바이오정보통계학과, 부교수.
E-mail: shjun@cju.ac.kr

행하고 있다. 새로운 기술발명에 대한 상당부분이 대학에서 이루어지고 있기 때문이다. 특히 '특허와 정보분석' 등에서 다루어지는 특허데이터의 분석은 출원, 등록된 기술특허를 다양하게 분석하여 향후 필요한 공백기술을 찾아내고 현재 개발되는 기술이 기존의 특허에 침해되는지 여부를 알게 한다. 효과적인 특허데이터의 분석을 통하여 기업은 연구개발에 대한 구체적인 계획을 세우고 불필요한 특허소송을 미리 막을 수 있게 된다. 점점 다양하고 섬세한 기술들이 특허로 등록됨에 따라서 고도의 특허데이터 분석기술이 요구된다. 그림과 빈도에 의존한 기존의 정량적 특허데이터분석에 이제는 다변량분석, 시계열분석 등 고급의 통계분석기법이 필요하게 되었다. 따라서 특허데이터분석 분야에 대한 통계전문가의 관심이 필요하게 되었다. 하지만 현재 특허데이터분석 분야에서 통계학 전문가의 활동은 거의 없는 실정이다. 그러므로 본 논문에서는 처음 특허가 등록되기 시작하여 데이터베이스에 저장된 1948년부터 2009년까지 등록된 모든 통계분석 관련기술의 특허동향을 살펴보고 검색된 특허데이터의 통계분석에 대한 통계분석기법의 적용방안에 대하여 알아본다. 이를 위하여 국내특허청과 미국특허청에 등록된 전체특허를 검색하여 비교분석한다. 아울러 특허데이터 자체에 대한 고급의 통계분석기법의 적용에 대한 문제점과 이에 대한 해결방안을 제안한다.

본 논문에서는 특허와 통계학의 연결을 2가지 관점에서 살펴본다. 우선 통계학 분야에서 특허의 활용이다. 현재까지 통계학 연구자들은 대부분의 연구결과를 논문지에 발표하였다. 이것은 새로운 연구에 대한 최초의미는 있지만 법적으로 해당 연구결과에 대한 독점적 사용권을 보장 받지는 못한다. 이에 비해 연구결과의 논문발표와 병행하여 특허출원 및 등록을 하게 되면 연구결과에 대한 배타적인 권리와 사용권이 보장되는 이점이 있다. 따라서 본 연구에서는 현재까지 출원, 등록된 통계학 관련 전체특허를 검색하여 어떠한 내용들이 특허로 등록되어 왔고 앞으로 이 분야에서 필요한 연구주제에 대하여 알아본다. 이와 함께 본 논문에서는 특허 분야에서 통계학의 활용방안도 연구된다. 연구개발의 기획, 특허권 침해여부, 유망기술의 발굴 등 기업의 특허경영에서 가장 우선적으로 고려되는 것 중의 하나는 특허분석이다. 현재는 표와 그래프 등 기술통계의 결과와 관련 분야 전문가의 주관적 경험에 의존하고 있다. 다양한 통계적 분석기법들이 적용된다면 훨씬 좋은 특허데이터분석 결과가 기대될 것이다. 그러므로 본 논문에서 특허데이터의 통계적 분석에 대한 가능성을 보이기 위하여 주성분분석, 실루엣 측도(Silhouette measure), K-means 군집화 기법을 이용한 특허데이터의 분석을 수행한다. 이와 같은 특허와 통계학의 연결을 통한 시너지 효과를 알아본다.

2. 특허와 특허데이터

연구개발의 결과는 특허로 출원, 등록되고 유지되어야만 법적으로 그 권리를 보호받을 수 있다. 우리나라를 포함한 전 세계 대부분의 나라들은 이와 같은 특허제도를 가지고 있다. 다음 그림 1은 우리나라의 특허출원 절차를 나타내고 있다 (특허청 정보기획팀 등, 2007).

특허를 받기위한 절차의 시작은 특허청에 해당연구기술을 출원하는 것이다. 출원 후 1년 6개월이 경과한 시점에서는 해당 기술을 공개하여 누구든지 그 내용을 볼 수 있게 한다. 이를 통하여 기술연구에 대한 중복투자를 막고 출원인에 대한 해당기술에 대한 우선적인 권리를 인정한다. 출원인은 5년 이내에 특허심사를 청구할 수 있다. 특허심사를 통하여 출원된 특허는 등록이 결정되거나 취소된다. 등록이 결정된 특허라 하더라도 무효의 원인이 발생하면 무효심판청구를 통하여 등록이 취소될 수 있다. 거절된 특허는 의견서와 보정서를 통하여 재심사를 받을 수 있다. 또한 특허가 거절되었을 때 이에 불복할 경우 특허심판, 특허법원 그리고 대법원의 심판청구를 통하여 이의신청을 할 수 있다. 하나의 연구개발 결과가 특허로 결정되기 위해서는 산업상 이용 가능성, 신규성(novelty) 그리고 진보성(inventive step)의 3가지 기준을 갖추고 있어야 한다. 현재 전 세계의 특허청은 자신의 국가에서 출원, 등록된 특허를 데이터베이스로 저장하고 있고 누구든지 자유롭게 접속하여 특허정보를 검색할 수

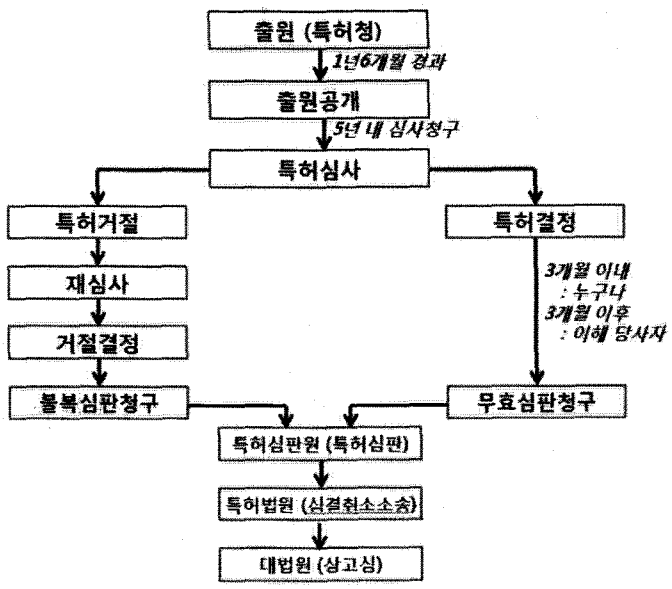


그림 1: *특허출원절차(우리나라)*

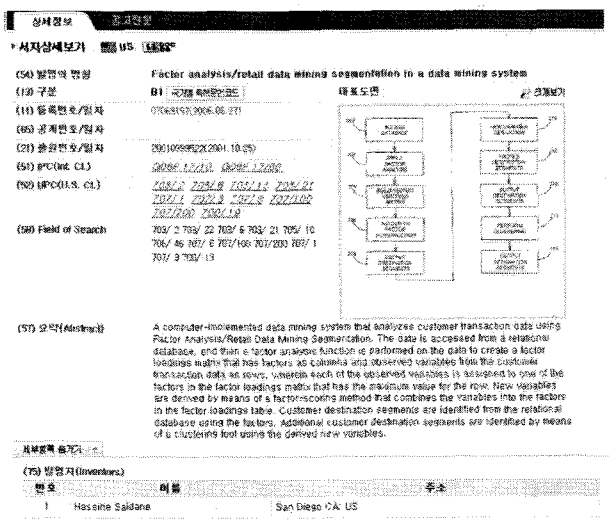
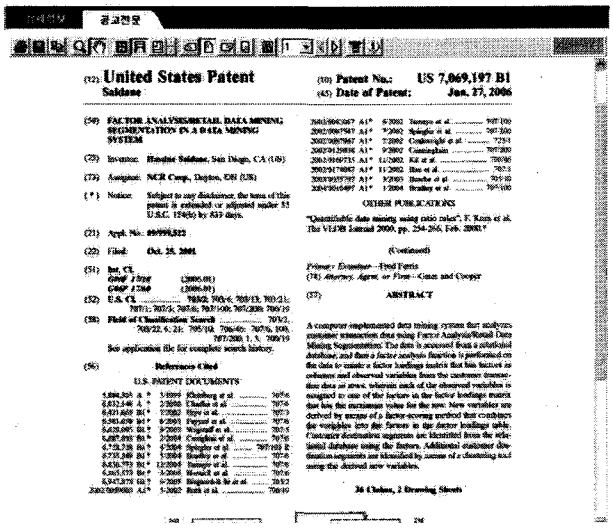


그림 2: *특허데이터의 예 - 상세정보 (출처: 특허정보검색서비스, http://www.kipris.or.kr)*

있게 하고 있다. 그러므로 특허 데이터베이스로부터 자신의 기술 분야에 해당되는 특허를 검색하고 분석하여 연구개발의 기획, 특허침해 여부의 사전판단, 기술예측 등 다양한 의사결정에 활용하고 있다. 이와 같은 특허 데이터베이스에 저장된 특허데이터는 통계학에서 다루어지는 일반적인 데이터와는 상이한 구조를 갖는다. 수치형(numeric) 또는 명목형(nominal)이 아닌 문서(document)의 데이터형태를 이루고 있다. 이와 같은 특허데이터의 특성 때문에 다양한 통계적 분석기법을 사용하여 수준 높은 분석결과를 기대하기에는 어려움이 따른다. 그림 2는 특허데이터의 일반적인 모습을 보여준다.

그림 2는 전체 특허데이터의 일부 내용을 보여주고 있다. 1개의 특허문서에는 발명의 명칭, 출원일



(12) **United States Patent**
Saldana

(10) **Patent No.: US 7,069,197 B1**
(45) **Date of Patent: Jan. 27, 2006**

(54) FACTOR ANALYSIS/RETAIL DATA MINING SEGMENTATION IN A DATA MINING SYSTEM

(75) Inventor: Ricardo Saldana, San Diego, CA (US)

(73) Assignee: NCR Corp., Dayton, OH (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 837 days.

(21) Appl. No.: 09/998,823

(22) Filed: Oct. 25, 2001

(51) Int. Cl.
G06F 17/30 (2006.01)
G06F 17/00 (2006.01)

(52) U.S. Cl. 704/10; 704/9; 704/13; 704/231; 707/1; 707/5; 707/6; 707/100; 707/200; 709/219

(58) Field of Classification Search 707/2, 707/22, 6, 23; 705/10; 704/9, 10, 13, 231; 707/200, 1, 5, 100; 709/219
See application file for complete search history.

(56) **References Cited**
U.S. PATENT DOCUMENTS

OTHER PUBLICATIONS

"Quantitative data mining using ratio rules", F. Korn et al. The VLDB Journal 2000, pp. 254-266, Feb. 2000.*

(Continued)

Primary Examiner—Fred Ferris
(74) *Attorney, Agent, or Firm*—Gates and Cooper

(57) **ABSTRACT**

A computer implemented data mining system that analyzes customer transaction data using Factor Analysis/Retail Data Mining Segmentation. The data is accessed from a established database and then a factor analysis function is performed on the data to create a factor loadings matrix that has factors as columns and observed variables from the customer transaction data as rows, wherein each of the observed variables is assigned to one of the factors in the factor loadings matrix that has the maximum value for the row. New variables are derived by means of a factor scoring method that combines the variables into the factors in the factor loadings table. Customer discriminate segments are identified from the established database using the factors. Additional customer discrimination segments are identified by means of a clustering tool using the derived new variables.

36 Claims, 2 Drawing Sheets

그림 3: *특허데이터의 예 - 공고전문 (출처: 특허정보검색서비스, http://www.kipris.or.kr)*

자, 공개일자, 등록일자, 발명자, 기술요약 등 해당기술에 대한 모든 정보가 포함된다. 실제 등록된 특허의 원본은 다음과 같은 문서형태로 되어 있다. 각국 특허청의 특허데이터베이스로부터 엑셀(excel), 텍스트(text) 등 다양한 파일형태로 제공받을 수 있다.

그림 3은 특허데이터의 일부이며 이들 이외에도 출원인을 포함한 인적사항, 출원인의 국적과 같은 지역정보, 인용특허, 관련특허, 대리인, 심사관 등의 서지적 데이터를 포함한다. 청구항(claim)과 명세서(specification)에는 해당기술에 대한 방대한 기술사항이 나열되어 있다. 이와 같은 특허데이터의 분석은 통계학자보다 산업공학, 경영공학 등의 전문가가 통계학의 분석기법을 이용하여 결과를 얻고 있다. 하지만 아직 시작 단계이다. 통계학이 적용되는 대부분의 응용분야와 마찬가지로 특허데이터의 통계분석에서도 가장 먼저 필요한 것은 특허분야에 대한 지식(domain knowledge)이다. 특허문서에 포함된 여러 종류의 날짜가 의미하는 것을 이해할 수 있어야만 시간에 따른 기술특허의 변화를 모형화 할 수 있고 인용특허를 이용한 특허들 간의 기술 연관성분석도 가능할 것이다 (유선희, 2004; 남영준과 정의섭, 2006). 해당 특허에 대한 가장 많은 정보를 갖고 있는 초록, 명세서 그리고 청구항의 통계적 분석을 위해서는 우선 텍스트 마이닝(text mining)의 전처리(preprocessing) 과정을 통하여 분석가능한 데이터구조로의 변환작업이 필요하다. 특허분석을 위한 텍스트 마이닝의 세부적인 작업은 다음과 같다 (Tseng 등, 2007).

표 1을 통하여 특허데이터의 분석에서 텍스트 마이닝이 많은 작업을 수행함을 알 수 있다. 첫 번째 텍스트 세분화(text segmentation) 작업에서는 특허기술이 나타나 있는 청구항(claim)과 명세서(description)의 내용을 정확히 추출한다. 두 번째부터 여섯 번째까지는 웹문서 등 일반적인 문서에도 적용되는 텍스트 마이닝 작업과 같다. 하지만 마지막 단계의 주제 맵핑(topic mapping)은 특허에 관한 지식을 가지고 있어야만 가능하다. 그러므로 텍스트 마이닝 작업을 통한 효과적인 특허분석을 위해서는 특허에 대한 지식이 꼭 필요하다. 하지만 분석가가 특허에 대한 수준 높은 지식을 갖추는 것은 쉽지 않을 뿐만 아니라 특허전문가의 차이에 따라 결과의 해석도 달라질 수 있기 때문에 보다 객관적인 특허분석의 방법론이 필요하게 된다. 이와 같은 요구에 따라 텍스트 마이닝 결과를 그대로 해석하지 않고, 이후 추가적인 통계분석을 수행하여 사용자의 특허지식에 크게 좌우되지 않는 좀 더 객관적인 특허분석을 얻는 연구가 필요하다. 물론 특허전문가의 지식이 완전히 배제되는 것은 불가능하지만

표 1: 특허분석을 위한 텍스트 마이닝 작업

단계별 기술 세분화	구체적인 작업
1) Text segmentation	특허문서로부터 제목, 요약, 청구항, 명세서 등의 내용을 세분화하여 추출
2) Text Summarization	자연어처리기법을 이용하여 문서를 요약하고 순위 등 간단한 통계량을 사용
3) Stopword and stemming	단순 기능어들을 제거
4) Key-word and key-phrase extraction	특허문서를 대표할 수 있는 핵심단어나 핵심어구를 추출
5) Term association	단어들 간의 동시발생(co-occurrence) 빈도를 이용하여 단어들 간의 연관성 파악
6) Topic clustering	문서-키워드 행렬형태로 특허데이터를 구조화 시켜서 군집화 수행
7) Topic mapping	특허에 대한 도메인지식을 바탕으로 군집화 결과를 적절하게 해석

표 2: 주성분분석을 이용한 특허분석

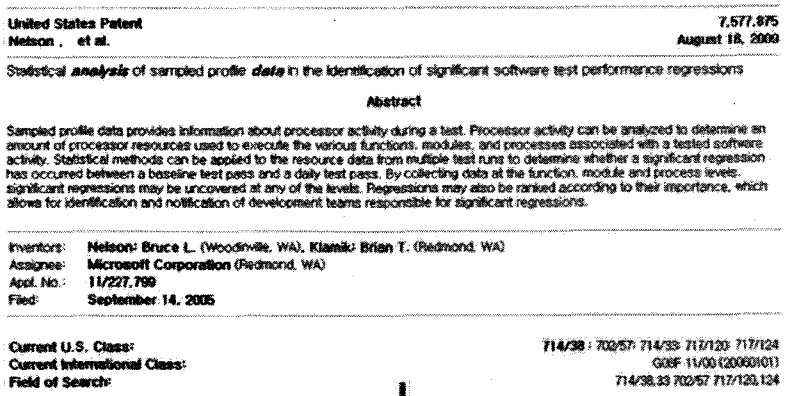
단계	단계별 처리
1) Patent retrieval and collection	주제에 맞는 특허를 검색하여 저장
2) Structured patent data	텍스트 마이닝을 이용하여 특허문서를 구조화 데이터로 변환
3) Principal component analysis	이전 단계에서 구축된 데이터셋을 이용하여 주성분분석을 수행
4) Patent mapping	제1주성분과 제2주성분을 이용하여 산점도를 그리고 이를 바탕으로 공백기술을 찾음

그 영향력에서 좀 벗어나려는 시도이다. 가장 최근의 연구는 주성분분석(principal component analysis; PCA)을 이용한 특허데이터의 분석이다 (Lee 등, 2009). 이 방법은 표 2와 같이 4단계의 과정을 거친다.

첫 번째 단계에서는 전 세계 특허청에서 제공하는 특허데이터베이스로부터 기술주제에 맞는 특허를 검색한다. 이 때 기술주제에 알맞은 특허를 검색하기 위한 검색이 결정이 필요하다. 검색된 특허문서에 대하여 통계적 분석이 가능한 구조화된 데이터로의 변환을 텍스트 마이닝의 전처리 과정을 통하여 두 번째 단계에서 수행한다. 다음 단계에서는 주성분분석을 이용하여 구조화된 특허데이터를 분석한다. 마지막 단계에서는 주성분분석의 결과를 이용하여 산점도를 그리고 이를 바탕으로 공백기술을 찾아낸다. 다음 그림 4는 위 두 번째 단계에서 검색된 특허문서에 대하여 행(row)은 개별특허문서로 열(column)은 키워드로 이루어진 데이터셋(data set)으로의 변환과정을 나타내고 있다.

구조화된 특허데이터의 키워드와 특허는 각각 통계학에서 다루는 데이터구조에서의 변수(variable)와 관측치(observation)에 해당한다. 하지만 이 데이터셋을 이용하여 곧바로 주성분분석을 수행하기에는 어려움이 있다. 왜냐하면 일반적으로 주성분분석은 관측치의 수가 변수의 수에 비해 많아야 하기 때문이다. 검색된 특허문서결과를 이용하여 그림 4의 특허 데이터셋을 구축한 결과를 살펴보면 보통 키워드가 적게는 수백개에서 많게는 수천개 이상이 된다. 하지만 경우에 따라 유효한 특허의 수는 이에 미치지 못하는 경우가 발생한다. 결과적으로 변수의 수가 관측치의 수보다 훨씬 큰 경우가 발생한다. 기존의 특허분석연구에서는 이 문제를 해결하기 위하여 유의어(synonym) 정보를 이용한다 (Feinerer 등, 2008). 대부분의 텍스트 마이닝의 유의어정보는 워드넷(WordNet) 데이터베이스를 사용한다 (Fellbaum 등, 1998). 다음은 R-Project의 'wordnet' 패키지를 이용하여 키워드에 대한 유의어를 찾는 방법이다 (Feinerer 등, 2008).

```
> library(wordnet)
> synonyms("technology")
[1] "applied science"    "engineering"    "engineering science"
[4] "technology"
```



United States Patent 7,577,875
Nelson , et al. August 18, 2009

Statistical ***analysis*** of sampled profile ***data*** in the identification of significant software test performance regressions

Abstract

Sampled profile data provides information about processor activity during a test. Processor activity can be analyzed to determine an amount of processor resources used to execute the various functions, modules, and processes associated with a tested software activity. Statistical methods can be applied to the resource data from multiple test runs to determine whether a significant regression has occurred between a baseline test pass and a daily test pass. By collecting data at the function, module and process levels, significant regressions may be uncovered at any of the levels. Regressions may also be ranked according to their importance, which allows for identification and notification of development teams responsible for significant regressions.

Inventors: Nelson; Bruce L. (Woodinville, WA); Klamik; Brian T. (Redmond, WA)
Assignee: Microsoft Corporation (Redmond, WA)
Appl. No.: 11/227,799
Filed: September 14, 2005

Current U.S. Class: 714/38 ; 702/57; 714/33; 717/120; 717/124
Current International Class: G06F 11/00 (20060101)
Field of Search: 714/38,33 702/57 717/120,124

	Keyword 1	Keyword 2	...	Keyword m
Patent 1				
Patent 2				
...				
Patent n				

그림 4: *구조화된 특허데이터 셋*

	applied science	engineering	engineering science	technology
Patent 1	2	4	1	3
Patent 2	1	2	3	1

	technology
Patent 1	10
Patent 2	7

그림 5: *유의어를 이용한 구조화된 특허데이터의 차원축소*

위 결과는 'technology'와 유사한 단어를 찾는 예제이다. 이를 이용하여 'applied science', 'engineering' 그리고 'engineering science' 키워드들은 키워드 'technology'로 나타낼 수 있다. 즉 그림 5와 같이 구조화된 특허데이터의 차원을 그 만큼 줄일 수 있게 된다.

유의어를 이용한 차원축소를 통하여 수백개 또는 수천개의 키워드는 수십개의 키워드 열로 줄일 수 있다. 키워드 기반의 구조화된 특허데이터의 분석을 위하여 사용 가능한 전처리 과정이다. 하지만 유의어에 의한 차원축소는 작업의 양과 질적인 면에서 어려움이 있다. 적당한 유의어를 찾는 과정은 생각보다 많은 시행착오와 노동력을 요구한다. 또한 유의어를 선정하고 대표 키워드를 결정하는 과정은 주관적이고 이 과정은 최종적인 데이터분석의 결과에도 영향을 미친다. 본 논문에서는 이와 같은 주관적이고 노동력을 요구하는 유의어 중심의 차원축소를 대신할 수 있는 하나의 객관적인 시도로서 SVD(singular value decomposition)를 이용한다 (Haykin, 1999).

표 3: 특허검색을 위한 후보검색어

검색주제	한글 검색어	영문 검색어
통계분석	통계, 정보, 데이터, 자료, 분석, 마이닝	statistic, information, data, analysis, mining

표 4: 특허검색을 위하여 사용된 검색식

검색주제	국내특허 검색식	미국특허 검색식
통계분석	TL = [(통계 + 정보 + 데이터 + 자료) * (분석 + 마이닝)]	TL = [(statistic + information + data) * (analysis + mining)]

표 5: 국내와 미국의 특허검색 결과

검색구분	총 검색건수	유효건수
국내특허	190	53
미국특허	999	340

3. 특허와 통계학의 연결

3.1. 통계학 분야에서 특허의 활용

본 논문에서는 지금까지 한국과 미국에서 출원, 등록된 특허를 모두 검색하여 지금까지의 '통계분석'에 관한 기술동향을 파악하고, 향후 필요한 공백기술을 제시한다. 이를 위하여 무료로 전 세계의 특허검색이 가능한 특허데이터베이스(DB)인 KIPRIS(Korea Intellectual Property Rights Information Service)의 특허정보검색서비스에서 제공하는 국내 출원 및 등록 특허와 미국 출원 및 등록 특허를 이용한다 (특허정보검색서비스, 2009). 특허DB로부터 원하는 특허를 검색하기 위해서는 정확한 검색식이 필요하다. 정확한 검색을 위한 후보검색어로 본 논문에서는 표 3과 같은 검색어를 고려한다.

표 3의 검색어를 바탕으로 통계분석 기술에 관한 특허검색을 위하여 사용된 검색식은 다음 표 4와 같다. 최종적인 검색식은 다양한 후보 검색식들 중에서 유효특허를 가장 많이 검색해 주는 것으로 결정한다. 유효특허는 통계분석이라는 주제에 합당한 특허를 나타낸다.

표 4에서 TL은 특허제목인 발명의 명칭(title)을 나타내고 +와 *는 각각 'or'와 'and' 연산을 나타낸다. 다음 표 5는 결정된 검색식을 이용하여 특허DB로부터 검색한 결과이다.

표 5의 검색결과는 2009년 11월 13일 현재, 1948년 6월 20일부터 2009년 11월 12일까지 출원, 등록된 특허를 대상으로 하였다. 총 검색건수는 키워드 검색식을 통하여 최초에 얻게 된 특허데이터 전체를 나타낸다. 하지만 이 결과가 본 논문의 연구주제인 통계분석기법에 대한 특허만을 포함하고 있지는 않다. 이는 모든 기술주제의 특허검색에서 나타나는 공통된 현상이다. 최초로 검색된 전체특허를 확인하여 기술주제에 적합한 특허만을 최종적으로 선택한 것이 유효특허이다. 기술분야에 따라 다르겠지만 대부분의 특허검색에서 유효특허건수에 대한 비율은 20% 내외로 나타나고 있다. 물론 검색식의 차이에 따라 유효건수의 비율은 조금씩 달라지게 된다. 본 논문의 검색식도 유효특허 건수의 비율을 가장 크게 하는 검색식으로 선정하였다. 국내특허의 경우 검색된 전체특허에서 최종적으로 선정된 유효특허는 모두 53건이다. 유효특허 비율은 27.89%로서 일반적인 기술분야에 비해 다소 높은 유효특허비율을 나타내고 있다. 이는 통계분석이 다른 기술분야에 비해 상대적으로 더 세부적이고 전문적이기 때문이다. 미국특허의 경우에는 34.03%의 유효특허비율을 나타내고 있다. 다음 그림 6은 지금까지 출원, 등록된 통계분석에 대한 특허 건수를 연도별로 나타낸다.

그림 6을 통하여 1948년 특허출원 이후, 2000년에 처음으로 통계분석에 대한 특허등록이 이루어졌음을 알 수 있다. 이는 다른 분야에 비해 현재 통계분석 기술에 대한 특허는 시작단계임을 알 수 있다. 2004년 이후에 점차로 통계분석 관련 기술특허의 등록이 증가하고 있음을 알 수 있다. 이에 비해 다음

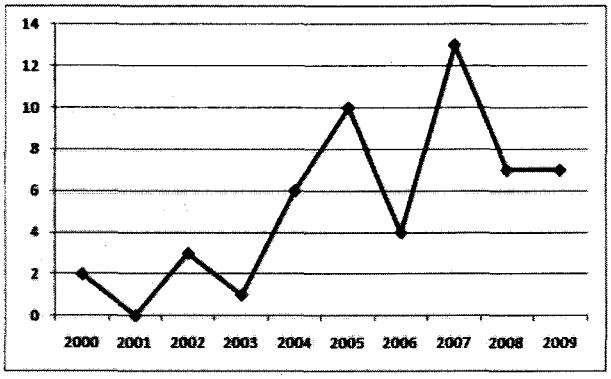


그림 6: *통계분석 기술에 대한 연도별 등록특허 건수(국내특허)*

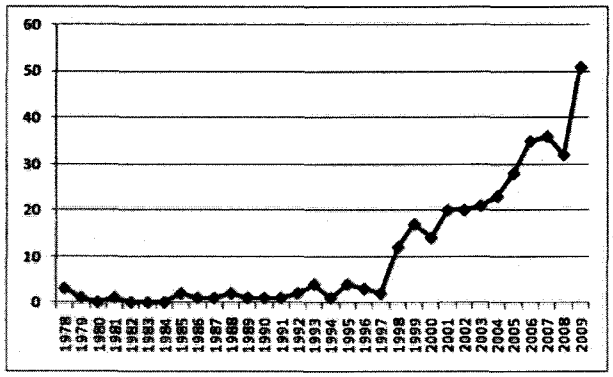


그림 7: *통계분석 기술에 대한 연도별 등록특허 건수(미국특허)*

의 미국의 경우를 보면 국내에 비해 활발한 특허등록이 이루어지고 있음을 알 수 있다. 하지만 다른 기술분야에 비해 미국의 경우도 여전히 시작단계임을 알 수 있다.

그림 7을 보면 1978년에 처음으로 통계분석 기술특허의 등록이 이루어졌지만 1998년에 이르러 등록특허가 증가하고 있음을 알 수 있다. 이후 빠른 증가세를 나타내고 있다. 국내와 미국의 통계분석 기술특허의 연도별 등록건수를 종합해 보면 다른 기술분야에 비해 통계분석 기술에 대한 특허출원을 초창기이고 아직도 출원, 등록될 수 있는 기술들이 매우 많이 존재할 수 있음을 알 수 있다.

유효특허로 선정된 통계분석기술에 관한 전체 국내특허를 살펴보면 몇 가지 특징이 있다. 우선 발명의 명칭이 '통계데이터 분석방법(method for analyzing statistical data)', '데이터 분석방법(data analysis method)', '데이터 마이닝 시스템 및 그 방법(data mining method and data mining system)'과 같이 매우 넓은 기술 주제를 포함하는 것부터 '데이터마이닝의 분류 의사 결정 나무에서 분산이 작은, 즉 순수한 관심 노드 분류를 통한 자료의 통계적 분류 방법(statistic classification method of data using one-sided purity splitting criteria for classification trees in data mining)', '데이터마이닝의 분류 의사 결정 나무에서 극단값을 가지는 관심 노드 분류를 통한 자료의 통계적 분류 방법(statistic classification method of data using one-sided extreme splitting criteria for classification trees in data mining)'과 같이 세부적인 주제를 갖는 기술까지 다양하게 등록되어 있음을 알 수 있다(부록A (1)~(2); (9)~(12)). 또한 2000년 이후부터는 통계학자들을 중심으로 한 특허가 조금씩 나타나고 있다(부록A (1)~(7)). 등록된 기술특허의 발명의 명칭은 미국특허도 국내특허와 비슷한 결과를 보인다. 'Data analysis method and apparatus

표 6: 통계분석에 대한 특허 기술분류 (국내특허)

원천기술	응용기술
A1 - 분석기법(알고리즘)	B1 - 인터넷, 정보기술(IT)
A2 - 데이터의 저장, 관리	B2 - 경영, 마케팅, 기술경영
A3 - 데이터 분석환경	B3 - 바이오, 의료, 헬스
	B4 - 제조, 건설, 교통

표 7: 통계분석에 대한 특허 기술분류 (국내특허)

기술분류		특허건수 (%)
원천기술	A1	14 (26.4)
	A2	3 (5.7)
	A3	7 (13.2)
응용기술	B1	9 (17.0)
	B2	9 (17.0)
	B3	8 (15.0)
	B4	3 (5.7)
전체		53 (100)

for data mining', 'Data analysis system', 'Data analysis apparatus', 'Data analysis system and method', 'Method and apparatus for data analysis' 등과 같이 기술의 주제가 광범위한 경우와 'Analysis of retail transactions using gaussian mixture models in a data mining system', 'Partial stepwise regression for data mining', 'System and method for biological data analysis using a bayesian network combined with a support vector machine' 등과 같이 매우 구체적인 통계분석기술에 대한 특허를 등록한 경우도 있다. 본 논문에서는 국내에 출원된 통계분석 특허에 대한 세부적인 기술조사를 위하여 다음 표 6과 같은 기술분류를 하였다.

본 연구에서는 통계분석을 위한 기술을 크게 원천기술과 응용기술로 구분하였다. 물론 한국통계학회, 한국과학재단, 등에서 제공하는 통계학의 세부분류도 고려될 수 있지만 이들 분류코드는 매우 세분화되어 있기 때문에 53건의 유효특허에 이들 세부분류기준을 적용하기에는 어려움이 있다. 본 연구에서는 통계분석을 하나의 기술분야로 고려하였고 또 검색된 유효특허를 고려하여 위 표와 같은 분류기준을 만들었다. 본 논문에서 제시하는 기술분류가 물론 절대적인 기준은 아니다. 기술분류가 이루어지는 목적과 확보된 특허데이터에 따라 그 기준을 정할 수 있을 것이다. 새로운 데이터분석기법, 대용량 데이터의 저장 및 관리, 효율적인 데이터분석을 위한 환경(interface)구축 등이 원천기술에 해당되고, 응용기술은 웹 데이터의 분석 및 정보기술에 바탕을 둔 통계분석방법, 경영, 마케팅, 기술경영(management of technology) 그리고 바이오, 의료, 헬스분야에서 발생하는 데이터의 통계분석 기술 등을 포함한다. 기타 제조, 건설, 교통 분야도 응용기술의 범주로 포함시켰다. 다음 표 7은 각 기술분야에 따른 전체 국내의 유효특허 분포이다.

전체 유효특허 중에서 원천기술에 해당되는 특허는 24건으로 45.3%를 차지하고 나머지 54.7%인 29건은 응용기술이다. 전체적으로 원천기술에 비해 응용기술이 조금 더 많이 출원, 등록되었다. 원천기술과 응용기술을 모두 포함한 비교에서는 새로운 분석기법(A1)에 대한 특허가 전체의 26.4%로 가장 많이 출원, 등록되었음을 알 수 있다. 응용기술에서는 인터넷 정보기술, 마케팅 경영 그리고 바이오 기술 등이 모두 비슷하게 출원, 등록되었다. 다음 표 8은 기술분류에 따른 연도별 특허건수이다.

원천기술과 응용기술의 전 분야에 대한 통계분석기술에 관한 특허는 2000년 이후부터 본격적으로 나타나고 있다. 연도항목에 표시되지 않은 연도는 특허출원이 1건도 없었다. 다음 표 9는 연도별 등록된 특허건수이다. 출원 후 등록까지의 심사기간이 있기 때문에 출원보다는 늦음을 알 수 있다.

표 8: 기술분류에 따른 연도별 출원건수 (국내특허)

연도	A1	A2	A3	B1	B2	B3	B4	합계
1996	1							1
1997				1			1	2
1999					2			2
2000	1							1
2001	2			3	1	1		7
2002	6			2	1	2		11
2003		1			2	1		4
2004			3	1				4
2005	2		2			2		6
2006		1	1	1	1	1	1	6
2007	2	1	1	1	2	1	1	9
합계	14	3	7	9	9	8	3	53

표 9: 기술분류에 따른 연도별 출원건수 (국내특허)

연도	A1	A2	A3	B1	B2	B3	B4	합계
2000	1			1				2
2002	1				1			2
2003				1				1
2004	1			1	2	1	1	6
2005	4			3	1	2		10
2006	2					2		4
2007	2	2	5	1	1	1	1	13
2008			2	1	2	1	1	7
2009	3	1		1	1	1		7
합계	14	3	7	9	9	8	3	53

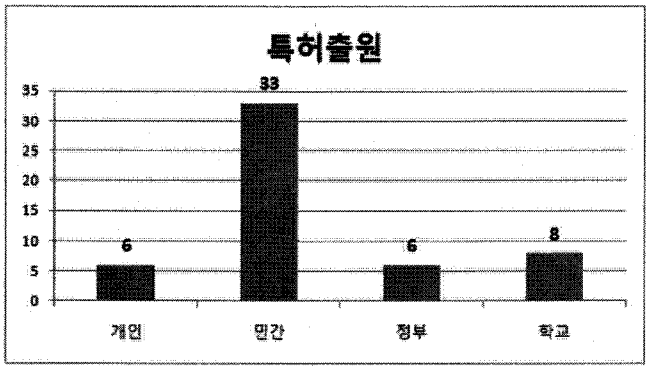


그림 8: *통계분석 기술에 대한 국내 출원인별 특허건수 (국내특허)*

2005년 이후 꾸준히 특허가 등록되고 있음을 알 수 있다. 다음 그림 8은 출원인별로 그 동안 등록된 특허를 나타낸다.

개인은 자연인인 개인자격으로 특허를 출원한 경우이고, 민간은 민간연구소를 포함한 사기업이며, 정부는 정부출연연구소와 공기업을 포함한 정부기관이다. 그리고 학교는 대학교를 나타낸다. 민간기업이 33건으로 전체 등록특허의 약 62%를 차지하고 있다. 다음으로는 학교가 15%의 특허를 출원, 등록하였다. 통계학과 교수를 포함한 석박사과정의 학생 등 통계학 전문가의 상당부분이 속해있는 학교

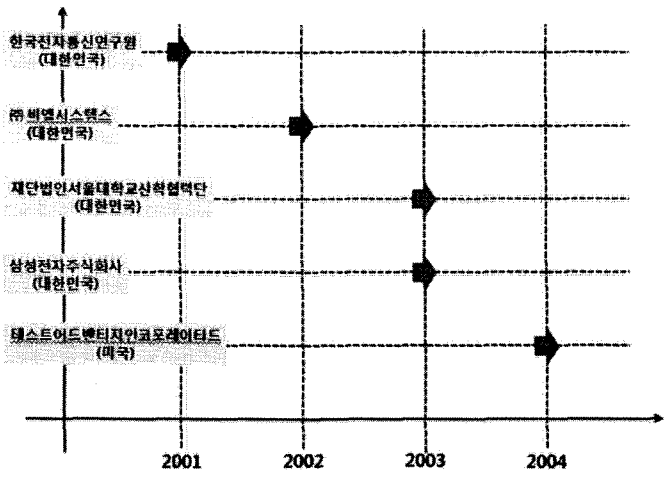


그림 9: *출원인 분석 (상위 5)*

에서의 특허출원이 활발히 이루어지지 않고 있음을 알 수 있다. 이는 대학교의 산학협력단 등에서의 특허출원이 공대 위주로 이루어지고 있고, 또한 공대교수이외의 교수들이 아직 특허에 대한 관심이 상대적으로 많지 않기 때문일 것이다. 다음 그림 9는 통계분석 기술관련 국내특허의 상위 5위의 출원인과 이들이 처음으로 특허를 출원을 출원한 시점을 나타내고 있다.

그림 9를 보면 국가연구기관, 대학교, 대기업, 중소기업 그리고 외국업체가 각각 1개씩 분포되어 있음을 알 수 있다. 이와 같은 결과를 통하여 통계분석 관련 새로운 기술에 대한 연구, 개발은 국가기관, 학교, 기업들에 의해 다양하게 이루어지고 있음을 알 수 있다. 특히 미국의 업체가 국내에 많은 특허를 출원하여 기술의 권리를 인정받으려고 함을 알 수 있다. 이는 국내 통계분석 관련 시장규모가 앞으로 지속적으로 커질 것임을 나타내고 있다.

3.2. 특허 분야에서 통계학의 활용

앞 절에서는 통계분석기술에 관한 국내의 전체 등록특허를 검색하여 통계분석에 대한 기술동향을 알아보았다. 다음으로 본 논문에서는 '통계분석' 기술에 관한 유효한 미국특허를 이용하여 통계적 분석기법을 이용한 특허데이터의 분석을 수행한다. 텍스트 마이닝의 전처리 과정을 거쳐서 SVD에 의한 주성분분석을 수행한 후에 최종적으로 점도표를 이용하여 특허지도를 만들었다. 구축된 특허지도(patent map)를 통하여 일차적으로 공백기술의 가능성이 확인되면 이후 추가적인 분석을 통하여 공백기술에 대한 객관적이고 구체적인 결과를 얻게 된다. 본 연구에서는 2003년까지의 특허데이터를 이용하여 공백기술을 예측하고 이것을 2004년 이후의 특허데이터에 적용하여 예측방법의 타당성을 평가한다. 다음 그림 10은 2003년까지 등록된 통계분석기술특허에 대한 주성분분석 결과를 이용하여 제1주성분과 제2주성분을 각각 가로축과 세로축으로 하여 그린 주성분 산점도이다.

그림 10을 통하여 주관적으로 전체특허를 4개의 군집으로 고려하면 우측 상단의 2개의 군집은 많은 특허가 몰려있다. 이에 비하여 좌측 하단의 군집은 몇 개의 특허가 있을 뿐 아직 특허가 거의 등록되지 못하는 부분이다. 즉 다른 군집들에 비해 상대적으로 군집의 밀도가 낮다. 이와 같은 군집이 공백기술을 나타낼 가능성이 있다 (Lee 등, 2009). 따라서 추가적인 분석을 위하여 본 논문에서는 군집분석을 수행하였다. 우선 객관적으로 적절한 군집수를 결정하기 위하여 실루엣 값(Silhouette width)을 이용하였다 (Rousseeuw, 1987). 실루엣 값은 선행연구에서 AIC(Akaike's information crite-

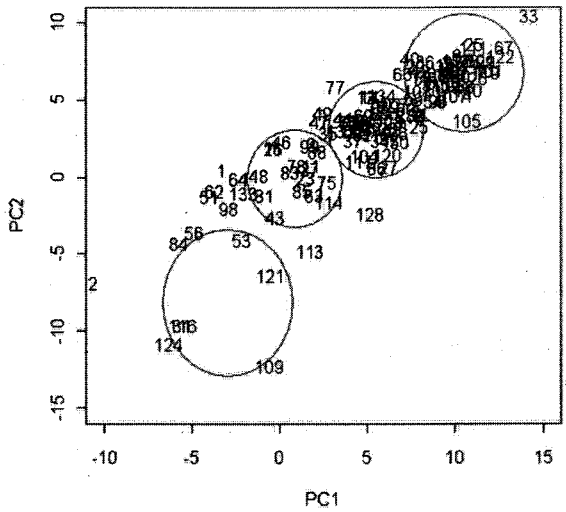


그림 10: *주성분 산점도를 이용한 통계분석기술에 대한 특허지도 (미국특허)*

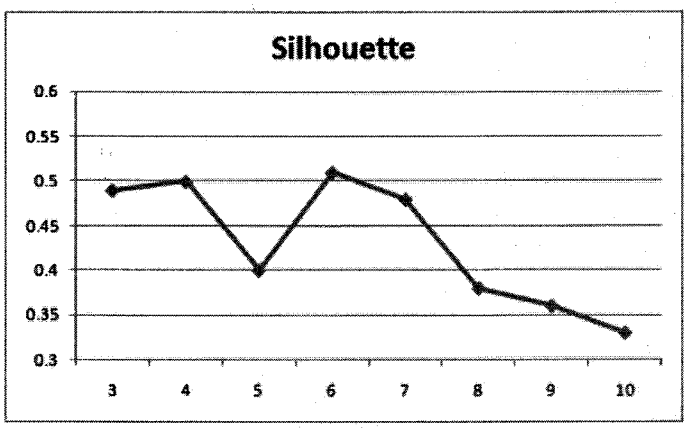


그림 11: *Silhouette 값을 이용한 최적 군집수 결정 (미국특허)*

rion)나 BIC(Bayesian information criterion) 등의 군집수 결정 측도들에 비해 효과적인 결과를 제시하고 있다 (Jun과 Lee, 2010). 이 방법에 의하면 가장 큰 실루엣 값을 갖는 경우의 군집수를 최적 군집수로 결정한다. 다음 그림 11은 실루엣 측도를 이용하여 2003년까지의 전체 특허데이터에 대한 군집수에 대한 실루엣 값의 분포이다. 가로축은 군집의 개수이고 세로축은 실루엣 값을 나타낸다.

그림 11의 결과를 이용하여 가장 큰 실루엣 값을 갖는 군집수 6으로 최적군집수를 결정하였다. 군집수를 6으로 하는 *K*-means 군집화를 수행한 결과는 다음과 같다. 적절한 *K* 값을 선택할 수 있는 경우에는 K-means 군집화가 우수한 군집결과를 제시한다는 기존의 많은 연구결과에 의해 본 논문에서도 *K*-means 군집화를 사용하였다 (Fang과 Ma, 2009; Chen 등, 2009; Kanungo 등, 2002).

표 10의 군집분석결과를 보면 군집1, 군집2, 군집4 그리고 군집5는 군집3과 군집6에 비해 많은 특허들을 가지고 있다. 따라서 이들 4개의 각 군집은 해당기술이 많이 특허로 나타났음을 알 수 있다. 하지만 군집3과 6은 아직 기술의 특허화가 제대로 이루어지지 못하고 있다. 군집3은 단지 2개의 특허가 1개의 군집을 형성하고 있다. 때문에 통계분석기술 분야에서 아주 특별한 경우(outliers)에 해당될 수

표 10: *K*-means 군집화를 이용한 특허분류 (미국특허)

군집	해당특허	건수(%)
1	4, 5, 6, 8, 9, 10, 15, 18, 20, 25, 26, 29, 32, 33, 36, 38, 40, 42, 52, 55, 58, 67, 68, 70, 71, 73, 79, 87, 90, 91, 95, 96, 97, 101, 103, 105, 107, 110, 111, 112, 115, 119, 122, 126, 129, 130	46(34.3%)
2	7, 12, 14, 17, 28, 39, 45, 50, 54, 59, 60, 66, 72, 74, 80, 89, 92, 94, 102, 106, 108, 118, 123, 125, 134	26(19.4%)
3	100, 131	2(1.5%)
4	1, 11, 16, 21, 23, 27, 43, 46, 48, 51, 53, 62, 63, 64, 75, 78, 81, 83, 85, 98, 113, 114, 133	23(17.2%)
5	2, 3, 13, 19, 24, 30, 34, 35, 37, 41, 44, 47, 49, 57, 61, 65, 69, 76, 77, 82, 86, 88, 99, 104, 117, 120, 127, 128	29(21.6%)
6	22, 31, 56, 84, 109, 116, 121, 124	8(6.0%)

표 11: 각 군집 별 상위 10개의 키워드 (미국특허)

군집	키워드
1	image, video, signal, insulation, control, device, wavelet, motion, filter, chromosome
2	user, stored, space, engine, management, record, relational, performance, interface, material
3	computer, time, pattern, node, automatic, integrate, communication, digital, learning, dynamic
4	association, evaluation, display, virtual, scanner, device, content, textual, chart, vibration
5	transaction, signal, corresponding, file, flight, transportation, game, physical, device, health
6	pattern, stream, web, adaptive, network, dynamic, motion, map, fraudulent, neural

있다. 이에 비하여 군집6은 8개의 특허가 포함되어 있고 본 논문에서는 이 군집이 통계분석기술에 대한 공백기술을 나타낸다고 판정하였다. 물론 군집3에 대한 분석도 의미가 있을 수 있지만 본 논문에서는 좀 더 많은 특허를 포함한 군집6에 대하여 공백기술 예측방법을 적용하였다. 각 군집이 나타내는 기술을 정의하기 위하여 본 논문에서는 포함된 특허들의 키워드를 이용하였다. 각 군집에 대한 상위 10개의 키워드는 다음 표 11과 같다.

통계분석기술에 대한 공백기술로 정의된 군집6의 상위 10개의 키워드를 살펴보면 신경망과 같은 인공지능 기법을 이용하여 인터넷 환경에서 발생하는 연속적인 자료를 패턴화, 시각화하는 기술에 대한 특허가 아직은 활발하지 않음을 알 수 있다. 추가적으로 지리정보의 분석과 비정상적인 패턴의 탐지 모형화기술 등도 통계분석 관련 공백기술로 정의할 수 있을 것이다. 2003년까지의 특허 데이터를 이용한 위의 분석결과를 이용하여 2004년 이후의 그룹6 기술을 해당하는 특허의 분포를 살펴보면 다음 그림 12와 같다.

2003년까지의 전체특허의 6%에 불과하던 그룹6의 특허비율이 2004년부터 2009년 11월까지는 전체특허(205건)의 27.8%(57건)으로 늘어났다. 특히 2009년은 11월 현재 17건으로 증가세가 매우 크게 나타났다. 이는 기술을 개발하고 출원 및 등록을 하는 기간을 고려하면 2003년에 미리 공백기술을 예측하여 개발하였다면 개발부터 출원 및 등록기간을 단축하여 다른 기업보다 먼저 새로운 기술에 대한 특허 권리를 확보할 수 있음을 알 수 있다. 이와 같은 공백기술 예측방법은 통계분석기술 뿐만 아니라 바이오, 정보기술 등 모든 기술분야에 적용할 수 있을 것이다. 다음 그림 13은 본 논문에서 이루어진 특허데이터의 통계분석에 대한 절차이다.

각국의 특허데이터베이스로부터 적당한 검색식을 만들어 필요한 특허데이터를 검색하고 이후 텍스트마이닝을 이용한 전처리를 통하여 통계적 분석이 가능한 분석데이터를 얻게 된다. 그림 13과 같이 본 논문에서는 주성분분석, 실루엣 측도, *K*-means 군집화 기법을 이용하여 특허데이터를 분석하였다. 최종적으로 분석결과를 이용하여 앞으로 필요한 공백기술을 예측하였다. 앞으로 다양한 통계적 분석기법들이 적용되면 더 세분화되고 정교한 특허데이터의 분석과 이를 바탕으로 한 최적의 의사결정이 가능할 것이다.

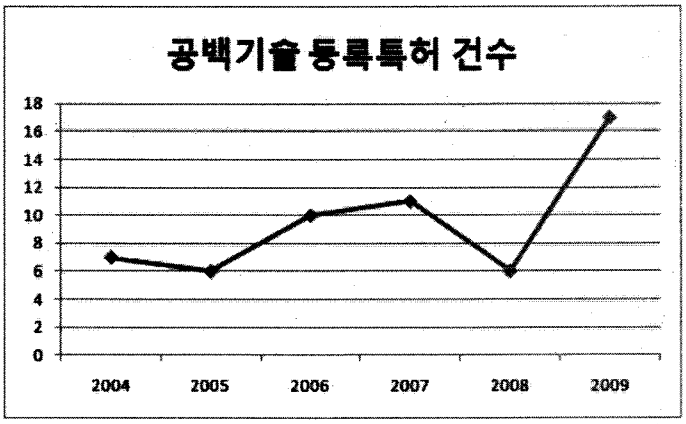


그림 12: *통계분석 기술관련 공백기술(그룹 6)의 연도별 등록특허 건수 (미국특허)*

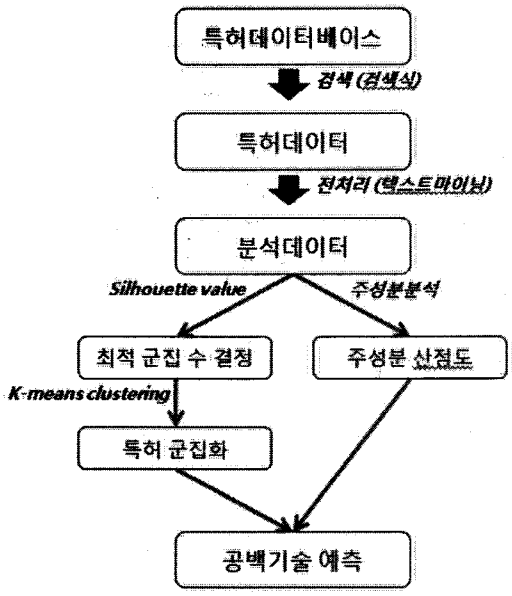


그림 13: *특허데이터의 통계분석절차*

3.3. 특허와 통계학의 구체적인 연결방안

본 논문은 특허와 통계학의 구체적인 연결방안에 대하여 연구하였다. 그 동안 통계학 분야에서 연구자는 주로 논문지를 통하여 자신의 연구결과를 발표하였다. 하지만 새로운 연구개발에 대한 창의적인 결과물에 대한 법적인 권리를 제대로 인정 받기에는 어려움이 있다. 하지만 논문지의 발표와 함께 특허출원을 병행한다면 자신의 연구결과에 대한 독점적이고 배타적인 권리를 행사할 수 있게 된다. 더 나아가 해당 특허가 기업체 등에서 사용하게 되면 특허사용로도 받게 된다. 또한 지금까지 출원, 등록된 통계학 분야의 특허를 검색하여 분석하게 되면 기업에서 실제로 필요로 하는 통계기법들에 대한 추세변화와 동향파악을 할 수 있게 된다. 이를 통하여 작게는 통계연구자가 자신의 새로운 연구분야를

Patent → Statistics	Patent ← Statistics
1. 통계학 연구결과의 적극적인 특허출원 2. 통계학 관련 특허분석을 통하여 통계학 관련 연구기술의 동향파악 3. 기업에서 필요한 통계방법론의 기획 및 연구개발	1. 기존의 빈도와 그래프 위주의 특허분석에서 다양한 통계분석기법의 적용 2. 통계적 분석을 위한 특허데이터의 새로운 전처리방법 개발 3. 통계적 자료분석에 기반한 특허경영(기술예측, 기술마케팅, 기술정책, ...)

그림 14: *특허와 통계학의 연결방안*

찾을 수 있고 크게는 국가경제에 이바지 할 수 있는 통계학 연구가 가능할 것이다.

이와 함께 특허 분야에서도 통계학의 기여가 기대된다. 특허분석의 목적은 대규모로 저장된 특허데이터를 통하여 기술의 개발추세와 동향을 파악하고 분석대상 기술의 과거와 현재의 특허데이터를 통하여 미래에 필요한 기술을 예측하여 기업의 향후 기술개발 방향을 정립하는 것이다. 효과적인 특허데이터의 분석은 국가와 기업의 올바른 특허경영을 가능하게 한다. 현재 대부분의 특허데이터의 분석은 특허의 서지적 사항과 특허빈도를 이용한 표와 그래프 등 기술통계와 관련분야 전문가의 주관적 지식에 의존하고 있다. 좀 더 다양한 통계적 분석기법의 적용이 필요한 상황이다. 또한 특허데이터의 전처리 과정부터 통계전공자가 통계적 데이터분석을 고려한 참여가 필요할 것이다. 현재 대부분의 전처리 과정은 전산학의 텍스트 마이닝에 의존하고 있지만 보다 다양한 통계적 전처리 과정이 필요할 것이다. 다음 그림 14는 특허와 통계학의 구체적인 연결방안을 정리한 그림이다.

4. 결론 및 향후 연구과제

통계학과 특허의 연결에 대하여 본 논문에서는 2가지 접근방안을 제시하였다. 첫째는 통계분석 기술에 대한 특허출원이다. 통계학과 대학교수와 대학원생을 포함한 통계전문가가 자신의 연구결과를 논문지에 투고할 뿐만 아니라 적극적으로 특허출원을 하여 자신이 노력해서 얻은 연구결과의 권리를 법으로 보호받고 이를 적극적으로 국가의 산업발전에 활용하는 것이다. 이를 위하여 본 논문에서는 지금까지 등록된 통계분석기술에 관한 전체 특허를 대상으로 기술동향을 알아보았다. 최근에 통계분석에 대한 원천기술에 대한 특허가 증가하고 있음을 알 수 있었다. 등록된 특허의 대부분이 기업에서 이루어졌지만 최근에 대학교수들에 의한 특허가 나타나고 있음도 아울러 확인할 수 있었다.

다음으로 통계학의 분석기법을 적용하여 다양한 기술분야의 특허데이터를 분석하고 공백기술을 찾아내어 기업의 연구개발에 필요한 중요한 지식을 찾아내는 것이다. 이를 통하여 기술개발의 중복투자를 막고 막대한 비용이 소요되는 특허소송을 방지할 수 있게 될 것이다. 본 논문에서는 SVD에 의한 주성분분석, 실루엣 값 그리고 K-means 군집화를 이용하여 공백기술을 예측하였다.

앞으로 좀 더 다양한 통계분석 기법들을 적용할 수 있는 새로운 특허문서의 전처리방법과 함께 특허데이터의 분석에 필요한 통계적 분석기법에 대한 연구가 활발히 이루어져야 할 것이다.

부록 A: 주요국내특허 (발명자, 발명의 명칭, 등록번호, 등록일자)

(1) 이영섭, 데이터마이닝의 분류 의사 결정 나무에서 분산이 작은, 즉 순수한 관심 노드 분류를 통한 자료의 통계적 분류 방법(statistic classification method of data using one-sided purity splitting criteria for classification trees in data mining), 10-0498651-0000, 2005.06.22.

(2) 이영섭, 데이터마이닝의 분류 의사 결정 나무에서 극단값을 가지는 관심 노드 분류를 통한 자료의 통계적 분류 방법(statistic classification method of data using one-sided extreme splitting criteria for classification trees in data mining), 10-0484375-0000, 2005.04.12.

(3) 김용대, 데이터 마이닝에서의 앙상블기법에 적용되는 연관성 규칙생성 장치 및 그 방법(Apparatus and method for rule generation in ensemble machines), 10-0497212-0000, 2005.06.15.

(4) 김용대, 데이터 마이닝을 위한 최적의 의사 결정 나무 선택 장치 및 그 방법(Apparatus and method for optimal decision tree selection), 10-0497211-0000, 2005.06.15.

(5) 박태성, 디엔에이 마이크로어레이 자료 분석 방법 및 그 시스템(METHOD AND SYSTEM FOR ANALYZING DNA MICROARRAY DATA), 10-0469608-0000, 2005.01.24.

(6) 김용대, 앙상블 모형을 이용한 데이터 마이닝 모형 구축 장치 및그 방법(Apparatus and method for construction model of datamining using ensemble machines), 10-0640264-0000, 2006.11.24.

(7) 박용태, 윤병운, 이성주, 특허 정보의 텍스트 마이닝(Text Mining)에 의한 기술 공백의 발견 방법과 그 시스템(the process and system for finding patent vaccum bytext mining), 10-0714267-0000, 2007.04.26.

(8) 박태성, 이은경, 마이크로어레이 데이터 분석 방법 및 시스템(METHOD AND SYSTEM FOR ANALYZING MICROARRAY DATA), 10-0817103-0000, 2008.03.20.

(9) 김영로, 김원영, 통계데이터 분석방법(Method for analyzing statistical data), 10-0343524-0000, 2002.06.25.

(10) 오성혁, 데이터분석방법(DATA ANALYSIS METHOD), 10-0273567-0000, 2000.09.04.

(11) 이원석, 데이터 마이닝 시스템 및 그 방법(Data Mining Method and Data Mining System), 10-0913027-0000, 2009.08.12.

(12) 이정한, 데이터 마이닝 방법(Method for Datamining), 10-0919684-0000, 2009.09.23.

부록 B: 주요미국특허 (발명자, 발명의 명칭, 등록번호, 등록일자)

(1) Erika Ayukawa, Toyohisa Morita, Akira Maeda, Yukiyasu Ito, Data analysis method and apparatus for data mining, 6510457, 2003.01.21.

(2) Kenneth K. Dickinson, Ann Arbor, Data analysis system, 5797796, 1998.08.25.

(3) Yoshinori Satou, Akira Maeda, Hideyuki Maki, Katsumi Omori, Data analysis apparatus, 5802254, 1998.09.01.

(4) Colin P. Sheppard, Data analysis system and method, 6026397, 2000.02.15.

(5) Harald Aagaard Martens, Jan Otto Reberg, Method and apparatus for data analysis, 5983251, 1999.11.09.

(6) Mikael Bisgaard-B hr, Scott Woodroofe Cunningham, Analysis of retail transactions using gaussian mixture models in a data mining system, 6947878, 2005.09.20.

(7) Martin Keller, Partial stepwise regression for data mining, 6895411, 2005.05.17.

(8) Jie Cheng, Chao Yuan, Bernd Wachmann, Claus Neubauer, System and method for biological data analysis using a bayesian network combined with a support vector machine, 7240042, 2007.07.03.

참고 문헌

나까야마 노부히로 (2001). <특허법>, 법문사.

남영준, 정의섭 (2006). 인용정보를 이용한 신 특허지수 개발에 관한 연구, <정보관리학회지>, **23**, 221–241.

유선희 (2004). 특허인용 분석을 통한 기술수명예측모델 개발에 관한 연구, <정보관리연구>, **35**, 93–112.

제대식, 이은철, 윤국섭 역 (2000). <지식경영과 특허전략>, 세종서적.

특허정보검색서비스, http://www.kipris.or.kr/kor/main/main.jsp

특허청 정보기획팀, 한국발명진흥회 정보활용지원팀 (2007). <특허와 정보분석(개정판)>, 성민.

황종환 (2001). <특허법>, 한빛지적소유권센터.

Chen, X., Yin, W., Tu, P. and Zhang, H. (2009). Weighted *k*-means algorithm based text clustering, *Proceedings of International Symposium on Information Engineering and Electronic Commerce*, 51–55.

Fang, C. and Ma, J. (2009). A novel *k*'-means algorithm for clustering analysis, *Proceedings of International Conference on Biomedical Engineering and Informatics*, 1–5.

Feinerer, I., Hornik, K. and Meyer, D. (2008). Text mining infrastructure in R, *Journal of Statistical Software*, **25**, 1–54.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*, Bradford Books.

Haykin, S. (1999). *Neural Networks A Comprehensive Foundation*, Prentice Hall.

Jun, S. and Lee, S., (2010). Empirical comparisons of clustering algorithms using silhouette information, *International Journal of Fuzzy Logic and Intelligent Systems*, (accepted).

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R. and Wu, A. Y. (2002). An efficient *k*-means clustering algorithm: Analysis and implementation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 881–892.

Lee, S., Yoon, B. and Park, Y. (2009). An approach to discovering new technology opportunities: Keyword-based patent map approach, *Technovation*, **29**, 481–497.

Rousseeuw, P. J. (1987). Silhouette: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, **20**, 53–65.

Tseng, Y. H., Lin, C. J. and Lin, Y. I. (2007). Text mining techniques for patent analysis, *Information Processing & Management*, **43**, 1216–1247.

2010년 1월 접수; 2010년 3월 채택

Patent and Statistics, What's the Connection?

Sunghae Jun[1,a], Daiho Uhm[b]

[a]Department of Bioinformatics & Statistics, Cheongju University
[b]Department of Statistics, Oklahoma State University

Abstract

A patent is a right of intellectual properties to an inventor or its assignee for a limited period under an international law. Not only in an invention of new machines, but it is competitive for using and creating technology in the world based on the patents. Most of the business models are good examples for patented technology, however a statistical analyzing model could be another one. In this paper we study and analyze the patents for the statistical analyzing and data mining models which are currently applied and registered, and suggest a statistical tool for analyzing and categorizing patent data. For this study all the patents in Korea and U.S. are listed and searched to sample the only cases concerning statistics.

Keywords: Patent, statistical analysis, patent data analysis.

[1] Corresponding author: Associate Professor, Department of Bioinformatics & Statistics, Cheongju University, Cheongju Chungbuk 360-764, Korea. E-mail: shjun@cju.ac.kr