

Social network analysis of keyword community network in IoT patent data

Do Hyun Kim^a · Hyon Hee Kim^a · Donggeon Kim^a · Jinnam Jo^{a,1}

^aDepartment of Statistics and Information Science, Dongduk Women's University

(Received April 1, 2016; Revised May 16, 2016; Accepted May 23, 2016)

Abstract

In this paper, we analyzed IoT patent data using the social network analysis of keyword community network in patents related to Internet of Things technology. To identify the difference of IoT patent trends between Korea and USA, 100 Korea patents and 100 USA patents were collected, respectively. First, we first extracted important keywords from IoT patent abstracts using the TF-IDF weight and their correlation and then constructed the keyword network based on the selected keywords. Second, we constructed a keyword community network based on the keyword community and performed social network analysis. Our experimental results showed while Korea patents focus on the core technologies of IoT (such as security, semiconductors and image process areas), USA patents focus on the applications of IoT (such as the smart home, interactive media and telecommunications).

Keywords: IoT patent analysis, patent keyword network, community detection, keyword community network

1. 서론

최근 다양한 인터넷상의 문서 수집이 용이해짐에 따라서 텍스트 문서를 분석하여 문서를 분류하거나 요약하는 등 텍스트 분석에 대한 관심이 높아지고 있다. 많은 관심을 받고 있는 트위터나 페이스북과 같은 소셜 미디어 분석의 경우, 사용자들이 직접 생성한 콘텐츠가 정제되지 않은 정보들을 포함하고 있기 때문에 과학 기술 동향 예측과 같이 정확한 정보를 필요로 하는 영역에서는 분석 데이터로 사용하기에 적합하지 않다. 이에 반하여 특허 정보는 출원 날짜, 특허 종류, 특허 요약문 등 정제된 형태로써 특정 기술 분야의 핵심 정보를 포함하므로 소셜 미디어상의 텍스트들과는 달리 분석 결과의 신뢰성을 보장할 수 있다. 또한 어떠한 이슈가 대중에게 널리 퍼지기 전에 선행되는 단계가 기술자와 개발자들의 특허 등록이라는 점에 주목한다면 등록된 최신의 특허 정보들을 분석하는 것은 현재의 기술 동향을 파악하고 향후 기술을 예측하는데 필수적이라고 할 수 있다.

본 연구에서는 위와 같은 특허 분석의 중요성에 기반 하여 최근 이슈가 되고 있는 사물인터넷(internet of things) 관련 특허에 대해 텍스트 분석을 실시하였다. 특허 정보 포털 사이트인 KIPRIS(<http://www.kipris.or.kr>)로부터 “Internet of Things”를 검색어로 하여 2015년을 기준으로 가장 최근에 등록

This research was supported by the Dongduk Women's University grant 2015.

¹Corresponding author: Department of Statistics and Information Science, Dongduk Women's University, 60 Hwarang-Ro 13-gil, Seongbuk-Gu, Seoul 02748, Korea. E-mail: jinnam@dongduk.ac.kr

혹은 출원된 특허를 한국 특허와 미국 특허로 나누어 각각 100개씩을 수집하였다. 먼저 특허 문서 집합에서 중요 단어를 추출하여 키워드 네트워크를 구축한 후 한국과 미국의 특허 키워드 네트워크 구조를 분석하였다. 다음으로 구축된 키워드 네트워크에 커뮤니티 탐지 알고리즘을 적용하여 키워드 커뮤니티를 파악하였다. 각 커뮤니티에 속한 단어들을 바탕으로 해당 기술을 정의하고 이 기술들 간의 연관성을 파악하기 위하여 키워드 커뮤니티 네트워크를 재구축한 다음 소셜 네트워크 분석을 실시하였다.

특히 본 연구는 특허 키워드 네트워크를 구축하고 키워드 네트워크의 커뮤니티 탐지 결과로 생성된 키워드 커뮤니티 네트워크 분석에 초점을 둔다. Choi와 Hwang (2014)은 키워드 네트워크를 구축한 다음 커뮤니티 분석을 시행하여 기술 간의 관계를 예측하였다. 특허 데이터를 분석에 이용한다는 점과 커뮤니티 탐지를 이용한다는 점이 본 연구와 유사하지만 본 연구는 커뮤니티 탐지 수행 후 생성된 키워드 커뮤니티들을 각각 한 개의 대표 기술로 보고 기술 간의 관계를 분석하기 위해서 새로운 키워드 커뮤니티 네트워크를 구축하여 분석하는 점에 주목한다. 즉, 키워드 커뮤니티 네트워크를 재구성하고 이에 대해 소셜 네트워크 분석을 적용하여 타 기술 분야와 연결이 강한 기술 분야 및 강한 연결 관계를 갖는 기술 분야들을 찾아내었다.

본 연구는 키워드 추출 및 키워드 네트워크 구축 부분과 커뮤니티 탐지 및 키워드 커뮤니티 네트워크 분석의 두 부분으로 구성된다. 먼저 특허의 영문 초록을 바탕으로 키워드 네트워크를 구축하였다. 이 과정에서 텍스트의 중요 키워드를 찾는 것이 성과를 좌우하는 것이라는 점에 주목하여 TF-IDF(Term Frequency-Inverse Document Frequency) 가중치를 기준으로 중요 키워드를 추출하였다. 선정된 키워드에 연관규칙을 이용하여 상관관계가 높은 동시에 TF-IDF 가중치가 높은 키워드들을 추출하여 새로운 데이터 셋을 형성하였다. 이 작업을 통해서 수집한 모든 특허 정보를 대상으로 중요 주제어들을 선정하고 전체 특허 데이터에 대한 키워드 네트워크를 형성하였다. 다음으로 생성된 전체 네트워크에 대하여 최대 연결된 서브 네트워크를 추출한 후 추출된 네트워크에 대해 커뮤니티 탐지를 수행하였다. 그리고 커뮤니티 탐지 결과를 통해 생성된 각 커뮤니티들을 하나의 노드로 전환하여 그 노드들 간의 연결 관계를 분석하는 키워드 커뮤니티 네트워크 분석을 실시하였다.

분석을 위해 사물 인터넷 관련 특허 중 하드웨어 분야 특허만을 대상으로 한국 특허와 미국 특허를 각각 100개씩 수집하였다. 하드웨어 분야 특허는 특허의 종류를 구분하는 International Patent Classification(IPC) 코드로 H그룹에 속하는 특허를 뜻한다. 한국 특허의 경우 2015년도에 출원 및 등록된 특허가 114개이고, 미국 특허의 경우 3,553개이므로 두 국가의 그래프 구조를 비교하기 위해서 최신 출원 및 등록 특허 100개를 각각 수집하여 분석하였다. 전체 네트워크 생성 후 가장 큰 네트워크만을 추출하여 분석에 이용한 결과 한국과 미국 두 네트워크 안에서 다른 노드와 많이 연결되어 있는 즉, 상호작용이 활발한 중요 키워드를 도출할 수 있었다. 네트워크 구조 분석 결과 한국 특허와 미국 특허간의 연결 중심성, 근접 중심성 계수나 커뮤니티 밀도 그리고 군집 계수의 차이는 크게 나타나지 않았다. 하지만 미국의 매개 중심성 평균값이 한국보다 약 3배 크다는 결과를 얻었다. 매개 중심성은 서로 다른 두 노드를 연결하는 역할을 정의하는 척도로서 매개 중심성이 크다고 하는 것은 다른 노드들의 상호작용에 큰 영향력을 발휘한다는 것을 의미한다.

커뮤니티 탐지 결과 한국의 경우 46개, 미국의 경우 44개의 커뮤니티가 산출되었다. 본 연구에서는 산출된 46개와 44개의 커뮤니티에서 각각 13개와 10개의 커뮤니티를 추출하여 커뮤니티 간 네트워크를 형성하고 커뮤니티 간의 연결 형태와 연결 정도를 분석하였다. 키워드 커뮤니티 네트워크 분석 결과에 따르면 한국의 경우 보안 기술 분야 그리고 미국의 경우 통신 분야의 특허가 두드러졌다. 또한, 한국의 경우는 특히 보안과 반도체 분야 그리고 보안과 이미지 프로세스 사이에 강한 연결이 나타났다. 미국의 경우 통신기술에 관련된 분야가 네트워크상에서 중추적인 역할을 했으며 통신기술 관련 특허와 대화형 매체 그리고 통신기술 관련 특허와 스마트 홈의 연결이 눈에 띄었다.

본 논문은 다음과 같이 구성된다. 2절에서 특허 분석과 네트워크 분석, 커뮤니티 탐지에 대한 관련 연구들을 살펴보고, 3절에서 데이터 구성과 연구 방법에 대해 자세히 설명한다. 4절에서 특허 분석 결과를 서술하고 마지막으로 5절에서 결론을 제시한다.

2. 선행 연구

Park (2014)은 발전 가능성이 큰 기술 발견을 위하여 특허 수의 점유 증가 값, 평균과 같은 수치형 데이터를 분석한 반면에 본 연구는 특허의 초록문에 텍스트 마이닝을 적용하여 분석 데이터를 구축하였다. Kang 등 (2015)은 특허 코드 간의 지지도와 향상도를 이용한 연관규칙을 사용하여 기술 간의 융합 추세를 파악하였다. 반면 본 연구에서는 분석 데이터 구축 시에 TF-IDF 알고리즘과 연관규칙 기법을 사용한 후에 얻은 새로운 데이터 셋에 대하여 네트워크 분석과 커뮤니티 분석을 통해서 유망한 기술을 파악하였다.

특히 특허의 초록문은 특허의 내용을 요약해놓은 핵심정보이기에 특허정보에 대해 텍스트 마이닝을 적용하여 주제를 추출하는 연구는 결과의 활용도가 주목된다. Choi (2013)은 본 연구에서 주제어 추출에 대해서 TF-IDF 방법을 이용하여 추출하는 것과 달리 장갑분야에 대해 출현 빈도수를 기반 하여 주제를 추출하고 주제어 간의 비계층적 군집분석 방법을 통해 군집별 기술적 특성을 파악하였다. You 등 (2015)의 연구는 본 연구와 마찬가지로 중요 키워드를 추출하는 데에 TF-IDF 기법의 사용을 주목하였다. 하지만 You 등 (2015)이 각 소설의 주제어를 찾는 방법의 성능에 대해 주목하였다면 본 연구는 전체적인 연결 관계 파악을 위해 수집한 모든 특허 정보를 대상으로 중요 주제어들을 선정하고 전체 특허 데이터에 대한 네트워크를 형성하여 분석하였다.

Choi와 Hwang (2014)은 네트워크 분석과 커뮤니티 탐지에 관한 연구로서 키워드 네트워크 특허 분석을 통해 어떠한 키워드가 네트워크상에서 두드러진 역할을 하며 키워드들의 경향이 어떠한지에 대해 주목하여 분석하였다. Kim 등 (2013)은 계층 군집화 방법을 이용하여 모바일 상호작용 데이터로부터 사용자 커뮤니티를 발견하는 방법을 제안하였다. 본 연구 또한 커뮤니티 탐지 분석을 실시하였지만 커뮤니티 탐지에 그치지 않고 이를 이용하여 커뮤니티를 노드로 전환한 후에 네트워크를 형성하여 커뮤니티의 특징과 커뮤니티 간의 관계를 도출하였다. 분석 결과 커뮤니티의 의미와 커뮤니티 네트워크가 나타내는 형태 및 그 기술 간의 연관 관계를 분석할 수 있었다.

3. 데이터 구성과 연구 방법

본 연구에서는 특허 검색 정보 서비스인 KIPRIS(<http://www.kipris.or.kr>)에서 제공하는 특허 정보를 수집하여 사용하였다. 한국과 미국의 사물 인터넷 관련 특허를 수집하기 위하여 “Internet of things”를 검색어로 사용하였으며 검색 조건을 전자 계열인 H그룹에 해당하는 H04와 H06에 대해 미국과 한국 2015년을 기준으로 등록되었거나 출원된 최신의 특허데이터를 각각 100개씩 총 200개의 자료를 수집하였다. 수집한 특허 자료에서 발명의 핵심 내용이 요약문에 나타나기 때문에 본 연구에서는 영문 초록만을 텍스트 분석에 이용하였다.

본 연구에서는 각 요약문을 분석함에 있어서 동사, 부사 그리고 형용사의 경우 명사보다 중요도가 떨어진다고 판단되어 각 문서에 대하여 명사추출을 시행하였다. 이때 프로그래밍 언어 R의 openNLP 패키지를 사용하여 명사추출을 시행하였다. TF-IDF 가중치 방법을 이용하여 가중치 값이 큰 상위 300개의 키워드를 선정하였다. Manning 등 (2008)에 따르면 TF-IDF 가중치는 다음 수식 (3.1)과 같이 나타낼 수 있다. 문서 j 에 속하는 키워드 i 의 가중치 $w_{i,j}$ 는 tf 값과 df 값의 곱이다. 여기서 $tf_{i,j}$ 는 문서 j 에 나

타난 키워드 i 의 빈도수이고, $df_{i,j}$ 는 키워드 i 를 포함하는 문서수의 역수에 로그를 취한 값이다.

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_{i,j}} \right). \quad (3.1)$$

단어의 빈도수만을 고려하면 가장 빈도수가 높은 키워드는 대부분 문서 집합에서 일반적으로 사용되는 키워드인 경우가 많다. 이러한 키워드를 제거하고 구체적인 키워드를 추출하기 위해서 키워드가 문서 집합에 등장한 횟수의 역수를 곱하면 문서 집합 전체에 많이 등장한 일반적인 키워드는 가중치 값이 줄어들게 되어 중요도가 높은 단어들을 선정할 수 있다.

그 후에 연관 규칙 알고리즘을 이용하여 추출된 300개 각 키워드에 연관성이 높은 키워드를 추출하였다. 그 후에 두 키워드 집합 간의 공통된 키워드 집합을 추출하였고 그와 연관성이 높은 키워드를 구해 새로운 데이터 세트를 형성하였다. 즉 생성된 데이터는 TF-IDF가 상대적으로 높은 점수인 동시에 다른 키워드와의 연관성도 높은 키워드로 구성된 것을 의미한다. 그 결과 한국은 2,476개의 행을 가진 데이터, 미국은 2,438개의 행을 가진 데이터가 생성되었다. 텍스트 마이닝을 위해 사용된 언어는 R이며, R에서 제공하는 tm 패키지를 이용하여 텍스트 분석을 실시하였다.

키워드 네트워크 구축과 분석을 위해서 R에서 제공되는 igraph 패키지를 사용하였다. TF-IDF 가중치가 높은 상위 300개의 중요 키워드를 추출하고, 이 중요한 키워드간 상관관계가 높은 연관 키워드를 추출하여 최종 키워드를 선정하였다. 중요 키워드 i 와 연관 키워드 j 간의 관계를 $\langle i, j \rangle$ 의 순서쌍으로 표현한 방향 그래프로부터 생성된 최대 컴포넌트를 분석 대상으로 정하였다. 여기서 컴포넌트란 서로 연결된 서브 그래프를 의미한다. 먼저 키워드 네트워크 분석 단계에서는 한국과 미국 간의 거시적인 특허 경향을 파악하기 위해서 연결 중심성, 매개 중심성, 그리고 근접 중심성의 centralization 값을 비교하였으며, density와 clustering coefficient 값도 비교하였다. 다음으로, 해당 컴포넌트를 커뮤니티 탐지 방법으로 세분화하고 개별 커뮤니티를 하나의 노드로 간주하여 키워드 커뮤니티를 정의한 다음 키워드 커뮤니티 네트워크를 구축하고 분석하였다. 각 키워드 커뮤니티 노드는 해당 키워드들을 해석하여 한 개의 대표 기술을 나타낸다. 키워드 커뮤니티 노드 간에 공통으로 등장하는 키워드들을 가중치로 하여 네트워크 분석을 하여 기술 간의 연결 관계를 찾을 수 있었다.

4. 분석 결과

4.1. 키워드 네트워크 분석

본 절에서는 제 3장에서 추출된 키워드를 각 노드로 하고 정의된 상관 계수 이상으로 동시에 사용된 키워드를 서로 연결하여 키워드 네트워크를 구성한 다음, 분석 결과를 제시하였다. 본 연구에서는 커뮤니티 간의 네트워크를 생성하여 커뮤니티 간 연결 양상을 확인하기 위한 연구이므로 고립되어 있는 지역적인 커뮤니티들을 제거하고 최대 컴포넌트에 대하여 분석을 실시하였다. Figure 4.1(a)와 (b)는 각각 한국과 미국의 키워드 네트워크의 최대 컴포넌트를 추출하여 시각화한 결과이다. 각 네트워크 그래프에 나타나는 키워드들의 크기는 연결 중심성 계수의 크기에 비례하여 나타난다. 다른 노드와 얼마나 많은 연결이 이루어졌는지 알 수 있는 연결 중심성은 그 값이 클수록 키워가 다른 키워드와 상호작용이 활발함을 의미한다. 따라서 해당 키워드를 전망 있는 분야에 관련된 키워드로 해석할 수 있다.

Table 4.1은 키워드 네트워크의 구조를 파악하기 위한 그래프 레벨 분석 결과이다. 먼저 소셜 네트워크 분석에서 가장 기본적으로 사용되는 척도 (Wasserman과 Faust, 1994)인 연결정도 중심성(degree centrality), 매개 중심성(betweenness centrality), 그리고 근접 중심성에 대하여 중앙성(centralization)을 살펴보았다. 중앙성이란 전체 네트워크의 형태가 중앙에 집중되어 있는지를 알 수 있는 척도로서 네트워크의 구조적 특성 파악이 가능하다. 다음으로 네트워크의 전역적인 특성인 밀도(density)를 측정하였

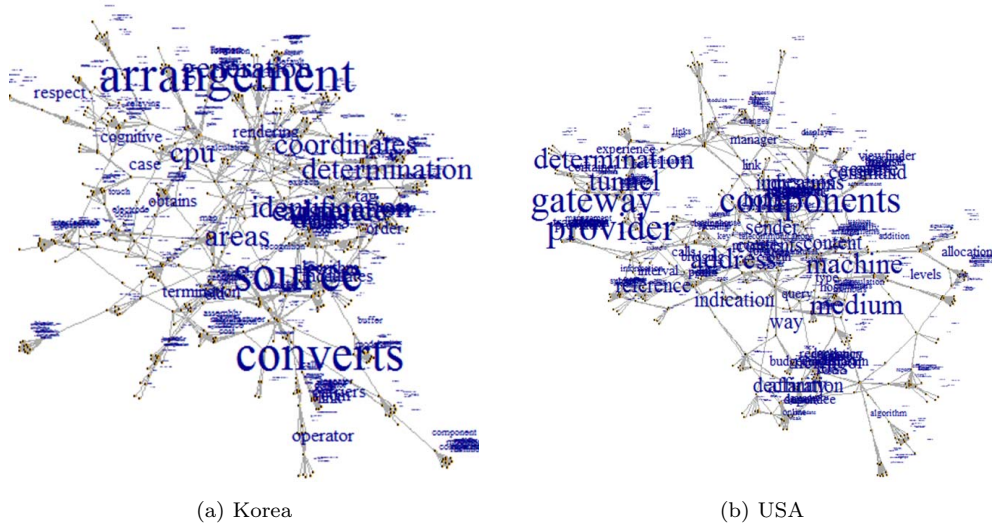


Figure 4.1. The largest patent keyword network.

Table 4.1. Graph-level analysis of the largest patent keyword network

Group	Centralization			Density	Clustering coefficient
	Degree	Closeness	Betweenness		
Korea	0.0099	0.0007	0.0052	0.0060	0.5107
USA	0.0100	0.0009	0.0152	0.0070	0.5798

으며, 밀도는 연결 가능한 엣지들의 수와 실제 연결된 엣지들의 수의 비율로서 네트워크가 얼마나 뭉쳐 있는지 알 수 있다. 마지막으로 군집 계수는 전체 네트워크에 있는 닫힌 세 노드쌍의 수를 생성 가능한 모든 세 노드쌍의 수로 나눈 값으로 키워드간의 추이성을 알 수 있다.

한국과 미국의 연결 중심성, 근접 중심성 계수나 커뮤니티 밀도 그리고 지역적 특성을 나타내는 군집 계수의 차이는 크게 나타나지 않았다. 하지만 예외적으로 미국의 매개 중심성(betweenness centrality)이 한국보다 상당히 높다는 것을 알 수 있다. 매개 중심성은 2개 노드를 연결하는 다리의 역할, 부 네트워크를 연결하는 역할 (Huh, 2012)을 의미한다. 즉, 한국보다 큰 매개 중심성 값을 가진 미국에서 노드들이 자신을 제외한 다른 노드들의 상호작용에 행사하는 영향력이 더 큰 것으로 판단된다. 미국의 경우, 매개 중심성이 큰 값의 단어들은 대체적으로 웹 분야, 연결과 상관이 있는 네트워크 관련 분야이다.

4.2. 키워드 커뮤니티 네트워크 분석

4.1절에서 구축된 키워드 네트워크에 커뮤니티 탐지를 수행하여 각 커뮤니티를 한 개의 노드로 하고 커뮤니티를 구성하는 공통 키워드의 수를 연결 가중치로 하여 새로운 네트워크를 구축하였다. 이를 키워드 커뮤니티 네트워크라고 하며 본 절에서는 키워드 커뮤니티 네트워크에 대한 분석 결과를 제시한다. 커뮤니티 탐지를 위해서 본 연구에서는 닫힌 길 커뮤니티(walktrap community) 알고리즘을 이용하였다. 닫힌 길 커뮤니티는 특정한 두 노드에서 시작된 2명의 독립적인 임의보행자가 일정시간 후에 특정 노드에 머물 확률이 얼마나 비슷한지를 계산하여 네트워크 내부의 커뮤니티를 찾아내는 방법이다 (Pons와 Latapy, 2005). 한국의 경우 46개의 커뮤니티가 산출되었으며 미국의 경우는 44개의 커뮤니티가 생성되었다.

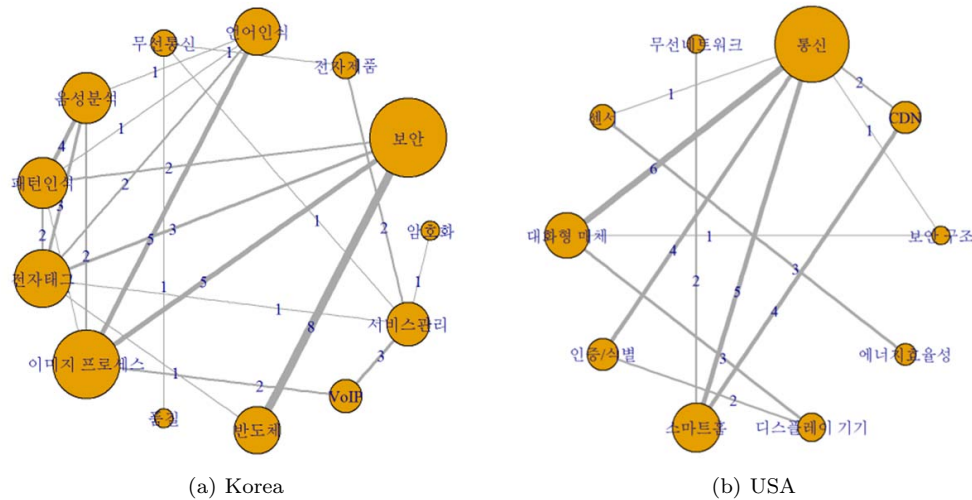


Figure 4.2. Keyword Community Network.

Algorithm 4.1. Construction of Keyword Community Network

1. Let i , which contains n rows data denotes the first column of keyword network data and j is last.
Key word network data is consisted of two columns $\langle i, j \rangle$ are formed by direction from i to j . i is critical keyword and j is correlated keyword.
 2. For $k = 1, \dots, n$:
 - (a) Compare i_k 's word to words of communities extracted by the number of words in the community.
 - (b) Allocate the name of community, which has i_k to i_k .
If i_k is included in non-extracted community then allocate "others".
 3. Repeat Step 2 on all j_k .
 4. Return the community network by using the new network data set, which is consisted of communities' name but "others".
-

본 연구에서 커뮤니티 탐지 결과를 키워드 커뮤니티 네트워크로 형성하기 위해 사용한 알고리즘은 Algorithm 4.1과 같다. $\langle i, j \rangle$ 인 2열로 이루어진 네트워크 데이터는 i 에서 j 로의 방향 네트워크를 형성하며 i 는 중요 키워드이고 j 는 연관 키워드를 의미한다. 각 열에 포함된 단어들을 추출된 커뮤니티에 포함된 단어들과 비교하여 해당 단어에 대응되는 단어가 속한 커뮤니티를 할당해주는 것이다.

예를 들어 "Internet"이라는 단어가 추출된 커뮤니티 A안에 있는 단어라면 네트워크 데이터의 "Internet"에는 "A"라는 값이 할당되는 것이다. 탐지된 커뮤니티 중에서 커뮤니티에 속한 키워드 수가 큰 것을 기준으로 노드를 추출하였으며 공통 키워드로 연결되는 노드들을 선정한 결과 한국은 13개의 노드가 추출되었고, 미국은 10개의 노드가 추출되었다.

키워드 커뮤니티 네트워크를 시각화한 결과는 Figure 4.2이다. 노드의 크기는 연결 정도 가중치가 적용된 연결 중심성에 비례해서 표현되었으며 노드 간 간선의 굵기는 노드간의 연결정도로 시각화하였다.

Table 4.2. Extracted keywords from keyword community network in Korea and USA

국가	기술영역	Word
한국	보안	authentication, identity, entity, terminal, comprising, network, feature, eye, left, perception, stereopsis, depth, images, image, process, adjusting, bundle, environments, panorama, point, calculation, camera, analysis, business, document, preview, setting
	이미지 프로세스	detects, extracts, gamut, mapping, consideration, contracts, sides, color, space, coordinates, rendering, pixel, defines, tests, weights, sub, colors, pixels, imaging, map, luminance, maps, incident, threshold, noise
	반도체	key, semiconductor, errors, ica, icb, icx, icy, properties, code, generation, arrangement, direction, array, beam, formation, listeners, region, transducers, sound, part
미국	통신	call, bridging, caller, phone, pool, queue, rating, calls, association, indication, address, web, browser, communication, portions, inputs, login, app, experiences, site, flow, portion, session, contents, form, proceeds, stages, telecommunications, transmission, power, receiver, schedule, topology, sender, channel, parameters, context, contexts, naf, pertaining, sessions, terminal, apparatus, page, server, addresses, interactions, switch, number, heartbeat, intervals, presence, text, stage, connection, offline, message, element, user
	대화형 매체	engine, interaction, layout, objects, scripting, interface, manner, functionality, manager, broker, attributes, configurer, origin, rules, translations, incoming, systems, communications
	스마트 홈	home, message, publishing, criteria, items, feature, mail, messaging, options, tablet, automation, indications, presents, profile, screens, smartphone, structure, zone, components, remainder, types, services, component

Figure 4.2(a)는 한국의 결과로 Figure 4.2(a)에서 보안 분야가 가중치 연결 중심성이 가장 크다. 그리고 Figure 4.2(b)에 해당하는 미국의 경우 통신 분야가 가장 큰 가중치 연결 중심성 값을 가진다. 연결 중심성이 높고 연결성이 강한 노드들을 선정하여 해당 노드들의 연관된 분야와 노드에 포함된 단어들을 나열한 것을 Table 4.2에서 확인할 수 있다. 노드에 속한 단어들을 통해 하드웨어 관련 사물인터넷의 분야를 유추하였다.

Figure 4.2(a)에서 볼 수 있는 바와 같이 한국 특히 데이터에서 연결중심성이 가장 큰 분야는 보안과 반도체 분야 그리고 보안과 이미지 프로세스 순이다. 이는 사물인터넷 환경의 급증으로 인해 반도체 보안 칩 등을 이용한 보안성이 요구되며 또한 이미지 프로세싱이 보안 분야에서 타겟 탐지와 같은 광범위한 분야의 감시에 많이 이용되는 사례들이 현실에 반영된 결과로 해석할 수 있다. 이러한 한국의 특히 결과는 사물인터넷을 사용하기 위한 기반기술에 주목하고 있다고 볼 수 있다.

미국 네트워크의 분석 시각화 결과는 비방향 그래프인 Figure 4.2(b)를 보면 알 수 있다. 통신과 대화형 매체 그리고 통신과 스마트 홈 순으로 연결 정도 가중치가 적용된 연결 중심성이 크다. 특히 미국 네트워크의 통신 분야 노드는 연결 중심성 뿐만 아니라 매개 중심성, 근접 중심성 또한 가장 컸다. 이를 통해 통신 분야가 미국의 하드웨어와 관련된 사물네트워크 분야의 중추적인 역할을 담당하여 미국 사물 인터넷 특히에서 큰 영향력을 가진다는 것을 볼 수 있다. 결과적으로 통신과 대화형 매체 그리고 통신과 스마트 홈 간의 연결이 강하다는 점을 통해 미국의 특히 분석 결과는 사물 인터넷 환경 구성 및 적용에 초점이 맞춰졌다고 볼 수 있다. 미국의 통신, 대화형 매체 그리고 스마트 홈에 해당되는 키워드들은 Table 4.2의 미국 부분을 통해 확인할 수 있다.

5. 결론

본 연구는 커뮤니티 탐지와 키워드 커뮤니티 네트워크 분석을 이용하여 하드웨어 관련 사물 인터넷 분야의 경향을 파악하였다. 먼저 하드웨어 분야 특허에 대해 한국과 미국의 특허 정보를 수집하였고 키워드 네트워크를 구축 및 분석하였다. 키워드 네트워크 구축을 위해 TF-IDF 가중치와 연관 규칙을 적용하여 새로운 데이터 셋을 형성하여 네트워크 구축에 이용하였다. 이는 기존의 키워드 네트워크에서 한 문서에 동시에 등장한 키워드들을 연결 관계로 보는 것과는 달리 다른 문서에서도 등장한 키워드들을 추출하여 모든 문서의 특허 정보를 표현하는 키워드 네트워크를 구축할 수 있도록 하였다.

분석 과정에서는 전체적인 네트워크를 생성한 후 최대 서브 그래프만을 추출하여 분석에 이용하였다. 키워드 네트워크의 구조를 파악하기 위한 그래프 레벨 분석 결과를 통해 연결 중심성, 근접 중심성 계수와 커뮤니티 밀도 그리고 군집 계수의 측면에서는 한국과 미국이 비슷했지만 예외적으로 미국의 매개 중심성이 한국보다 약 3배 높다는 결과를 얻었다.

또한 커뮤니티 탐지 결과 각 커뮤니티들 안에 포함된 키워드들을 통해 관련 분야를 파악하고 키워드 커뮤니티 네트워크를 형성하여 분야 간 연결 양태를 분석하였다. 그 결과 한국의 경우 보안과 반도체 영역 그리고 보안과 이미지 프로세스의 결합이 강했다. 한국의 결과는 사물 인터넷 환경의 증가로 산업, 의료 기기 등을 인터넷에 연결시키는 경우가 급증하여 반도체 보안칩과 같은 장치 또는 타겟 탐지 등에 이용되는 이미지 프로세스의 응용을 통해 보안성을 확보하고자하는 사물인터넷 분야의 현실을 반영하는 것이라고 해석할 수 있다.

미국의 경우 대체적으로 통신기술 분야가 네트워크에서 매우 중요한 역할을 하였다. 특히 커뮤니티 네트워크를 시각화한 결과 통신기술 관련 특허와 대화형 매체 그리고 통신기술 관련 특허와 스마트 홈 사이의 강한 결합을 보였다. 이는 미국에서 출원 및 등록된 하드웨어 관련 사물 인터넷 특허들이 통신 관련 분야와 연관이 있을 확률이 높다는 것을 의미한다. 사물인터넷에서 통신 분야는 기존의 사람들 간의 정보를 공유하는 방식이 사람 대 사람, 사람 대 사람 간의 통신으로 확장된 것을 의미한다. 통신을 중심으로 스마트 홈과 태블릿, 스마트 폰과 같은 대화형 매체와의 연결이 이루어진 미국의 연결 형태를 통해 미국은 편리한 생활을 지원하는 총체적인 사물인터넷 환경에 주목한다고 판단된다. 즉 2015년도 사물인터넷 특허 분석 결과 한국은 사물 인터넷의 기반 기술에 관련된 특허가 주로 출원 및 등록되고 있으며, 미국은 응용 기술과 관련된 특허가 주를 이루고 있다고 볼 수 있다. 또한 미국의 사물인터넷 기술은 우리의 삶에 직접적인 영향을 미치는 사물 인터넷 환경을 조성하는 방향으로 발전하고 있으며, 한국은 네트워크 보안, 반도체, 그리고 이미지 프로세싱 분야와 같이 사물 인터넷 인프라를 구축하는 기반 기술 방향으로 발전할 것으로 기대된다.

References

- Choi, J. S. (2013). Keyword-based patent trend analysis using statistical analysis: the case of Armor technology, *The Journal of Intellectual Property*, **8**, 223–252.
- Choi, J. and Hwang, Y. S. (2014). Patent keyword network analysis for improving technology development efficiency, *Technological Forecasting & Social Change*, **83**, 170–182.
- Huh, M. H. (2012). *Introduction to Social Network Analysis*, Freedom Academy, Seoul.
- Kang, J. H., Kim, J. C., Lee, J. H., Park, S. S., and Jang, D. S. (2015). A patent trend analysis for technological convergence of IoT and wearables, *Journal of Korean Institute of Intelligent Systems*, **25**, 306–311.
- Kim, P., Kim, K., and Kim, S. (2013). A user community detection from mobile interaction big data, *HCI KOREA 2013*, 224–227.
- Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*, Cambridge University Press, Cambridge.

- Park, J. H. (2014). A study for exploration of the promising technology through patent analysis (Master's Thesis), Hongik University, Seoul, Korea.
- Pons, P. and Latapy, M. (2005). Computing communities in large networks using random walks, *Computer and Information Sciences*, **3733**, 284–293
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge.
- You, E. S., Choi, G.-H., and Kim, S.-H. (2015). Study on extraction of keywords using TF-IDF and text structure of novels, *Journal of The Korea Society of Computer and Information*, **20**, 121–129.

키워드 커뮤니티 네트워크의 소셜 네트워크 분석을 이용한 사물 인터넷 특허 분석

김도현^a · 김현희^a · 김동건^a · 조진남^{a,1}

^a동덕여자대학교 정보통계학과

(2016년 4월 1일 접수, 2016년 5월 16일 수정, 2016년 5월 23일 채택)

요약

본 연구에서는 한국과 미국의 사물 인터넷 관련 특허 초록을 수집하여 키워드 네트워크 및 키워드 커뮤니티 네트워크를 구축하고 네트워크 분석을 실시하였다. 먼저 TF-IDF 가중치를 적용하여 중요 키워드를 추출하고 이 중요 키워드와 상관관계가 높은 키워드들을 재추출하여 핵심 키워드를 선정하였다. 선정된 키워드를 중심으로 키워드 네트워크를 구축한 다음 네트워크 탐지를 시행하여 키워드 커뮤니티 네트워크를 재구축하여 기술 간의 연결 관계를 분석하였다. 본 연구에서 생성한 키워드 커뮤니티 네트워크는 특허의 내용을 예측할 수 있을 뿐만 아니라 커뮤니티 간의 연결 관계를 분석함으로써 기술 간의 연관 관계도 파악할 수 있다. 키워드 커뮤니티 네트워크 분석 결과 한국은 보안, 반도체, 이미지 프로세스와 같은 사물 인터넷의 기반 기술 분야의 특허가 중요한 특허 기술로 나타난 반면 미국의 경우 스마트 홈, 대화형 매체 그리고 통신 등과 같은 사물 인터넷 환경, 응용 분야의 기술이 중요한 기술로서 자리 잡고 있음을 알 수 있다.

주요용어: 사물인터넷 특허 분석, 키워드 커뮤니티 네트워크, 소셜 네트워크 분석

이 논문은 2015년도 동덕여자대학교 학술연구비 지원에 의하여 수행된 것임.

¹교신저자: (02748) 서울시 성북구 화랑로 13길 60, 동덕여자대학교 정보통계학과.

E-mail: jinnam@dongduk.ac.kr