

융합기술의 동향분석을 위한 의미론적 특허분석 접근 방법 —토픽모델링과 교차영향분석의 활용—*

정병기** · 김정옥*** · 윤장혁****

- I. 서 론
- II. 배경연구
 - 1. 토픽모델링(Topic modeling)
 - 2. 교차영향분석(Cross-impact analysis)
 - 3. 정보 엔트로피(Information entropy)
- III. 연구방법
 - 1. 데이터 수집 및 유효특허 선별
 - 2. 세부기술 정의
 - 3. 기술 분야별 영향관계 산출
 - 4. 기술 확산 네트워크 생성 및 분석
- IV. 사례연구: 나노기술(Nanotechnology)
 - 1. 데이터 수집 및 데이터 수집 및 전처리(Preprocessing)
 - 2. 세부기술 정의
 - 3. 기술 분야별 영향관계 산출
 - 4. 기술 확산 네트워크 생성 및 분석
- V. 결론 및 추후연구

* 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2015R1A1A1A05027889). 본 논문은 2016년 제11회 대학(원)생 지식재산 우수논문공모전의 수상작을 개선 및 발전시킨 논문임.

** 건국대학교 산업공학과.

*** 건국대학교 산업공학과.

**** 건국대학교 산업공학과.

초 록

기술발전의 양상이 다양화 및 복합화됨에 따라, 개별 기술들 간의 경계가 점차 희미해진 융합기술이 나타나게 되었다. 융합기술을 분석함에 있어, 정적인 특허 분류코드를 활용하는 기존 연구방법들은 새로이 나타나는 기술들을 정의하고 이들 간의 파급관계를 분석하는 데 한계점을 지닌다. 따라서 본 연구는 대량의 특허에 대해 토픽모델링과 교차영향분석을 활용하여 융합기술을 분석하는 의미론적 분석 접근법을 제안한다. 제안된 방법은 토픽모델링을 특허 텍스트에 적용하여 분석 대상기술 내에 존재하는 잠재적인 세부 기술분야를 도출한 후, 교차영향분석을 활용하여 발굴된 세부기술들 간의 파급관계를 담고 있는 기술확산 매트릭스를 생성한다. 다음으로, 본 연구는 생성된 기술확산 매트릭스를 이용하여 융합기술 내에 존재하는 기술분야 간의 관계 시각화 및 정보엔트로피를 이용한 기술분야 간의 확산관계 분석을 실시한다. 본 논문에서는 융합 범위가 넓고 응용이 활발한 나노기술에 대한 사례연구를 통해 제안된 방법의 활용성을 파악한다. 본 연구는 기존의 특허 분류코드를 이용한 연구에서는 불가능했던 기술들의 정의와 이들 간의 파급관계를 파악할 수 있도록 하며, 따라서 빠르게 변화하는 기술에 대한 효율적 모니터링을 가능하게 하는 전문가 지원도구로 활용될 수 있을 것이다.

주제어

융합기술, 기술 모니터링, 토픽 모델링, 교차영향 분석, 나노기술, 특허분석

I. 서 론

융합 기술이란 기존의 단일 기술만으로는 해결이 힘든 문제를 극복하거나, 기술 간의 시너지 효과를 위하여 기술들이 결합되어 보다 큰 가치를 만들어내는 신기술을 말한다.¹⁾ 기술발전의 복잡도가 높아짐에 따라 기술 간 융합을 통한 새로운 차원의 기술개발에 대한 수요가 점점 증대되고 있다.²⁾ 이미 국가별, 연구 주제별로 다양한 기술 융합연구가 이뤄지고 있으며, 투자의 규모와 영역 역시 넓어지고 있는 실정이고, 우리나라도 6대 신기술(IT, BT, NT ST, ET, CT)을 선정하여 적극적인 연구개발을 진행 중이다.³⁾

기술 자체의 R&D뿐만 아니라 기술융합의 흐름이나 특징을 분석, 예측하려는 다수의 연구가 진행되어 왔으나 대부분 전문가의 직관이나 설문조사에 의존하고 있어서 결과가 정성적이고 비용과 시간이 크게 소요되는 등 몇 가지 문제점을 지니고 있다.⁴⁾ 이러한 문제점을 해결하기 위해 특허를 활용하여 융합기술을 분석하려는 시도가 있었다. 이러한 연구들은 국제 특허분류(International Patent Classification: IPC) 혹은 미국 특허분류(United States Patent Classification: USPC)의 동시 출현을 통하여 융합기술을 분류하였다.⁵⁾⁶⁾⁷⁾ 그러나 특허 분류코드를 활용한 분류방법에는 몇 가지 한계점이 존재한다. 먼

1) Shmulewitz, A. & R. Langer, and J. Patton, "Convergence in biomedical technology," Nature Biotechnology, Vol. 24(2006).

2) 조용래, 김의석, "특허 네트워크와 전략지표 분석을 통한 기업 기술융합 전략 연구," 지식재산연구(The Journal of Intellectual Property) 제9권 제4호(2014).

3) 송성수, "한국 과학기술정책의 특성에 관한 이론적 고찰," 과학기술학연구(Journal of Science & Technology Studies) 제2권 제1호(2002).

4) 강희종 · 엄미정 · 김동명, "특허분석을 통한 유망융합기술의 예측," 기술혁신연구(Journal of Technology innovation), 제14권 제3호(2006).

5) 박현석 · 서원철 · 윤장혁, "한국 국가 R&D의 학제적 특성 분석," 한국문헌정보학회지(Journal of The Korean society for library and information science) 제46권 제4호(2012).

6) Verspagen, B, "Measuring intersectoral technology spillovers: estimates from the European and US patent office databases," Economic Systems Research, 9(1997).

7) 정오뜸 · 고시근 · 서원철, "기술적 파급효과 도출을 통한 기술융합 분석 연구," 지식재산연구(The Journal of Intellectual Property) 제10권 제4호(2015).

저, 특허분류 코드의 할당이 전문가의 정성적인 판단에 의존한다. 따라서 유사한 특허라도 판단에 따라 서로 다른 코드가 할당될 수 있다. 또한 특허 내에서 융합기술을 표현하기 위해 융합된 개별 기술의 분류코드를 나열하여 표현하기 때문에, 온전한 의미의 융합기술을 표현하기 힘들다. 마지막으로 빠르게 생성되고 변화하는 기술 융합 흐름에 비해 분류체계의 변화속도는 이를 따라잡지 못하여 기존에 존재하지 않던 전혀 새로운 특성을 지닌 융합 기술의 경우 기술을 분류하기 힘든 경우가 존재한다. 이를 보완한 CPC 등의 발전된 특허분류 체계가 개발, 적용되고 있지만 정적인 분류체계의 특성상 기술이 발달함에 따라 여전히 같은 문제가 발생한다.

본 연구에서는 이러한 문제점을 극복하기 위하여 융합기술 내의 특허문서 내용을 기반으로 토픽모델링을 통해 세부 기술 분야를 분류한다. 그리고 분류된 기술 분야 간의 교차영향분석을 통해 융합기술 내에 존재하는 세부 기술 분야 간의 기술 확산을 분석하여 기술 확산 네트워크를 생성한다. 마지막으로 생성된 기술 확산 네트워크를 다양한 방법으로 분석하여 최종적으로 융합 기술 내의 기술 파급 분석과 핵심기술 파악, 기술 융합의 패턴 분석 방법을 제시한다.

본 연구에서 제시하는 방법을 통해 특허데이터를 기반으로 신속하고 정량적인 융합기술의 동향을 분석할 수 있다. 특히 기존에 특허분류코드를 활용한 분석 방법에 비해 내용기반의 유연한 분류방법을 사용함으로써 새로운 기술의 출현에 빠르게 대응할 수 있다. 또한 생성된 기술 확산 네트워크를 다양한 방법으로 분석하여 연구기획이나 기술기획에 있어 다양한 관점에서 시사점을 도출하여 연구개발의 효율성 증대가 기대된다.

이 글의 2장에서는 본 연구의 배경연구인 토픽모델링과 교차영향분석을 설명하고, 3장에서는 본 연구에서 제시하는 분석방법에 대해서 설명한다. 4장에서는 융합 범위가 넓고 응용이 활발한 나노기술(Nanotechnology: NT)에 대하여 본 연구에서 제시한 방법의 사례연구를 진행한다. 마지막 5장에서는 본 연구의 결론 및 추후연구에 대해 기술한다.

II. 배경연구

1. 토픽모델링(Topic modeling)

문서는 단어들의 조합으로 이루어지고 각 단어들의 조합은 해당 문서가 말하는 주제를 나타낸다. 또한 하나의 문서는 여러 개의 주제를 표현할 수 있고 문서들의 집합 안에서는 유사한 주제의식을 공유하고 있는 문서가 존재할 수 있다. 만약 같은 주제의식을 공유한다면 유사한 단어가 출현할 것이다. 토픽모델링은 문서 집합에서 맥락과 관련된 단서들을 이용하여 의미론적으로 단어들을 군집화하여 주제를 추론하는 방법이다. 특히, 개별 문서가 주제에 속할 확률을 이용하여 문서를 분류할 수 있다.

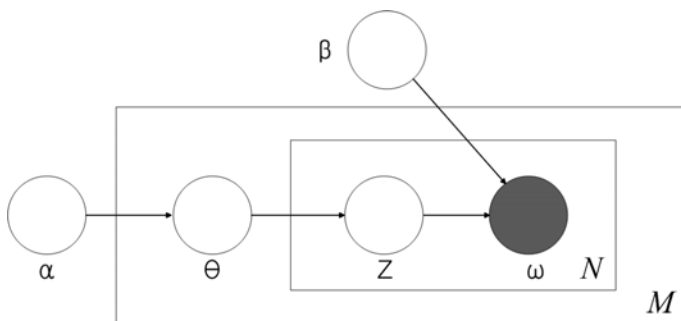
여러 가지 토픽모델링 알고리즘 중 본 연구에서 사용하는 LDA(Latent Dirichlet Allocation) 모형은 기존모형이 문서수준의 확률모형이 존재하지 않는 것을 보완하기 위해 제시된 모형으로 디리클레(Dirichlet)분포 기반의 확률적 토픽모델링 알고리즘이다.⁸⁾

LDA모형은 사람이 글을 쓰는 방식과 유사한 방식으로 진행되는 생성적 모형(Generative model)의 일종으로 다음과 같은 절차로 진행된다(그림 1). 그림1에서 w 는 하나의 문서를 의미한다. 또한 z 는 토픽들의 번호를 의미하는 것으로 단어는 토픽으로부터 생성되며 해당 문서가 어떠한 토픽비율(θ)을 가질 것인지는 디리클레 분포로부터 생성되는데, 이때 해당 분포의 파라미터는 α 이다. 또한 β 역시 디리클레 분포의 파라미터인데, β 는 해당 토픽이 어떤 단어들로 구성될지를 결정한다. 결국 단어들의 분포와 토픽들의 분포를 추정하여 각 문서를 생성하는 것이다. 이 과정의 반복을 통해 주어진 모든 문서들에 대하여 적절한 토픽의 분포와 생성된 토픽에서 적절한 단어의 분포를 알 수 있다.

8) Blei, D. M. & A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," the Journal of machine Learning research, Vol.3(2003).

LDA모형은 기존의 알고리즘들과 비교할 때 성능이 가장 우수한 것으로 알려져 있다.⁹⁾ LDA모형을 통해 단어들이 각 주제별로 속할 확률을 계산하고, 계산된 단어의 분포를 바탕으로 주어진 문서 집합이 각 주제에 속할 확률을 분석함으로써 개별 문서가 어떤 주제를 다루고 있는지 예측할 수 있다.

〈그림 1〉 LDA 토픽모델링의 도해적 설명



- ✓ Choose $\theta_i \sim \text{Dir}(\alpha)$, where $i \in \{1, \dots, M\}$
- ✓ Choose $\varphi_k \sim \text{Dir}(\beta)$, where $k \in \{1, \dots, K\}$
- ✓ For each word position i, j where $j \in \{1, \dots, N_i\}$, and $i \in \{1, \dots, M\}$
 - Choose a topic $z_{ij} \sim \text{Multinomial}(\theta_i)$.
 - Choose a word $w_{ij} \sim \text{Multinomial}(\varphi_{z_{ij}})$.

2. 교차영향분석(Cross-impact analysis)

교차영향분석(Cross-impact analysis)는 미래사건들 간의 영향관계를 설명하기 위해 개발된 개념으로 Gordon과 Helmer에 의해 처음 제안되었다.¹⁰⁾¹¹⁾

9) Chiru, C.-G. & T. Rebedea, and S. Ciotec., "Comparison between LSA-LDA-Lexical Chains," WEBIST, Vol.2(2014).

10) Brauers, J., and M. Weber, "A new method of scenario analysis for strategic planning," Journal of forecasting, Vol.7(1988).

11) Turoff, M, "An alternative approach to cross impact analysis," Technological forecasting and

Delphi method와 같은 기존의 예측 방법은 미래 사건의 상호 영향을 고려하지 않는 방법이었다. 이와 같은 단점을 개선하고자 새로 제시된 개념이 교차영향분석이다.¹²⁾

교차영향분석의 기본 개념은 사건A에 의해 영향을 받는 사건B가 있을 때, 사건B의 초기 발생가능성은 사건A의 발생에 따라 변한다는 것이다. 이때의 초기 발생가능성은 사건들 간의 영향관계에 따라 수정될 수 있다.¹³⁾ 이러한 영향관계를 파악하기 위해선 초기 확률과 조건부 확률이 필요하고 이를 통해 전문가들에 의해 판단된다. 그러나 판단해야 할 사건의 수가 많을 때는 각 사건들을 쌍대비교 하여 영향관계를 분석하고 이를 교차영향척도로 나타내어 영향의 방향과 강도를 알아낼 수 있다.¹⁴⁾ 이를 통해 각 사건 혹은 변수들 사이에 어떤 상호작용이 일어나는지를 알아내기도 하며, 사건 혹은 변수들의 관계가 종속적인지 독립적인지 알 수 있다.

3. 정보 엔트로피(Information entropy)

정보 엔트로피(Information entropy)는 Claude E. Shannon에 의해 1948년도에 정보이론의 일부로 처음 제시된 것으로, 논리 엔트로피(Logical entropy) 혹은, 섀넌 엔트로피(Shannon's entropy)라고도 불린다.¹⁵⁾ 여러 종류의 요소로 구성되어 있는 군집에서 k번째 요소의 군집 내 비율(확률)일 때 정보 엔트로피(H)는 <식 1>과 같이 계산된다.

social change, Vol.3(1972).

12) Asan, U. & C. E. Bozdağ, and S. Polat, "A fuzzy approach to qualitative cross impact analysis," Omega, Vol.32(2004).

13) 김진한 · 김성홍, "교차영향분석의 적용을 통한 국내 IT 환경 시나리오에 대한 연구," 경영과학 (Korean mangement science review) 제21권 제3호(2004).

14) 손영석 · 김익, "T.A.I.D.A 시나리오 플래닝 방법과 교차영향분석을 이용한 U-Eco City 개발과 미래유망기술의 R&D 전략 수립에 관한 연구," 대한건축학회 논문집(Journal of The architectural institute of korea planning & design) 제24권 제11호(2008).

15) Kannathal, N. & M. L. Choo & U. R. Acharya, and P. Sadasivan, "Entropies for detection of epilepsy in EEG," Computer methods and programs in biomedicine, Vol.80(2005).

$$H = - \sum p_k \log p_k \quad \langle \text{식 1} \rangle$$

정보 엔트로피는 불확실성, 혹은 다양성의 지표로 널리 사용되는 것 중 하나로, 정보처리 뿐만 아니라 생물학 등에도 널리 사용된다.¹⁶⁾ 정보 엔트로피가 높다는 것은 주어진 군집이 다양한 요소로 균등하게 구성되어 있다는 것을 의미하고, 이는 불확실성이 높은 것으로 해석된다. 즉, 자료들이 균등분포에 가깝게 널리 분포되어 있을수록 높은 정보 엔트로피를 가지고, 좁게 분포되어 있을수록 낮은 정보 엔트로피를 가진다. 정보 엔트로피는 그 자체에 대한 연구뿐만 아니라 다양한 연구에서 응용되어 왔는데, 자료포락분석(Date envelopment analysis: DEA)과 함께 효율성을 연구한 사례가 있고,¹⁷⁾¹⁸⁾ 생물학에서 어떤 생태계의 종의 다양성 지표로 사용된 연구도 존재하며,¹⁹⁾ 제품다양성의 척도로 사용한 연구도 존재한다.²⁰⁾

본 연구에서는 기술과급네트워크에서 노드별로 전파와 흡수에 대한 엔트로피 분석을 통해 하나의 세부 기술분야가 인접한 기술분야들에 대해 지니는 융합성을 측정하는 지표로 사용된다.

III. 연구방법

본 연구는 융합기술의 기술 확산 분석을 위해 대량의 특허집합에 대하여

16) Ricotta, C. "Bridging the gap between ecological diversity indices and measures of biodiversity with Shannon's entropy: comment to Izsák and Papp," Ecological Modelling, Vol.152(2002).

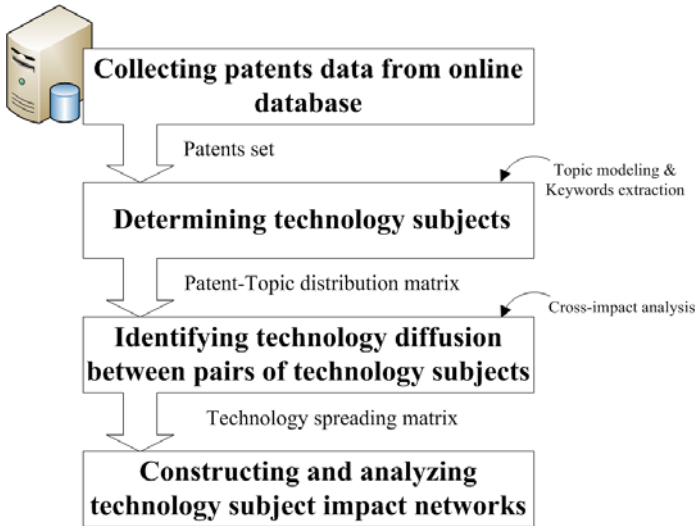
17) Bian, Y., and F. Yang, "Resource and environment efficiency analysis of provinces in China: A DEA approach based on Shannon's entropy," Energy Policy, Vol.38(2010).

18) Soleimani-Damaneh, M., and M. Zarepisheh, "Shannon's entropy for combining the efficiency results of different DEA models: Method and application," Expert Systems with Applications, Vol.36(2009).

19) Ricotta, C., "Bridging the gap between ecological diversity indices and measures of biodiversity with Shannon's entropy: comment to Izsák and Papp," Ecological Modelling, Vol.152(2002).

20) Straathof, S. M. "Shannon's entropy as an index of product variety," Economics Letters, Vol.94(2007).

〈그림 2〉 연구 절차



토픽모델링과 교차영향분석을 사용하여 세부기술 수준에서 지식의 흐름을 분석한 연구이다. 본 연구는 데이터수집, 기술 분야 파악, 기술 분야별 영향 관계 산출 및 기술 확산 네트워크 분석 등 총 네 단계를 거치게 된다. 본 연구에서 제시하는 방법론의 자세한 절차는 〈그림2〉와 같다.

1. 데이터 수집 및 유효특허 선별

우선 기술 확산효과 분석을 위해 분석 대상이 되는 기술 분야에 대해 온라인상의 특허 데이터베이스로부터 대량의 특허 데이터를 수집한다. 특히 본 연구에서는 미국 특허청인 USPTO(United States Patent and Trademark Office)로부터 신규성과 진보성이 인정된 등록특허를 사용하여 연구를 진행한다. 그리고 본 연구방법은 키워드 추출과 토픽모델링 등의 텍스트 마이닝 기법을 사용하므로, 특허 데이터를 수집할 때 반드시 제목, 초록 등의 텍스트 데이터가 포함되어야 하며, 각 특허를 구분할 수 있는 출원번호, 등록번호

등의 정보 또한 필수적으로 포함되어야 한다. 그리고 본 연구는 수집된 전체 특허집합에 대해서 노이즈 특허 필터링 등의 유효특허 선별 과정을 거쳐 분석 기술 분야와 직접적으로 연관되는 유효특허를 선별한다.

2. 세부기술 정의

선별된 전체 특허 집합에서 키워드를 추출하고 추출된 키워드 중, 실제로 분석대상이 되는 기술 분야와 직접적인 상관관계가 존재하는 키워드를 선별한다. 최종적으로 선별된 키워드는 토픽모델링에서 특허 집합 내에 존재하는 기술 분야를 결정하는 데 활용된다. 본 연구에서 사용되는 확률 기반의 토픽모델링 알고리즘인 LDA모형은 사용자가 임의로 토픽 수를 결정해 주어야 하는데, 적절한 토픽 수를 결정하는 방법은 초기에는 기술 분야에 대한 사전조사를 통해 해결하였으나 이후에 여러 연구자들에 의해 정량적인 방법이 연구되었다.

본 연구에서는 키워드들을 가장 잘 분리하는 토픽 수를 결정하는 방법을 사용한다. 키워드들을 가장 잘 분리하는 방법은 토픽-키워드 분포를 이용하여 각 토픽 간 유사도의 평균을 계산한 뒤에 토픽 간 유사도의 평균값이 최하점에 도달했다가 다시 상승하는 지점인 반등점을 사용하거나 평균값이 안정화되기 시작하는 Elbow point를 이용하는 방법을 사용할 수 있다.²¹⁾

본 과정의 결과물로서 각 기술 분야가 개별 키워드들로부터 기여되는 정도를 확률로 나타낸 기술 분야-키워드 분포와 개별 특허가 각각의 기술 분야에 속하는 정도를 확률로 나타낸 특허-기술분야 분포를 얻을 수 있다. 기술 분야-키워드 분포는 각 토픽의 속성을 결정하는 데 사용되고, 특허-기술분야 분포는 다음 단계에서 기술 분야 간의 영향관계를 산출할 때 사용된다.

21) Wang, B. & S. Liu & K. Ding & Z. Liu, and J. Xu, "Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology," *Scientometrics*, Vol.101(2014).

3. 기술 분야별 영향관계 산출

기술 분야별 영향관계 산출에는 앞선 과정인 세부기술을 정의하는 과정에서 토픽모델링을 통해 획득한 특허-기술분야 분포와 영향관계분석이 사용된다. 그러나 LDA모형은 디리클레 분포를 기반으로 생성된 확률모형이기 때문에 LDA모형을 이용한 토픽모델링 결과로 얻을 수 있는 특허별 기술 분야 분포에서 개별 특허는 모든 기술 분야에 대해 0보다 큰 값을 가진다. 따라서 영향관계 파악을 위해 특허-기술분야 분포에 대하여 적절한 Threshold를 결정하여 전체 기술 분야에 대해 해당 문서가 실제로 속하는 기술 분야로 이진화(Dichotomize)한다. 이를 통해 각 특허 내에서 공동으로 출현하는 기술 분야를 알 수 있으며 이진화된 특허-기술분야 분포 표는 교차영향분석의 데이터로 활용된다. 교차영향분석을 통해서 기술 분야 간 상호 영향관계를 파악할 수 있고, 최종적으로 분야 간 기술 확산 네트워크를 생성할 수 있다. 기술 분야 간의 영향관계는 〈식 2〉와 같이 계산되는데, 이는 기술 분야 A가 기술 분야 B에 주는 영향으로 해석된다.

$$Impact(Tech_A, Tech_B) = \frac{N(Tech_A \cap Tech_B)}{N(Tech_A)} \quad \langle \text{식 2} \rangle$$

4. 기술 확산 네트워크 생성 및 분석

생성된 기술 확산 네트워크는 융합기술의 특허 내에서 분류된 세부기술 간의 영향관계를 담고 있다. 특히, 네트워크는 매트릭스 형태로 표현될 수 있는데, 매트릭스 형태로 전환했을 때, 하나의 행은 해당 세부기술에서 다른 세부기술로 전파(Out)되는 정도를 의미하고, 하나의 열은 외부에서 해당 세부기술로 흡수(In)되는 정도를 의미한다. 이를 이용하면 여러 가지 관점에서 기술간 영향관계를 분석할 수 있다. 특히 네트워크를 시각화하여 전반적인 기술흐름을 파악할 수 있고 세부기술별로 흡수와 전파되는 정도와 각 기술

별 특징을 구분할 수 있다.

특히 다양성을 나타내는 지표 중 통신이나 전자회로, 생물학 등 여러 분야에서 이용되는 정보 엔트로피를 이용하면 전파와 흡수의 정보 엔트로피를 2차원상에 배치하고, 각 축의 평균을 이용하여 사분면으로 분할할 수 있다(그림 3). 예를 들어, 기술 분야 A가 다른 기술 분야들에 주는 영향을 이용하여 산출된 정보 엔트로피 값이 높을 경우 기술 분야 A가 다른 분야들에 고루 영향을 미치며 동시에 영향의 강도가 높다는 것을 의미하며, 정보 엔트로피 값이 낮을 경우 기술 분야 A가 다른 기술 분야에 주는 영향이 편중되거나 미약함을 의미하게 된다. 이러한 기술 분야별 정보 엔트로피의 특성을 활용하여 사분면에서 기술의 위치에 따라 해당 기술의 잠재적인 특성을 정의할 수 있을 것이다. 우선 흡수의 엔트로피는 낮고 전파의 엔트로피는 높은 기술들은 다른 기술에서 파급되는 것보다 본연의 기술을 다른 기술들로 전달하는 경우가 많으므로 상대적으로 원천기술이나 기초기술, 기반기술에 가까울 가능성이 크다. 또한 흡수와 전파의 엔트로피가 모두 높은 기술들은 여러 기술들에게 영향을 받아서 발전하였지만, 첨단기술은 아닌 이미 상당히 성숙한 융합기술 또는 일종의 중개자의 역할을 하는 기술들로 예상할 수 있으며, 흡수와 전파의 엔트로피 모두 낮은 사분면에 위치하는 기술들은 개발과 응용 모두 활발하지 않기 때문에 해당 기술영역 내에서 비주류에 속하는 기술들이 위치할 가능성이 크다. 마지막으로 흡수의 엔트로피는 높지만 전파의 엔트로피는 낮은 기술들은 다양한 기술에게 영향을 받고 있지만, 해당 기술이 다른 기술에 영향을 주는 정도는 낮으므로, 해당 영역 내에서 성숙되지는 못하였지만 최근 가장 활발하게 발전하고 상태에 있는 응용기술 또는 첨단기술이 위치할 가능성이 클 것으로 생각된다.

〈그림 3〉 기술 확산 엔트로피 맵의 분석 예

(1) Low in & High out : 기초기술	(2) High in & High out : 상용화 된 융합기술
(3) Low in & Low out : 외면기술	(4) High in & Low out : 첨단(비 상용화) 융합기술

IV. 사례연구: 나노기술(Nanotechnology)

본 연구는 특허데이터를 통해 세부기술을 파악하고 동시에 세부기술 간의 확산 네트워크 분석을 통해 다양한 관점에서 대상기술을 분석한다. 본 연구에서는 나노기술(NT: Nanotechnology)을 대상으로 방법론의 검증을 실시하였다. 나노기술은 육안으로 구분할 수 없는 나노 크기의 물질을 이용하여 물체를 만들고 조작하는 기술을 통칭하는 말로서, 1959년 미국의 물리학자 리처드 파인만이 나노기술의 가능성을 주장한 이래로 지속적으로 급속하게 발전해온 유망한 기술 분야이다.²²⁾

나노기술은 전자, 생명, 화학 등에서 포괄적으로 응용되고 있는데, 기존의 물질을 나노미터 수준으로 가공하여 소형화하는 Top-down 방식과 나노 단위에서 물질들을 조작하여 새로운 성질과 기능을 가진 물질을 만드는 Bottom-up 방식으로 구분할 수 있다.²³⁾ 나노기술의 특징으로는 높은 기술 집약도, 경제성 등을 들 수 있다. 또한, 나노기술을 통해 기존의 물리, 화학, 생물 등과 같은 기초과학분야들을 횡적으로 연결하여 시너지 효과를 내고 새로운 영역을 구축할 수도 있다.²⁴⁾ 위와 같이 나노기술은 그 적용범위가 다양하기 때문에 융합의 정도가 높아서 기술의 정밀한 분할 및 구체적인 확산 효과 파악이 어렵다. 따라서 본 연구에서 제시하는 방법인 토픽모델링을 이용한 의미론적 분할과 교차영향분석을 활용한 확산 분석 대상으로 적절하다고 판단하였다.

22) Siegel, R. W., "Creating nanophase materials," Scientific American, Vol.275(1996).

23) Spatz, J. P. & V. H. Chan & S. Mößner & F. M. Kamm & A. Plettl & P. Ziemann, and M. Möller, "A combined top-down/bottom-up approach to the microscopic localization of metallic nanodots," Advanced Materials, Vol.14(2002).

24) Roco, M. C. & R. S. Williams, and P. Alivisatos, "Nanotechnology Research Directions: IWGN Workshop Report: Vision for Nanotechnology in the Next Decade," Springer Science & Business Media(2000).

1. 데이터 수집 및 전 처리(Preprocessing)

나노기술에 대한 분석을 실시하기 위해 나노기술과 관련된 특허를 수집해야 한다. 본 연구에서는 나노기술과 관련된 미국특허 중에서 특허로서 신규성과 진보성이 인정된 등록 특허만을 연구의 대상이 되는 데이터로 선정하였다.

한국특허정보원에서 제공하는 온라인상의 특허 데이터베이스인 키프리스에서 미국 특허상표청(USPTO)에 등록된 특허를 대상으로 하여 나노기술과 관련된 포괄적인 검색식으로 특허 수집을 실시하였다(표 1).

수집된 특허 데이터들의 초록을 대상으로 나노기술과 연관성을 가지는 키워드를 선별하여 키워드사전을 구축하고 선정된 키워드가 출현하지 않는 특허는 노이즈 특허로 판단하여 제거한다. 이를 위해 모든 특허들의 초록에 대해 시맨틱 텍스트 분석 서비스인 AlchemyAPI(<http://www.alchemyapi.com>)를 활용하여 키워드를 추출하여 초기 키워드사전을 구축한 후, 텍스트 분석의 용이성을 위해 출현횟수가 1회 이하인 키워드들을 제거하였다. 다음으로 기술적 의미를 지니지 못하거나 너무 일반적인 용어, 예를 들어 ‘system’, ‘invention’ 등과 같은 키워드들을 제거하여, 최종적으로 87,446건의 특허와 22,072개의 키워드가 본 연구의 분석대상으로 선정되었다.

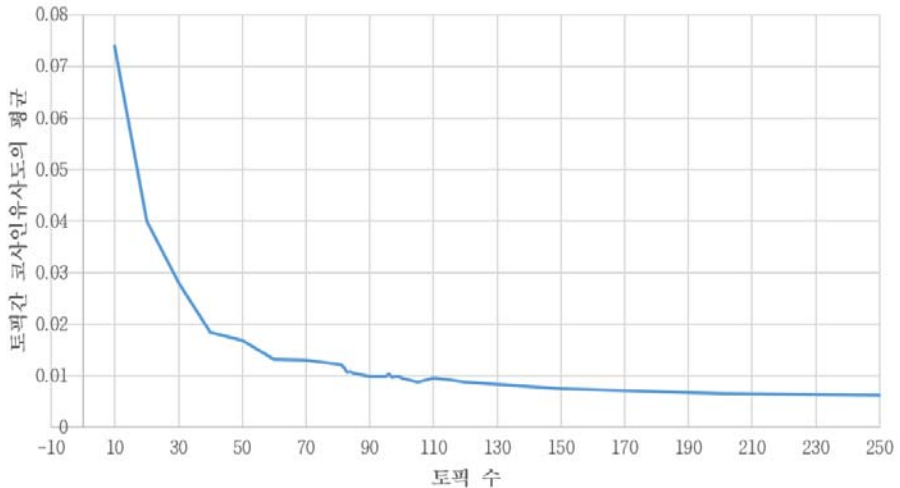
〈표 1〉 나노기술(Nanotechnology)의 특허 검색식

특허 검색식
(nano + submicro + (sub ^1 micro) + ultrafine + (ultra ^1 fine))

2. 세부기술 정의

분석 대상으로 선정된 87,446개의 특허와 구축된 키워드사전을 이용하여 LDA 토픽모델링을 실시한다. 토픽모델링의 결과로 추출되는 특허 내의 토픽들은 나노기술의 세부 기술 분야로 볼 수 있다. 또한 본 연구에서는 LDA

〈그림 4〉 토픽 수별 토픽 간 코사인유사도의 평균



모형의 적절한 토픽 수를 결정하기 위하여, 각 토픽 내의 키워드를 가장 잘 분리하는 토픽 수를 탐색적으로 발견하는 방법을 활용한다. 본 연구에서는 적절한 토픽 수를 발견하기 위해 3장 2절에서 언급한 Elbow point 방법을 이용하였으며, 토픽간 유사도는 코사인유사도를 이용하여 계량화되었다. 코사인 유사도는 두 벡터 간의 유사도를 측정하는 지표로써, 두 벡터가 공간상에서 이루는 각도의 코사인값을 이용하여 0~1사이의 값으로 산출된다(식 3). Elbow point 분석결과 토픽 간 키워드의 코사인 유사도 평균은 반등값을 가지는 형태는 아니었으며, 하락을 유지하며 안정화되는 형태임이 확인되었다. 실제로, 토픽모델링을 실시했을 때에 토픽 수가 105개를 시작으로 하여 코사인 유사도의 평균이 안정화되기 시작하였으며, 따라서 본 연구는 토픽 수가 105개일 때를 Elbow point로 판단하였다(그림 4). 이후 결정된 105개의 토픽에 대하여 각 키워드가 기여되는 정도인 토픽-키워드 분포를 통하여 각 토픽이 의미하는 기술분야를 정의하였다(표 2). 예를 들어, 기술분야 ‘Signal’의 경우 기여도가 높은 키워드인 ‘signal’, ‘frequency’, ‘receiver’와 같은 키워드들을 종합하여 정의될 수 있다.

$$Cosine\ Similarity = \cos(\theta) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad \langle \text{식 3} \rangle$$

〈표 2〉 나노기술의 주요 기술 분야와 해당 키워드

기술 분야	주요 키워드 및 기여도
Insulation	Electrode(0.361), Insulating(0.018), Gate electrode(0.018), Electrode Layer(0.010)
Substituent	Alkyl(0.066), Substitute(0.033), Atoms(0.023), Hydrogen(0.021)
Signal	Signal(0.181), Frequency(0.050), Match(0.012), Receiver(0.012)
Photo Resist	Pattern(0.197), Resist(0.052), Photo(0.043), Lithography(0.021)
Composition	Composition(0.313), Composition comprising(0.017), Decomposition(0.006), Cosmetic composition(0.003)
Receptor	Derivative(0.060), Receptor(0.048), Derivatives(0.044), Inhibit(0.026)
Semiconductor2	Memory(0.129), Memory device(0.031), Memory cell(0.029), Storage(0.023)
Crystal	Crystal(0.233), Crystalline(0.068), Liquid crystal(0.043), Nano crystal(0.032)
Medical	Biologic(0.041), Cancer(0.032), Tumor(0.02), Therapeutic(0.017)

3. 기술 분야별 영향관계 산출

본 단계에서는 앞선 단계를 통해 획득한 특허-기술분야 확률분포를 이용한다(표 3). 본 단계에서 사용하는 특허와 기술 분야 간의 확률분포는 분석대상으로 선정된 87,446건의 나노기술 관련 특허와 105개의 세부기술에 대해 각각의 특허문서가 어떤 세부기술을 얼마나 이야기하고 있는지에 대한 수치를 0~1 사이의 확률 값으로 나타낸 분포이다. 따라서 교차영향분석을 위해 Threshold를 선정하여 이진화한다(표 4). 특허-기술분야 분포를 이진화하면 각 특허 내에서 어떤 기술 분야가 공동으로 출현했는지 알 수 있고, 이를 통해서 전체 특허 내에서 기술 분야들의 융합 동향을 알 수 있다. 본 연구에서

는 다양한 Threshold를 이용하여 특허별 소속 기술분야를 이진화하고 특허별 기술분야 도출 결과에 대한 적합성 분석을 실시하였으며, 최종적으로 Threshold가 0.10일 경우 가장 적합할 것으로 판단하였다. Threshold를 사용하여 이진화한 이후 교차영향분석을 사용해서 세부 기술의 기술 확산을 분석하였다(표 5). 나노기술의 세부기술에 대한 분석 결과, 총 11,025개의 경우의 수에서 자기 자신에게 향하는 105개의 경우를 제외하고 1의 영향도를 가진 관계는 존재하지 않았고 2904개의 관계가 영향을 미치지 않는 것으로 나타났다. 또한 8,016개의 관계가 0초과, 1미만의 영향관계가 있는 것으로 나타났다. 특히 그 중에서 Semiconductor1이 Semiconductor2에 0.430793으로 가장 높은 영향을 주는 것으로 나타났고, Semiconductor4가 Fabric에 0.000281로 가장 낮은 영향을 주는 것으로 나타났다.

〈표 3〉 특허-기술 분야 분포 매트릭스의 일부

Grant No.	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8	Topic9
7687112	0.005666	0.008986	0.007454	0.006815	0.009242	0.005666	0.00541	0.012945	0.006687
8580946	0.003807	0.003895	0.008548	0.003895	0.005124	0.003895	0.027424	0.003807	0.003895
7879417	0.004491	0.004809	0.126838	0.004809	0.004915	0.004491	0.004491	0.007248	0.004809
8216670	0.003690	0.003778	0.00369	0.00369	0.003952	0.00369	0.00369	0.228009	0.005781
7658897	0.042552	0.008071	0.008071	0.008071	0.046171	0.008071	0.008261	0.008071	0.008071
6720553	0.003943	0.003526	0.003776	0.014514	0.005025	0.018593	0.003526	0.005025	0.007938
8241759	0.006789	0.00554	0.005540	0.050119	0.007163	0.005415	0.00529	0.005914	0.005540

〈표 4〉 0.1을 기준으로 이진화된 특허-기술분야 분포 매트릭스의 일부

[illegible]

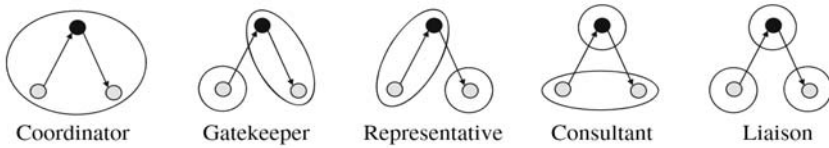
〈표 5〉 기술확산 매트릭스의 일부

Topics	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8	Topic9
Topic1	1	0.052271	0.003428	0.004284	0.001714	0.001714	0.007712	0	0
Topic2	0.029483	1	0.000967	0	0	0	0.001450	0	0
Topic3	0.005848	0.002924	1	0.004386	0.007310	0.021930	0.011696	0	0
Topic4	0.008052	0	0.004831	1	0.003221	0.003221	0.020934	0.001610	0
Topic5	0.003067	0	0.007669	0.003067	1	0.001534	0	0.027607	0.001534
Topic6	0.002112	0	0.015839	0.002112	0.001056	1	0.006336	0.003168	0
Topic7	0.003686	0.001229	0.003276	0.005324	0	0.002457	1	0.011466	0.007781
Topic8	0	0	0	0.000873	0.015721	0.002620	0.024454	1	0.002620
Topic9	0	0	0	0	0.004854	0	0.092233	0.014563	1

4. 기술 확산네트워크 생성 및 분석

본 절차에서는 앞선 단계에서 생성된 기술 확산 매트릭스에서 0을 초과하고 1 미만의 영향관계가 있는 8,016개의 관계를 대상으로 나노기술 내의 기술 확산을 분석한다. 다양한 방법을 적용할 수 있지만 본 사례연구에서는 전체 영향관계에서 상위 1% 영향도를 이용하여 2차원 상의 네트워크로 시각화 하였다(그림 6). 또한 네트워크 클러스터 간의 연결 형태를 고려하여 각 노드의 특성을 구분한 선행연구를 참고하여 중요 노드들에 대해 조정자(Coordinator), 문지기(Gate-keeper), 대리자(Representative), 상담가(Consultant), 연락책(Liaison) 등으로 구분하여 추가적인 분석을 실시한다.²⁵⁾ 조정자란 한 클러스터 내에서 기술전달을 중개하는 역할을 하는 노드를 말하며, 한 클러스터에서 다른 클러스터로 기술이 전달될 때 반드시 거치는 노드는 문지기과 대리자인데, 문지기는 전달받는 클러스터의 노드, 대리자는 전달하는 클러스터의 노드를 말한다(그림 5). 또한 G문지기과 비슷하지만 상담가는 자신이 받은 기술지식을 같은 클러스터의 다른 노드로 전달하는 역할이고, 연락책은 자신이 받은 기술지식을 전혀 다른 클러스터로 전달하

25) Gould, R. V., and R. M. Fernandez, "Structures of mediation: A formal approach to brokerage in transaction networks," Sociological methodology, Vol.19(1989).

〈그림 5〉 네트워크 내에서 노드별 역할 구분²⁶⁾

는 경우이다(그림 5). 네트워크상에서 노드의 크기는 전체 87,446건의 특허 내에서 해당 기술분야의 토픽량이고 링크의 두께는 각 기술 분야의 영향강도이다.

상위 1% 영향관계를 기준으로 구성된 네트워크의 연결강도와 기술내용 등을 보았을 때, 총 네 개의 기술 클러스터가 존재함을 알 수 있다. 첫 번째 클러스터는 Manufacturing으로 Battery, Solar battery, Charge injection, Insulation 등 물질이나 제품의 제조에 나노기술이 응용되는 기술들로 구성되어 있다. 실제로 이 클러스터를 구성하는 기술 분야 중 하나인 Insulation에는 “Photovoltaic device and method for manufacturing thereof(US grant no 8802969)” 등 대부분 나노 수준에서 절연처리를 하여 생산하는 방법을 나타내는 특허가 속한다. 또한 Insulation은 클러스터 내에서 Nano array와 Micron stone, Battery, Terminal의 영향을 받고 있다.

하지만 Insulation에게 영향을 받는 기술 분야는 다른 클러스터에 속하는 Semiconductor²뿐이다. 따라서 Insulation은 Manufacturing 클러스터 내의 기술을 Semiconductor 클러스터에 전달하는 대리자의 역할을 하고 있다. 즉, Semiconductor² 분류에 속하는 기술들이 Insulation의 방법으로 구현되는 경우가 많다고 생각할 수 있다. 또한 Transistor²는 Insulation과 같이 한 방향으로의 전달이 아니라 Manufacturing과 Semiconductor 클러스터 양쪽으로 기술을 파급하고 있다. 따라서 Transistor²는 연락책의 역할을 하고 있다. 실제로 Transistor²에 강하게 속하는 특허는 “Signal transmitting

26) Lim, Hyojeong, and Yongtae Park, “Identification of technological knowledge intermediaries,” *Scientometrics* 84,3(2010).

Display, Lithography 등 반도체 제품이나 반도체의 제조기법 등으로 구성되었다. 특히 본 클러스터는 Manufacturing 클러스터와 달리 중심이 되는 기술이 존재하는데, Semiconductor4가 그 역할을 하는 것으로 파악된다. 특히, Semiconductor4는 전체 링크의 1%수준에서 12개 기술 분야에게 파급을 받고 있지만 직접적으로 전파하고 있는 기술은 전무하다. 따라서 Semiconductor4는 주변의 다른 기술들을 응용한 응용기술의 성격을 강하게 나타낼 것이라고 예측할 수 있다. 실제로 Semiconductor4에 강하게 속하는 특허는 “Method of manufacturing a semiconductor device with different lattice properties(US grant No 7151019)”와 같이 다양한 기술들을 이용하여 복합적인 반도체를 생성하는 방법에 대한 특허들인 것으로 파악되었다. 또한 Lithography기술은 Semiconductor 클러스터와 Material 클러스터 사이에서 문지기의 역할을 한다. 즉, 상위 1% 영향도 기준으로 Material 클러스터의 기술은 Lithography를 통해서만 Semiconductor 클러스터로 전달된다. Lithography는 반도체 생산에 사용되는 기술로 화학물질을 통해 실리콘 웨이퍼에 반도체를 새기는 기술로써 반도체와 관련된 대표적인 기술이다 (Moreau, 2012). 세 번째 클러스터는 Polymer, Ionic, Solvent, Peptide, Substitution 등 소재나 물질 관련기술들로 구성된 Material 클러스터이다. Material 클러스터의 토픽량은 Photographic(1248.79), Peptide(1110.27)와 Polymer(1144.15)가 가장 발전이 많이 된 기술이지만 기술 간의 파급방향까지 고려하면 Semiconductor 클러스터의 Semiconductor4와 같이 중심이 되는 기술은 없다. 그러나 뚜렷하게 역할이 정해진 기술은 존재하는데, Peptide가Material의 기술을 Nanomedicine에 전파하는 대리자의 역할을 한다. 즉, 상위 1% 영향도 기준으로 Material 클러스터 내의 기술은 Peptide를 통해 Nanomedicine으로 전파된다. 마지막 클러스터는 Medical2, Medical3, Substance delivery 등 의료분야에 활용되는 기술들을 중심으로 구성된 Nanomedicine이다. 나노 의학(Nanomedicine)은 질병과 관련하여 치료, 진단, 예방을 위한 목적으로 분자수준에서 설계 및 제조된 의약품, 의료기기 등을 모두 총괄하는 말이다.²⁷⁾ 전체 나노기술의 1% 영향관계 수준에서

Nanomedicine 클러스터는 총 7개의 기술로 구성되어 있는데, 중심의 역할을 하는 기술 없이 상대적으로 단순하게 연결된 것을 볼 수 있다. 특히 그중에서 다른 클러스터에서 가장 멀리 떨어져 있는 Substance delivery는 다른 클러스터와 가장 관련이 적은 분야인데, “Sustained release beads and suspensions including the same for sustained delivery of active ingredients(US grant No 9271938)”와 같이, 분자수준에서 특정 성분을 피부 내로 침투시키는 기술들로 구성되어 있는, Nanomedicine의 독창적인 기술이다.

상위수준의 영향관계를 통한 나노기술의 전반적인 분석 외에, 본 연구에서는 각 기술별로 융합의 정도를 비교하는 지표로 정보 엔트로피를 제시한다. 기술 확산 매트릭스를 통해 각 기술별로 전파하고 흡수하는 강도를 확인할 수 있는데, 이를 〈식 1〉에 대입하여 기술별로 전파와 흡수의 엔트로피를 계산한다. 본 과정을 거쳐 최종적으로 산출된 나노기술 내의 세부기술들의 정보 엔트로피는 〈표 6〉과 같다. 전체 105개 기술들의 흡수 엔트로피의 평균은 0.902919로, 가장 높은 흡수 엔트로피는 가진 기술은 ‘Semiconductor4(1.518731)’이었고, 가장 낮은 흡수 엔트로피를 ‘Micron stone (0.285523)’이었다. 전파 엔트로피는 0.938493의 평균값을 가지고, 가장 높은 전파 엔트로피를 가진 기술은 ‘Oxide(1.095797)’이고 가장 낮은 전파엔트로피를 가진 기술은 ‘Medical2(0.647802)’로 나타났다. 산출된 흡수/전파의 엔트로피 값과 그 평균을 이용하여 본 연구에서 제시하는 기술들의 융합정도를 파악할 수 있는 시각화 자료인 기술 확산 엔트로피 지도를 만들 수 있다.

나노기술의 세부기술들의 전파와 흡수 엔트로피를 2차원상에 매핑하고, 흡수와 전파 평균인 0.902919와 전파 엔트로피의 평균인 0.938493으로 2차원을 4분면으로 분할하였다(그림 7). 흡수 엔트로피는 낮고 전파 엔트로피는 높은 영역에는 30개의 세부기술이 위치하였는데, 예상했던 것과 같이 Solvent, Composition, Substitution, Extract 등 화학, 전자 등 나노기술 응용분야들의 기초기술이었다. 두 엔트로피가 모두 높은 영역에는 Semiconductor4,

27) Leonard, B., “Cancer nanotechnology: going small for big advances: using nanotechnology to advance cancer diagnosis, prevention and treatment,” DIANE Publishing(2009).

Glass, Solar battery 등 25개의 기술 분야가 위치했는데, 충분히 상용화가 진행되었다고 판단되는 기술들을 중심으로 구성되어 있었다. 흡수와 전파 엔트로피 모두 낮은 영역에는 25개의 기술들이 위치하는데, 평균선 근처에 위치한 일부 기술들은 오롯이 이 영역에 위치했다고 보기 어렵지만 그 외에는 실제로 핵심기술이 아니고 발전이 되지 않은 기술들이거나 ‘Silver halide film’와 같이 외면받고 있는 기술들이 위치하였다. 마지막으로 흡수 엔트로피가 높고 전파 엔트로피가 낮은 영역에는 최신의 융합기술이 위치할 것으로 예측하였는데, 실제로 Fuel cell, Nano-tube, Plasma 등 아직 생소하고, 실용화되지 않은 첨단기술들을 중심으로 25개의 세부기술이 위치하였다.

〈표 6-1〉 기술 분야별 흡수/전파의 정보엔트로피

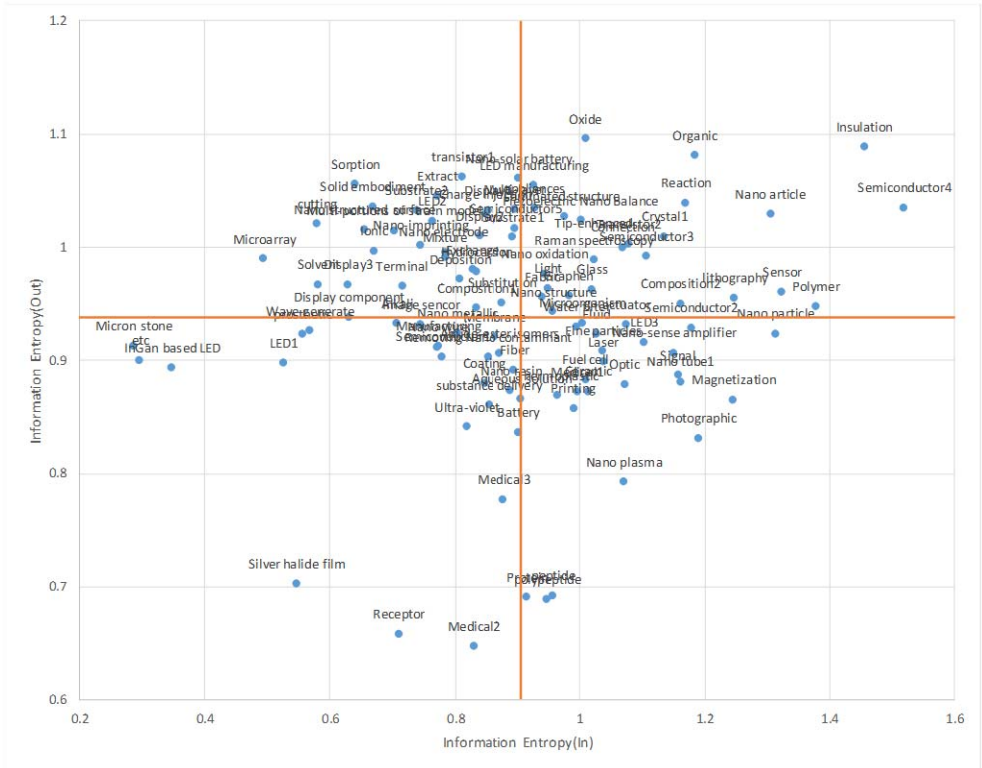
기술 분야	흡수	전파	기술 분야	흡수	전파
Semiconductor ⁴	1.518731	1.035165	Multi-layer	0.894319	1.033777
Insulation	1.455307	1.088549	Fiber	0.894131	0.891988
Polymer	1.377944	0.94796	Substrate ¹	0.890812	1.009368
Sensor	1.323068	0.960767	Nano resin	0.888084	0.873585
Nano particle	1.313535	0.923913	Medical ³	0.876997	0.777589
Nano article	1.305225	1.029682	Substitution	0.873539	0.951434
lithography	1.247343	0.955564	Amide-ester isomers	0.870814	0.906786
Magnetization	1.245435	0.865676	Membrane	0.862473	0.921031
Photographic	1.190132	0.831423	substance delivery	0.853363	0.860865
Organic	1.183683	1.081476	Removing Nano contaminant	0.852374	0.903222
Semiconductor ²	1.177511	0.929379	Display ¹	0.851659	1.032989
Reaction	1.168572	1.039365	Charge injection	0.847002	1.029265
Composition ²	1.160452	0.950536	Coating	0.84576	0.880107
Nano tube ¹	1.160122	0.881692	Display ²	0.837629	1.010104
Signal	1.1571	0.887022	Hydrocarbon	0.833079	0.978599
Nano-sense amplifier	1.15	0.906969	Composition ¹	0.832903	0.946469
Crystal ¹	1.135341	1.009827	Medical ²	0.828324	0.647802
Semiconductor ³	1.105699	0.992189	Exchange	0.827334	0.980481
LED ³	1.102871	0.916343	Ultra-violet	0.818076	0.841558
Transistor ²	1.07765	1.003462	transistor ¹	0.810256	1.062873
actuator	1.073355	0.932252	Deposition	0.806749	0.972097

Optic	1.071007	0.879546	Nano metallic	0.802478	0.924457
Nano plasma	1.069723	0.792912	Nano electrode	0.78295	0.99606
Connection	1.068543	1.000337	Mixture	0.782737	0.991323
Laser	1.037467	0.899219	Semiconductor1	0.778445	0.90391
Fine particles	1.036658	0.908755	Manufacturing	0.772618	0.912994
Fluid	1.026254	0.923126	Extract	0.770791	1.045944
Tip-enhanced Raman spectroscopy	1.021889	0.988943	Nano wire	0.769943	0.912392
Glass	1.018594	0.963065	LED2	0.762148	1.023589
Ceramic	1.013847	0.873023	Image sensor	0.743645	0.931639
Fuel cell	1.008751	0.883804	Nano-imprinting	0.743359	1.002235
Oxide	1.008572	1.095797	Substrate2	0.736894	1.032666
Microorganism	1.004028	0.93322	Terminal	0.714084	0.96558

〈표 6-2〉 기술 분야별 흡수/전파의 정보엔트로피

기술 분야	흡수	전파	기술 분야	흡수	전파
Piezoelectric Nano Balance	1.000829	1.023789	Receptor	0.708976	0.658702
Medical1	0.995012	0.872571	Alkali	0.705684	0.932649
Water filter	0.993741	0.93011	Multi-portions of strain model	0.701759	1.014824
Printing	0.989591	0.857882	Ionic	0.668605	0.997016
Graphene	0.981791	0.957118	Solid embodiment	0.666559	1.036395
Laminated structure	0.97509	1.027462	Nano structured surface	0.653311	1.015865
Thermoplastic	0.964179	0.869195	Sorption	0.638597	1.055936
peptide	0.956551	0.692208	Display component	0.628598	0.938595
Nano structure	0.956504	0.9432	Display3	0.627517	0.967269
Light	0.948905	0.964078	Solvent	0.580774	0.967227
polypeptide	0.946945	0.689438	cutting	0.579148	1.020896
Nano oxidation	0.942932	0.976553	Wave generate	0.56703	0.926224
Fabric	0.939459	0.956796	prosthetic	0.555135	0.923695
appliances	0.927377	1.034539	Silver halide film	0.546858	0.702891
LED manufacturing	0.925267	1.054933	LED1	0.525273	0.898084
Protein	0.915015	0.691191	Microarray	0.493152	0.990019
Aqueous solution	0.905381	0.86634	InGaN based LED	0.346838	0.893948
Battery	0.900897	0.836594	etc.	0.295897	0.900298
Nano-solar battery	0.900045	1.061347	Micron stone	0.285523	0.913176
Semiconductor5	0.895554	1.016979	전체 평균	0.902919	0.938493

〈그림 7〉 나노기술의 기술 확산 엔트로피 맵



V. 결론 및 추후 연구

본 연구는 기존 특허와 분류코드를 활용한 융합기술의 분석 연구의 한계점에 대하여 토픽모델링과 교차영향분석을 이용한 융합기술 분석 방법을 제시한다. 특히 기존의 연구에서 활용된 단순 서지정보가 아닌 다량의 특허를 수집하여 해당 문서의 내용을 기반으로 한 의미론적 기술 분류를 실시하며, 특허 내에서 기술 분류의 공동 출현을 이용하여 교차영향분석으로 기술 분야 간의 영향관계를 계산한다. 최종적으로 교차영향분석을 통해 작성된 기술 확산 매트릭스는 관점에 따라 다양한 기법으로 분석될 수 있다. 또한 본

연구에서 제시한 방법론의 검증에 위해 나노기술의 특허를 대상으로 사례연구를 실시하였고, 분석방법으로는 상위 영향관계를 이용한 네트워크 시각화와, 정보 엔트로피를 이용한 기술 확산 엔트로피 지도를 제시하였다. 본 연구는 정적인 특허분류체계를 이용하여 특허기반으로 융합기술을 분석한 기존의 연구와 달리 토픽모델링을 통해 유연한 기술 분류를 적용하고 교차영향분석을 통해 각 기술 분야 간의 파급을 정량적으로 비교분석이 가능하도록 하였다. 따라서 본 연구에서 제시하는 방법을 통해 기존의 분류코드를 활용한 분석의 한계점인 기술의 변화에 따라 유연한 대응 및 분석이 가능하다. 따라서 본 연구에서 제시하는 방법이 기존의 방법과 복합적으로 사용될 경우 융합기술과 관련된 R&D의 비중이 점차 높아지는 현실에서 R&D기획의 효율성을 높이는 수단으로 사용될 수 있다.

하지만 본 연구는 몇 가지 현실적인 한계점을 가진다. 우선, 텍스트데이터의 전 처리 과정에서 키워드 사전을 구축할 때 키워드 사전이 달라짐에 따라 같은 특허를 사용하여도 기술 분류가 변할 수 있다. 또한 적절한 토픽 수를 선정하는 방법에 따라 기술 분야의 개수가 변화한다. 따라서 해석하고자 하는 분야별로 적절한 키워드의 수준과 토픽 수에 대한 후속연구가 필요하다. 다음으로, 본 연구는 정보분석적 측면에서 나노기술의 세부 기술분야를 도출하고 이들 간의 영향의 분석을 시도하였기 때문에, 향후 나노기술 분야 전문가의 검토를 통해 분석된 결과가 실제 현상을 얼마나 잘 설명하는지를 세밀히 검증할 필요가 있을 것이다. 또한 기술 분야에 따라 특허-기술 분야 분포를 이진화할 때의 적절한 Threshold를 선정하는 방법에 대한 후속연구가 필요하며 추후 일반적으로 기술의 파급을 측정하기 위해 사용되는 특허의 인용정보가 본 연구에서 제시하는 방법론에 반영된다면 보다 직관적이고 정확한 해석이 가능할 것으로 보인다. 마지막으로 본 연구의 방법론에 시계열적 요소들을 반영하여 시간에 따른 기술파급의 동태를 파악하면 정밀한 분석이 가능한 것으로 판단된다.

참고문헌

〈국내 학술지〉

- 강희종 · 엄미정 · 김동명, “특허분석을 통한 유망융합기술의 예측,” 기술혁신연구 (Journal of Technology innovation) 제14권 제3호(2006).
- 김진한 · 김성홍, “교차영향분석의 적용을 통한 국내 IT 환경 시나리오에 대한 연구,” 경영과학(Korean mangement science review) 제21권 제3호(2004).
- 박현석 · 서원철 · 윤장혁, “한국 국가 R&D의 학제적 특성 분석,” 한국문헌정보학회지 (Journal of The korean society for library and information science) 제46권 제4호 (2012).
- 손영석 · 김억, “T,A,I,D,A 시나리오 플래닝 방법과 교차영향분석을 이용한 U-Eco City 개발과 미래유망기술의 R&D 전략 수립에 관한 연구,” 대한건축학회 논문집 (Journal of The architectural institute of korea planning & design) 제24권 제11호(2008).
- 송성수, “한국 과학기술정책의 특성에 관한 시론적 고찰,” 과학기술학연구(Journal of Science & Technology Studies) 제2권 제1호(2002).
- 조용래 · 김의석, “특허 네트워크와 전략지표 분석을 통한 기업 기술융합 전략 연구,” 지식재산연구(The Journal of Intellectual Property) 제9권 제4호(2014).
- 정으뜸 · 고시근 · 서원철, “기술적 파급효과 도출을 통한 기술융합 분석 연구,” 지식재산연구(The Journal of Intellectual Property) 제10권 제4호(2015).

〈해외 학술지〉

- Asan, U. & C. E. Bozdağ, and S. Polat, “A fuzzy approach to qualitative cross impact analysis,” *Omega*, Vol.32(2004).
- Bian, Y., and F. Yang, “Resource and environment efficiency analysis of provinces in China: A DEA approach based on Shannon’s entropy,” *Energy Policy*, Vol.38(2010).
- Blei, D. M. & A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *the Journal of machine Learning research*, Vol.3(2003).
- Brauers, J., and M. Weber, “A new method of scenario analysis for strategic planning,” *Journal of forecasting*, Vol.7(1988).
- Chiru, C.-G. & T. Rebedea, and S. Ciotec, “Comparison between LSA-LDA-Lexical Chains,” *WEBIST*, Vol.2(2014).
- Gould, R. V., and R. M. Fernandez, “Structures of mediation: A formal approach to brokerage in transaction networks,” *Sociological methodology*, Vol.19(1989).

- Kannathal, N. & M. L. Choo & U. R. Acharya, and P. Sadasivan, "Entropies for detection of epilepsy in EEG," *Computer methods and programs in biomedicine*, Vol.80(2005).
- Leonard, B., "Cancer nanotechnology: going small for big advances: using nanotechnology to advance cancer diagnosis, prevention and treatment," *DIANE Publishing*(2009).
- Lim, Hyojeong, and Yongtae Park, "Identification of technological knowledge intermediaries," *Scientometrics* 84,3(2010).
- Moreau, W. M. "Semiconductor lithography: principles, practices, and materials," *Springer Science & Business Media*, (2012).
- Ricotta, C. "Bridging the gap between ecological diversity indices and measures of biodiversity with Shannon's entropy: comment to Izsák and Papp," *Ecological Modelling*, Vol.152(2002).
- Roco, M. C. & R. S. Williams, and P. Alivisatos, "Nanotechnology Research Directions: IWGN Workshop Report: Vision for Nanotechnology in the Next Decade," *Springer Science & Business Media*(2000).
- Shannon, C. E., "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol.5(2001).
- Shmulewitz, A. & R. Langer, and J. Patton, "Convergence in biomedical technology," *Nature Biotechnology*, Vol.24(2006).
- Siegel, R. W., "Creating nanophase materials," *Scientific American*, Vol.275(1996).
- Soleimani-Damaneh, M., and M. Zarepisheh, "Shannon's entropy for combining the efficiency results of different DEA models: Method and application," *Expert Systems with Applications*, Vol.36(2009).
- Spatz, J. P. & V. H. Chan & S. Mößmer & F. M. Kamm & A. Plettl & P. Ziemann, and M. Möller, "A combined top-down/bottom-up approach to the microscopic localization of metallic nanodots," *Advanced Materials*, Vol.14(2002).
- Straathof, S. M., "Shannon's entropy as an index of product variety," *Economics Letters*, Vol.94(2007).
- Turoff, M. "An alternative approach to cross impact analysis," *Technological forecasting and social change*, Vol.3(1972).
- Verspagen, B., "Measuring intersectoral technology spillovers: estimates from the European and US patent office databases," *Economic Systems Research*, Vol.9(1997).
- Wang, B. & S. Liu & K. Ding & Z. Liu, and J. Xu, "Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology," *Scientometrics*, Vol.101(2014).

A Semantic Patent Analysis Approach to Identifying Trends of Convergence Technology: Application of Topic Modeling and Cross-impact Analysis

Jeong Byeongki, Kim Jungwook & Yoon Janghyeok

As technologies have been increasingly complex and dynamic, barriers among different technologies are becoming blurred. Although studies based on patent classification codes have been proposed to analyze trends of convergence technologies, their limitations are caused by the difficulty to deal with newly emerging or converging technological topics. As a remedy, this paper proposes a semantic approach to identify trends of convergence technologies, by applying topic modeling and cross-impact analyses to massive patents related to a convergence technology. The proposed approach applies topic modeling analysis to the textual data of the patents to determine semantic technology subjects within a technology for analysis, and then a cross-impact analysis to identify technological diffusion relationships among pairs of the technology subjects. Using the technology spreading matrix produced by the two techniques, this paper generates a network visual to show the interactive relationship among sub-technology subjects in the convergence technology and carries out technological diffusion analyses using the information entropy. This approach is illustrated using the patents related to nanotechnology, which is a representative technology with a variety

of convergence applications. This approach allows experts to identify the appearance and relationship of new technologies or converged technologies that may not be recognized by patent classification codes, so it is expected that this approach would assist technology experts in monitoring sub-technologies and their technological spreading within a convergence technology field.

Keyword

Convergence Technology, Technology Monitoring, Topic Modeling, Cross-Impact Analysis, Nanotechnology, Patent Analysis