Evaluation of the Industrial and Social Impacts of Academic Research Using Patents and News Articles

Shumpei Iinuma[a] Hiroshima City University, Japan

Satoshi Fukuda[a] Hiroshima City University, Japan

Hidetsugu Nanba[a] Hiroshima City University, Japan

Toshiyuki Takezawa[a] Hiroshima City University, Japan

Abstract

In scientometrics and citation analysis, several measures for evaluating the industrial relevance or the impact of academic research fields have been proposed. What seems to be lacking, however, is that these measures could not evaluate the recent industrial relevance or impact of each field, because most of them rely on citations of research papers in patents and vice versa. In this paper, we attempt to evaluate the industrial and social impact of research fields using document classification techniques. Our method classifies research papers and news articles using systems including the International Patent Classification (IPC) and the KAKEN classification index. Then it evaluates the industrial and social impact of each field by comparing the number of research papers with the number of patents or articles in the IPC categories and the projects funded in each KAKEN category. In addition to the classification, we extracted key phrases of technologies to capture trends.

Keywords: scientometrics, text classification, patent, research paper, news article

1. Introduction

In scientometrics and citation analysis, researchers attempt to evaluate the impact of academic research or its industrial relevance. Such evaluations are intended to discover salient work or maintain academic quality; they are also leveraged to select research topics suitable for funding. The main evaluation method is peer review: other researchers in the field assess justifiability and evaluate the work from the academic aspect. However, the social and economic effects of research activities have gained increased importance, and it is necessary to capture these effects, which are difficult to evaluate with traditional peer review or citation analysis. To address this problem, evaluation of academic research should consider other aspects as well as the academic.

In this study, we classified research papers and news articles using several classification systems and then we evaluated the impact of each research field by comparing the number of papers with the number of patents or articles for each category. First, we classified research papers using the International Patent Classification (IPC) and examined the number of publications for each IPC category for a comparison with the number of patent applications. We also applied our automatic classification method to news articles to find research fields that attract social interest. We adopted a patent classification system because patents can be regarded as achievements of academic research and technical development. In addition to this examination, we tried to capture the trends in technologies by extracting phrases of elemental technologies from patents, research papers, and news articles.

A classification system for research fields plays an important role in peer review by research funding agencies, because reviewers are assigned according to the classification system. The KAKEN classification index was designed to classify projects by the KAKEN research fund in Japan. In addition to the analysis using patents and news articles, we classified research papers into the KAKEN categories, and compared the number of publications in each with the number of projects funded by the KAKEN research fund.

The remainder of this paper is organized as follows. Section 2 describes related work. Section 3 explains our method for the classification of research papers and news articles. Section 4 reports on the experiment, and discusses the results. We present some conclusions in Section 5.

2. Related Work

The purpose of this study is to evaluate the industrial and social impacts of academic research. This section describes related prior work. For the purposes of this study, we applied an automatic document classification method to documents that belong to different genres (research paper, patent, news article). This section also considers studies on cross-genre information access and automatic document classification.

2.1 Measurement of the Industrial Impact of Academic Research

Attempts to evaluate the industrial contribution of academic research have been based mainly on scientometric

[a] Graduate School of Information Sciences, Hiroshima City University, 3-4-1, Ozuka-Higashi, Asa-Minami-Ku, Hiroshima, 731-3194, Japan {iinuma,fukuda,nanba,takezawa}@ls.info.hiroshima-cu.ac.jp

analysis of relationships between industrial and academic research. A typical approach is to analyze citations between patents and research papers, as represented by Narin et al. [1]. They examined citations between research papers and patents in five countries: the United States, the United Kingdom, the former West Germany, Japan, and France. They regarded research papers and patents as science and technology respectively, and analyzed the impacts of science in each country on technology inside and outside the country. Such analyses using citations between patents and research papers have been done from various aspects, e.g., targeting specific fields such as laser medicine [2], space engineering [3], or focusing on publications and patents in specific locations [4].

However, a problem is that analyzing citations cannot always enable evaluation of recent impact. Cutting-edge work is not yet cited because propagation of knowledge takes time. In this study, we adopt an automatic document classification technique to solve this problem.

2.2 Cross-Genre Information Access and Automatic Document Classification

A great deal of effort has been devoted to cross-genre information access and document classification. An example is the task of technical trend research at NTCIR-3, which aims to retrieve patents related to a news article given as an input [5]. One of the points of this task was to take account of the difference of terms used in research papers and patents. For example, “resident” appears frequently in news articles, but not in patents. Therefore, if terms are weighted using inverse document frequency, the importance of a given term differs greatly between patents and news articles. Terminology differences between genres are also observable in research papers and patents. There are terms that appear only in patents, such as “document-editing device” which becomes "word processor" in research papers. Nanba et al. proposed a method to translate scholarly terms into patent terms using citations between patents and research papers or leveraging thesaurus automatically built from patents [6]. This method is useful when users search both research papers and patents in a particular field.

Another example is the Patent Mining Task of the Seventh and Eighth NTCIR Workshops [7, 8]. The task's goal was the creation of technical trend maps from a set of research papers and patents in a particular research field. The task was composed of the following two subtasks.

- Research paper classification: Classification of research papers in terms of the IPC. The aim of this subtask was to enable comprehensive collection of patents and research papers in the particular field.

- Technical trend map creation: Extracting expressions of elemental technologies and their effects from research papers and patents.

Most participant groups employed the k-nearest neighbor (k-NN) method in the subtask of research paper classification. This was because the IPC was designed minutely so that it has 30,855 categories at the lowest level. The number of training documents was also large, about four million. In this study, we examine the trends in research papers and news articles using the classification technique employed in the first subtask above.

Fukuda et al. proposed a method for research paper classification in terms of the KAKEN classification index [9]. They improved classification accuracy by focusing on the expressions of elemental technologies and their effects, and they also employed the k-NN method. We employed their method to classify research papers, then compare the number of documents with the selection of projects funded by the KAKEN research fund.

3. Classification of Research Papers and News Articles

The points we would like to make clear are as follows.

- Differences exist between the numbers of patent applications, research papers, and news articles in categories defined through the IPC system.

- Gaps exist between the numbers of research papers and of projects funded by the KAKEN research fund in many research fields.

For the purpose of the first analysis, we classified research papers and news articles into the IPC system using a classifier based on the k-NN method. The IPC system is a global standard hierarchical patent classification system, which is organized as a five-level hierarchy: section, class, subclass, main group, and subgroup. Figure 1 and Table 1 show examples for the IPC code “G06F 1/32”. In this study, we classified the documents into categories at the third level according to the 8th version of the IPC system (subclass level; 643 categories).

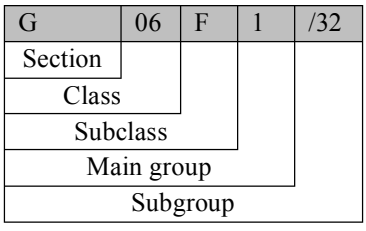

Figure 1. Hierarchy of the IPC system using the example “G06F 1/32”

Table 1. IPC code example for "G06F 1/32"

IPC code	Description
G	Physics
G06	Computing; calculating; counting
G06F	Electrical digital data processing
G06F 1	Details of data-processing equipment not covered by groups G06F 3/00 to G06F 13/00
G06F 1/32	Means for saving power

Table 2. Examples from the KAKEN classification index

Area	Discipline	Research field
Informatics	Computing	Computer system, software, information network, etc.
	Human informatics	Cognitive science, soft computing, human interface and interaction, etc.
Social science	Economics	Theoretical economics, finance, economic history, etc.
	Psychology	Social psychology, educational psychology, clinical psychology, etc.

The classifier employs a patent retrieval system that indexes nouns, verbs, and adjectives, and adopts Okapi BM25 as its similarity measure [10]. First, the top k documents $\{d_1, d_2, \dots, d_k\}$ with highest similarity (k-NN) are retrieved for a research paper or a news article given as input. When the system computes the similarity between a research paper and a patent, it uses the title of the research paper, author's name, and publication name to compare with the patent's specification field and the inventor's name. The similarity is computed as $sim\big(title(p), spec(d)\big) + sim\big(author(p), inventor(d)\big) + sim(venue(p), spec(d))$, where p, d refer to the research paper and the patent respectively. The full text of research papers was not available with our dataset. For news articles, the system computes the similarity between a news article and a patent as $sim(body(a), spec(d))$ where a, d refer to the news article and the patent, respectively. Then, the system calculates $score(c)$ for the IPC codes of the retrieved documents. Here, $score(c)$ can be regarded as a measure of the likelihood that the input document has label c (IPC code). Finally, the input document is classified into the IPC code with the highest score. Our system uses the following ranking method (the Listweak method) [11]:

$$score_{Listweak}(c) = \sum_{i=1}^{k} occur(c, d_i) sim(q, d_i) r^i \quad (1)$$

where *occur* $occur(c, d_i)$ returns 1 if document d_i has IPC code c, otherwise it returns 0. $sim(q, d_i)$ denotes BM25 similarity between input q and d_i (a document retrieved by the patent retrieval system). In addition, r^i is a penalty factor against documents with lower rank, r is set to 0.95 in our system. r was determined by preliminary experiments.

For the classification into the KAKEN classification index, we used the method proposed by Fukuda et al., which focuses on the expressions of elemental technologies and their effects [9]. Their method is also based on the k-NN method. The KAKEN classification index is a classification system designed by the KAKEN research fund, and it is used to determine resource distribution. The index is organized as a three-level hierarchy: Area, Discipline, and Research Field, and it has been modified as new research fields arise. Table 2 shows examples from the index.

The classification method proposed by Fukuda et al. is characterized by its indexing module. They set weights on words, using a key phrase list extracted in advance. Some studies have shown that adjusting the weights on words appearing in documents is effective for classification [12, 13]. Fukuda et al. extracted expressions of elemental technologies and their effects from research papers as key phrases. They used the information extraction method developed in their prior study [14]. The extraction method is based on machine learning and formulated as a sequence-labeling problem. An example of key phrase extraction is given in Figure 2. The information extracted is defined as follows:

- **TECHNOLOGY**: Expressions about algorithms, materials, tools, and data used in studies;
- **EFFECT**: Pairs of ATTRIBUTE and VALUE;

```
Through <TECHNOLOGY>closed-loop feedback
control</TECHNOLOGY>, the system could
<EFFECT><VALUE>minimize</VALUE> the
<ATTRIBUTE>power loss</ATTRIBUTE></EFFECT>
```

Figure 2. Example of key phrase extraction. (Translated from Japanese)

They employed the support vector machine approach, which obtained higher precision than the conditional random field [15] approach. They conducted an experiment using the dataset for the NTCIR-8 Patent Mining Task. They obtained recall and precision scores of 0.276 and 0.539, respectively. Key phrases extracted were used to set the weights on word frequencies appearing in documents, and they studied the weights' effect on the similarity when the system retrieves k-NN documents. Fukuda et al. also adopted the Listweak method, which we described earlier in this section. In addition to their classification method, we leveraged the key phrase extraction technique itself to examine which technologies were mentioned frequently in each document set: patents, research papers, or news articles. Gao et al. built a model to calculate the technology life cycle based on various patent-related indicators, such as the number of unique inventors and citations [16]. Phrases extracted in this

study could also be useful as indicators of life cycles of technologies.

4. Experiments

First, we built the document classifiers, then classified research papers and news articles into the IPC system and counted the number of documents for each IPC code. We also classified research papers in terms of the KAKEN classification index, and compared the number of documents with the number of projects funded by the KAKEN research fund.

4.1 Building and Evaluating the Classifier

We built classifiers for research papers using the IPC system. Table 3 shows the patent data used for the k-NN method described in the previous section. All documents in the Japanese published unexamined patent applications and Unite States Patents have IPC codes that are manually assigned. Each set of documents is used in the classifier for research papers written in Japanese or English respectively. Next, we evaluated the classifiers using the test collections for the research paper classification subtask at NTCIR-7, -8. The task contains four subtasks:

- Japanese subtask: Classification of Japanese research papers using patent data written in Japanese.
- English subtask: Classification of English research papers using patent data written in English.
- Cross-lingual subtasks (J2E, E2J): Classification of Japanese research papers using patent data written in English, and vice versa.

We evaluated the classifier for Japanese research papers using the test collection for the Japanese subtask. Similarly, we evaluated the classifier for English research papers using the test collection for the English subtask. Each classifier was evaluated by the precision of the first ranked classification code. It should be noted that we classified research papers without using their abstracts, although the documents in the test collections contain title, authors, source, and abstract.

Table 3. Document data used for classifiers

Data	Period	Num. of docs	Language
Japanese published unexamined patent application	1993-2012	6,910,194	Japanese
U.S. Patent	1993-2012	2,895,149	English

4.2 Classification into the IPC System and the KAKEN Classification Index

Table 4 shows the data that we classified for the analysis. We used JST[1] Scientific and Technological Data during the period 2003-2012, articles from Yomiuri News that mention development or practical application of technologies (1993-2012), and news articles from TechCrunch[2], which is a news website, focused on information technology (June 2005-December 2013). It should be noted that JST Data is bibliographic data, therefore, abstracts or the full text of research papers were not available. We used title, authors' names, and publication name for classification. News articles from Yomiuri and TechCrunch contained the headline and body of each article. The number of documents in the Japanese published unexamined patent applications for each IPC category was used for a comparison (1993–2012). We classified JST Scientific and Technological Data and Yomiuri News articles into the IPC system using the classifier for research papers written in Japanese. In the same way, articles from TechCrunch were classified using the classifier for research papers written in English. In addition, we classified JST Data in terms of the KAKEN classification index, using the k-NN based classifier developed by Fukuda et al. Their classifier obtained a precision score of 0.827 on classification by the title of research papers.

Table 4. Document data

Data	Language	Classification system	Num. of docs
JST Scientific and Technological Data (bibliographic data)	Japanese	IPC	850k
		KAKEN	6,533,269
Yomiuri News (Articles about development or practical application of technologies were selected.)	Japanese	IPC	8,674
TechCrunch (IT news)	English	IPC	120,596

[1] Japan Science and Technology Agency: http://www.jst.go.jp/

[2] http://www.techcrunch.com

4.3 Extraction of Key Phrases

We extracted expressions of elemental technologies from the data described in previous subsections including Japanese Patents, JST Data, and Yomiuri News. We focused on documents belonging to a specific field that has a large population. The object field was chosen according to the results of the classification described in previous subsections. The purpose of the extraction was to find active technologies and capture technological trends through these documents. We employed the method developed by Fukuda et al. to extract and count expressions. In addition to this analysis, we examined the numbers of expressions about specific technologies for each year. We chose technologies that frequently appear in both Japanese Patents and JST Data. Then we compared the trends in patents and research papers.

5. Results

As described in Section 4.1, the classifiers were evaluated using the test collections for the task at NTCIR-7, -8. For the results of evaluation we obtained precision scores of 0.815 (at $k = 300$) and 0.656 (at $k = 300$) on the Japanese and English paper classification, respectively. Examples of classified documents are shown in Figure 3 and 4. As can be seen from these figures, both documents were classified correctly.

```
<doc>
<title>Evaluation on Center Line Extraction of
Brain Vessels from MRA Images</title>
<authors>Matsumoto N., Fujii T., Jiang H., Sugou
N., Mito T., Shibata I.</authors>
<source>IEICE technical report. 100(596),
117-122, 2001-01-18</source>
</doc>
```

Figure 3. Example of a research paper classified as A61B (Diagnosis; surgery; identification).

```
<doc>
<title>Yahoo Acquires SkyPhrase</title>
<body>
Yahoo has acquired SkyPhrase, a startup that
builds natural language processing technology,
the company revealed today in a blog post. ...
to help continue its goal of "making computers
deeply understand people's natural language and
intentions."
</body>
</doc>
```

Figure 4. Example of a news article from TechCrunch, classified as G06F (Electrical digital data processing).

Table 5 shows the number of documents in JST Scientific and Technological Data captured through the IPC system. The number of documents classified as G06F (Electrical digital data processing), A61K (Preparations for medical, dental, or toilet purposes), and C12N (Microorganisms or enzymes) are relatively large. It is clear that there are a large number of studies in these fields, as seen through the IPC system.

Table 5. The number of research papers for each IPC code (top 10).

IPC	Description	Num. of docs
G06F	Electrical digital data processing	94,943
A61K	Preparations for medical, dental, or toilet purposes	65,059
C12N	Microorganisms or enzymes	62,034
G01N	Investigating or analyzing materials	49,391
H01L	Semiconductor devices	42,070
H04N	Pictorial communication, e.g., television	18,847
G02F	Optical operation	17,428
A61B	Diagnosis; surgery; identification	15,845
C01B	Nonmetallic elements	15,613
A01G	Horticulture	12,728

Next, Table 6 shows the numbers of applications in Japanese Patents, research papers (JST Data), and news articles (Yomiuri News) for each IPC category. Some categories such as A23L (Food, foodstuffs) and A61K (Preparations for medical, dental, or toilet purposes) appear at a relatively higher rank in terms of the number of news articles. Specifically, A01G (Horticulture) appears in the top 10 list of news articles, but not of other genres. According to the IPC, Section A is defined as "Human necessities." This result indicates that research and development on daily necessities tends to attract public concern. Comparing patents and research papers, there is some variability in IPC categories appearing in the results. Thus we see that the briskness of academic research, as captured by the number of publications, does not always lead to patent applications. IPC categories such as G06F (Electrical digital data processing), A61K (Preparations for medical, dental, or toilet purposes), and H04N (Pictorial communication, e.g., television) were ranked in the top 10 with regard to all genres. From these results we can say that research in these fields is likely to lead to patent applications, which we regard as an achievement of research and development, and tends to attract social concern. While we do not show the results in tables, we classified articles from TechCrunch using the IPC system. It should be understood that about 90% of these articles were classified into G06F (Electrical digital data processing) or G06Q (Data processing system or methods).

Table 6. Comparison of patents, research papers, and news articles

Japanese Patent		JST Data (research papers)		Yomiuri News	
IPC	Description	IPC	Description	IPC	Description
H01L	Semiconductor devices	G06F	Electrical digital data processing	G06F	Electrical digital data processing
G06F	Electrical digital data processing	A61K	Preparations for medical, dental, or toilet purposes	G06Q	Data processing system or methods
H04N	Pictorial communication, e.g., television	C12N	Microorganisms or enzymes	A23L	Food, foodstuffs
G03G	Electrography	G01N	Investigating or analyzing materials	A61K	Preparations for medical, dental, or toilet purposes
G11B	Information storage	H01L	Semiconductor devices	H04N	Pictorial communication, e.g., television
G02B	Optical element, systems, or apparatus	H04N	Pictorial communication, e.g., television	C12N	Microorganisms or enzymes
B41J	Typewriters; selective printing machines	G02F	Optical operation	G01N	Investigating or analyzing materials
A61K	Preparations for medical, dental, or toilet purposes	A61B	Diagnosis; surgery; identification	H04M	Telephonic communication
G01N	Investigating or analyzing materials	C01B	Nonmetallic elements	A01G	Horticulture
H01M	Processes of means, e.g., batteries	H01M	Processes of means, e.g., batteries	G09B	Educational or demonstration appli-ances

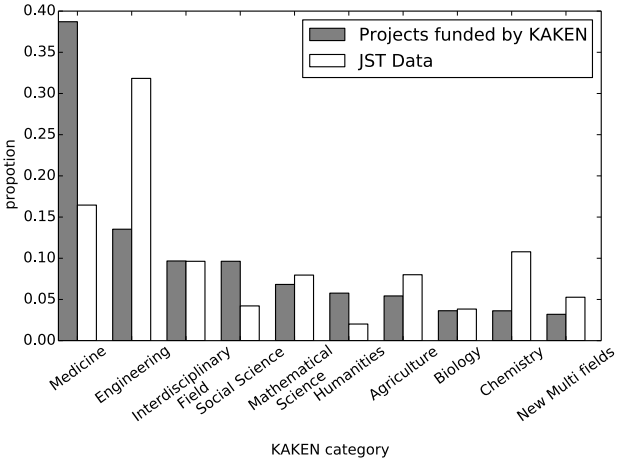

Figure 5. Proportions of the number of documents in each KAKEN category

Figure 5 shows proportions of the number of documents in each KAKEN category. It shows the proportions for each data set, KAKEN-funded projects and JST Data. The results were obtained from the classification of research papers (JST Data) using the KAKEN classification index. Each document was classified at the third level of the index. The figure shows the result of classification at the highest level (Area level), because the number of categories at the third level is so large. While the proportion of "Engineering" is highest in the JST Data, "Medicine" makes up the largest proportion of KAKEN-funded projects. It is clear that research funds such as KAKEN put higher value on "Medicine," even if the briskness of research on "Engineering" is the highest as the number of documents indicates.

Next, we describe the results of key phrase extraction. We counted the number of appearances of key phrases (TECHNOLOGY) extracted from documents classified as "G06F." We used this category because the number of documents belonging to it is relatively large (see Tables 5 and 6). Table 7 shows the 10 most frequent phrases tagged as TECHNOLOGY in each document set. The sources of phrases are documents written in Japanese, but the table shows translations. They were extracted from sections describing "effect of invention" in patents, titles of research papers, and headline and body of news articles. As can be seen from the Japanese Patents column, more general terms appear in patents than in research papers (JST Data). The reason for this is not hard to see: the terms used in patents are often more abstract or creative than those used in research papers, because they are intended to widen the scope of claims. We will show the results of analysis focusing on some specific phrases later. Because nonexperts write most news articles and their readers are also nonexperts, phrases extracted from news articles are also more abstract. Some of them are familiar in daily life, e.g., e-mail, microwave, and mobile computer. With regard to research papers, the top two in the list are related to machine learning: neural networks and genetic algorithms. We can observe the high research activity in these areas.

In the same way, we count the number of appearances of phrases tagged as ATTRIBUTE, which we described in Section 3. We targeted news articles to study social trends: what kind of effects people expect from technologies. Table 8 shows the phrases that frequently appeared in news articles. The five most frequent phrases were cost, precision, safety, consumed electricity, and capability. It seems reasonable that examining effects of technologies enables us to see social expectations from technologies. However, object technologies are not mentioned in Table 8. We explain some of these object technologies. For example, take "cost," which has been mentioned in news articles

Table 7. Top 10 frequent phrases about TECHNOLOGY in documents classified as G06F, "Electrical digital data processing." (Translated from Japanese)

Japanese Patent	JST Data	Yomiuri News
Database	Neural network	E-mail
Program	Genetic algorithm	Motor
System	Finite element method	Software
Network	Graphics processing unit	Cryptographic technology
Computer	Molecular dynamics simulation	Electronic newspaper
Information processing	FPGA (Field programmable gate array)	Microwave oven
Processor register	First principles calculation	Monoelectron element
Application	GIS (Geographical information system)	Electronics
Processor	Artificial neural networks	Glass
Cache memory	Self-organizing map	Mobile computer

about various technologies' effects on "cost," e.g., a new method for Freon destruction or reuse of waste materials. With regard to "consumed electricity," the number of articles related to televisions or computers was relatively large.

Table 8. Ten most frequent phrases of ATTRIBUTE in news articles classified as G06F. (Translated from Japanese)

Phrase (ATTRIBUTE)	Frequency
Cost	50
Precision	47
Consumed electricity	43
Capability	43
Durability	35
Fuel consumption	30
Reliability	25
Quality	25
Performance	20
Production cost	17

To capture the trends in patents and research papers, we chose three phrases: Neural Network, Carbon Nanotube, and Optical Fiber. We chose these because these phrases appeared frequently in the Japanese Patents and research papers datasets. While we used documents belong to "G06F" in the previous analysis, all of the documents were employed in this analysis. The number of phrases extracted from news articles is relatively small in comparison with other datasets, so we used only patents and research papers. We counted the number of appearances for each extracted phrase year by year. It should be noticed that we count only extracted phrases, which is different from just counting all appearances of the phrases in documents. First, Figure 6 shows the frequency for Neural Network. The number of uses of Neural Network as a technology decreases with respect to research papers. While the frequency of the phrase is high in research papers, the frequency in patents is low. We may say that Neural Network itself did not contribute to industry, but it has attracted the interest of researchers. Figures 7 and 8 show the frequency for Carbon Nanotube and Optical Fiber, respectively. There is a common pattern of increase and decrease in both patents and research papers, especially in Figure 8 (Optical Fiber). There are common peaks in 2006 and 2009. It is fair to say that the extent of academic research affects development of technology (patents) and vice versa.

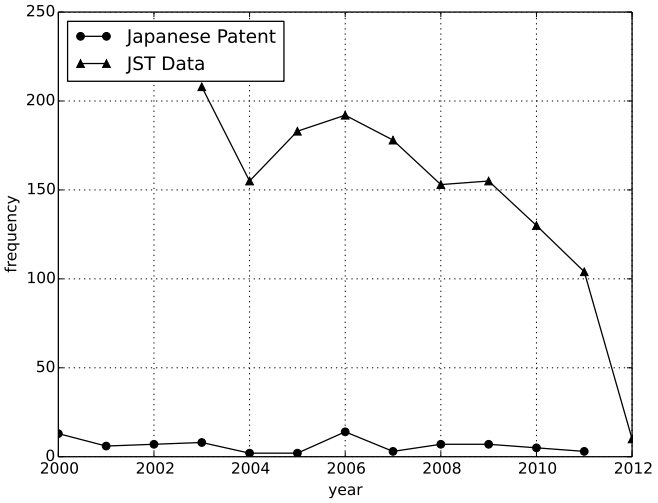

Figure 6. Frequency plot for "Neural Network."

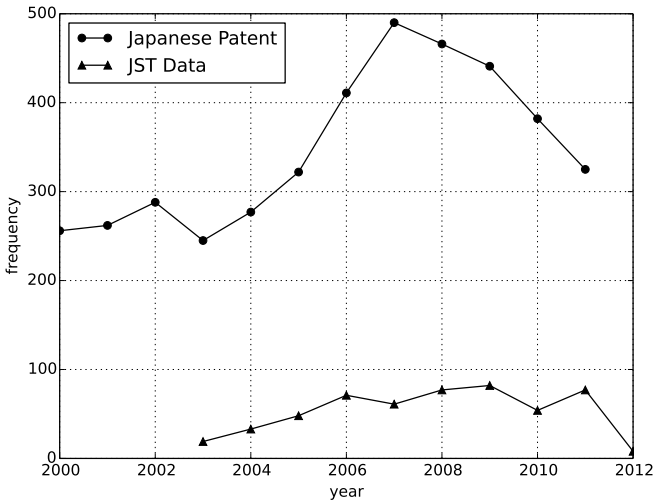

Figure 7. Frequency plot for "Carbon Nanotube."

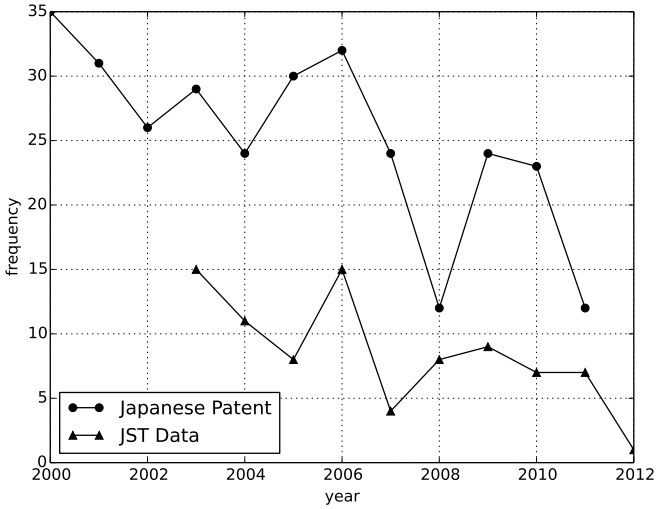

Figure 8 Frequency plot for "Optical Fiber."

5. Conclusion

In this study, we demonstrated that classification using various systems enables comparison between research fields from alternative perspectives. We classified research papers and news articles using the IPC system and then compared the numbers of documents with those of patent applications for each field. The results showed that the number of news articles mentioning developments and practical applications in fields related to food or medicine is relatively large, so these fields tend to attract social concern. In addition we classified research papers in terms of the KAKEN classification index which was designed by the KAKEN research fund, and compared the proportions of documents in each category with those of funded projects by KAKEN. From the results, we identified some gaps between the briskness of research and the number of projects funded by KAKEN research fund.

To examine social concerns, we used news articles. It should be realized that most of the articles about information technology (TechCrunch) were classified into categories related to data processing. For further analysis, classification at a finer level is required. As described in the previous section, we achieved a precision score of 0.656 on the classification of English research papers at the subclass level. Given the limitations of our classifiers, it is difficult to classify at the group level. The classifiers built in this study used only title, authors, and source of publications, because JST Data do not include the abstract or full text. If the abstracts or full text of research papers were used, classification performance should improve. In addition to the data-specific problem, the content difference and terminology difference between patents and news articles should be taken into account. In news articles, technologies tend to be described by focusing on superficial characteristics such as practical applications, although patents mainly describe the technologies themselves. Addressing these problems should lead leads to alignment of news articles with specific patents, and should enable further analysis.

In addition to the document classification, we employed an information-extraction technique to capture technology trends in industry, academia, and society. We studied patents, research papers, and news articles, respectively. As a result, we showed the top 10 frequently mentioned technologies used in each document set, and the activity in research fields. However, phrases extracted from patents were abstract and creative; therefore, we could not see specific technologies in patents as we could in research papers. To discover more-specific technologies, we should prepare some filters against generic terms. We also examined the frequency for each year, focusing on specific phrases. The results showed that activity in academic research affects development of technology, and vice versa. It should be also added that the extracted phrases including elemental technologies and their effects could be used as indicators for life cycle analysis of technologies. For example, we could detect the reason for peaks by using phrases tagged as EFFECT, including VALUE and ATTRIBUTE. These studies help us to see expectations in technologies. We plan like to study this aspect further in our future work.

Acknowledgments

This research owes much to the scientific and technological data of the Japan Science and Technology Agency. I would like to thank JST for providing the data.

References

[1] F. Narin, D. Olivastro, and K. A. Stevens, "Bibliometrics / Theory, Practice and Problems," *Evaluation Review,* vol. 18, pp. 65-76, 1994.

[2] E. C. M. Noyons, A. F. J. van Raan, H. Grupp, and U. Schnoch, "Exploring the Science and Technology Interface: Inventor-author Relations in Laser Medicine Research," *Research Policy,* vol. 23, pp. 443-457, 1994.

[3] U. Schmoch, N. Kirsch, W. Lay, E. Plescher, and K. O. Jung, "Analysis of Technical Spin-off Effects of Space-related R&D by Means of Patent Indicators," *Acta Astronautica,* vol. 24, pp. 353-362, 1991.

[4] D. Coronado and M. Acosta, "The Effects of Regional Scientific Opportunities in Science-technology Flows: Evidence from Scientific Literature Cited in Firms' Patent Data," in *ERSA conference papers, ERSA'03*. European Regional Science Association, 2003.

[5] M. Iwayama, A. Fujii, N. Kando, and A. Takano, "Overview of Patent Retrieval Task at NTCIR-3," in *Working Notes of the 3rd NTCIR Workshop Meeting, Part III: Patent Retrieval Task,* pp. 1-10, 2002.

[6] H. Nanba, T. Takezawa, K. Uchiyama, and A. Aizawa, "Automatic Translation of Scholarly Terms into Patent Terms Using Synonym Extraction Techniques," in *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, 2012.

[7] H. Nanba, A. Fujii, M. Iwayama, and T. Hashimoto, "Overview of the Patent Mining Task at the NTCIR-7 Workshop," in *Proceedings of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-lingual Information Access,* pp. 325-332, 2008.

[8] H. Nanba, A. Fujii, M. Iwayama, and T. Hashimoto, "Overview of the Patent Mining Task at the NTCIR-8 Workshop," in *Proceedings of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-lingual Information Access,* pp. 293-302, 2010.

[9] S. Fukuda, H. Nanba, T. Takezawa, and A. Aizawa, "Classification of Research Papers Focusing on Elemental Technologies and Their Effects," in *Proceedings of the 6th Language & Technology Conference, LTC'13,* pp. 366-370, 2013.

[10] S. E. Robertson and S. Walker, "Some Simple Effective Approximations to the 2-poisson Model for Probabilistic Weighted Retrieval," in *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'94,* pp. 232-241, 1994.

[11] T. Xiao, F. Cao, T. Li, G. Song, K. Zhou, J. Zhu, and H. Wang, "KNN and Re-ranking Models for English Patent Mining at NTCIR-7," in *Proceedings of the 7th NTCIR Workshop Meeting,* pp. 333-340, 2008.

[12] L. S. Larkey, "Some Issues in the Automatic Classification of U.S. Patents," *Working Notes for the AAAI-98 Workshop on Learning for Text Categorization*, pp. 87-90, 1998.

[13] C. J. Fall, A. Torcsvari, K. Benzineb and G. Karetka, "Automated Categorization in the International Patent Classification," in *Proceedings of the ACM SIGIR Forum*, pp. 10-25, 2003.

[14] S. Fukuda, H. Nanba, and T. Takezawa, "Extraction and Visualization of Technical Trend Information from Research Papers and Patents," in *Proceedings of the 1st International Workshop on Mining Scientific Publications, collocated with JCDL 2012*, 2012.

[15] H. Nanba, T. Kondo, and T. Takezawa, "Automatic Creation of a Technical Trend Map from Research Papers and Patents," in *Proceedings of the 3rd International CIKM Workshop on Patent Information Retrieval* (PalR'10), pp. 11-15, 2010.

[16] L. Gao, A. L Porter, J. Wang, S. Fang, X. Zhang, T. Ma, W. Wang, L. Huang, "Technology Life Cycle Analysis Modeling Based on Patent Documents," *Fourth International Seville Conference on Future-Oriented Technology Analysis (FTA) FTA and Grand Societal Challenges-Shaping and Driving Structural and Systemic Transformations SEVILLE*, pp. 12-13, 2011.

Shumpei Iinuma is a graduate student in the Graduate School of Information Sciences at Hiroshima City University. He received his bachelor of Information Sciences from Hiroshima City University in 2014. His research interest is in automatic text summarization targeting technical documentation, such as research papers. Evaluation of the quality of research papers can be seen as a subtask of research paper summarization, because a set of salient research papers should be detected as the summarization objects.

Satoshi Fukuda is a Ph.D. student at the Graduate School of Information Sciences, Hiroshima City University. He graduated from the Faculty of Information Sciences, Hiroshima City University in 2011, and the master's program at the Graduate School of Information Sciences, Hiroshima City University in 2013. His research interests include natural language processing and information retrieval.

Hidetsugu Nanba is an associate professor in the Graduate School of Information Sciences at Hiroshima City University. Dr. Nanba's research focus is natural language processing, text data mining, and information retrieval. He received his Ph.D. from Japan Advanced Institute of Science and Technology in 2001. He is a member of the Association for Computing Machinery (ACM) and the Association for Computational Linguistics (ACL), Information Processing Society of Japan (JSAI), the Association for Natural Language Processing, and Information Processing Society of Japan (IPSJ).

Toshiyuki Takezawa is a professor in the Graduate School of Information Sciences at Hiroshima City University. Dr. Takezawa's research interests include natural language processing, speech translation and dialogue systems. He received his B.E., M.E., and D.Eng. degrees in electrical engineering from Waseda University in 1984, 1986, and 1989. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), Information Processing Society of Japan (IPSJ), Japanese Society for Artificial Intelligence (JSAI), Acoustical Society of Japan (ASJ), and the Association for Natural Language Processing.