https://doi.org/10.5392/JKCA.2018.18.10.443

CPC 기반 특허 기술 분류 분석 모델

A Study of CPC-based Technology Classification Analysis Model of Patents

채수현, 김장원
국립군산대학교 소프트웨어융합공학과

Soo-Hyeon Chae(shchae@kunsan.ac.kr), Jangwon Gim(jwgim@kunsan.ac.kr)

요약

최근 들어 지식재산권의 확보는 기업의 기술 경쟁력 확보를 위해 점점 더 중요하게 되었다. 특히 특허는 기업의 핵심 기술 및 요소 기술을 포함하고 있기 때문에 특허 분석을 통한 기업 가치 측정 및 경쟁 기술 분야 분석 등의 연구가 활발히 진행되고 있다. 국제특허분류(IPC)를 기반으로 다양한 특허 분석 연구가 진행되었으나, IPC는 최신의 기술 분야를 포함하고 있지 않으며 기술의 상세 분류가 충분하지 않아 기술 분류 정확도가 낮아진다. 이를 보완하기 위해 최신의 기술 분야를 포함하고 상세한 기술 분류를 위한 선진특허분류(CPC)가 개발되었으나 이러한 특징을 고려한 특허 분석 연구가 아직 미흡하다. 본 논문에서는 CPC의 상세 분류체계를 이용하여 특허에 포함된 기술 분류 분석 모델을 제안한다. CPC의 상세 분류체계간의 연관관계 중요도 및 효율성을 고려하여 출원인의 특허를 분석하여 핵심 기술 분류 추출을 통해 기존 IPC 기반의 방법보다 상세하고 정확한 분석이 가능하다. 기존의 IPC 기반의 특허 분석 방법과 비교 평가를 통해 제안 모델이 출원인의 핵심 기술 분류를 분석함에 있어 더 좋은 성능을 보임을 확인하였다.

■ 중심어 : | 선진특허분류 | 국제특허분류 | 특허 분석 | 특허 데이터 마이닝 |

Abstract

With the explosively increasing intellectual property rights, securing technological competitiveness of companies is more and more important. In particular, since patents include core technologies and element technologies, patent analysis researches are actively conducted to measure the technological value of companies. Various patent analysis studies have been conducted by the International Patent Classification(IPC), which does not include the latest technical classification, and the technical classification accuracy is low. In order to overcome this problem, the Cooperative Patent Classification(CPC), which includes the latest technology classification and detailed technical classification, has been developed. In this paper, we propose a model to analyze the classification of the technologies included in the patent by using the detailed classification system of CPC. It is possible to analyze the inventor's patents in consideration of the relation, importance, and efficiency between the detailed classification schemes of the CPCs to extract the core technology fields and to analyze the details more accurately than the existing IPC-based methods. Also, we perform the comparative evaluation with the existing IPC based patent analysis method and confirm that the proposed model shows better performance in analyzing the inventor's core technology classification.

■ keyword : | IPC | CPC | Patent analysis | ARM | ANP | DEA |

* 이 논문은 2017학년도 군산대학교 신임교수 연구비 지원에 의하여 연구되었음.
접수일자 : 2018년 08월 29일
수정일자 : 2018년 09월 27일
심사완료일 : 2018년 10월 19일
교신저자 : 김장원, e-mail : jwgim@kunsan.ac.kr

I. 서 론

과학기술이 발전함에 따라서 기술 및 시장의 변화가 빠르게 이루어지고 있으며, 국가 또는 기업 간의 지식재산과 관련된 경쟁이 치열해지고 있다. 특히 정보통신기술 분야 및 첨단기술 기반 산업의 경우 지식재산 경쟁우위를 지속적으로 선점하기 위해 기술의 독점 및 사용 권리가 보장되는 특허를 지속적으로 생산하고 있다. 유망한 특허권의 확보가 기업의 시장가치를 상승시킨다는 결과도 발표되었다[1]. 또한 4차 산업 시대에 들어서면서 다양한 기술들의 확산 및 융합의 요구사항이 증가되어 기업들의 신규시장으로의 진출은 확산되고 있다. 이러한 상황에서 신규시장에 첫 발을 들여놓게 되는 기업들은 해당 시장을 선점한 회사들과의 특허 분쟁 위협에 노출될 수 있으므로 목표하는 시장의 특허 정보를 사전에 분석하여 지식 재산의 장벽의 높고 낮음에 대한 정확한 판단이 요구된다. 즉 사전에 철저한 지식 수립 전략을 수립하여 선행 기술을 분석하고 강력한 특허 포트폴리오를 구축하여 특허 분쟁을 대비하는 것이 필요하다[2]. 그러므로 시장을 선도하는 핵심 기술 및 기업 별 요소 기술들을 상세히 분석할 필요가 있다. 이를 위해 다양한 특허 분석 기법들이 연구되었으며, 주로 텍스트 마이닝 기법과 자연어 처리를 이용하여 특허 기술 용어를 분석하는 방법[3][4], 그리고 특허 발명자 정보를 기반으로 한 특허 발명 네트워크를 분석하는 연구들이 수행되었다[5][6]. 그렇지만 특허가 발명되기 위해서는 신규성, 진보성 등이 있어야하기 때문에 특허 내 표현되는 기술 용어의 경우 유사한 기술에 대한 권리를 주장함에도 서로 다른 용어로 표현되는 특성이 있다. 따라서 논문, 보고서 등의 과학기술문헌과 다르게 특허 기술 용어에 대한 검색의 재현율이 낮다. 검색 재현율을 높이기 위한 다양한 연구가 진행되었으며, 그 중에서 특허에 포함된 분류 체계 코드를 기반으로 특허 분류 및 특허 검색에 대한 연구가 진행되었다. 특허 분류체계는 신규 특허에 대한 심사과정에서 신규 특허의 신규성 및 진보성을 기존 특허와 비교 평가하기 위해 사용되며, 특허 기술 발전의 추이를 분석하기 위한 기초 통계 정보 도출을 위해 사용된다. 이러한 분류 체계로 1968년 제정된 IPC(International Patent Classification) 체계가 사용되며 62개 국가에서 채택하여 사용되었다. 그렇지만 미국, 유럽, 일본의 경우 서로 다른 분류 체계를 국가별로 개발해서 사용하고 있기 때문에 특허 기술에 대한 글로벌 트렌드의 통합 비교에 어려움이 발생한다. IPC가 개발된 이래로 최근 10여 년간 기술 개발 속도는 점차 증가되어 기존의 분류 체계로 정의하지 못하는 최신 기술들이 등장하게 되어 신규 및 상세 기술에 대응하는 분류가 부족하게 되었다. 그러므로 특허 분류 및 검색 시 효율성이 저하되는 문제가 증가되었다. 또한 재현율이 낮은 특허의 특성 때문에 IPC 분류체계를 참고하여 신규 특허를 심사하는 심사관의 경우 연간 출원되는 출원량 보다 심사하는 연간 처리량이 적은 상황에 이르게 되었다. 따라서 이러한 상황을 극복하고 특허 문헌에 대한 분류품질을 향상시켜 과거에 수행된 분류 결과를 재활용 하고 분류 체계의 상호 연동을 위해 선진특허분류(CPC, Cooperative Patent Classification)가 2013년 제안되었다. CPC는 새로운 기술을 반영하고 기존 IPC의 97.7%를 포함하여 호환이 가능하다. 그 결과 CPC를 사용하는 국가별(미국, 유럽 등) 특허 비율은 2013년 31.1%에서 2016년 70.9%로 증가하고 있다. 국내에서는 2015년 1월 이후 모든 신규출원 특허에 CPC가 부여되고 있으며 2018년 7월 현재 한국특허정보원(Kipris)에서 제공되는 약 111만 건의 국내 특허에 CPC가 부여되고 있지만 CPC에 대한 연구가 아직 미흡한 상황이다. 따라서 본 논문에서는 출원인의 핵심 기술 분류 체계를 분석하기 위해 CPC 기반 특허 기술 분류 분석 모델을 제안한다. 기존 IPC 기반의 핵심 기술 분류 도출 방법과 비교 분석을 통해 CPC의 유용성을 보이며, CPC를 통해 기업의 핵심 기술 분류를 보다 상세하고 정확하게 추출할 수 있음을 보인다. 이를 위해 본 논문에서는 IPC 기반의 연구인 Kim[7]의 접근 방법을 CPC에 적용하여 IPC와 CPC의 비교 평가를 수행한다. 본 논문의 기여도를 나열하면 다음과 같다. 첫째, 지속적으로 활용되고 있는 CPC 기반 핵심 기술 분류 분석 모델을 정의하고 구현하였다. 둘째, 최신 기술 및 선도적인 연구 및 개발을 추진하는 정부출연연구소에서 발명된 특허에 제안모델

을 적용해 연구소의 핵심 기술 분류에 대한 분석이 가능함을 보인다.

본 논문의 구성은 다음과 같다. 2장은 관련연구로 특허 분석 연구의 필요성 및 분류 체계 기반 특허 분석 연구에 대해 기술한다. 3장에서는 본 논문에서 제안하는 CPC 기반 핵심 기술 분류 분석 모델에 대해 자세하게 기술한다. 4장에서는 기존 IPC 기반의 모델과 제안 모델과의 비교 평가를 실시하고 이를 기반으로 출원인들의 핵심 기술 분류에 대한 비교 분석 결과를 보인다. 마지막으로 5장에서는 결론에 대해 기술하고 본 논문이 갖는 한계점 및 향후 연구에 대하여 기술한다.

II. 관련 연구

특허에 표현된 기술용어들은 기본적인 특허가 가지는 신규성, 진보성 측면 때문에 동일 기술용어 또는 유사한 기술임에도 서로 다른, 또는 새로운 기술로 표현되어서 특허에 대한 기술용어 검색 시 기술용어의 다양성 및 공통 기술용어 사전의 부재로 검색에 대한 재현율이 낮다. 그 결과 기업들의 핵심 기술 및 요소 기술의 도출, 기업의 핵심 기술 개발 추이 등의 분석 정확도가 낮다. 이를 해결하기 위한 방법으로 텍스트 분석 기법 대신 특허의 분류 체계를 이용한 연구들이 수행되었다. Kim은 IPC의 출현 빈도와 사회네트워크 분석을 기반으로 특정 기술 분야의 핵심 기술을 추출하고, 특허의 패밀리정보와 피인용정보를 활용하여 핵심 특허를 도출하는 방법을 제안한다[8]. 핵심 기술을 추출하기 위한 세부적인 기술 분야를 선정하고, IPC를 서브 클래스(sub-class) 수준까지 정제한 뒤 출현 빈도와 사회네트워크 분석의 매개 중심성 및 중개 중심성 값을 기반으로 IPC별 우선순위를 통해 핵심 기술을 선정한다. Lee는 기술이전이 된 특허들의 IPC를 기반으로 연관규칙 분석 및 사회네트워크 분석을 통해 기술 추이 및 기술 관계를 분석하여 기술이전 대상 기술의 파악 및 어떠한 기술들이 융합되었는지 추출하는 방법을 제안하였다[9]. 이를 위해 공공기관 및 대학의 특허 중 기술이전이 된 특허들의 IPC를 서브 클래스 수준으로 정제한 뒤 출현 빈도 및 연관규칙 분석을 적용해 산출된 결과 값과 사회네트워크 분석으로 도출된 기술의 중복 결과를 고려하여 기술 관계 및 기술 융합을 추출한다. Kim은 특허의 IPC에 데이터 마이닝 및 다기준 의사결정 접근법을 적용해 특허의 핵심 기술을 도출하는 방법을 제안했다[7]. 핵심 기술 도출을 위해 특정 영역 특허들의 IPC를 메인그룹(Main Group) 수준으로 정제한다. IPC 기반 연관규칙 행렬을 도출하여 기술상호관계 값을 추출하고 ANP(Analytic Network Process)기법을 적용하여 기술 중요도를 계산한다. 도출된 기술 중요도 값은 DEA(Data Envelopment Analysis)방법을 통해 핵심 기술을 도출한다. 이처럼 IPC를 기반으로 다양한 연구들이 수행되었으나, 최근 기술의 발전 속도가 가속화됨에 따라 기 정의된 IPC로는 기술에 대한 상세한 표현이 부족할 뿐만 아니라 새로운 기술에 대응할 수 없으며, 서로 다른 나라에서 출원되는 특허들의 기술 정보를 통합 관리할 수 없는 문제가 발생하였다. 이에 대안으로 CPC가 개발되었다. CPC는 IPC 체계를 기반으로 다양하고, 세부적으로 확장된 체계로 국내에서는 2015년 이후 발명 특허부터 CPC가 부여되었으나 최근에는 2013년 이전의 특허에도 CPC가 부여되어 과거 개발된 특허에 대한 분류 및 정보 검색에 CPC를 활용할 수 있다. 따라서 IPC 기반의 기존 선행연구들과 CPC 기반 분석 결과의 차이가 발생함이 연구되었다[10].

III. CPC 기반 핵심 기술 분류 분석 모델

본 논문에서는 CPC를 기반으로 한 핵심 기술 분류 분석 모델을 제안한다. [그림 1]은 핵심 기술 분류 도출을 위한 구성요소의 개요를 나타내며 특허 데이터 수집, 특허 데이터 정제 및 처리, 특허 데이터 분석, 그리고 핵심 기술 분류 도출 및 비교 평가의 4단계로 구성된다.

1. 특허 데이터 수집

한국특허정보원(Kipris)에서는 특허청이 보유하고 있는 국내외 지식재산권 관련 정보를 데이터베이스로 구축하여 검색 및 열람할 수 있는 Open API를 제공한다.

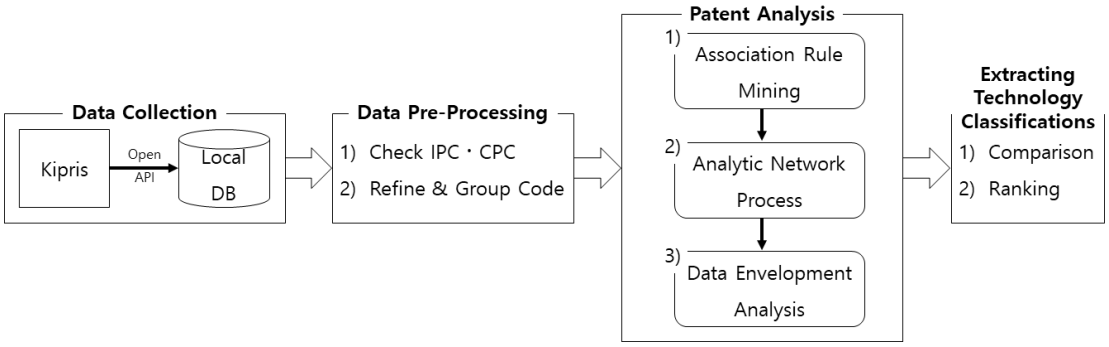

그림 1. 특허의 핵심 기술 분류 분석 모델

출원인의 핵심 기술 분류 분석을 위해 출원번호, IPC, CPC, 출원인 및 출원연도 정보 등을 수집하며 이때 IPC와 CPC를 포함하지 않는 특허는 실험 대상에서 제외한다.

2. 데이터 정제 및 처리

CPC 기술 분류는 [그림 2]와 같이 섹션(section), 클래스(class), 서브클래스(sub-class), 메인그룹(main-group), 서브그룹(sub-group), 서브 그룹 이하의 계층적 구조로 기술을 분류하며 코드 형태로 기술 분류를 표현한다. 본 연구에서 CPC 코드 기반의 출원인의 핵심 기술 분류를 도출하고 이를 분석하기 위해 [표 1]과 같이 정제 과정을 수행한다. 한 개의 특허는 한 개 이상의 CPC 분류 코드를 가지고 있으며 각 분류 코드는 서브 그룹 이하까지 상세하게 표현될 수 있다. 따라서 분류 체계 간의 상호 연관관계를 추출하기 위해 빈도수가 낮은 서브그룹 이하의 코드는 제거하여 서브그룹 수준으로 CPC 코드를 정제한다. 예를 들어, 특허 A에 포함된 'G06F 17/30067', 'G06F 15/17343'과 같은 코드는 서브그룹 수준인 'G06F 17/30', 'G06F 15/17'과 같이 정제한다. 또한, 특허 B와 같이 한 개의 특허에 중복 표현된 'G06Q 50/30'과 같은 코드를 제거한다. 그리고 특허 C의 'G05Q 50/24'와 같이 중복 코드 제거 과정을 통해 한 개의 코드만을 가지는 특허는 연관규칙이 생성되지 않으므로 실험 대상에서 제외한다.

그림 2. CPC 계층 구조

표 1. CPC 코드 정제 일부 예

특허	CPC 분류 코드
특허 A	{G06F 17/30~~067~~ → G06F 17/30, G06F 15/17~~343~~ → G06F 15/17}
특허 B	{G06F 17/30, G06Q 50/30, ~~G06Q 50/30~~}
~~특허 C~~	{G05Q 50/24, ~~G05Q 50/24~~}

3. 핵심 기술 분류 도출

제안모델을 통해 출원인의 핵심 기술 분류 분석을 위해 연관규칙 분석(ARM), 기술 분류 중요도 분석(ANP), 기술 분류의 효율성 분석(DEA) 과정을 수행한다.

3.1 연관규칙 분석 (ARM)

기술 분류들 간의 연관관계를 분석하기 위해 CPC 기반 연관규칙 분석(Association Rule Mining, ARM) 기

법을 적용한다. 연관규칙 도출에 사용되는 트랜잭션은 개별 특허로 대응하고 항목집합은 코드들의 집합으로 대응하며, 코드들의 빈발 항목집합은 지지도(Support) 값으로 수식 (1)과 같이 계산한다. 최소지지도 임계값(min_sup)을 만족한 집합들에 대해 수식 (2)와 같이 빈발 항목집합의 연관성 값인 신뢰도(Confidence)를 계산한다. 그리고 추출된 빈발 항목집합의 향상도(Lift) 값을 수식 (3)과 같이 계산하여 기술분류 코드들의 연관관계를 파악한다. 기술 분류를 나타내는 코드들의 연관관계를 계산하기 위해 2개의 분류 코드를 가진 항목집합을 대상으로 연관관계를 도출하며 항목집합 중에서 동일한 분류 코드를 가지는 쌍들의 연관관계는 항상 같으므로 1로 계산한다. [표 2]는 기술 분류 코드의 지지도 행렬을 나타낸다. T_i는 서브그룹 수준의 CPC 분류 코드를, $s(T_i, T_j)$는 지지도 값을 나타낸다.

$$Support: s(X \rightarrow Y) = \frac{n(X \cup Y)}{N} \quad (1)$$

$$Confidence: c(X \rightarrow Y) = \frac{n(X \cup Y)}{n(X)} \quad (2)$$

$$Lift: Lift(X, Y) = \frac{c(X \rightarrow Y)}{s(Y)} \quad (3)$$

출원인의 특허에 포함된 CPC 코드의 연관관계에 대한 지지도(수식 1), 신뢰도(수식 2), 향상도(수식 3)를 기반으로 [표 2]와 같은 지지도 행렬을 도출하여 CPC 코드별 연관관계 값을 ANP기법의 입력 값으로 사용한다.

표 2. 지지도 행렬

	T_1	T_2	…	T_n
T_1	1	$sup_{(T1,\ T2)}$	…	$sup_{(T1,\ Tn)}$
T_2	$sup_{(T2,\ T1)}$	1	…	$sup_{(T2,\ Tn)}$
…	…	…	1	…
T_n	$sup_{(Tn,\ T1)}$	$sup_{(Tn,\ T2)}$	…	1

3.2 기술 분류 중요도 측정 (ANP)

ANP는 다기준 의사결정 분석방법(MCDM, Multi-Criteria Decision Marking)중 하나로 각 의사결정 요소를 네트워크 구조로 만들어 의사결정 요소 간 복잡한 상호관계를 분석하여 문제를 해결하는 방법이다[11]. 본 논문에서는 ANP 기법에서 정의하고 있는 의사결정 요소를 CPC 코드들로 대응하여 핵심 기술 분류를 분석한다. 연관규칙 분석을 통해 도출된 지지도 행렬을 통해 변수 간 상관관계 분석, 네트워크 모형 구축, 그리고 쌍대 비교를 하여 분류 코드별 중요도 값을 도출한다. 이를 위해 각 연관관계 행렬에 대해 행의 합이 1이 되도록 정규화하여 초기 대행렬을 계산하고, 두 기준 집합의 가중치 값을 곱하여 가중 대행렬(Weighted Supermatrix)을 계산한다. 그리고 가중 대행렬의 각 변수 값이 일정한 값에 수렴할 때까지 무한대로 곱하여 변수 간의 중요도를 나타내는 수렴 대행렬(Limited Supermatrix)을 계산한다. 그 결과 기술 분류 코드들 간의 연관관계 기반의 기술 분류에 대한 중요도 값을 도출할 수 있다. 도출된 중요도 값은 출원인의 핵심 기술 분류를 도출하기 위한 입력 값으로 사용된다.

3.3 핵심 기술 분류 도출 (DEA)

ANP 기법을 통해 추출된 기술 분류 코드의 중요도 값은 DEA모형에 입력되어 출원인별 핵심 기술 분류를 도출한다. DEA는 조직의 효율성을 평가하기 위해 경영 및 산업공업 분야에서 주로 사용되는 기법 중의 하나로 투입변수와 산출변수를 활용하여 평가대상이 되는 의사결정단위(DMUs, Decision Making Units) 간 효율성을 상대적으로 측정하는 비모수적 접근방법이다[12]. DEA는 투입과 산출의 관계가 규모에 상관없이 일정한 비율에서 기술적 효율성을 측정하는 불변규모수익(CRS, Constant Returns to Scale)모형과, 투입과 산출의 관계가 규모에 따라 변할 때 기술적 효율성을 측정하는 가변규모수익(VRS, Variable Returns to Scale)모형이 있으며, 모형을 제안한 저자의 이름을 따서 CCR(Charnes, Cooper and Rhodes)모형과 BCC(Banker, Charnes and Cooper)모형이라고도 한다. 또한 산출을 고정시킨 상태에서 투입을 최소화하는 투입지향(input-oriented)모형과 투입을 고정시킨 상태에서 산출을 최대화하는 산출지향(output-oriented)모형으로 구분된다. 투입지향모형은 비경쟁적인 환경에서 사용되며, 산출지향모형은 경쟁적인 환경에서 사용된다. 그러므로 투입지향모형은 공공분야처럼 비경쟁적인 환경에서 사용하고, 산출

지향모형은 일반 기업과 같이 최대 성과를 내는 경쟁적인 환경에서 사용한다[13]. DEA의 결과로 도출되는 효율성 값은 0에서 1사이의 값을 가지며 값이 1에 가까울수록 대상 변수들 간의 상대적 효율성이 높으므로 이를 출원인의 핵심 기술 분류로 추천할 수 있다. 본 논문에서는 각각의 기술 분류를 나타내는 CPC 기술 분류 코드의 중요도를 종합적으로 평가하여 CPC 코드들의 상대적인 효율성을 측정하고 효율성점수가 높은 코드를 핵심 기술 분류로 추출한다. 이를 위해 ANP를 통해 계산된 출원인의 CPC 코드별 지지도, 신뢰도, 향상도를 산출요소로 사용하며 투입요소를 고정시켜 투입요소와 산출요소의 관계가 가변규모수익인 BCC모형과, 산출요소를 최대화하는 산출지향(output-oriented)모형을 이용하여 출원인별 핵심 기술 분류 코드를 도출한다.

Ⅳ. 실험 결과

4.1 실험 데이터

본 논문에서 제안하는 모델의 실험을 위해 Kipris에서 2013년 이후 발명된 특허 데이터를 수집하였으며, 최신 및 원천 기술을 주로 연구하는 연구소들에 대한 특허 기술 분류를 분석하기 위해 국가과학기술연구회 산하 정부출연연구소 중에서 특허 건수 기준, 최대, 최소, 중간 값을 가지는 3개 출원인(연구소)을 선정하였다. [표 3]은 데이터 정제 및 처리에 따른 각 출원인별 특허의 개수를 나타내며 [표 4]는 핵심 기술 분류 도출을 위해 정제 및 처리를 통해 추출된 각 출원인별 고유한 IPC와 CPC의 개수를 나타낸다.

표 3. 실험 데이터 (IPC, CPC 포함된 특허 수)

특허의 개수	IPC	CPC
출원인 A	13,541	12,679
출원인 B	1,211	1,197
출원인 C	295	256

표 4. 실험 데이터 집합의 건수(IPC, CPC)

출원인	출원인 A		출원인 B		출원인 C	
코드	IPC	CPC	IPC	CPC	IPC	CPC
개수	601	853	92	129	19	16

4.2 연관규칙 분석 (ARM)

각 출원인별 실험 데이터를 대상으로 연관규칙 분석을 수행한다. 이를 위해 출원인의 IPC, CPC 코드에 대해 지지도, 신뢰도, 향상도 값을 계산하여 연관관계 행렬을 도출한다. [표 5]는 출원인 A에 대한 연관관계 행렬의 일부로 지지도 값을 기준으로 상위 값을 가지는 4개의 CPC 코드에 대한 결과 값을 나타낸다. 예를 들어, CPC 'H04B 7/06'→'H04B 7/04'의 지지도 값 0.009를 가지며 모든 행렬의 대각 원소의 값은 1을 가진다.

표 5. 출원인 A의 CPC 상위 4개 코드의 지지도 행렬

CPC	H04B 7/06	H04B 7/04	H04L 27/26	H04W 72/12	…
H04B 7/06	1	0.009	0.001	0.001	…
H04B 7/04	0.009	1	0.002	0.001	…
H04L 27/26	0.001	0.002	1	0	…
H04W 72/12	0.001	0.001	0	1	…
…	…	…	…	…	…

4.3 기술 분류 중요도 측정 (ANP)

각 출원인의 지지도, 신뢰도, 향상도 행렬을 이용하여 CPC 기술 분류에 대한 중요도를 측정한다. [표 5]의 각 행의 합을 구해 그 행의 모든 원소를 나누어 각 행의 원소의 합이 1이 되도록 정규화하여 [표 6]과 같이 초기 대행렬을 계산한다. 기술 분류 집합 간의 가중치가 존재하지 않으므로 초기 대행렬이 곧 가중 대행렬이며, 각 행이 일정한 값으로 수렴하도록 하여 [표 7]과 같이 수렴 대행렬을 계산한다. 수렴 대행렬의 개별 원소의 값이 곧 기술 분류 중요도이며, [표 7]의 'H04B 7/06' 기술 분류 코드는 0.009의 중요도 값을 가진다.

표 6. 출원인 A의 CPC에 대한 지지도 대행렬

CPC	H04B 7/06	H04B 7/04	H04L 27/26	H04W 72/12	…
H04B 7/06	0.96	0.008	0.001	0.001	…
H04B 7/04	0.009	0.96	0.001	0.001	…
H04L 27/26	0.001	0.001	0.94	0	…
H04W 72/12	0.001	0.001	0	0.97	…
…	…	…	…	…	…

표 7. 출원인 A의 CPC에 대한 지지도 수렴 대행렬

CPC	H04B 7/06	H04B 7/04	H04L 27/26	H04W 72/12	…
H04B 7/06	0.009	0.009	0.009	0.008	…
H04B 7/04	0.009	0.009	0.009	0.008	…
H04L 27/26	0.009	0.009	0.009	0.008	…
H04W 72/12	0.009	0.009	0.009	0.008	…
…	…	…	…	…	…

4.4 핵심 기술 분류 도출 (DEA)

각 출원인의 핵심 기술 분류를 도출하기 위해 DEA 기법에서 BCC모형과 산출지향모형을 적용한다. 출원인의 모든 기술 분류 코드의 지지도, 신뢰도, 향상도를 산출요소로 대응하고, 투입요소를 10으로 고정하여 IPC와 CPC 기술 분류 각각에 대한 상대적 효율성 점수를 측정한다. 오픈 소스 R 통계 도구에서 제공하는 rDEA 패키지(version: 1.2-5)를 이용해 실험하였다[14]. [표 9-11]은 실험 대상 전체 코드 중에서 실험을 통해 효율성 값이 1인 핵심 기술 분류와 각 코드의 지지도, 신뢰도, 향상도를 나타낸다. 예를 들어, [표 9]에서 나타나는 출원인 A의 특허들에 대한 핵심 기술 분류 코드는 IPC 코드인 경우 'A61B 17/34'등이 있으며 CPC 코드의 경우 'A61B 6/50'등과 같다.

4.5 실험 평가

제안 모델을 적용하여 출원인별 핵심 기술 분류를 추출하고 IPC 분류 코드와 CPC 분류 코드에 대한 비교 평가를 수행하였다. [표 8]은 제안 모델을 통해 도출된 핵심 기술 분류 코드에 대한 수치를 나타낸다. 이는 출원인 별로 효율성이 가장 높은 1의 값을 나타내는 IPC, CPC 코드에 대한 수치이며, 전체 개수는 핵심 기술 분류로 도출된 코드의 수, 중복 개수는 IPC와 CPC에서 동시에 중복된 기술 분류이며, 고유한 개수는 중복된 코드를 제외한 IPC와 CPC 각각의 고유한 기술 분류의 개수이다. 비율은 각 코드에 대해 핵심 기술 분류로 도출된 코드 중 고유한 코드의 비율을 나타낸 것이다. 각 출원인의 IPC와 CPC 별 도출된 핵심 기술 분류들이 상이함을 [표 8]을 통해 확인할 수 있다. 또한 [표 3]과 [표 8]에 나타나는 각 출원인별 특징에서 IPC를 포함한 특허의 개수보다 CPC가 포함된 특허의 개수가 더 적었음에도 CPC로 표현된 핵심 기술 분류의 수는 IPC보다 더 많음을 확인 할 수 있다. 다시 말해서 특허 기술을 명세하고 분류함에 있어 CPC가 IPC보다 유용함을 확인 할 수 있다. 또한 [표 9, 10]의 결과를 통해 출원인 A, B에 대한 기술 분류 섹션 중 IPC를 통해 표현되지 못한 섹션(C, Y 섹션)이 CPC 기술 분류를 통해 나타남을 확인할 수 있다. 이는 곧, 새로운 분야의 기술을 포함하고 있는 특허에 사용된 기술 중 IPC로 분류되지 않았던 기술이 CPC 분류 코드를 통해 분류되는 것을 확인할 수 있다.

표 8. IPC와 CPC의 핵심 기술 분류 수치 비교

출원인	출원인 A		출원인 B		출원인 C	
코드	IPC	CPC	IPC	CPC	IPC	CPC
전체 개수	29	46	18	24	9	11
중복 개수	7		17		8	
고유한 개수	22	39	1	7	1	3
비율(%)	75.9	79.6	5.6	29.2	11.1	27.3

V. 결론 및 향후 연구

최근 정보통신 기술의 발전과 함께 관련된 시장의 기술 변화가 급속도로 진행됨에 따라 기업은 특허 발명을 통해 시장에서의 경쟁력 확보에 많은 노력을 기울이고 있다. 기존 시장 분석 및 경쟁 기술 등의 분석을 위해 특허 분석 연구가 중요하게 되었다. 그렇지만 재현율이 낮은 특허 문헌의 특성에 따라 분석 정확도 향상을 위한 연구가 수행되었고, 그 중의 하나로 코드 기반의 특허 분류 및 검색 연구가 진행되었다. 분류체계 기반의 특허 분석 연구 또한 제안되었으나 CPC 기술 분류에 대한 연구가 미흡하다. 본 논문에서는 CPC 기술 분류 체계를 이용하여 특허의 핵심 기술 분류 분석 모델을 제안하고, IPC 기반의 기존 연구 방법과 비교하여 CPC 기술 분류 체계의 유용성 및 활용성을 검증하였다. 이를 위해 세 개의 출원인을 대상으로 IPC와 CPC 기술 분류 각각에 대한 핵심 기술 분류를 도출하였다.

표 9. 출원인 A의 IPC와 CPC 비교 결과

IPC	기술 분류 효율성			CPC	기술 분류 효율성		
	지지도	신뢰도	향상도		지지도	신뢰도	향상도
A61B 17/34	0.5	0.5	0.485714286	A61B 17/34	0.5	0.5	0.515151515
A61B 6/03	0.5	0.5	0.342105263	A61B 6/50	0.5	0.5	0.558823529
A61N 1/05	0.5	0.5	0.5	A61B 6/52	0.5	0.5	0.441176471
A61N 1/36	0.5	0.5	0.5	A61N 1/05	0.5	0.5	0.47826087
A63B 71/06	0.5	0.5	0.423076923	A61N 1/36	0.5	0.5	0.52173913
B60W 40/08	0.5	0.5	0.483870968	A61N 5/10	0.5	0.5	0.521126761
B60W 50/08	0.5	0.5	0.516129032	C09K 11/77	0.5	0.5	0.48
G01N 33/49	0.5	0.5	0.644444445	G01N 33/68	0.5	0.5	0.571428571
G01N 35/08	0.5	0.5	0.355555556	G01R 29/08	0.5	0.5	0.581818182
G01S 19/20	0.5	0.5	0.55	G01R 29/10	0.5	0.5	0.418181818
G01S 19/23	0.5	0.5	0.45	G02F 1/15	0.5	0.5	0.6
G06Q 20/16	0.5	0.5	0.441176471	G06K 9/34	0.5	0.5	0.555555556
G06Q 20/32	0.5	0.5	0.558823529	G06N 3/04	0.5	0.5	0.393939394
G08B 3/10	0.5	0.5	0.483870968	G06N 3/08	0.5	0.5	0.606060606
G08B 5/22	0.5	0.5	0.516129032	G06T 1/20	0.5	0.5	0.564102564
G08C 23/04	0.5	0.5	0.407407407	G06T 1/60	0.5	0.5	0.435897436
G09B 5/02	0.5	0.5	0.416666667	G08B 3/10	0.5	0.5	0.515151515
H01S 5/02	0.5	0.5	0.647058824	G08B 5/22	0.5	0.5	0.484848485
H01S 5/06	0.5	0.5	0.352941177	G10L 17/02	0.5	0.5	0.485714286
H03F 1/32	0.5	0.5	0.631578947	G10L 17/04	0.5	0.5	0.514285714
H03F 3/24	0.5	0.5	0.368421053	H01G 9/20	0.5	0.5	0.446428571
H04B 1/10	0.5	0.5	0.5	H01J 35/06	0.5	0.5	0.555555556
H04B 1/16	0.5	0.5	0.5	H01J 35/14	0.5	0.5	0.444444444
H04H 20/02	0.5	0.5	0.4	H01L 41/11	0.5	0.5	0.392857143
H04H 20/51	0.5	0.5	0.428571429	H01P 1/22	0.5	0.5	0.541666667
H04H 20/59	0.5	0.5	0.6	H02N 2/18	0.5	0.5	0.607142857
H04N 7/20	0.5	0.5	0.571428571	H03H 17/02	0.5	0.5	0.571428571
H04Q 9/04	0.5	0.5	0.592592593	H03H 17/06	0.5	0.5	0.428571429
H04R 19/04	0.5	0.5	0.535714286	H03H 7/24	0.5	0.5	0.458333333
				H03M 1/06	0.5	0.5	0.571428571
				H04H 20/51	0.5	0.5	0.423076923
				H04L 12/40	0.5	0.5	0.6
				H04L 47/11	0.5	0.5	0.227272727
				H04L 47/12	0.5	0.5	0.772727273
				H04L 49/30	0.5	0.5	0.397260274
				H04L 49/35	0.5	0.5	0.602739726
				H04M 3/42	0.5	0.5	0.563636364
				H04N 13/33	0.5	0.5	0.48
				H04N 13/34	0.5	0.5	0.52
				H04N 7/20	0.5	0.5	0.576923077
				H04W 4/16	0.5	0.5	0.436363636
				H04W 4/90	0.5	0.5	0.615384615
				H04W 76/50	0.5	0.5	0.384615385
				Y02B 20/18	0.5	0.5	0.52
				Y02B 70/32	0.5	0.5	0.433333333
				Y02E 10/54	0.5	0.5	0.553571429

표 10. 출원인 B의 IPC와 CPC 비교 결과

IPC	기술 분류 효율성			CPC	기술 분류 효율성		
	지지도	신뢰도	향상도		지지도	신뢰도	향상도
A47J 47/02	0.5	0.5	0.4444444	A01G 31/02	0.5	0.5	0.5
A61K 35/61	0.5	0.5	0.5	A23D 9/02	0.5	0.5	0.5
A61K 38/01	0.5	0.5	0.5	A23L 2/68	0.5	0.5	0.555555556
B01J 13/02	0.5	0.5	0.6363636	A47J 47/02	0.5	0.5	0.444444445
B01J 19/10	0.5	0.5	0.5	A61K 35/61	0.5	0.5	0.5
B07C 5/34	0.5	0.5	0.6	A61K 38/01	0.5	0.5	0.5
B07C 5/36	0.5	0.5	0.4	A61K 8/63	0.5	0.5	0.5
B65D 77/04	0.5	0.5	0.3636364	B01J 13/02	0.5	0.5	0.583333333
B65D 81/18	0.5	0.5	0.6363636	B01J 19/10	0.5	0.5	0.5
C07K 14/78	0.5	0.5	0.5	B07C 5/34	0.5	0.5	0.6
C08J 3/22	0.5	0.5	0.5555556	B07C 5/36	0.5	0.5	0.4
C08J 5/18	0.5	0.5	0.5714286	B65D 77/04	0.5	0.5	0.4
C09K 5/06	0.5	0.5	0.3636364	B65D 81/18	0.5	0.5	0.6
F24F 13/28	0.5	0.5	0.4444444	C07K 14/78	0.5	0.5	0.5
G01N 21/55	0.5	0.5	0.5555556	C08J 3/22	0.5	0.5	0.555555556
G01N 21/59	0.5	0.5	0.4444444	C08J 5/18	0.5	0.5	0.571428571
G06Q 30/02	0.5	0.5	0.4545455	C08L 23/02	0.5	0.5	0.5
G06Q 30/06	0.5	0.5	0.5454545	C09K 5/06	0.5	0.5	0.416666667
				F24F 13/28	0.5	0.5	0.5
				G01N 21/01	0.5	0.5	0.555555556
				G01N 21/55	0.5	0.5	0.636363636
				G06Q 30/02	0.5	0.5	0.454545455
				G06Q 30/06	0.5	0.5	0.545454546
				Y02P 60/21	0.5	0.5	0.5

표 11. 출원인 C의 IPC와 CPC 비교 결과

IPC	기술 분류 효율성			CPC	기술 분류 효율성		
	지지도	신뢰도	향상도		지지도	신뢰도	향상도
G06F 9/38	0.5	0.5	0.545454545	G06F 13/40	0.5	0.5	0.5
G06F 9/46	0.5	0.5	0.454545455	G06F 21/31	0.5	0.5	0.5555556
G06N 3/02	0.5	0.5	0.444444445	G06F 9/38	0.5	0.5	0.5
G06N 3/08	0.5	0.5	0.533333333	G06F 9/46	0.5	0.5	0.5
G08G 1/01	0.5	0.5	0.555555556	G06N 3/02	0.5	0.5	0.4444444
H04L 12/24	0.5	0.5	0.642857143	G06N 3/08	0.5	0.5	0.5333333
H04L 12/46	0.5	0.5	0.357142857	G08G 1/01	0.5	0.5	0.5555556
H04N 21/23	0.5	0.5	0.6	H04L 12/46	0.5	0.5	0.5
H04N 21/24	0.5	0.5	0.4	H04L 41/08	0.5	0.5	0.5
				H04N 21/23	0.5	0.5	0.6
				H04N 21/24	0.5	0.5	0.4

실험 결과 출원인에서 CPC 기술 분류를 통해 IPC 분류에서 도출되지 않았던 새로운 기술 분류가 도출됨을 확인하였다. 또한 제안 모델을 통해 각 출원인별 핵심 기술 분류를 도출하여 발명된 특허에 사용된 기술들을 대표하는 대표 기술 분야를 도출하였다. 따라서 출원인의 핵심 기술 분류 코드에는 기존의 IPC 기술 분류로 분류되지 않았던 새로운 기술 분류가 포함될 수 있으며 CPC 기술 분류가 유용함을 재확인 하였다. 추후 제안 모델의 분석 결과를 통해 출원인의 대표적인 기술을 분류하고 대표 기술을 상세하게 도출하는 정보로 사용할 수 있을 것이다. 본 논문의 한계는 제안 모델을 기반으로 출원인별 핵심 기술 분류를 도출하였지만 제목, 청구항 등에서 나타나는 상세 기술들과 연계가 부족하다. 따라서 텍스트 마이닝 기법을 이용하여 CPC 기술 분류와 기술명 매핑을 통한 핵심 기술도출 연구를 향후과제로 한다.

참 고 문 헌

[1] Z. Griliches, A. Pakes, and B. H. Hall, *The Value of Patents as Indicators of Inventive Activity*, Economic Policy and Technological Performance, Cambridge: Cambridge University Press, contents Pub, 1987.

[2] 전기억, "특허 빅데이터를 활용한 R&D 전략과 특허분쟁 대응," 전자공학회지, 제41권, 제9호, pp.64-72, 2014.

[3] 남성현, 김광수, "유사특허의 청구항 기반 특허분석법," 대한산업공학회 추계학술대회 논문집, pp.2248-2252, 2015.

[4] 노태연, 정유진, 윤병운, "LSA와 연관규칙 마이닝 기반의 특허 서술형태 분석을 통한 선택적 기술정보 도출," 대한산업공학회 춘계공동학술대회 논문집, pp.2949-2966, 2016.

[5] 안성준, 김종찬, 이준혁, 박상성, 장동식, "특허정보를 활용한 기업 기술동향분석," 한국지능시스템학회 학술발표 논문집, 제25권, 제1호, pp.13-14, 2015

[6] 이원상, 손소영, "빅데이터 기술을 활용한 대용량 삼극특허 분석 기반의 기술융복합 패턴 예측," 대한산업공학회 추계학술대회 논문pp.1153-1170, 2013.

[7] 김철현, "핵심 기술 파악을 위한 특허 분석 방법: 데이터 마이닝 및 다기준 의사결정 접근법," 대한안전경영과학회지, 제16권, 제1호, pp.213-220, 2014.

[8] 김현우, 김종찬, 이준혁, 박상성, 장동식, "핵심 기술 및 특허 추출을 위한 IP 마이닝에 관한 연구," 한국지능시스템학회 논문지, 제25권, 제4호, pp.392-397, 2015.

[9] 이경래, 이준석, 박상성, "IPC코드 기반의 기술이전 특허의 기술관계," 한국지능시스템학회 학술발표 논문집, 제25권, 제2호, pp.181-182, 2015.

[10] 강지호, 김종찬, 이준혁, 박상성, 장동식, "IoT와 Wearables 기술융합을 위한 특허동향분석," 한국지능시스템학회 논문지, 제25권, 제3호, pp.306-311, 2015.

[11] T. L. Saaty, "Decision making with the analytic hierarchy process," J. of services sciences, Vol.1, No.1, pp.83-98, 2008.

[12] 서창적, 이정식, "DEA 모형을 활용한 전자소매점포의 상대적 효율성 평가 - S전자 소매점포를 대상으로," 서비스경영학회지, 제15권, 제1호, pp.243-268, 2014.

[13] 조남권, 김규환, 이석진, "DEA를 통한 중소·중견기업의 R&D 효율성 분석," 지식재산연구, 제13권, 제2호, pp.207-236, 2018.

[14] J. Simm, G. Besstremyannaya, and M. J. Simm, Package 'rDEA', 2016. https://github.com/jaak-s/rDEA

저 자 소 개

채 수 현(Soo-Hyeon Chae) 준회원

- 2013년 3월 ~ 현재 : 군산대학교 소프트웨어융합공학과 학부생

<관심분야> : 텍스트 마이닝, 빅데이터 분석

김 장 원(Jangwon Gim) 정회원

- 2005년 2월 : 상명대학교 컴퓨터소프트웨어공학과(공학사)
- 2008년 2월 : 고려대학교 컴퓨터학과(이학석사)
- 2012년 8월 : 고려대학교 컴퓨터·전파·통신공학과(공학박사)
- 2013년 3월 ~ 2017년 3월 : 한국과학기술정보연구원 선임연구원
- 2017년 4월 ~ 현재 : 국립군산대학교 소프트웨어융합공학과 조교수

<관심분야> : 텍스트 마이닝, 빅데이터 분석, 메타데이터, 공공데이터