

논문의 균집화를 통한 과학기술정책 현안 도출 관련 연구

심위* · 권오진** · 배영철*** · 문영호****

I. 서론

과학기술정책은 한 국가의 과학기술, 나아가 산업 발전의 방향을 제시하는 이정표 역할을 수행한다(Dasgupta Parthaa & Paul A. David, 1994). 정책의 수립에는 대중의 니즈, 학술적 근거, 제반 환경 분석 등 다양한 사전 조사와 준비가 선행되어야 하며, 따라서 정책 관련 연구를 통해 관련 현안을 도출하는 것은 급변하는 과학기술 R&D와 산업 환경에 대한 대비책으로 활용될 수 있다(Gilbert S. Omenn, 2006).

과학기술정책 관련 연구는 다양한 측면에서 시도되어 왔다. Leonard L. Lederman(1987)은 Funding sources를 통해 여러 국가의 과학기술정책을 비교하며 유사성과 차이성을 도출하였고, Schoening, NC et al.(1998)은 미국, 영국, 대만과 한국을 대상으로 과학기술정책이 산업에 끼치는 영향력을 분석하였다. 개별 국가의 과학기술정책 관련 연구 역시 다양한 방법으로 이루어져 왔으며(Watanabe, C & Clark, T, 1991; Sanzmenendez, L., 1995; Uzun, A., 2006; Huang, Cui et al., 2015), 데이터를 매개로 한 과학기술정책 연구도 활발히 진행되고 있다. D. Crouch, J. et al.(1986)은 영국 Advisory Board for the Research Councils의 해류와 단백질 결정학 분야 데이터를 통해 정책 결정과 평가에 있어 과학계량학적 방법의 장점을 역설하였고, Koensraad Debackere, Wolfgang Glänzel(2004)은 과학기술정책 수립에 있어 서지계량학적 방법을 통한 지원의 가능성을 시사하였다. 최근 연구로는 1949년부터의 중국 과학기술정책 데이터베이스를 사용하여 서지계량학적 방법으로 정책의 흐름을 도출한 Huang, Cui et al.(2015), 수정 델파이 기법을 사용하여 미래 이슈와 과학기술정책의 효용성 분석을 시도한 Parker, M. et al.(2014) 등이 있다.

이처럼 과학기술정책을 대상으로 다양한 관점에서의 연구가 이루어지고 있으나, 논문 데이터베이스의 균집화를 통한 정책 현안 도출 연구는 아직 미미한 실정이다. 따라서 본 연구에서는 인용 관계를 통해 특정 학술지에 실린 논문을 균집화한 후, 현재 이슈가 되고 있는 과학기술정책의 현안을 살펴보고자 한다.

II. 데이터 및 연구방법론

1. 데이터

본 연구의 주제인 과학기술정책 현안 도출을 위해 사용한 데이터는 다음과 같다. 해당 주제를 살펴봄에 있어 가장 적합한 성격의 학술지라 할 수 있는 Research Policy 저널의 2011~2014년 전체 논문 서지정보를 수집하였으며, 각 논문의 고유번호, 출판연도, 제목, 요약, 인용정보 등의 필드를 재구축하였다.

* 심위, 과학기술연합대학원대학교 박사과정, 02-3299-6150, sw@kisti.re.kr

** 권오진, 한국과학기술정보연구원 책임연구원, 02-3299-6097, dbajin@kisti.re.kr

*** 배영철, 전남대학교 교수, 061-659-7315, ycbae@chonnam.ac.kr

**** 문영호, 한국과학기술정보연구원 책임연구원, 02-3299-6090, yhmoon@kisti.re.kr

<표 1> 논문 데이터셋의 서지정보

eid	Title	Abstract	Publication Year	Reference List
e1	Introduction to...	Some of the re...	2011	r1;r2;r3;...
e2	The impact of...	This paper cont...	2012	r1;r3;r4...
e3	The mecha...	This paper invest...	2011	r2;r5;r6...
e4	Do competitiv...	This paper analiz...	2014	r4;r7;r8...
e5	Rickety numb...	The most popu...	2012	r9;r10;r11...
e6	The implications...	Concerns have be...	2013	r1;r12;r13...
e7	Divergent paths...	This paper inve..	2011	r5;r12;r14...
.
.
.

총 545편의 논문 데이터셋이 구축되었으며, 이를 바탕으로 Bibliographic coupling(M. M. Kessler, 1963) matrix를 도출하였다. Bibliographic coupling matrix는 논문의 참고문헌을 바탕으로 해당 논문들의 동시발생 빈도를 나타내는 행렬이라 할 수 있다. 즉 하나의 논문과 다른 논문이 같은 참고문헌을 공유하고 있을 때 둘 간의 관계가 형성된 것으로 간주하고 이를 수치화한 것이다.

<표 2> Bibliographic coupling matrix의 예시

Co-occurrence (eid * eid) Cross with : Reference	e1	e2	e3	e4	e5	...
e1	9	2	0	0	2	...
e2	2	15	0	1	0	...
e3	0	0	25	0	1	...
e4	0	1	0	7	0	...
e5	2	0	1	0	25	...
.
.
.

이 과정에서 다른 논문과 어떠한 관계도 형성되지 않은 논문(Isolated nodes) 257편을 분석 대상에서 제외시킨 후, 다른 논문과 참고문헌을 공유하고 있는 논문 288편을 최종 분석 대상 데이터셋으로 설정하여 분석을 수행하였다.

2. 연구방법론

논문 데이터셋을 통하여 참고문헌을 매개로 한 동시발생행렬을 도출한 후, 과학기술정책의 현안을 좀 더 명확하고 용이하게 살펴보기 위해 논문의 군집화(Clustering)를 수행하였다. 군집화를 위해 Modularity Clustering 기법 중 Louvain method(Vincent D. Blondel et al., 2008)를 사용하였으며, Louvain method는 다음 단계를 거쳐 군집을 생성한다.

- 1단계 : 각 노드에 대해 클러스터를 할당한 후, 특정 클러스터를 인접한 클러스터에 편입시킨다. 이 때

계산한 modularity 값이 종전의 값보다 커지면 현재의 클러스터를 유지하고, 종전의 값보다 작다면 다시 원래대로 회귀한다. 이러한 계산을 특정 클러스터의 모든 인접 클러스터에 대해 수행하며, 또한 모든 클러스터에 대해 수행한다. 이 때 modularity의 변화량을 구하는 공식은 다음과 같다.

$$\Delta Q = \left[\frac{\Sigma_{in} + k_{i,in}}{2m} - \left(\frac{\Sigma_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\Sigma_{in}}{2m} - \left(\frac{\Sigma_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right]$$

- Σ_{in} : 특정 클러스터 내부에 연결된 링크 weight의 총합
- Σ_{tot} : 특정 클러스터에 인접한 링크 weight의 총합
- k_i : 노드 i 에 연결된 링크 weight의 총합
- $k_{i,in}$: 특정 클러스터 내 노드 i 에 연결된 링크 weight의 총합
- m : 전체 네트워크에 존재하는 링크 weight의 총합

모든 클러스터에 대해 편입을 반복하며 modularity의 변화량을 계산하고, 더 이상 modularity 값이 증가하지 않을 때 개별 노드에 대한 1단계 군집화를 종료한다.

· 2단계 : 1단계에서 생성된 군집을 하나의 노드로 간주하고, 해당 군집에 대하여 1단계의 계산을 되풀이한다. 즉, 클러스터의 수를 더욱 줄이는 방향으로 클러스터링이 진행되며, 마찬가지로 더 이상 modularity 값이 증가하지 않을 때 연산을 종료한다. 이 때 modularity 값을 구하는 수식은 다음과 같다.

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

- A_{ij} : i 와 j 를 연결하는 링크 weight
- k_i : $\left(= \sum_j A_{ij} \right)$ 노드 i 와 연결되어 있는 링크 weight의 총합
- c_i : i 가 포함되어 있는 클러스터
- $\delta(c_i, c_j)$: $c_i = c_j$ 일 때 1, $c_i \neq c_j$ 일 때 0의 값을 가지는 함수
- m : $\left(= \frac{1}{2} \sum_{i,j} A_{ij} \right)$ 전체 네트워크에 존재하는 링크 weight의 총합

이를 단순화하여 다음과 같이 표현할 수 있다(한치근, 조무형, 2012).

$$Q = \sum_i (e_{ij} - a_i^2)$$

- e_{ij} : 클러스터 i 와 j 를 연결하는 링크의 비율
- a_i : 클러스터 i 에 연결되어 있는 링크의 비율

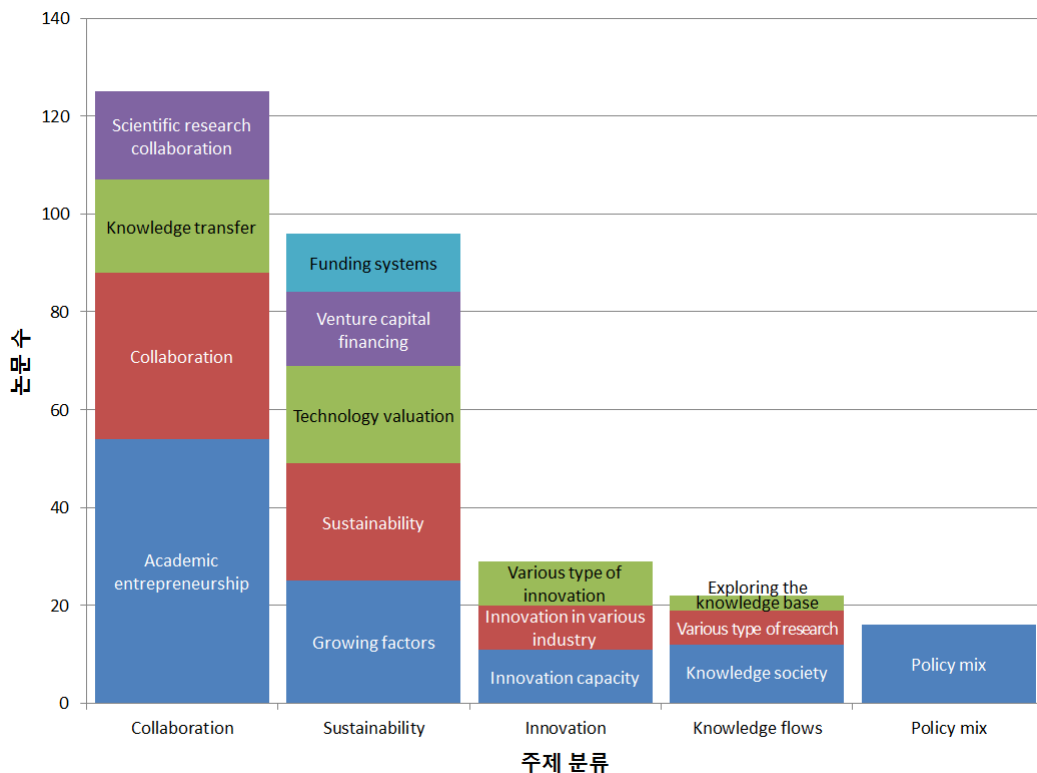
이상의 방법론을 도식화한 그림은 다음과 같다.

총 16개의 군집이 형성되었으며, 각 군집에 포함된 논문의 제목과 초록을 통해 해당 군집의 성격을 해석하였다.

<표 3> 16개 논문 군집의 해석

Cluster Number (# of Papers)	Cluster Name	Keywords
1 (54)	Academic entrepreneurship	Breaking the Ivory Tower
		University IPR regulations
		The impact of academic patenting
		Bayh-Dole act
		Information-sharing in academia
2 (34)	Collaboration	Commercialized inventions in university
		Co-opetition
3 (25)	Growing factors	Exploration versus exploitation
		Triple helix
		Market failure
4 (24)	Sustainability	Proactive R&D management
		Transitions
5 (20)	Technology valuation	niches
		Patent citations networks
6 (19)	Knowledge transfer	Essential patent claims
		International scientist mobility
		developing country
7 (18)	Scientific research collaboration	the Mexican farmers and academic researchers
		Interdisciplinary research
8 (16)	Policy mix	Open source in the firm
		Impact of environmental policies
9 (15)	Venture capital financing	Missions-oriented RD&D institutions in energy
		High-tech start-ups
10 (12)	Innovation capacity	Patent trolls
		Sources of innovation
		A comparison between the manufacturing and service sectors
11 (12)	Funding systems	the effect of employee diversity on innovation
		A quantitative assessment
		Coordination modes
12 (11)	Knowledge society	User organisations as co-producers
		The emerging knowledge base
13 (9)	Various type of research	The destabilisation of existing regimes
		The difference between men and women in research collaborations
		Knowledge originating in other technological domains
14 (9)	Exploring the knowledge base	Mobilizing for change
		Knowledge flows
15 (7)	Innovation in various industry	Core literature of innovation
		Electricity sector liberalisation
		Knowledge flows in the solar photovoltaic industry
		nanotechnology
16 (3)	Various type of innovation	The thin film transistor-liquid crystal display industries
		The U.S. auto industry
		The importance of institutional characteristic
		non-high-tech vs. high-tech SMEs

가장 많은 논문으로 이루어져 있는 군집은 “Academic entrepreneurship”이며, 이는 현재 정책적인 산·학 연계의 필요성이 대두되고 있음을 시사한다. “Collaboration” 군집의 Triple Helix 역시 산·학·연의 협력을 제언하는 의미의 키워드라 할 수 있으며, 더불어 Co-opetition, Exploration versus exploitation는 기업 간 협력의 형태를 여러 관점에서 보여주고 있다. 도출된 군집은 또 다시 몇 가지 주제로 나누어볼 수 있는데, “Academic entrepreneurship”과 “Collaboration”, “Knowledge transfer”, “Scientific research collaboration” 등의 군집은 주체 간 관계 형성을 통한 상생의 의미를 가진다. “Growing factors”, “Sustainability”, “Technology valuation”, “Venture capital financing”, “Funding systems” 등의 군집은 새로운 먹거리 창출을 위한 성장동력 발굴에 초점을 맞추고 있으며, “Innovation capacity”, “Innovation in various industry”, “Various type of innovation” 등 혁신 관련 군집도 살펴볼 수 있다. “Knowledge society”, “Exploring the knowledge base”, “Various type of research” 군집은 지식의 흐름을 통한 사회적 변화를 주제로 삼고 있으며, “Policy mix” 군집을 통해 다양한 정책 수립과 방향 설정 역시 현재 과학기술정책 분야의 이슈로 자리잡고 있음을 알 수 있다.



(그림 3) 군집의 재분류

본 연구에서는 논문의 Bibliographic coupling과 Modularity clustering 기법 중 하나인 Louvain method을 통해 과학기술정책의 현안을 도출하고자 하였다. 분석 결과, 총 16개의 군집이 형성되었으며 이를 5개의 분류로 재구성하여 현재 정책적 이슈가 되고 있는 사안들을 살펴보았다. 가장 많은 연구가 이루어지고 있는 분야는 Collaboration이며, Sustainability, Innovation, Knowledge flows, Policy mix 등이 그 뒤를 잇고 있다. 이는 곧 지식의 교류와 협력을 통해 혁신을 이루고, 이를 바탕으로 지속 가능한 성장을 이끌어가야 하며, 다양한 정책을 통해 그 성장이 뒷받침되어야 한다는 것으로 요약해볼 수 있을 것이다.

참고문헌

- 한치근, 조무형(2012), “대규모 네트워크에서 Modularity를 이용한 향상된 커뮤니티 추출 알고리즘”, *Journal of Korean Society Internet Information*, 13(3): 75-82.
- Cui Huang, Jun Su, Xiang Xie, Xuanting Ye, Zhang Li, Alan Porter, Jiang Li(2015), “A bibliometric study of China’s science and technology policies: 1949–2010”, *Scientometrics*, 102(2): 1521-1539.
- D. Crouch, J. Irvine, B. R. Martin(1986), “Bibliometric analysis for science policy: An evaluation of the United Kingdom's research performance in ocean currents and protein crystallography”, *Scientometrics*, 9(5-6): 239-267.
- Dasgupta Parthaa, Paul A. David(1994), “Toward a new economics of science”, *Research Policy*, 23(5): 487-521.
- Gilbert S. Omenn(2006), “Grand Challenges and Great Opportunities in Science, Technology, and Public Policy”, *Science*, 314(5806): 1696-1704.
- Koenraad Debackere, Wolfgang Glänzel(2004), “Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key”, *Scientometrics*, 59(2): 253-276.
- Leonard L. Lederman(1987), “Science and Technology Policies and Priorities: A Comparative Analysis”, *Science*, 237(4819): 1125-1133.
- M. M. Kessler(1963), "Bibliographic coupling between scientific papers", *American Documentation*, 14(1): 10-25.
- Parker M, Acland A, Armstrong HJ, Bellingham JR, Bland J, Bodmer HC, et al.(2014), “Identifying the Science and Technology Dimensions of Emerging Public Policy Issues through Horizon Scanning”, *PLoS ONE*, 9(5): e96480.
- Sanzmenendez, L.(1995), “Policy choices, institutional constraints and policy learning: The Spanish science and technology policy in the eighties”, *International Journal of Technology Management*, 10(4-6): 622-641.
- Schoening, NC, Souder, WE, Lee, J, Cooper, R(1998), “The influence of government science and technology policies on new product development in the USA, UK, South Korea and Taiwan”, *International Journal of Technology Management*, 15(8): 821-835.
- Uzun, A(2006), “Science and technology policy in Turkey. National strategies for innovation and change during the 1983-2003 period and beyond”, *Scientometrics*, 66(3): 551-559.
- Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre(2008), “Fast unfolding of communities in large networks”, *Journal of Statistical Mechanics*, P10008.
- Watanabe, C & Clark, T(1991), “Inducing technological innovation in Japan – The mechanisms of Japan industrial-science and technology policies”, *Journal of scientific & industrial research*, 50(10): 771-785.