論文 / 著書情報
Article / Book Information

題目(和文)	技術経営と政策の科学への計量書誌分析の応用：モノのインターネット関連技術を対象とした事例研究
Title(English)	Application of Bibliometrics for Management of Technology and Science Policy: Case study of Internet of Things related technologies
著者(和文)	高野泰朋
Author(English)	Yasutomo Takano
出典(和文)	学位:博士(技術経営), 学位授与機関:東京工業大学, 報告番号:甲第10577号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:梶川 裕矢,安藤 真,橋本 正洋,日髙 一義,坂田 一郎,仙石 愼太郎 ,辻本 将晴
Citation(English)	Degree:Doctor (Management of Technology), Conferring organization: Tokyo Institute of Technology, Report number:甲第10577号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	要約
Type(English)	Outline

Application of Bibliometrics for Management of Technology and Science Policy: Case study of Internet of Things related technologies

by

Yasutomo Takano

Department of Innovation
Graduate School of Innovation Management
Tokyo Institute of Technology

Outline

Abstract

Bibliometrics is a methodological research field that studies quantitatively the bibliographic material. Technology forecasting and detecting emerging technology using bibliometrics are basis for policy making. However, it contributed less in the federal or academic science policy context: tools are adopted to a small extent, the degree of usage is limited, and they are not practical so far. Therefore, the objective of this thesis is to propose new applications of bibliometrics for management of technology and science policy. Four methods are proposed: method to enhance maximum component of patent network, method to detect innovation seeds although research trends are complicated, method to assess commercialization possibility of innovation seeds, and method to detect plausible applications of innovation seeds. IoT-related technologies are undertaken as analysis target to validate the effectiveness of proposed methods. They allow R&D managers and policy makers to focus on new applications of innovation seeds with possibility to be commercialized in the future by using the integrated method.

Acknowledgement

Since coming to the Tokyo Institute of Technology, I have been incredibly fortunate to have the support of an outstanding network of mentors, collaborators, and friends, who beyond their contributions to this report, helped me expand the horizons of my knowledge.

I am especially grateful to Dr. Yuya Kajikawa, my thesis supervisor, for his outstanding mentorship. I am deeply thankful to Kajikawa Sensei being so generous with his time, for his always-helpful feedback, and most of all, for pushing me to ask questions and pursue ideas far beyond.

I appreciate all thesis committee members: Dr. Makoto Ando, Dr. Ichiro Sakata, Dr. Kazuyoshi Hidaka, Dr. Masahiro Hashimoto, Dr. Masaharu Tsujimoto, and Dr. Shintaro Sengoku. I show my appreciation for them who spend their time to read, check my thesis and give great feedbacks and comments on it.

I was blessed with a nice and friendly environment and colleagues. I am grateful to all of them. This doctoral course would not have been the same without amazing group of fellow students in Kajikawa Lab. Intensive and candid discussion in the seminar accelerated my growth. Above all, I am especially grateful to Mr. Ichiro Okabe, Mr. Cristian Mejia, and Mr. Mizutomo Takeuchi for many technical discussions and intellectual suggestion through working group activities on patent, network theory, GLM, and so on. For camaraderie, I would like to show my gratitude for Dr. Kotaro Tamura for insightful suggestions on the application of graph theory to sociology. I also thank to Mr. Masaki Hattori who supported data collection, and Mr. Osamu Niikawa who greatly supported linkage analysis. Finally, I show my best appreciation to Takenaka Scholarship Foundation, fellowships from Japan Society for Promotion of Science, Mr. Akio Tanaka, and my family, who always supported me.

Until master student, I wanted to be somebody. I wished to do something great, but did not know what to do, and I knew that I could not to do. How to find what to do to be somebody? This was my question. This question started after finished research of antenna for HAYABUSA2 (an asteroid explorer). To find some hits to answer the question, I spent long time while I did not do research at all for Ando-sensei, which I always felt sorry. I was suffered from trade off between staying one point to accumulate achievements and exploring other points where I believed it would be better. This might be similar to make sand mountain in the park. You have sand on your hand, and drop the sand from your hand to the ground. You can choose the position of your hand. If you keep staying same position, the sand mountain might get height. If you keep moving different positions, the sand mountain might get wide. I was a boy who tried to get a wide mountain while feeling inferiority complex by looking tall mountains made by my friends. In June 2013, I was in a class offered by Kajikawa-sensei, and I finally find a hit. This thesis might have started since that very moment. This thesis shows some methodologies to set target of technological option, which I think that my experience to find the answer my question could be reflected. I know that I could not give best performance, but I could keep facing my research as a whole. I heard that this attitude is called “grit”. Just after the next day of the final exam, I read a book “Grit: The Power of Passion and Perseverance” written by Dr. Angela Duckworth. Then I found following sentence: “The challenge of writing is to see your horribleness on page, to see your terribleness, and then to go to bed and wake up the next day and take that horribleness and that terribleness and refine it and make it … not so terrible and not so horrible. And then go to bed again and come the next day and refine it a little bit more and make it not so bad. And then go to bed the next day and do it again and make it … maybe average. And then one more time, you know, if you’re lucky, maybe you get to good. And if you’ve done that, that’s a success. - *Ta-Nehisi Coates, 2015 MacArthur Fellow.*” Every time when I get feedback from Kajikawa-sensei, I found my horribleness on page, to see my terribleness, and then tried to change the situation to get closer to the better situation. It was quite hard for me. However, I personally could enjoy the whole process of writing this thesis since I know that there is his love and responsibility behind. I do not know whether my thesis reached to a success. I am in between a sense of achievement and some regrets on my ability, but I would like to allow myself because my challenge is just about to start.

1. Introduction

1.1. Bibliometrics in the field of Management of Technology

Management of technology (MOT) links engineering, science, and management disciplines to plan develop, and implement technological capabilities to shape and accomplish the strategic and operational objectives of an organization [1]. There are some terminologies such as technology management (TM), engineering management (EM), technology innovation management (TIM). MOT has expanded with great speed over the last two decades since the US National Research Council (1987) mentioned about MOT [2]. Today, many countries give "a lot of importance to this field with a strong emphasis on the connection with research & development (R&D)" [3-5].

MOT is a multidisciplinary research [2]. Currently the domain could be divided into 45 subdomains. Most of them are theoretical researches (e.g., open innovation, new product development, technology transfer) and quite few are methodological researches (e.g., product portfolio management, roadmap, design structure matrix). Among these sub domains, ***bibliometrics*** is the most active research target and its most-cited research is relatively new (see Appendix). It is a methodological research field that studies quantitatively the bibliographic material [6]. By applying bibliometrics, it is possible to detect emerging technologies computationally (e.g., text mining, citation network analysis). Objective result could be obtained in short time compared to expert-based approach (e.g., interview, delphi) [7].

1.2. Bibliometrics in the field of Science of Science Policy

The Science of Science Policy (SoSP) is an emerging interdisciplinary field of research [8] and a community of practice that seeks to "provide a scientifically rigorous and quantitative basis for science policy" [9]. Previously, several bibliometric tools or methods have been proposed for SoSP. For an example of tools for SoSP, STAR METRICS (Science and Technology for America's Reinvestment—Measuring the EffecTs of Research on Innovation, Competitiveness, and Science) [10, 11] has been developed. The goal of STAR METRICS is allow the scientific community to understand and describe how, where, and why investments yield value to science and to the nation [8]. The program has "created several pilot tools to respond to the needs of agency officials, policy makers, and researchers" [8]. The particular focus has been to examine researchers' grant proposals, and thus the system applies topic modeling [12] to analyze the proposals. As methods for SoSP, Moed and Halevi [13, 14] summarized types of research outputs, impacts and indicators are surveyed. Some example are "indicators based on publications and citations in peer-reviewed journals and books" that represent "contribution to scientific-scholarly progress: creation of new scientific knowledge." The another is "(inter)national collaborations including co-authorships; participation in emerging topics." They are indicators of "integration in (inter)national scientific-scholarly networks and research teams." These tools or methods are good for obtaining technology options. However, the evaluation of researches or projects tends to be *ex-post*. Therefore, the method to assess and select the specific technology options *ex-ante* is required. For instance, Internet of Things (IoT) is a quite new and an emerging technology. The assessment is still under the progress. IoT is a broad technology [15] and its concrete application, value, and feasibility are still obscure. However, IoT has started to be recognized as an essential emerging technology among various stakeholders. In industry, promotional organizations such as the Industrial Internet Consortium (IIC) [16] have been founded. It is estimated that the market of the IoT will have a total potential economic impact of US$ 3.9-11.1 trillion per year by 2025 [17]. There are several other reports [18, 19], but in general they have described the IoT as having strong impact on industries and economies in the near future. Academia has also focused on IoT, and the number of papers including the terms "Internet of Things" or "IoT" has been increasing exponentially. A total of 197 papers with these terms were published before 2010, and the number increased to 1,870 by 2014 in the Web of Science Core Collection. Because the potential is huge, governments consider IoT as one of the main technologies that should be strategically promoted. For instance, the Council for Science, Technology, and Innovation (CSTI) devoted much attention to IoT in the fifth Science and Technology Basic Plan (fiscal 2016-2020) [20]. Another indication is that IoT became a main topic in the World Economic Forum in January 2016 [21]. Therefore, IoT has been highly expected in the society, and thus, it is important to comprehend IoT trends.

1.3. Previous literature of Bibliometrics for MOT and SoSP

In the subsection 1.1, bibliometrics was extracted as a most active subdomain of MOT. In this subsection, previous literature of the subdomain for MOT and SoSP is described. Watts and Porter [22] utilized various bibliometric methods: technology life cycle indicators, innovation context indicators, product value chain indicators. In their paper, the role of technology forecasting was introduced: help management make better decisions. Behkami and Daim [23] proposed a mixed-method approach: content analysis, cluster analysis, keyword analysis, social network analysis. The method was for technology forecasting, and authors introduced it as a tool to improve policy planning and implementation. Lee and Su [24] studied the network and knowledge map that cover country, institute, paper and keyword as network actor. The map could allow potential quantitative applications in the policy science context such as for research and development (R&D) allocation, and for understanding of future research opportunity. Kajikawa et al. [25] applied citation network analysis for energy research to track emerging research domains. They mentioned the importance for R&D managers and policy makers to notice emerging research domains because growth more reliably depends on the application of new science and technology. Furukawa et al. [26] proposed a method to analyze chronological changes in research topics. Authors emphasized that policy makers and administrators at the national levels attention to strategic R&D investment in emerging technologies enhance industrial competitiveness. Haegeman et al. [27] summarized approaches in Future-oriented Technology Analysis (FTA). Demands of FTA especially from policy makers were introduced because they are in search for effective polices. Many researchers have been developed bibliometric methods to detect emerging technologies due to the considerable and growing interest on it, especially from the policy-maker [28]. However, the definition was differs among researchers. Hence, Rotolo et al. [28] reviewed previous literature and define what is an emerging technology. In their paper, methods to detect emerging technologies were also summarized.

Details of previous literature of bibliometrics for MOT and SoSP are shown in Table 1.1. Recapitulate, researchers regard ***technology forecasting*** and ***detecting emerging technology*** using bibliometrics as a basis ***for policy making***.

1.4. Previous Bibliometrics studies for IoT-related technologies

There is some previous literature that has applied bibliometrics to forecast or comprehend technological trends. Table 1.2 summarizes the IoT-related literature that has applied bibliometric methods. These studies are divided into two categories based on document type: article-based studies and patent-based studies. Article-based studies use article papers obtained from such databases as Science Citation Index (SCI), Social Sciences Citation Index (SSCI), and Conference Proceedings Citation Index (CPCI)-Science/Social Sciences and Humanities. Previous article-based studies studied IoT-related article-based studies include information technology, IoT, RFID, and pervasive or ubiquitous computing with citation analysis [29-32] and co-word analysis [33]. For an example, Yan et al. [33] utilized co-word analysis to explore papers in the field of IoT to examine the scientific development in the area. 758 papers were divided into seven clusters that represent the intellectual structure of IoT: “IoT and Security”, “Middleware”, “RFID”, “Internet”, “Cloud computing”, “Wireless sensor networks” and ”6LoWPAN.” The extracted clusters are too broad to comprehend the current status of IoT domain. Patent-based studies use patents obtained from such databases as the State Intellectual Property Office of the People’s Republic of China (SIPO), EPO Worldwide Patent Statistical Database, also known as PATSTAT, United States Patent and Trademark Office (USPTO), Japan Patent Office (JPO), and World Intellectual Property Organization (WIPO). Previous studies apply citation analysis [34], K-means and S-curve [35], citation network analysis [36, 37], text mining [37], non-exhaustive overlapping clustering [38], and fuzzy ontological knowledge document clustering [39]. Scope of analysis in patent-based studies on technologies relating IoT includes only

RFID as shown in Table 1.2. For an example, Trappey et al. [35] combined patent content clustering (K-means) and technology life cycle forecasting to find a niche space of RFID technology development in China. After the clustering, they detected that "RFID frequency and waves is still in the early growth stage with the most potential for further development" by observing the number of patents in the domain. In summary, previous bibliometric studies in IoT-related technologies tend to only focus on a part of IoT related technologies, and they were ***not comprehensive***. Data utilized for the analysis is either only papers or only patents. Moreover, these studies offer technology options but ***assessment of the options tends to be out of scope***.

Table 1.1. Summary of previous literature of bibliometrics for MOT and SoSP

Method	Relation with SoSP	Reference
Utilize various bibliometric methods: technology life cycle indicators, innovation context indicators, product value chain indicators	"***Technological forecasting*** purports to provide timely insight into the prospects for significant technological change. Such information should ***help management make better decisions*** with regard to strategic corporate planning, R&D management, product development, investment in new process technology, production and marketing, purchasing of new technology, and so forth."	Watts & Porter [22]
A mixed-method approach: content analysis, cluster analysis, keyword analysis, social network analysis	"Due to the rapid pace of change in technology and its impact on society, there is an increasing demand for use of ***Technology Forecasting*** methods to ***improve policy planning*** and implementation."	Behkami & Daim [23]
The network and knowledge map: choosing different information for the network actor (i.e. country, institute, paper and keyword)	"The ***knowledge maps*** obtained quantitatively allow other potential quantitative applications, e.g. ***1) R&D resource allocation, 2) research performance evaluation, 3) understanding of future research opportunity, and 4) potential collaborator or competitor identification***."	Lee & Su [24]
Track emerging research domains in energy research by using citation network analysis	"In today's increasingly knowledge-based economy, growth more reliably depends on the application of new science and technology. Therefore, ***for R&D managers and policy makers***, ***noticing emerging research domains*** among numerous academic papers has become a significant task."	Kajikawa et al. [25]
Method to analyze chronological changes in research topics: using a text-mining technique for proceedings papers and conference sessions	"Strategic research and development (R&D) investment in ***emerging technologies*** an effectively enhance industrial competitiveness, attracting significant attention not only from industrial and business executives but ***also from policy makers and administrators*** at the national level."	Furukawa et al. [26]
Summarize quantitative and qualitative approaches in Future-oriented Technology Analysis (FTA)	"In ***FTA*** one of the main arguments for more use of quantitative approaches may come from ***policy-makers***, who, in search for effective policies, tend to aim at reducing or at least understanding the uncertainties that can alter the outcomes of their policies and try to quantify them ex ante."	Haegeman et al. [27]
Review literature to define "emerging technologies" and summarize methods to detect emerging technologies (indicators and trends, citation analysis, co-word analysis, overlay mapping, hybrid)	"There is considerable and growing interest in the ***emergence of novel technologies***, especially from the ***policy-making*** perspective."	Rotolo et al. [28]

Table 1.2. Previous bibliometrics studies for IoT-related technologies

	IoT-related target	Database	Methods
Paper-based	Information technology [29]	SCI, SSCI, CPCI, etc.	(Direct) Citation network analysis (centrality measures, path analysis, cluster analysis, flow vergence model)
	IoT [33]	SCI-Expanded, SSCI	Co-word analysis
	RFID [31]	SCI, SSCI	Citation analysis
	Pervasive / ubiquitous computing [32]	SCI	(Co-) Citation network analysis (subject category/country/ journal/author network, timeline)
	RFID [30]	SCI	Citation analysis
Patent-based	RFID [34]	PATSTAT	Citation analysis (citation to patents/non-patent literature/ pioneering innovations, and technology cycle time)
	RFID [35]	SIPO	K-means, S-curve
	RFID [36]	USTPO	Citation network analysis (centrality measures)
	RFID [38]	USPTO, JPO	Non-exhaustive overlapping clustering
	RFID [37]	USTPO	Text mining, citation network analysis
	RFID [39]	WIPO	Fuzzy ontological knowledge document clustering

1.5. Issues of Bibliometrics

Through the literature review, it was figured out that bibliometrics is an important research area, and its methods or tools have been developed for MOT and SoSP due to the demands of technology forecasting or detecting emerging technologies. It has been already applied for IoT-related technologies domain. However, the bibliometrics has issues. Regardless of such development, it less contributed in the federal or academic science policy context in the real world. Current and potential toolkit for science and innovation policy were identified in a federal research roadmap [40]. Table 1.3 shows the summary of bibliometrics related tools in the report. *Relevance to Vision* is "the degree to which the element provides a significant contribution to resolving one or more of the 10 Scientific Challenges identified by the ITG." *Breadth of Use* is "the extent of the adoption of the element in the Federal or academic science policy context." *Scientific Rigor* is "the quality of the scientific foundation of the element, in terms of publications, scientific openness, size of community and reproducibility." *Maturity* of the method or tool is "the degree to which the element is used in the Federal or academic science policy context." *Access to input* is "the practicality of using the element to develop the empirically based platform for decision making that is the goal of science policy." Each bibliometrics related tools were evaluated in those terms: green means "currently fulfilled with respect to the criterion," yellow is "on the way to fulfilling the criterion," and red stands for the status of "there are substantial gaps." As the Table 1.3 shows, bibliometrics related tools have issues especially in terms of Breadth of Use, Maturity and Access to Input. It other words, bibliometrics tools are adopted to a small extent, the degree of usage is limited, and they are not practical so far. Macilwain [41] described this situation as follows: "yet ***we lack the scientific tools and data*** to support our faith in publicly funded science or to help policy makers weigh future investments."

Table 1.3. Current status of bibliometrics related tools

- 🟩 Currently fulfilled with respect to the criterion
- 🟨 On the way to fulfilling the criterion
- 🟥 There are substantial gaps

MODELS/TOOLS		**Potential value**			**Potential cost**	
Deterministic models		Relevance to Vision	Breadth of Use	Scientific Rigor	Maturity	Access to Input
Bibliometrics related tools	Network analysis	🟩	🟥	🟩	🟥	🟥
	Visual analytics	🟩	🟥	🟩	🟨	🟥
	Science mapping	🟩	🟥	🟥	🟥	🟥
	Scientometrics	🟩	🟩	🟥	🟨	🟥

1.6. Previous approaches according to the process flow of Bibliometrics

To narrow down the issues of bibliometrics, previous approaches according to the process flow of bibliometrics are summarized in this subsection. Börner et al. [42] showed process flow for mapping knowledge domains. The processes are "(1) data extraction, (2) definition of unit of analysis, (3) selection of measures, (4) calculation of a similarity between units, (5) ordination, or the assignment of coordinates to each unit, and (6) use of the resulting visualization for analysis and interpretation". In this thesis, I generalized these procedures so that it is able to cover broad bibliometric methodologies. The generalized process flow is setting target, data retrieval, data preparation, data processing, and interpretation. Previous approaches are summarized as Fig 1.1 according to the process flow of bibliometrics.

1.6.1. Setting target

Normally, analysis target is set by human-based. It depends on objective of researchers. Sometimes, socially important elements such for realizing sustainability society are nominated as the analytic targets. Some examples are water purification, sustainable energy [43], biomass and bio-fuels [44]. On the other hand, emerging technologies or buzz words listed at Gartner's hype cycle [45] are also selected because there is no consensus on such targets in the early stage of diffusion, and thus, researchers try to apply bibliometrics to comprehend those domains or to try to detect what are the next research fronts. Some past examples were electrical conducting polymer nanocomposite [46], organic LEDs [47], polymer electrolyte fuel cells [48]. Novel biotechnologies such as induced pluripotent stem cell (iPS) or cutting edge ICT technologies such as artificial intelligence, big data, robot and IoT could be the next targets.

1.6.2. Data retrieval

Basically, networks are created from a collection of documents retrieved from database(s); the information retrieval strategy may have an impact on the size and connectivity of the network. It is needed to retrieve of the adequate dataset with appropriate quantity and quality. To improve the quantity of data, Al-Shboul and Myaeng [49] used Wikipedia for query phrase expansion. Cross language retrieval [50] and retrieval from multiple databases [14] has been also proposed. On the other hand, a noisy dataset (i.e., documents with little relatedness between them) might result in sparse networks that do not capture the full characteristics of the field in analysis. This concern is addressed by the information retrieval research, which looks for methods to increase the relatedness of documents retrieved. Therefore, to improve the quality of data, some previous studies focused on the improvement of the algorithms in the search engines [51] or on strategies for query refinement [52].

1.6.3. Data preparation and Data processing

Once the corpus of documents is obtained, several strategies can be used to establish the links between the documents. Frist, in the data preparation step, relationship of data were selected. Then, in the data processing step, the relationship is analyzed by applying algorithm(s). These construction methods are mainly based on text-based, category-based, and citation-based.

Text-based

The performance of analysis depends on the text preparation: selection of text fields, using words or phrases, the cleaning strategy of the text corpus. Then, to be able to classify the data, bag-of-words model (n-grams, unigram) [53] is commonly applied. In the model, “a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity” [54].

When selecting clustering algorithm for text-based analysis, researchers have mainly focused on two trends: one relying on text similarity measures and the other applying probabilistic models. In a large-scale experiment, Boyack et al. [55]evaluated the accuracy of nine text similarity measures for clustering 2.15 million biomedical articles. The PMRA (PubMed Related Articles) measure proved to be the best, followed by BM25 and a topic model-based measure. PMRA is a similarity measure specific to the PubMed interface; thus, its application in other fields is limited. The other two are common methods in text mining literature. Even though the aforementioned experiment was comprehensive, there was no general agreement. For instance, Hamedani and Kim [56] used a dataset of more than 1 million computer science articles to compare several similarity measures; their experiment showed that there was no significant difference between BM25 and the classic cosine similarity of tf-idf weighting. In another experiment, document classification using the tf-idf yielded better performance when the unit of comparison was phrases instead of single terms [57].

Other text-based methods rely on probabilistic models that can be estimated from text corpus. Yau et al. [58] explored the possibilities of classifying academic articles from seven research domains using topic models. Four topic model algorithms were compared, and the hierarchical Dirichlet process scored the highest performance. Gretarsson et al. [59] also exploited topic models for network creation. Their focus was more on improving the analysis of the document through visualization than on performance evaluation or clustering. However, There is still no significant evidence that will make researchers how to prepare the text and how to choose one method over the other (e.g., tuning of parameters in the algorithms).

Category-based

There are several types of category such as scientific discipline [60], international patent class (IPC) [61, 62] and Medical Subject Headings (MeSH) [63]. By using the categories, it is possible to represent the global science structure.

Raflos et al. [60] utilized ISI Web of Science subject categories that consisted of more than 250 subject areas in science, social sciences, and arts & humanities [64]. The categories linked by cosine similarity of co-citations patterns between journals. However, sometimes there is no appropriate ISI WoS subject categories for emerging topic.

IPC is the standard for patent classification. It divides technology into eight sections with approximately 70,000 subdivisions that consists of Arabic numerals and letters of the Latin alphabet [65]. In network analysis, IPC codes have been used to create networks by linking two or more patents whenever they share an IPC code. Leydesdorff [66] used this strategy to analyze a dataset of 138,751 patents, mapping networks at different levels of the IPC code for the whole dataset and for country subsets. The results, however, were not positive. Networks based on IPC made a poor representation of the technology space, and the use of other methods such as text-based networks was suggested.

MeSH terms are "provided on the basis of intensive indexing articles of more than 5,500 leading journals for the Medline/PubMed databases, a service of the National Institute of Health (NIH)" [63]. There are 27,883 descriptors in 2016 MeSH with over 87,000 entry terms that assist in finding the most appropriate MeSH Heading with 232,000 Supplementary Concept Records [67].

Leydesdorff et al. [63] linked MeSH terms by cosine similarity (based on the co-occurrence of MeSH terms at the publication level) to gain a bibliometric perspective on the dynamics of medical innovations. However, MeSH terms are only available for the medical publications, and can not apply for other desciplines.

Citation-based

Three types of citation networks can be constructed. The simplest one is direct citation, which refers to the establishment of the linkage from a document to all the cited references in it; this is repeated for each document in the dataset, resulting in a network of cited references [68]. The other two methods are bibliographic coupling and co-citation. Each possible pair of documents in a dataset is compared and bibliographically coupled when they share one or more cited references. The number of references they share is the strength of this connection [69]. Finally, co-citation connects all references cited within a document; the more two documents are cited together (co-occur) in other documents, the stronger their relation gets [70]. Details on the construction process and characteristics of these methods can be found in Shibata et al. [71]. Once a network is created by any of the three methods, tightly connected groups of documents can be identified and clustered. As stated by Waltman and Van Eck [72], intuitively, direct citation networks may convey the strongest relationship between the documents compared to co-citation and bibliographic coupling because they need a third intermediary document to establish the linkage, being indirect methods for network creation. To know which method represents research domains more accurately, previous research studies have compared the performance of these networks under different conditions. Shibata et al. [71] compared the three methods when analyzing research domains of gallium nitride, complex networks, and carbon nanotubes. In all cases, direct citation performed best based on different evaluations such as size of clusters, speed, and topological relevance. Clusters observed from the direct citation network suggested a high content similarity. Klavans and Boyack [73], who studied rather large datasets, recently conducted the most comprehensive citation network analysis to date by using a corpus of 48.5 million documents, which basically included all indexed articles in Scopus from 1996 to 2012 and other sources, and obtained the same results as those of Shibata et al. Although most of their works have favored co-citation and bibliographic coupling in the past, they now motivate future studies on citation networks to focus on direct citation.

Then, documents are connected by the citations among them and network is obtained. The network is divided into groups by clustering algorithm. There are a lot of clustering algorithms. Some examples are Newman [74], Louvain [75], and Random walk [76]. Newman's algorithm is a famous algorithm in the field of social network analysis. It discovers tightly knit clusters in large network within a feasible time period. These groups of papers were then mapped and visualized.

In contrast to networks of academic papers, patent networks may lose a considerable amount of nodes by neglecting the unconnected components [74]. When comparing the gap between academic research and patents of solar cell technologies, Shibata et al. [74] observed that, even though an increasing number of patents were filed through the years, the largest component of the patent network did not grow and remained small, connecting only 30% of the 333,207 patents retrieved. One of the reasons is the difference of citation behavior among countries. It might have a disadvantage in that they could be considered not large enough to represent a technology field adequately.

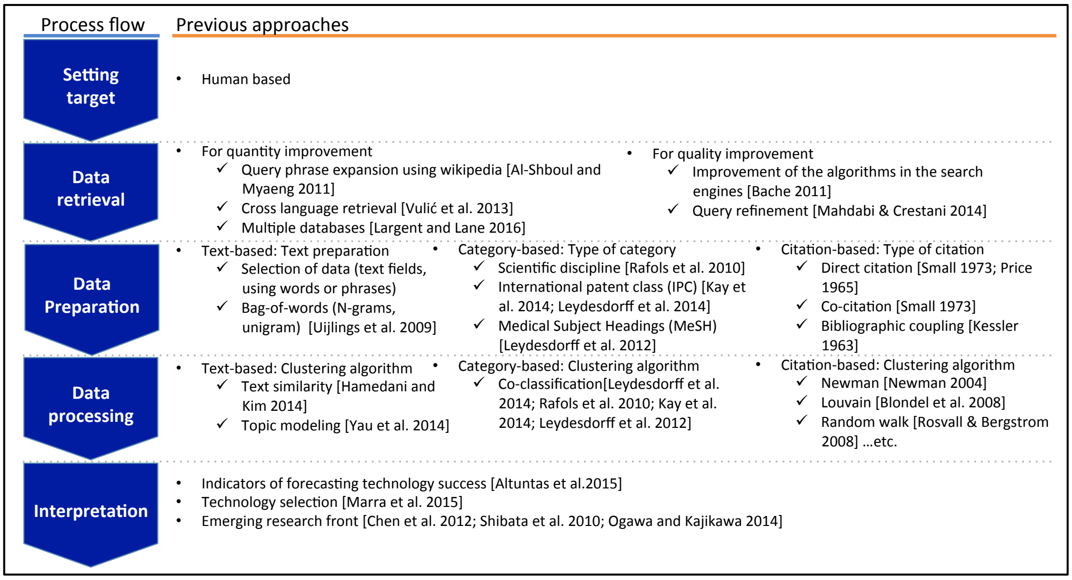
1.6.4. Interpretation

After the data processing, technology options are visualized. The outputs are suitable for comprehending the landscape of the target domain.

Some researchers use the average publication year of the papers and patents information to evaluate the level of activeness of research fields [48, 74, 79]. Chen et al. [79] studied smart grid field by applying bibliographic coupling with Girvan-Newman algorithm. While other studies on solar cell [74] and polymer electrolyte fuel cells [48] utilized direct citation with Neman algorithm. When the average year of the publications inside the clusters is younger, they are detected as emerging research fronts.

The other researchers utilized indicators to interpret the technology options. Altuntas et al. [80] discussed four indicators of forecasting technology success: technology life cycle (TLC), technology diffusion speed, patent power, and expansion potential. TLC was measured by S-curve made by accumulated number of patents according to time, and divided into initiation stage, growth stage, and saturation stage. Technology diffusion speed was defined by average number of forward citations per patent. Average number of IPC per patent is used as proxy for patent power. Finally, the total numbers of different IPC codes for the targets indicated expansion potential. These indicators were measured for thin film transistor-liquid crystal display, flash memory system, and personal digital assistant were evaluated, and personal digital assistant was selected as appropriate investment target. Meanwhile, Marra et al. [81] analyzed patent citations networks of renewable energy published between 2000 and 2013. They applied Hubs and Authorities algorithm [82] that are "are eigenvector centralities in the weighted undirected graphs constructed from a direct graph by means of bibliographic coupling and co-citation" [83]. Then, the most important contributions in the field of renewable energy were identified.

Even though these studies tried to interpret the technology options, there are limitations. First, there is a risk of missing the extraction of important clusters or group of documents when focus only on too much simple indicators (e.g., the average publication year) since the number of documents in some areas is not increasing or decreasing monotonously. Second, the dataset used in the analysis tends to cover only one domain with only one type of documents: either papers or patens. For instance, it is required to explore new applications using technologies of the domain that give solutions for other domains. Moreover, excepts the studies such above, most of previous works tend to only focus on comprehending intellectual structure or obtaining technology options of a domain. Hence, further interpretation is needed to assess the technology options.



Process flow	Previous approaches		
Setting target	• Human based		
Data retrieval	• For quantity improvement ✓ Query phrase expansion using wikipedia [Al-Shboul and Myaeng 2011] ✓ Cross language retrieval [Vulić et al. 2013] ✓ Multiple databases [Largent and Lane 2016]	• For quality improvement ✓ Improvement of the algorithms in the search engines [Bache 2011] ✓ Query refinement [Mahdabi & Crestani 2014]	
Data Preparation	• Text-based: Text preparation ✓ Selection of data (text fields, using words or phrases) ✓ Bag-of-words (N-grams, unigram) [Uijlings et al. 2009]	• Category-based: Type of category ✓ Scientific discipline [Rafols et al. 2010] ✓ International patent class (IPC) [Kay et al. 2014; Leydesdorff et al. 2014] ✓ Medical Subject Headings (MeSH) [Leydesdorff et al. 2012]	• Citation-based: Type of citation ✓ Direct citation [Small 1973; Price 1965] ✓ Co-citation [Small 1973] ✓ Bibliographic coupling [Kessler 1963]
Data processing	• Text-based: Clustering algorithm ✓ Text similarity [Hamedani and Kim 2014] ✓ Topic modeling [Yau et al. 2014]	• Category-based: Clustering algorithm ✓ Co-classification[Leydesdorff et al. 2014; Rafols et al. 2010; Kay et al. 2014; Leydesdorff et al. 2012]	• Citation-based: Clustering algorithm ✓ Newman [Newman 2004] ✓ Louvain [Blondel et al. 2008] ✓ Random walk [Rosvall & Bergstrom 2008] …etc.
Interpretation	• Indicators of forecasting technology success [Altuntas et al.2015] • Technology selection [Marra et al. 2015] • Emerging research front [Chen et al. 2012; Shibata et al. 2010; Ogawa and Kajikawa 2014]		

Fig. 1.1. Previous approaches according to the process flow of Bibliometrics

1.7. Scope and Objective of the thesis

Previous approaches according to the process flow of bibliometrics were summarized in last section to narrow down the issues of bibliometrics. In summary, there is no consensus on text-based analysis and there are disadvantages on category-based analysis such as availability. On the other hand, citation-based analysis reached consensus: direct citation is best [71-73]. Through the literature review, it was figured out that bibliometrics is an important research area in the context of MOT and SoSP where there are large demands for technology forecasting or detecting emerging technologies. Number of researches have been conducted not only to obtain technology options but also to assess the options by utilizing citation-based analysis that reached consensus in bibliometrics field. However, there are still problems that have to be overcome. First, when it comes to the patent network, the connectivity is sparse [74]. With little amount of data that could be used for data processing, the result would be not enough. It is required to improve connectivity of network so that it will be able to represent the technology field adequately. Second, the interpretation of bibliometric output needs to be improved. Previous approaches tend to rely on simple indicators that premise incremental changes, and tend to rely on dataset of single domain which means the analysis is closed only inside the technology field. There is demand to have methods to evaluate various aspects of the technology options by considering the potential of commercialization or the usages for the other domains. Such methods contribute to assess and select the technology options ex-ante.

Therefore, the scope of this thesis is to offer the solutions for above issues of citation-based bibliometircs. The objective of this thesis is to propose new applications of bibliometrics for Management of Technology and Science Policy. IoT-related technologies are undertaken as analysis target to validate the effectiveness of proposed methods.

The overall structure of this thesis takes the form of seven chapters, including this introductory chapter (Fig. 1.2). In chapter 2, the analysis targets, Internet of Things (IoT) related technologies are introduced. Also this chapter presents databases used in this thesis. Chapter 3 proposes an unconnected component inclusion technique (UCIT) for patent citation analysis. The method generates a cluster solution that includes unconnected and connected components of a direct citation network, enabling a more complete analysis of the technology fields. I observed that UCIT increased the number of nodes especially in relatively small networks. Chapter 4 applies research classification schema (RCS) to detect innovation seeds in the field of IoT related technologies. Subsequently, chapter 5 continues with the exploration of the research areas of IoT related technologies, for which there are opportunities for commercialization in the near future. Potential applications of these technologies in diverse systems are discussed. Following, chapter 6 explores plausible social issues to which IoT solutions can be applied. Semantic linkages between extracted research areas and social issues using WEHAB framework (i.e. water, energy, health, agriculture, and biodiversity) were investigated. On the basis of the results, I discuss potentials and challenges for IoT to solve social issues. Finally, chapter 7 concludes by providing a brief summary of the findings of this thesis. Then, implications, limitations, and future works are discussed.

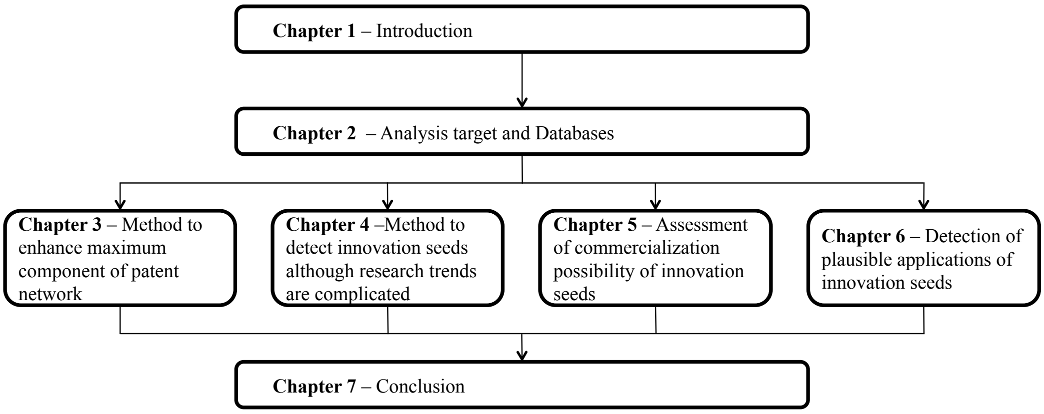


Fig. 1.2. General structure of this thesis

References

[1] Task Force on Management of Technology, National Research Council (US). Cross-Disciplinary Engineering Research Committee, and National Research Council (US). Manufacturing Studies Board. Management of technology: The hidden competitive advantage. National Academies, 1987.

[2] Lee, Hakyeon. "Uncovering the multidisciplinary nature of technology management: journal citation network analysis." Scientometrics 102.1 (2015): 51-75.

[3] Merigó, José M., et al. "Academic research in innovation: a country analysis." Scientometrics (2015): 1-35.

[4] Audretsch, David B., and Maryann P. Feldman. "R&D spillovers and the geography of innovation and production." The American economic review 86.3 (1996): 630-640.

[5] Teece, David J., Gary Pisano, and Amy Shuen. "Dynamic capabilities and strategic management." Strategic management journal (1997): 509-533.

[6] Broadus, Robert. "Toward a definition of "bibliometrics"." Scientometrics 12.5-6 (1987): 373-379.

[7] Kostoff, Ronald N., and Robert R. Schaller. "Science and technology roadmaps." IEEE Transactions on engineering management 48.2 (2001): 132-143.

[8] Largent, Mark A., and Julia I. Lane. "STAR METRICS and the Science of Science Policy." Review of Policy Research 29.3 (2012): 431-438.

[9] SoSP Central. "Science of science policy", 2016. [Online]. Available: http://scienceofsciencepolicy.net [Accessed: 9-Nov-2016]

[10] Lane, Julia. "Let's make science metrics more scientific." Nature 464.7288 (2010): 488-489.

[11] Lane, Julia, and Stefano Bertuzzi. "Measuring the results of science investments." Science 331.6018 (2011): 678-680.

[12] Newman, David, et al. "Evaluating topic models for digital libraries." Proceedings of the 10th annual joint conference on Digital libraries. ACM, 2010.

[13] Moed, H. F., & Halevi, G. The multidimensional assessment of scholarly research impact. Journal of the American Society for Information Science and Technology, 66.10 (2015): 1988–2002.

[14] Daraio, Cinzia, et al. "Data integration for research and innovation policy: an Ontology-Based Data Management approach." Scientometrics 106.2 (2016): 857-871.

[15] Whitmore, Andrew, Anurag Agarwal, and Li Da Xu. "The Internet of Things—A survey of topics and trends." Information Systems Frontiers 17.2 (2015): 261-274.

[16] Industrial Internet Consortium. "The Industrial Internet Reference Architecture Technical Paper." Retrieved in April (2016).

[17] Mckinsey.com, "Unlocking the potential of the Internet of Things", 2016. [Online]. Available: http://www.mckinsey.com/insights/business_technology/the_internet_of_things_the_value_of_digitizing_the_physical_world. [Accessed: 30- Jan- 2016].

[18] Joseph Bradley, J. Barbier, and D. Handler, Embracing the Internet of everything to capture your share of 14.4 trillion. CISCO White Paper. 2013. Available at: http://www.cisco.com/web/about/ac79/docs/innov/IoE_Economy.pdf. Accessed 29 August 2016.

[19] P. Middleton, P. Kjeldsen, J. Tully. Forecast: The Internet of Things, Worldwide, 2013. Gartner Research. 2013. Available at: https://www.gartner. com/doc/2625419/forecast-internet-things-worldwide. Accessed 29 August 2016.

[20] Council for Science, Technology and Innovation Cabinet Office, Government of Japan, Report on The 5th Science and Technology Basic Plan. Tokyo: Government of Japan, 2015, pp. 11-18 [Online]. Available: http://www8.cao.go.jp/cstp/kihonkeikaku/5basicplan_en.pdf. [Accessed: 30- Jan- 2016].

[21] World Economic Forum, "The Internet of Things Is Here", 2016. [Online]. Available: http://www.weforum.org/events/world-economic-forum-annual-meeting-2016/sessions/the-internet-of-things-is-here. [Accessed: 30- Jan- 2016].

[22] Watts, Robert J., and Alan L. Porter. "Innovation forecasting." Technological forecasting and social change 56.1 (1997): 25-47.

[23] Behkami, Nima A., and Tugrul U. Daim. "Research forecasting for health information technology (HIT), using technology intelligence." Technological Forecasting and Social Change 79.3 (2012): 498-508.

[24] Lee, Pei-Chun, and Hsin-Ning Su. "Quantitative mapping of scientific research—The case of electrical conducting polymer nanocomposite." Technological Forecasting and Social Change 78.1 (2011): 132-151.

[25] Kajikawa, Yuya, et al. "Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy." Technological Forecasting and Social Change 75.6 (2008): 771-782.

[26] Furukawa, Takao, et al. "Identifying the evolutionary process of emerging technologies: A chronological network analysis of World Wide Web conference sessions." Technological Forecasting and Social Change 91 (2015): 280-294.

[27] Haegeman, Karel, et al. "Quantitative and qualitative approaches in Future-oriented Technology Analysis (FTA): From combination to integration?." Technological Forecasting and Social Change 80.3 (2013): 386-397.

[28] Rotolo, Daniele, Diana Hicks, and Ben R. Martin. "What is an emerging technology?." Research Policy 44.10 (2015): 1827-1843.

[29] Prabhakaran, Thara, Hiran H. Lathabai, and Manoj Changat. "Detection of paradigm shifts and emerging fields using scientific network: A case study of Information Technology for Engineering." Technological Forecasting and Social Change 91 (2015): 124-145.

[30] Chao, Chia-Chen, Jiann-Min Yang, and Wen-Yuan Jen. "Determining technology trends and forecasts of RFID by a historical review and bibliometric analysis from 1991 to 2005." Technovation 27.5 (2007): 268-279.

[31] Liao, Wei-Pang, Tom MY Lin, and Shu-Hsien Liao. "Contributions to Radio Frequency Identification (RFID) research: An assessment of SCI-, SSCI-indexed papers from 2004 to 2008." Decision Support Systems 50.2 (2011): 548-556.

[32] Zhao, Rongying, and Ju Wang. "Visualizing the research on pervasive and ubiquitous computing." Scientometrics

86.3 (2011): 593-612.

[33] Yan, Bei-Ni, Tian-Shyug Lee, and Tsung-Pei Lee. "Mapping the intellectual structure of the Internet of Things (IoT) field (2000–2014): a co-word analysis." Scientometrics 105.2 (2015): 1285-1300.

[34] Karvonen, Matti, and Tuomo Kässi. "Patent citations as a tool for analysing the early stages of convergence." Technological Forecasting and Social Change80.6 (2013): 1094-1107.

[35] Trappey, Charles V., et al. "Using patent data for technology forecasting: China RFID patent analysis." Advanced Engineering Informatics 25.1 (2011): 53-64.

[36] Hung, Shiu-Wan, and An-Pang Wang. "Examining the small world phenomenon in the patent citation network: a case study of the radio frequency identification (RFID) network." Scientometrics 82.1 (2010): 121-134.

[37] Lee, Sungjoo, et al. "Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping." Technological Forecasting and Social Change 76.6 (2009): 769-786.

[38] Trappey, Charles V., Amy JC Trappey, and Chun-Yi Wu. "Clustering patents using non-exhaustive overlaps." Journal of Systems Science and Systems Engineering 19.2 (2010): 162-181.

[39] Trappey, Amy JC, et al. "A fuzzy ontological knowledge document clustering methodology." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 39.3 (2009): 806-814.

[40] Valdez, Bill, and Julia Lane. "The science of science policy: a federal research roadmap." Policy, O. o. S. a. T. Washington, DC, Executive Office of the President (2008).

[41] Macilwain, C. (2010). Science economics: What is science really worth. Nature, 465, 682–684.

[42] Börner, Katy, Chaomei Chen, and Kevin W. Boyack. "Visualizing knowledge domains." Annual review of information science and technology 37.1 (2003): 179-255.

[43] Kostoff, Ronald N., et al. "Literature-related discovery (LRD): water purification." Technological Forecasting and Social Change 75.2 (2008): 256-275.

[44] Kajikawa, Yuya, and Yoshiyuki Takeda. "Structure of research on biomass and bio-fuels: A citation-based approach." Technological Forecasting and Social Change 75.9 (2008): 1349-1359.

[45] Fenn, Jackie, and Mark Raskino. "Understanding Gartner's hype cycles." Gartner Inc., Stamford (2008).

[46] Lee, Pei-Chun, and Hsin-Ning Su. "Quantitative mapping of scientific research—The case of electrical conducting polymer nanocomposite." Technological Forecasting and Social Change 78.1 (2011): 132-151.

[47] Kajikawa, Yuya, and Yoshiyuki Takeda. "Citation network analysis of organic LEDs." Technological Forecasting and Social Change 76.8 (2009): 1115-1123.

[48] Ogawa, Takaya, and Yuya Kajikawa. "Assessing the industrial opportunity of academic research with patent relatedness: A case study on polymer electrolyte fuel cells." Technological Forecasting and Social Change 90 (2015): 469-475.

[49] Al-Shboul, Bashar, and Sung-Hyon Myaeng. "Wikipedia-based query phrase expansion in patent class search." Information Retrieval 17.5-6 (2014): 430-451.

[50] Vulić, Ivan, Wim De Smet, and Marie-Francine Moens. "Cross-language information retrieval models based on latent topic models trained with document-aligned comparable corpora." Information Retrieval 16.3 (2013): 331-368.

[51] Bache, Richard. "Patent retrieval–A question of access." World Patent Information 33.4 (2011): 345-351.

[52] Mahdabi, Parvaz, and Fabio Crestani. "Patent query formulation by synthesizing multiple sources of relevance evidence." ACM Transactions on Information Systems (TOIS) 32.4 (2014): 16.

[53] Uijlings, Jasper RR, Arnold WM Smeulders, and Remko JH Scha. "Real-time bag of words, approximately." Proceedings of the ACM international Conference on Image and Video Retrieval. ACM, 2009.

[54] Deepu S, Pethuru Raj and S.Rajaraajeswari. "A Framework for Text Analytics using the Bag of Words (BoW) Model for Prediction." Proceedings of International Journal of Advanced Networking & Applications (IJANA) , 2016.

[55] Boyack, Kevin W., et al. "Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches." PloS one 6.3 (2011): e18029.

[56] Hamedani, Masoud Reyhani, and Sang-Wook Kim. "On Computing Similarity in Academic Literature Data: Methods and Evaluation." International Conference on Web-Age Information Management. Springer International Publishing, 2014.

[57] Zhang, Yi, et al. "Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research." Technological Forecasting and Social Change 105 (2016): 179-191.

[58] Yau, Chyi-Kwei, et al. "Clustering scientific documents with topic modeling." Scientometrics 100.3 (2014): 767-786.

[59] Gretarsson, Brynjar, et al. "Topicnets: Visual analysis of large text corpora with topic modeling." ACM Transactions on Intelligent Systems and Technology (TIST) 3.2 (2012): 23.

[60] Rafols, Ismael, Alan L. Porter, and Loet Leydesdorff. "Science overlay maps: A new tool for research policy and library management." Journal of the American Society for information Science and Technology 61.9 (2010): 1871-1887.

[61] Kay, Luciano, et al. "Patent overlay mapping: Visualizing technological distance." Journal of the Association for Information Science and Technology65.12 (2014): 2432-2443.

[62] Leydesdorff, Loet, Duncan Kushnir, and Ismael Rafols. "Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC)." Scientometrics 98.3 (2014): 1583-1599.

[63] Leydesdorff, Loet, Daniele Rotolo, and Ismael Rafols. "Bibliometric perspectives on medical innovation using the Medical Subject Headings of PubMed." Journal of the American Society for Information Science and Technology 63.11 (2012): 2239-2253.

[64] Web of Sciene. "WoS Subject Categories." [Online] Available:
http://incites.isiknowledge.com/common/help/h_field_category_wos.html [Accessed: 9-Nov-2016]

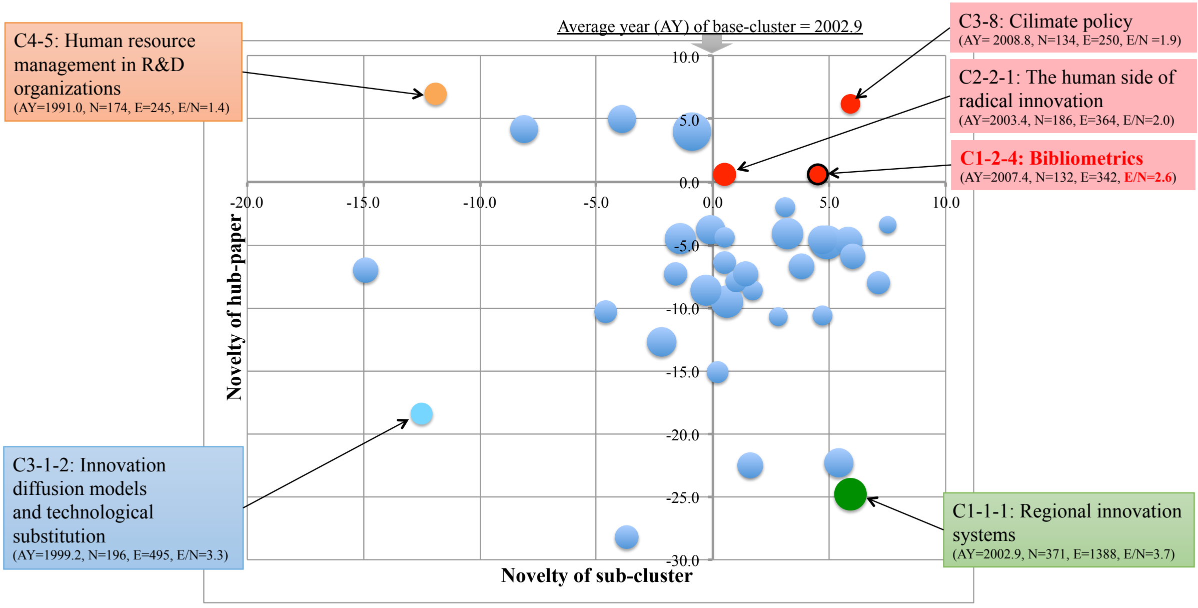
[65] WIPO. [Online] Available: http://www.wipo.int/classifications/ipc/en/preface.html [Accessed: 9-Nov-2016]

[66] Leydesdorff, Loet. "Patent classifications as indicators of intellectual organization." Journal of the American Society for Information Science and Technology 59.10 (2008): 1582-1597.

[67] NIH. [Online] Available : https://www.nlm.nih.gov/mesh/ [Accessed: 9-Nov-2016]

[68] De Solla Price, Derek J. "Networks of scientific papers." Science 149.3683 (1965): 510-515.
[69] Kessler, Maxwell Mirton. "Bibliographic coupling between scientific papers." American documentation 14.1 (1963): 10-25.
[70] Small, Henry. "Co‐citation in the scientific literature: A new measure of the relationship between two documents." Journal of the American Society for information Science 24.4 (1973): 265-269.
[71] N. Shibata, Y. Kajikawa, Y. Takeda, and K. Matsushima, "Comparative Study on Methods of Detecting Research Fronts Using DifferentTypes of Citation," J. Am. Soc. Inf. Sci. Technol., vol. 60, no. 3, pp. 571–580, 2009.
[72] Waltman, Ludo, and Nees Jan Eck. "A new methodology for constructing a publication‐level classification system of science." Journal of the American Society for Information Science and Technology 63.12 (2012): 2378-2392.
[73] Klavans, Richard, and Kevin W. Boyack. "Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?." arXiv preprint arXiv:1511.05078 (2015).
[74] Shibata, Naoki, Yuya Kajikawa, and Ichiro Sakata. "Extracting the commercialization gap between science and technology—Case study of a solar cell." Technological Forecasting and Social Change 77.7 (2010): 1147-1155.
[75] Newman, Mark EJ. "Fast algorithm for detecting community structure in networks." Physical review E 69.6 (2004): 066133.
[76] Blondel, Vincent D., et al. "Fast unfolding of communities in large networks." Journal of statistical mechanics: theory and experiment 2008.10 (2008): P10008.
[77] Rosvall, Martin, and Carl T. Bergstrom. "Maps of random walks on complex networks reveal community structure." Proceedings of the National Academy of Sciences 105.4 (2008): 1118-1123.
[78] Chen, Ssu-Han, et al. "Detecting the temporal gaps of technology fronts: A case study of smart grid field." Technological Forecasting and Social Change 79.9 (2012): 1705-1719.
[79] Altuntas, Serkan, Turkay Dereli, and Andrew Kusiak. "Forecasting technology success based on patent data." Technological Forecasting and Social Change96 (2015): 202-214.
[80] Marra, Marianna, et al. "The value of indirect ties in citation networks: SNA analysis with OWA operator weights." Information Sciences 314 (2015): 135-151.
[81] J. Kleinberg, Authoritative sources in a hyperlinked environment, J. Assoc. Comput. Mach. 46 (May) (1999) 668–677.
[82] U. Brandes, T. Willhalm, Visualization of bibliographic networks with a reshaped landscape metaphor, in: D. Ebert, P. Brunet, I. Navazo (Eds.), Joint
[83] Eurographics – IEEE TCVG Symposium on Visualization, 2002.

Appendix



Appendix. Research field of MOT
(AY is average year, N is the number of nodes, E is number of edges, E/N is the average number of citation in the domain)

RCS, the method shown in chapter 4, was applied in the field of MOT. Three clusters were classified as change-maker domain that is the domain with an active research target and its most-cited research is relatively new. Among the clusters classified as change-maker, researchers focus on bibliometics the most judging from the average number of citations.

2. Conclusion

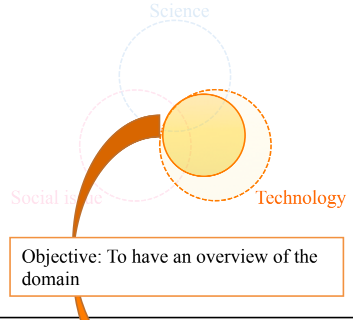
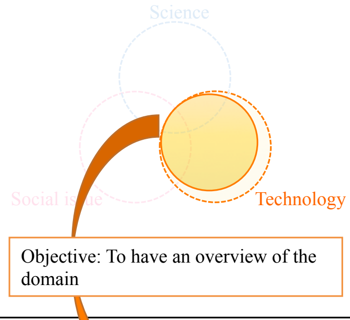
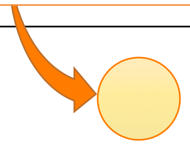
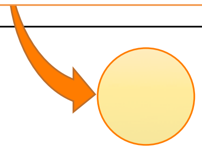
This chapter concludes by providing a brief summary of the findings and contributions of this thesis. Limitations and future works are also discussed.

2.1 Summary

Chapter 1 summarized the bibliometrics studies and their limitations. In the context of Management of Technology and Science Policy, researchers have contributed to develop bibliometric methodologies or tools to support policy making. However, in practice, such methodologies were not utilized because they tend to only extract technology options, and the assessment of those options remains out of scope. To overcome that situation, a method to assess and select the specific technology options ex-ante is required. Therefore, the objective of this thesis was to propose new applications of bibliometrics for Management of Technology and Science Policy. Problems of previous bibliometric methods were summarized, and solutions (new methodologies) were proposed (through Chapter 3 to 6). Then, IoT-related technologies are introduced in chapter 2 as analysis target to validate the effectiveness of the proposed methods.

Chapter 3 touched upon the issue of connectivity of patent citation networks. They are normally sparse because of the difference of citation behavior among countries. The existence of small networks is not wrong, but they might do not well represent the overview of the target domain. To solve this issue, patent family analysis has been proposed for increasing their connectivity. However, even though an increase in the connectivity is observed the performance continue to be limited. Hence, an unconnected component inclusion technique (UCIT) for patent citation analysis was proposed. The method generates a cluster solution that includes unconnected and connected components of a direct citation network. It was observed that UCIT increased the number of nodes especially in relatively small networks (MAX +1859%). The contribution is summarized in Table 2.1. Dotted line stands for the input dataset (facets), and the orange circle represents the patent network of IoT-related technologies created by direct citation (nodes). UCIT enables a better overview of domains thanks to the larger number of nodes with higher quality. Additionally, I analyzed how the clusters changed by adding unconnected patents to the citation network and identified four types of clustering phenomenon. UCIT can be used by patent officers, R&D managers, and policy makers when they want to understand the technology landscape better.

Table 2.1. Contribution of Chapter 3: able to have better overview of the domain

	Conventional methods	(Ch. 3) ***Proposed method***
Dataset	One dataset	
	• Patents of a domain	
Top view		
R&D area of focus		
Related works	Ebola research [1]; Microelectromechanical systems [2]; Five drive train technologies [3]	IoT-releted technologies

From chapter 4 to chapter 6, contributions were made to improve the interpretation of bibliometric outputs. Previous approaches tend to rely on simple indicators that premise incremental changes, and tend to rely on dataset of a single domain which means that the analysis cannot go beyond that technology field.

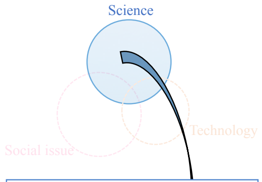
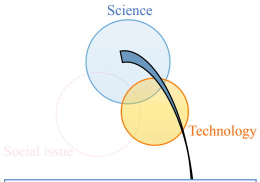
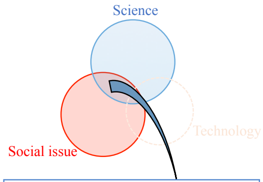
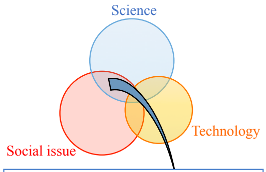
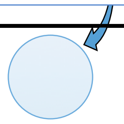
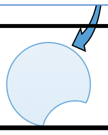
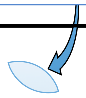
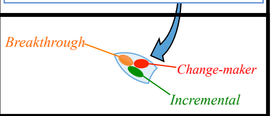
Chapter 4 tackled the limitation of having a simple indicator for detecting the emerging research fronts. Previous literature have utilized average publication year of the research field as the proxy. However, it is not matched to describe the phenomenon of R&D trends: number of papers is not increasing or decreasing monotonously. Thus, a new method called the research classification schema (RCS) is proposed to extract the important specific research fields, even if research trends are complicated. RCS utilizes two measures of publication profile: the novelty of sub-cluster, and the novelty of hub-paper. The effectiveness of RCS was demonstrated through a case study on one technical domain, where significant technologies previously not noticed by an expert, were detected. However, it is required to be validated by other targets to test the robustness of RCS. Hence, in chapter 4, IoT-related technologies were selected as case study, and RCS was expanded so that it could be used in multiple technologies simultaneously. IoT-related technologies were classified into four categories: Change-maker, Breakthrough, Matured, Incremental. Characteristic sub-clusters in each quadrant were evaluated. It was figured out that these characteristic sub-clusters were plotted in the right quaderants. Expanded RCS is especially useful for researchers and R&D managers to seek innovation seeds because it can compare multiple alternatives at once, and also it brings insights that were not revealed in the conventional method and also improved the effectiveness.

Chapter 5 used papers and patents of IoT-related technologies to find promising research areas which have commercialization opportunities. Previous literature of bibliometrics on IoT utilized only one dataset: either only papers or patents, hence the implications have been limited. My study is the first in comparing research areas of IoT-related technologies by analyzing massive quantities of papers and patents at the same time. In addition, I further developed forecasting methodology introduced by Shibata et al. [11] by dividing clusters into two layers: application level and technology level. Through the analysis, the differences among the technologies were figured out. To validate usefulness of the result, the output was compared to a public government document that contained IoT technology policy. There was good agreement between extracted clusters of IoT application level and the public document. The public document listed the 11 system's enabling technologies, but it does not included concrete technologies information. For instance, they just mentioned that artificial intelligence could contribute to applications of these systems. The presented analysis in this thesis could elucidate more concrete options. In addition, it was detected that the public document lacked some important technologies compared with extracted clusters of IoT technology level.

Chapter 6 utilized three types of datasets (papers of social issues, papers, and patents of IoT-related technologies as solution for the social issues) to interpret technological options more effectively. Previous studies of IoT tend to focus on industry applications such as supply chain or manufacturing. These applications have been already widely recognized. There is a large amount of demands to seek new applications of IoT. Meanwhile, WEHAB agenda has been proposed and discussed globally, to face issues related to water, energy, healthcare, agriculture, and biodiversity. However, there were no technological deep discussion in the agenda and solutions toward the social issues were vague. Therefore, chapter 6 explored plausible social issues within the WEHAB agenda where IoT can be applied. Semantic linkages between extracted research areas and social issues were investigated. The result showed that healthcare issues have the highest opportunities and potentials to be solved. 51.4% of social issues could be solved by IoT-related technologies—not now, but in the future. 31.4 % of technologies of companies have opportunities and potentials to solve social issues. Some extracted linkages were explored as examples. By using extracted linkages derived from dataset until 2014, there were identified some novel applications before their appearance in 2015. Therefore, it was confirmed that my proposed method could potentially forecast new applications ex-ante.

Moreover, a new bibliometric methodological framwork is obtained by combining the methods proposed in chapter 4-6. Table 2.2 summarized the contribution of the newly proposed method. One domain is the network created by direct citation. The blue circle stands for "science", consisting of journal articles of IoT-related domains, and the orange circle represent "technology", conformed by patents of IoT-related domains. Normally, the orange circle is smaller than the blue one as discussed in chapter 3: connectivity of patent network is smaller than that of the paper network. The red circle acts as "social issue" made of papers of WEHAB. Overlapped portion of the Venn diagram means high text similarity among clusters. Previous literature tended to utilize only one dataset to have an overview of the domain. It contributed to have technology options of the domain. When it comes to the methods using two datasets, it is possible to also assess its industrial opportunity (by removing the overlapped portion of Venn diagram), or to find linkage between a target domain and another domain of the solution (focus on the overlapped portion of Venn diagram). The proposed method utilized three datasets, and it is possible to obtain hybrid output of the above two results. Therefore, it is able to detect new applications of innovation seeds with possibility to be commercialized in the future. This not only enrich the interpretation of bibliometric analysis, but also can narrow down R&D area of focus more effectively than conventional methods (see "Area of R&D focus" in the Table 2.2)

Table 2.2. Contribution of Chapter 4-6: able to narrow down R&D area of focus effectively

Chapter	Conventional methods			(Ch. 4~6) ***Proposed method***
Dataset	One dataset	Two datasets		Three datasets
	• Papers of a domain	• Papers of a domain • Patents of domain of the Papers	• Papers of a domain as target • Papers of another domain as solution	• Papers of a domain as target • Papers of another domain as solution • Patents of domain of the Papers
Top view	 Objective: To have an overview of the domain	 Objective: To assess industrial opportunity of the domain	 Objective: To find linkage between a target domain and the solution	 Objective: To detect new applications of innovation seeds with possibility to be commercialized in the future
R&D area of focus				
Related works	Antenna [4]; Sustainability [5]; Bio fuel [6]; Organic LED [7]; RFID [8-10]	Solar cell [11]; Polymer electrolyte fuel cells [12]	Aging society & Robot [13]	IoT-releted technologies & WEHAB

2.2 Limitations and Future works

The improvement of the integrated method is a future direction. Some limitations of this thesis should be overcome to improve the quality of the output of integrated method.

2.2.1. Cluster classification related

In this thesis, the analysis targets were IoT-related technologies, thus the technology domains were near each other. However, when it comes to the assessment of the domains that have quite different features (e.g., IoT and biotechnology, material and law), the position of the sub-clusters plotted in RCS might not be accurate because they do not have common context of development. Hence, it is needed to apply calibration clusters in each domain of RCS to have more accurate positioning. Another task is to consider the type of documents. From the experiments, the position tends to be correct when the hub-paper corresponds to journal article or proceedings paper. Whereas, the position could be not appropriate when the hub-paper is review, survey, book, or news because the objective of these document is not to report cutting edge findings, but to summarize current state-of-the-art.

2.2.2. Semantic linkage related

The status of "commercialized in the future" depends on the domain. Some of the clusters classified as "already exist" could be "commercialized in the future" in other domains. It might be better not to narrow down paper clusters only for those with commercialization possibility in the dataset, rather it could be safe to explore the combination of the semantic linkages and assess the condition by considering context, usage, and meaning from users' point of view.

Another task is to investigate features of cluster that have plausible linkage between solution (technology) and target (social issue, demand). In this thesis, we utilized average value of text similarity. However, the results and outputs of proposed integrated methods highly depend on the text similarity threshold. Instead of the average, another more optimized threshold may exist, and thus, it is needed to discuss how to figure out such threshold.

Finally, once the linkages are extracted successfully, it could be better to hold workshop to obtain plausible application with tacit knowledge of experts.

2.2.3. Data related

In this thesis, papers, and patents were utilized. However, using different database(s) might give new perspective. For instance, news articles, handbooks, SNS (e.g., twitter) could be the next material. By utilizing multiple data sources, it would be possible to narrow down R&D area of focus more effectively.

References

[1] Fajardo-Ortiz, J. Ortega-Sánchez-de-Tagle, and V. M. Castaño, “Hegemonic structure of basic, clinical and patented knowledge on Ebola research: a US army reductionist initiative.,” J. Transl. Med., vol. 13, no. 1, p. 124, 2015.

[2] Hu and W. Liu, “Nano/micro-electro mechanical systems: a patent view,” J. Nanoparticle Res., vol. 17, no. 12, p. 465, 2015.

[3] Nakamura, S. Suzuki, and Y. Kajikawa, “The effect of patent family information in patent citation network analysis : a comparative case study in the drivetrain domain,” Scientometrics, vol. 104, no. 2, pp. 437–452, 2015.

[4] Takano, Yasutomo, Yuya Kajikawa, and Makoto Ando. "Trends and Typology of Emerging Antenna Propagation Technologies: Citation Network Analysis." International Journal of Innovation and Technology Management (2016): 1740005.

[5] Kajikawa, Yuya, et al. "Creating an academic landscape of sustainability science: an analysis of the citation network." Sustainability Science 2.2 (2007): 221-231.

[6] Kajikawa, Yuya, and Yoshiyuki Takeda. "Structure of research on biomass and bio-fuels: A citation-based approach." Technological Forecasting and Social Change 75.9 (2008): 1349-1359.

[7] Kajikawa, Yuya, and Yoshiyuki Takeda. "Citation network analysis of organic LEDs." Technological Forecasting and Social Change 76.8 (2009): 1115-1123.

[8] Prabhakaran, Thara, Hiran H. Lathabai, and Manoj Changat. "Detection of paradigm shifts and emerging fields using scientific network: A case study of Information Technology for Engineering." Technological Forecasting and Social Change 91 (2015): 124-145.

[9] Zhao, Rongying, and Ju Wang. "Visualizing the research on pervasive and ubiquitous computing." Scientometrics 86.3 (2011): 593-612.

[10] Liao, Wei-Pang, Tom MY Lin, and Shu-Hsien Liao. "Contributions to Radio Frequency Identification (RFID) research: An assessment of SCI-, SSCI-indexed papers from 2004 to 2008." Decision Support Systems 50.2 (2011): 548-556.

[11] N. Shibata, Y. Kajikawa, and I. Sakata, “Extracting the commercialization gap between science and technology — Case study of a solar cell,” Technol. Forecast. Soc. Chang., vol. 77, no. 7, pp. 1147–1155, Sep. 2010.

[12] Ogawa, Takaya, and Yuya Kajikawa. ""Assessing the industrial opportunity of academic research with patent relatedness: A case study on polymer electrolyte fuel cells."" Technological Forecasting and Social Change 90 (2015): 469-475."

[13] V. Ittipanuvat, K. Fujita, and I. Sakata, “Finding linkage between technology and social issue : A Literature Based Discovery approach,” J. Eng. Technol. Manag., 2013.