Science Intensity of industry by using linked dataset of science, technology and industry

By Kenta Ikeuchi (RIETI and NISTEP), Kazuzuki Motohashi[1] (University of Tokyo, NISTEP and RIETI), Ryuichi Tamura (Hitotsubashi University and NISTEP) and Naotoshi Tsukada (NISTEP)

Abstract

This paper presents new indicators measuring science intensity of industry, by linking scientific paper database (Science), patent information (Technology) and economic census data (Industry) in Japan. The new indicators reflect science and technology linkage embodied in human capital at academic researchers, which cannot be measured by a traditional indicator, such as non patent literature (NPL) citations by patents. We have found that science intensity increases over these 10 years, while the NPL indicator does not change. In addition, the variation of science intensity by industry decreases over time, showing that scientific knowledge becomes important inputs to industrial innovation in general, instead of benefiting particular "science based industries". Our study reconfirms that public support to science is an important policy to promote industrial innovation, and policies for promoting interactions between academia and industry are needed to further exploitation of scientific findings by firms.

[1] Correspondent author: motohashi@tmi.t.u-tokyo.ac.jp

1. Introduction

Scientific foundation becomes more and more important for industrial innovation process. A typical example is found in pharmaceutical industry where the genome science has changes its R&D process substantially. Miniaturization of LSI fabrication process requires understanding of nano-scale physicality of its materials. Furthermore, an advance in information technology has significant impact on society and economy, and “big data” analysis contributes to scientific understanding of business and management activities. Since science sectors involving universities and public research institutes are heavily subsidized by public money, policy interests in measuring the scientific contents in industrial innovation and performance are growing to understand the economic impact of pubic R&D under severe constraint in public spending in general.

Traditionally, the degree of science base, or science intensity of industry has been measured by non-patent literature (research paper) citation to the patent (Narin and Noma, 1985; Schmoch, 1997). This indicator captures to what extend patent (technology for industrial use) are based on the scientific contents in research paper. It is shown that the science linkage various over technology area, and that the intensity is particularly high in biotechnology field (van Looy et, al, 2002).

Alternatively, the science and technology linkage can be captured by patent-publication pairs, i.e., an identical content of research output/invention found in both patent and research paper. This required simultaneous disclosures of research results both in a patent and a research paper (Lissoni et al., 2013) or text mining technique to see the degree of overlap between two kinds of literature in terms of the contents (Magerman et al., 2015). This information gives exact match between science and technology, but only limited numbers of samples are available so that it is not suitable for aggregated indicators of science and technology linkage in at macro level.

In this paper, we propose a conceptual framework to understand science linkages indicators and present some alternative indicators, based on a novel dataset of science, technology and industry linkage. More specifically, we linked the data of research papers (Scopus by Elsevier) and patent data (IIP patent database) at author/inventor level to see how academic discipline and technology field interlinked at individual (academic) researcher level. This dataset provides the linkage between science and technology embodied at human capital (academic inventors), and both industry citation of the patent invented by such academic inventors and joint patent inventions of firms and such academic inventors reflect new channels of scientific knowledge flow from academia to industry from those measured by conventional indicators such as non patent literature citation of patents.

Furthermore, the concordance between technology field of patent data (IPC) and industry classification has been created by linked dataset of patent datasets (IIP patent data) and Japanese economic census data at firm level. At the end, we have developed the concordance tables of academic field (science), patent classification (technology) and industrial performance (industry) to see how scientification of industry and economy is progressed over time.

The next section is explaining the methodology of the dataset, then the conceptual framework and the methodology of new indicators are presented. We show also the trend of scientification of industry over these 15 years based on new indicators. Finally, this paper concludes with summary of new findings and some policy implications.

2. Dataset construction methodology

2-1. Author/Inventor level linkage of Scopus and IIP Patent database

A major work of this section is disambiguation of academic inventor from patent database. We use IIP Patent database, covering all patent application information to JPO (Goto and Motohashi, 1997). From patent information, the information of name and address of inventor is available. However, there is no information of identification of inventor, showing the identical inventor shown up across multiple patents. The name of inventor is strong information, but we need to disambiguate the different persons with the same name.

We apply the methodology of Li el. al (2014) for disambiguation of inventors in USPTO patents. Their methodology is originally based the Authority disambiguation approach developed by Torvik et al. (2005) and Torvik and Smalheiser (2009). We disambiguate all Japanese inventors of patents applied between 1995 and 2013, derived from the IIP patent database. We exclude non-Japanese inventors, whose name does not contain Chinese character (*Kanji*) and/or whose address is outside Japan. 12.4 million inventor-patent records are remained for analysis, which contains 1.2 million unique combinations of inventor’s name and address, and applicant’s name.

The methodology consists of four steps. (1) Blocking: Inventor-patent records are divided into several subsets according to inventors' name and similarity is compute between pairs of records within each block. (2) Training sets: We construct matched and non-match training set as pairs of matched and non-matching full inventor names defined as "rare", respectively. Using a telephone directory in 2000-2012, we define a list of "rare" names which appear only once or never appear in the telephone directory. (3) Ratio: We define a "similarity profile (vector)" $\mathbf{x} = (x_1 \quad \cdots \quad x_n)$, which represents the degree of similarity of inventor and patent attributes between two inventor-patent records, for all inventor-patent record pairs within blocks. As the inventor attributes, inventor's name and address are used and for patent attributes applicant's name and ID, main technology class at four digit level of the International Patent Classification (IPC) and the list of co-inventors' name are used. Applicant names and IDs are both normalized using the *NISTEP Dictionary of Corporate Names* and the *NISTEP Dictionary of Names of Universities and Public Organizations*, both developed by the *National Institute of Science and Technology Policy* and publicly available from its website[2]. Inventor address attribute is also normalized and divided into prefecture (*to-do-hu-ken*), city (*shi-ku-cho-son*), district (*chi-mei*) and street (*ban-chi* and *go*) using a commercialized geo-coding software provided by Kokusai Kogyo Co., Ltd., *Address-normalizing converter and geocoding tool*. We then calculate the likelihood "ratio" for each similarity profile from the training set as the ratio of times that a similarity profile appeared in the match set to the non-match set. (4) Pairwise matches: the (posterior) probability of a match between inventor-patent records based on Bayes theorem using the similarity profile and the corresponding likelihood ratios. Following to Li et. al (2014) we set the prior probability as the inverse of the number of pairs in the block. Minimum threshold for probability matching pair is set to 0.5. More detailed explanation for the data and method for the patent inventor disambiguation are described in Appendix 1.

Table 1 represents the results of inventor disambiguation and its estimation accuracy. We identify 1.71 million inventors from 12.4 million inventor-patent records, which means that the average number of patents per inventor is 7.1. We then check the precision of our inventor disambiguation results with the *KAKEN Database of Grants-in-Aid for Scientific Research* developed by the National Institute of Informatics. In the KAKEN database, all receivers of the public research fund from the *Japan Society for the Promotion of Science* (JSPS) are registered and a reliable identifier for

[2] http://www.nistep.go.jp

each researcher is available. On twelve thousand inventor-patent instances of six thousand inventors extracted from the KAKEN database, we calculate splitting and lumping error of our disambiguation result following to Li et. al (2014) and results show that the splitting error is 2.41% and the lumping error is 0.29%. These indicate that our results are better than Li et. al (2014) which reports 3.26% for splitting error and 2.34% for lumping error.

Table 1 Results of disambiguation of patent inventor

Disambiguation methods	(1) Modified Li et al.(2014) Algorithm	(2) Name Match	(3) Name-Address-Applicant Match
# of inventor-patent records	12,397,820	-	-
# of disambiguated inventors	1,709,880	-	-
# of KAKEN records (inventor-patent records)	11,958	11,974	11,974
# of KAKEN inventors	5,984	5,992	5,992
# of disambiguated KAKEN inventors	6,221 (96.2%)	5,973 (100.3%)	7,835 (76.4%)
# of KAKEN inventors with splitting error	233 (3.89%)	2 (0.03%)	1,227 (20.50%)
# of KAKEN inventors with lumping error	14 (0.23%)	42 (0.70%)	6 (0.10%)
# of KAKEN records with splitting error	288 (2.41%)	2 (0.02%)	2,233 (18.67%)
# of KAKEN records with splitting error	34 (0.28%)	65 (0.54%)	8 (0.08%)

From the inventor disambiguation result, we extract 62,983 inventors as academia. We then match these academic inventors with the authors of scientific papers. List of scientific papers are derived from the Elsevier Scopus database. From the list of scientific papers, we use the papers written by authors of which country of affiliation is Japan. Although matching is performed on inventor/author name and affiliation, both inventor and applicant name recorded in Japanese in the IIP patent database but the corresponding information are recorded in English in the Scopus database. IIP patent database, however, can be easily mapped with PATSTAT database. We replace original inventor name recorded in Japanese with the information on corresponding record in the PATSTAT. For the affiliation information, we use the *NISTEP Dictionary of Names of Universities and Public Organizations*, and its converter for the Scopus database provided by NISTEP. Among more than 9.7 million author-affiliation-paper instances in the Scopus database, we could successfully map the identifier of affiliation developed by NISTEP for 5.3 million Japanese instances. As a result, 30,432 inventors (48.3%) among 62,983 academic inventors are successfully matched with the authors in the Scopus database on inventor/author name and identifier of applicant/affiliation. Since the Scopus' author ID and the disambiguated inventors, however, are not completely matched each other, we combine inventor/author IDs iteratively until they are uniquely matched. Finally, almost 2,000 inventor IDs are integrate into the other and we obtain 28,433 matched inventors/authors.

Figure 1 illustrates the number of inventors and academic authors in Japan during the period from 2008 to 2011 based on the matching results. There were 563 thousand inventors and there were 382 thousand authors publishing a paper on an academic journal included in Scopus database. Among the patent inventors, 30.5 thousand inventors are affiliated to academic institutions and 15.6 thousand inventors published at least one paper on Scopus journals.

Figure 1. Patent inventors and academic authors active in 2008-2011 in Japan

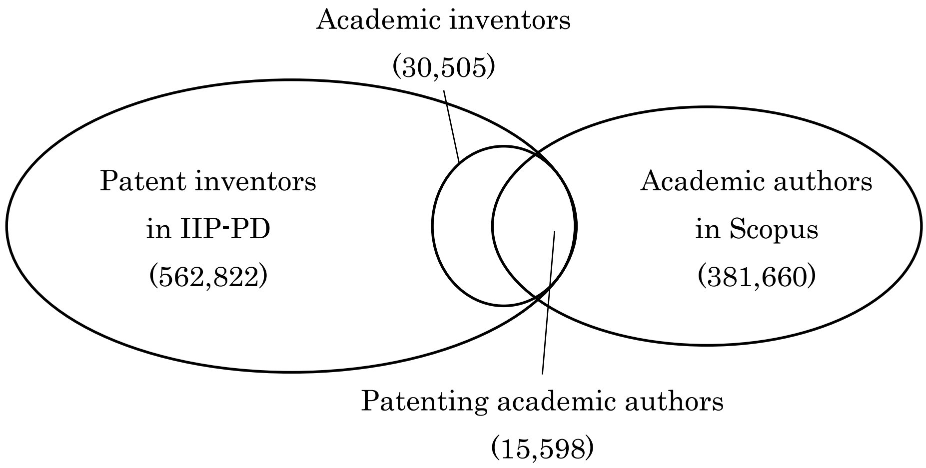


Table 2.1.2 shows that the proportion of academia in inventors increased from 3.2% in the period 2000-2003 to 5.4% in the period 2008-2011. Proportion of academic authors

with patent invention also increased from 3.0% in 2000-2003 to 4.1% in 2008-2011. Furthermore, proportion of academic authors in total inventors has doubled from 1.4% to 2.4% during the 12 years.

Table 2. Patent inventors and academic authors active in 2000-2011 in Japan

	2000-03	2004-07	2008-11	Total
# of authors [A]	316,031	355,936	381,660	739,372
# of all inventors [B]	673,927	623,849	562,822	1,229,027
# of academic inventors [C]	21,437	31,421	30,505	53,446
Proportion of academia in inventors [C / B]	3.2%	5.0%	5.4%	4.3%
# of patenting authors [D]	9,532	15,726	15,598	26,333
Proportion of inventors in authors [D / A]	3.0%	4.4%	4.1%	3.6%
Proportion of authors in inventors [D / B]	1.4%	2.5%	2.8%	2.1%

2-2. Firm level linkage of IIP patent database and Economic Census

We link the patent information from IIP Patent Database with establishment census data at organization level. As in Section 2-1, we focus on non-individual patent applications in which applicant addresses are in Japan. The patent applicants consist of organizations including firms, public institute, and universities. The number of the applications from 1964 to 2013 is 10,253,009, and the total number of applications in the period is 11,038,633. As for the establishment census, the following five data sets are used: the *Enterprise and Establishment Census* published in 2001, 2004, and 2006, and the *Economic Census of Japan* published in 2009 and 2012. We link the application data with each of the census data sets. This approach allows us to find the linkage from an applicant firm that did not exist at the time when either one of the census surveys was conducted. Table 3 shows the total number of establishments in each census data set and the breakdown by establishment type defined as follows: (1) the head office of a firm with multiple establishments (Headquarter) (2) a branch of a firm with multiple establishments (Branch), and (3) a single unit establishment (Single Est.).

Table 3: Number of Establishments by Type

Survey Year	Headquarter	Branch	Single Est.	Total

2001	229,436	1,185,929	4,722,947	6,138,312
2004	262,994	1,141,894	4,323,604	5,728,492
2006	228,664	1,255,827	4,238,068	5,722,559
2009	287,715	1,375,189	4,193,038	5,855,942
2012	270,634	1,296,421	3,855,672	5,422,727

As we focus on patent applications by non-individual Japanese applicants, they are applied from any one of establishments in Japan. Noting that patent applications are usually managed by an entire organization instead of an individual establishment, we establish a link from an applicant to the establishment where the head office exists. To do so, we make use of a unique organization identifier which is assigned to all the establishments the organization owns. To sum up, our methodology links applicant information with a headquarter of multi-establishment firm, or a single-establishment firm.

In implementing the linking methodology outlined above, we employ name and address information that are available both in applicant records of IIP Patent Database and establishment records of a census. Several issues arise when using these pieces of information, which are described as follows. First, names and addresses of applicants may contain spelling errors, and their format may be different between an applicant record and an establishment record. To get around a false negative problem in which the same entities are judged as different due to these notational variations, we develop a series of text processing programs to convert from the raw name/address data to its standardized representation. Second, both applicant and establishment addresses undergo changes by the consolidation of local administrative units such as municipalities. To cope with address changes of this kind, we again use the commercialized software from *Kokusai Kogyo* to convert the original addresses to the latest address format (as of 2014). Lastly, while an applicant address is written in a single line, an establishment address in a census has been recorded as a collection of five geographical components (prefecture, city or wards, district, street, and any others that follow such as a building name or a room number). In order for these different address formats to be comparable, we develop a text-scanning program to break a single line of an applicant address into those five parts. We then define a list of a prefectural name, a city (ward) name, a district name, and the street part to be the standardized representation of address by which applicant and establishment addresses are compared. With these standardized names and addresses, the basic strategy in our linking methodology is to regard them as primary information to establish a one-to-one link from an applicant to an establishment. In order for this strategy to work accurately and efficiently, we implement the linking procedure consisting of the following two steps.

In the first step, any establishments whose names exactly match with, or include, the applicant name are collected to form a set of "candidate" establishments. In the second step, the candidate set is refined by the extent to which the four components (prefecture, city or ward, district, and street) of the applicant address match with those of the address of a candidate establishment. The procedure makes a link from the applicant to an establishment whose address matches with the applicant address at the finest level. We then look at the organization identifier of the establishment, and relink to the headquarter establishment having the same organization identifier if it is a branch. The procedure finishes by linking the applicant with an organization that occupies the establishment.

For each applicant in a patent application, the applicant is successfully matched with an organization if a single headquarter establishment of the organization is linked in the second step of the linking procedure. We also regard as successful the case where more than two establishments are found in the second step, because we think it to be rare that any different organizations of the same name are found even in the largest geographical unit of the address components (prefecture). In this case, the establishment of the largest employment is linked and its owner is matched. The procedure fails to match if the candidate establishment set in the first step is empty, or even the prefectural part of establishment addresses does not match with that of the applicant address in the second step.

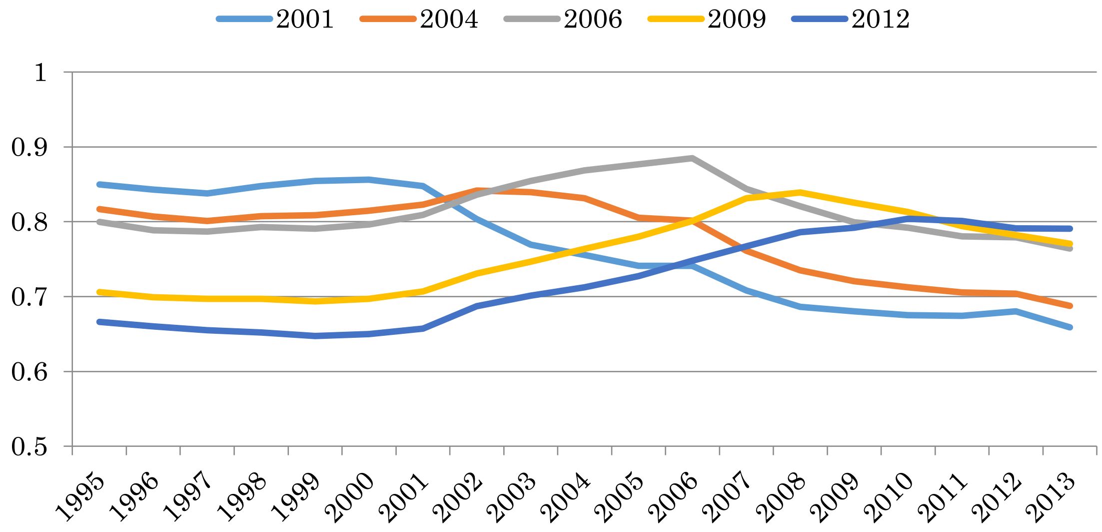
The results from the linking procedure for patent applications between 1964 and 2013 are shown in Table 4. In any of census data sets, the linking procedure finds that about 2% of organizations applied for patents. The rate of patent applications whose applicants have been found in the census data ("Matching Rate") is highest for 2001 census data, and is decreasing for the later census data sets. This may reflect the fact that the applicants that did not exist by the later census years are not included in the corresponding data sets. Meanwhile, these data sets should include applicants that appeared after the previous survey years.

Table 3: The Result of Patent-Organization Linkage
(for patent applications in 1964-2013)

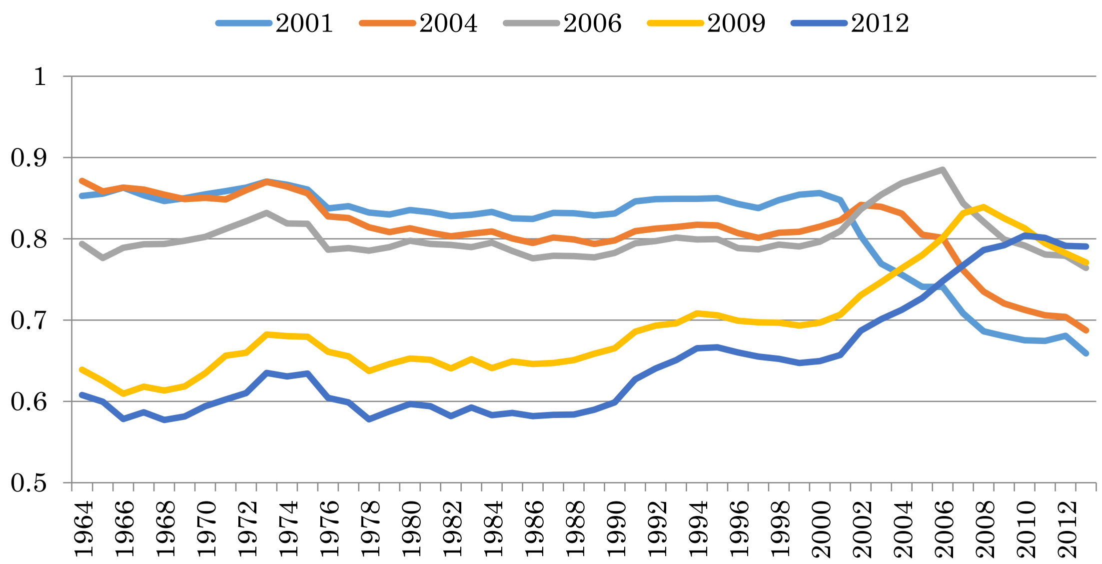
Census Year	2001	2004	2006	2009	2012
The Number of Organizations in Census	5,336,971	5,686,451	4,897,132	4,769,171	4,608,794
with Patent Applications	91,697	91,041	91,392	92,821	90,394
Percentage	1.72%	1.60%	1.87%	1.95%	1.96%
The Number of Patent Applications	10,253,009	10,253,009	10,253,009	10,253,009	10,253,009
Matched with Census	9,325,159	9,227,935	9,167,304	9,097,932	8,540,617
Matching Rate	90.95%	90.00%	89.41%	88.73%	83.30%

To see this more in detail, we observe the matching rates by application year. Figure 2 shows the result. As seen in Figure 2 (a), the peak of yearly matching rates for each census data set locates around the year when the census is published. Therefore, our linking procedure works well for patents applied around the census year.

Figure 2: Temporal Performance of the Linking Procedure
(a) For Applications between 1995 and 2013



(b) For Applications in the Full Period (1964-2013)



If we see the results in the whole application period as shown in Figure 2 (b), it is observed that the matching rates are reasonably high for patent applied in 1960s to 1990s. It is suggested that many of patents in this period are applied by organizations that are still active in the later census years.

Lastly, we assess the quality of the linking procedure. As shown in the figure above, the matching rates are high for patents applied around the census years, and can be low for patent applications away from the census years. Therefore they may not be a consistent indicator for the quality of the linking procedure. Instead, we look at patent applications and applicants that are failed to establish a link for *all* of the census data sets. These applicants include organizations which had applied for patents and did not exist before 2001, those which existed only between the census years, or those mistakenly judged as failures by errors in the implementation. Table 4 shows these failure cases generated in the linking procedure. While considerable numbers of patent applicants (56,746) have not been found in the census data, their applications only account for about 5.25% of total patent applications. Therefore, it can be concluded that most of patent applications are successfully linked to establishments included in either one of the census data sets.

Table 4: Patent Applications and Applicants Failed in the Linking Procedure for All of the Census Data Sets. Note that the applicants are identified by their names and address.

The total number of patent applications	11,038,633
Failed with all census data sets	579,429
Percentage	5.25%
The number of Failed Applicants	56,746

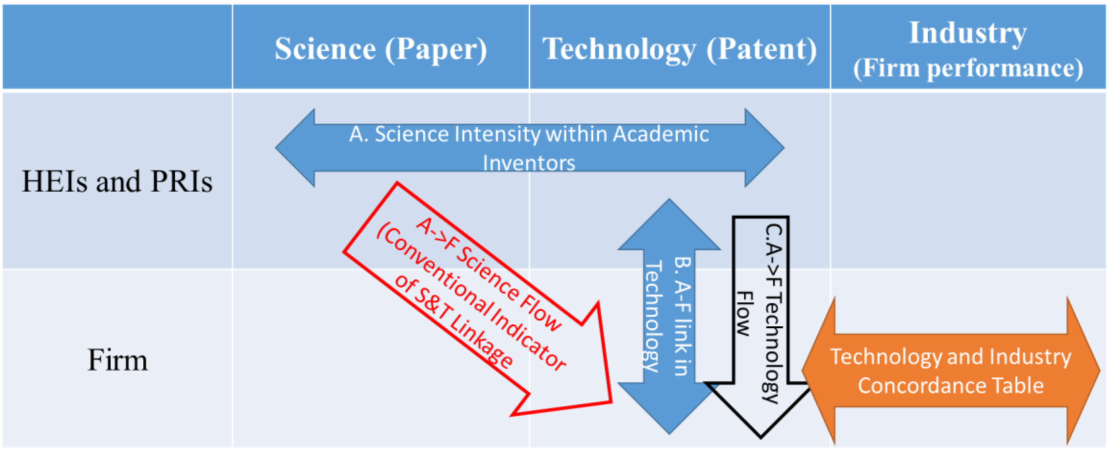
3. New indicators of science and industry linkage

3-1. Framework of indicators

Figure 1 illustrate various channels where scientific knowledge at higher educational institutes (HEIs) and public research institutions (PRIs) are used in firms. A conventional approach of science linkage is to look at non patent literature citations by patents, which can be indicated as “A->F Science Flow (science flow from Academic to Firm)” in Figure 3.[3]

Figure 3: Famework of indicators

[3] Non-patent literature include not only scientific papers, but non-scientific materials such as technical documents, but this information is often used for science linkage index. In addition, there are some scientific papers, published by industry researchers (outside HEIs and PRIs sectors), but its portion to total publications are very small. Therefore, we treat all NPL as a scientific output from academic (HEIs and PRIs) sector in Figure 1.



While NPL citation of patent information captures disembodied knowledge flow, there are other channels where scientific knowledge is used in industry sectors. Here, we propose two alternative science linkage indicators, reflecting knowledge flow via academic inventors. As it is shown later, substantial numbers of academic researchers in HEIs and PRIs, writing scientific papers, also apply patents. In this case, scientific knowledge is converted to technology for industrial use (patent) within the researcher by herself (Part A of Figure 1). Such technological contents can be flown to industry via either academic-industry collaboration in patents (joint patent applications) (Part B of Figure 1) or industry patent citation to the academic inventor patents (Part C of Figure 1).

Therefore, the alternative scientific linkage indicators are developed along the lines of A->B and A->C. In addition, all of three types of indicators (NPL citation one and two alternative indicators) by technology class of patents can be converted to the ones by industry, by using technology and industry concordance table.

3-2. Science and technology linkage by academic inventors

In this section, the methodology of science and technology linkage indicators by academic inventors (Part A of Figure 1) is presented. The original data is both the vector of paper count by academic fields (Rs) and the vector of patent count by technology fields (Pt) for individual auther/inventor (i). The number of paper per patent by technology field is calculated by

$$L_t^{\text{Paper}} = \frac{\sum_s W_{st}^{Paper}}{\sum_{it} P_{it}}$$

Where W_{st}^{Paper} is a matrix of number of paper by academic field "s" and technology field "t", calculated by

$$W_{st}^{\mathrm{Paper}} = \sum_{i} \left(R_{is} \times \frac{P_{it}}{\sum_t P_{it}} \right)$$

The table 5 shows the results of "L" (column C). In addition, the number of academic patens (column B) as well as the total number of patents (column "A") are also presented to see the presence of patents invented by academic inventors. The technology field is defined by WIPO classification, based on IPC code.

(Table 5)

While the total number of patent decreased from 2004 (column A), the number of academic patent increased for the period of 2004-2007 (column B). It should be noted that national universities of Japan became independent agency in 2004, then they have been allowed to have patent application as an institutional applicant since then. However, the number stopped increasing after 2008. In contrast, the number of paper per patent at academic inventor (column C) shows persistent increase. This reflect scientific contents involved at academic invention (science intensity to technology at academic inventors increases over time. This indicator is particularly high at life science technology, such as "pharmaceutical", "biotechnology" and "analysis of biological materials". However, the variation by technology is rather small, as is shown that low science intensity technologies such as "textile and paper machine" and "civil engineering" has certain values, more than one fifth of the highest value found in "pharmaceutical".

The table 6 shows the results of W for the period of 2000-2003 and 2008-2011. There are some academic disciplines contributing scientification of technologies, such as "biochemistry", "chemical", chemical engineering", "material science", "engineering" and "physics". In addition, it is clearly shown that the intensity of science (research paper) to patents increases over time, by comparing two tables. In addition to the academic disciplines mentioned above (showing significant contributions in the period of 2000-2003), "computer science" and "mathematics" become to be major contributors to science and technology linkage.

(Table 6a and 6b)

3-3. Comparison of three types of S-T linkage indicators

In order to compare with NPL citation indicator, the results of "L" above, showing # of papers per patent by academic inventor, should be multiplied by the share of academic form joint invention patents to industry patents (part B of Figure 1), or academic patent citations by industry patents (part C of Figure 1). The two alternative science linkage indicators reflect the knowledge flow of science at academia to firms by joint invention activities for A*B and the knowledge flow of science at academic to firms by patent citation for A*C, respectively.

Table 7 shows the results of two indicators mentioned above as well as the number of NPL citation by patent (a traditional science linkage indicator) by technology class. Here, NPL citations by technology are calculated by the count of NPL citations by patent manuscript by text mining (Tamada et. al, 2006), since the applicant citation information is not available in the front page information of JPO patents. It is shown that our NPL citation information gives similar distribution by technology to the ones by USPTO patents, used in existing studies (Narin and Norma, 1985; van Looy et, al, 2002), in Appendix 2.

In general, a common pattern is found in the distribution of indicator by technology, such as "biotechnology" has the highest science intensity for all three indicators. However, the time trend is quite difference, that is, two alternative indicators increase over time, while NPL indicator is stable for these 10 years. More detail look at the patterns of the science intensity by joint patent application (A*B) and one by academic patent citation (A*C), the former stop increasing after 2008, while the latter show more persistent increase over time period. The incorporation of Japanese national university in 2004 maybe the reason why joint patent inventions between academia and industry increased after 2004, which leads to subsequent increases in industry patent citations to academic patents, by more interactive invention process between these two parties.

(Table 7)

We convert these indicators by technology to the ones by industry, by using technology industry concordance tables. The concordance tables are created by linked data of IIP patent database and economic census of Japan, described in the section 2-2. We use the data for three time period, (1) 2001 census data linked with the patents of 2000-2003, (2) 2006 census data linked with the patents of 2004-2007 and (3) 2009 census data linked with the patents of 2008-2011. Using the matrix of industry "i" and technology "t" of patent, C_{it}^{Patent}, the three types science linkage indicator by technology "t" (Sit) can be converted to the one by industry "i", (SIi) as follows

$$SI_i = C_{it}^{Patent} \cdot SI_t$$

Finally, the value of SIi is normalized by the number of employees by industry. Since all three types of indicators reflect different channel of scientific knowledge flow, they can be added up to the aggregated science linkage indicators. The Figure 4a and 4b presents this aggregated indicators for the period of 2000-2003 and one of 2008-2011, respectively.

(Figure 4a) and (Figure 4b)

In the period of 2000-2003, the total science linkage is driven by NPL citation, disembodied scientific knowledge spillover by academic papers. However, more and more academic researchers has been involved with patenting activities, and the science linkages involving academic patenting contributed to almost half of total in the period of 2008-2011. In addition, the variation of total science linkage intensity decreased over time. In the period of 2000-2003, "chemicals" shows the highest value, leading the other sectors by great margin, while "ICT machinery" and "technology services" almost caught up with "chemicals" in the period of 2008-2011. In addition, there are many other sectors relatively uprising as compared to these high ranking sectors between these two periods.

4. Discussion and Conclusion

This paper presents new indicators measuring scientification of industry, by linking scientific paper database (Science), patent information (Technology) and economic census data (Industry) in Japan. The new indicators reflect some mechanism of science linkage with technology and industrial activities, which cannot be measured by non patent literature (NPL) citation of patent, capturing pure disembodied knowledge flow. That is, the linkage of scientific publications and patents at researcher level allows us to look at co-occurrence of science and technology embodied at human capital, which leads to scientific knowledge flow to industry subsequently.

New indicators of science linkage in Japan show increasing trend over these 10 years, while a traditional indicator (NPL citation) is stagnated during this period. More precisely, the indicator of co-inventions of academia industry increased after 2004, and its speed slowed down after 2008,. In contrast, he indicator based on industry citation of academic patens kept increasing over the whole period of 2000-2011. One of reason behind such trends is the institutional reform of academic sector in Japan, i.e., incorporation of national universities in 2004. In addition, the Japanese Government introduced various polices stimulating university industry collaborations from the late 1990's, such as TLO (Technology Licensing Organization) Promotion Law and Japanese Baye Dole Act (Motohashi and Muramatsu, 2012). All of such policy actions induce academic sectors (both HEIs and PRIs) to work together with industry which involves patenting activities.

The government policies are not only factors behind the trend of science linkage of industry. Growing importance of scientific inputs in industrial innovation affects as well. The 21st century started with the news of completing analysis of human genome sequence. Big-data analysis allows scientific understanding on business and economics activities, such as purchasing behavior and production process at factory. In our analysis, science linkage of industry is found not only in science based industry, such as pharmaceuticals and electronics, but also in many other industries. Actually the variation of the total science intensity index by industry decreases in these 10 years. Studies on the taxonomy of innovation suggest sectoral differences in its characteristics, and science based industry is one of the categories (Pavitt, 1984; Breschi and Malerba, 1997). However, our study has shown that scientific inputs become general inputs to almost all industries, and this trend can be called as "science based economy", for non science based industries as well.

Hence, public spending on science sectors should be supported, as scientific findings contribute to industrial innovations for over all economies, instead of benefiting limited number of science based industries. In addition, further interactions of academia and industry should be promoted, since direct interactions between them become more important source of science linkage than disembodied knowledge flow from science to industry, captured by NPL citations. Some concrete examples of such policies include promoting corporate research center inside universities and university based startups.

References

Breschi, S. and F. Malerba, 1997, Sectoral innovation systems: technological regimes, Schumpeterian dynamics and spatial boundaries, in (C. Edquist ed.) *Systems of Innovation, Technologies, Institutions and Organization,* pp. 130-156, London:

Cassel

European Patent Office, *The Worldwide Patent Statistical Database (PATSTAT)*, Autumn 2015 edition.

Goto, A. and K. Motohashi (2007), Construction of a Japanese Patent Database and a first look at Japanese patenting activities, *Research Policy*, 36(9), 1431-1442

Li, GC., R Lai, A D'Amour, DM Doolin, Y Sun, VI Torvik, ZY Amy, L Fleming (2014), Disambiguation and co-authorship networks of the US patent inventor database (1975–2010), *Research Policy* 43 (6), 941-955

Lissoni, F, Montabio, F. and Zirulia, L., 2013, Inventorship and authorship as attribution rights: an enquiry into the economics of scientific credit, *Journal of Econojic Behavior and Organization*, 95: 49-69

Magerman T., Van Looy, B and Debackere, K., 2015, Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology, *Research Policy*, 44(9): 1702-1713

Narin F. and Norma E., 1985, Is technology becoming science? *Scientometrics*, 7: 368-381

Pavitt, K. (1984), Sectoral patterns of innovation: toward a taxonomy and a theory, *Research Policy*, 28(6): 643-660

Schmoch, U., 1997, Indicators and relations between science and technology, *Scientometrics*, 38(1): 103-116

Tamada, Schumpeter, Yusuke Naito, Fumio Kodama, Kiminori Gemba, and Jun Suzuki (2006) "Significant difference of dependence upon scientific knowledge among different technologies," *Scientometrics*, 68(2) pp.289-302.

Van Looy, Zimmermann, E., Veugelers, R., Verbeek, A. and Debackere, K., 2002, Do science-technology interactions pay off when developing technology? An exploratory investigation of 10 science-intensive technology domain, KU Lueven Research Report 0244

Table 5: Science intensity by technology field by academic inventors

		# of all patent / 1000 (A)			# of academic patent / 1000 (B)			# of paper per patent by academia (C)		
WIPO Technology Classification		2000-03	2004-07	2008-11	2000-03	2004-07	2008-11	2000-03	2004-07	2008-11
1	Electrical machinery, apparatus, energy	144.74	148.06	144.48	2.80	3.24	3.24	3.91	6.36	6.00
2	Audio-visual technology	110.96	110.01	79.67	1.24	1.33	1.07	1.83	3.28	4.04
3	Telecommunications	84.54	84.23	62.51	1.49	1.66	1.44	1.73	3.67	4.56
4	Digital communication	50.12	56.55	57.09	0.82	1.05	1.18	1.54	3.79	3.85
5	Basic communication processes	19.52	18.93	16.19	0.32	0.40	0.40	2.93	5.55	6.70
6	Computer technology	154.14	124.31	101.58	2.39	2.86	2.45	2.36	5.37	6.31
7	IT methods for management	0.00	15.69	18.72	0.00	0.21	0.31	-	4.73	6.85
8	Semiconductors	91.62	99.21	91.41	3.02	3.54	3.30	4.59	6.25	7.24
9	Optics	135.16	133.27	109.67	2.59	2.42	2.29	3.63	6.11	5.95
10	Measurement	74.57	77.03	69.57	2.95	3.91	3.42	4.56	7.01	8.41
11	Analysis of biological materials	15.96	18.01	16.00	1.28	1.84	1.52	5.74	8.65	10.75
12	Control	35.37	31.51	26.75	0.52	0.66	0.59	2.04	5.69	7.42
13	Medical technology	55.91	65.59	61.40	1.37	2.66	2.48	6.06	8.47	10.19
14	Organic fine chemistry	49.51	50.46	44.20	2.56	3.65	3.30	6.85	9.14	10.35
15	Biotechnology	26.40	25.86	23.90	3.80	4.93	4.12	8.71	10.62	13.33
16	Pharmaceuticals	15.91	17.60	15.37	0.64	1.27	1.14	9.03	11.49	14.29
17	Macromolecular chemistry, polymers	42.94	39.04	34.79	1.47	1.60	1.46	3.99	6.35	5.02
18	Food chemistry	15.17	13.38	11.20	0.41	0.50	0.41	4.26	6.72	7.49
19	Basic materials chemistry	39.60	35.16	31.35	0.93	1.35	1.30	4.72	5.65	6.09
20	Materials, metallurgy	43.34	36.54	33.82	2.81	3.12	2.71	5.14	7.50	8.40
21	Surface technology, coating	36.55	34.06	30.32	1.36	1.63	1.36	5.81	6.35	6.31
22	Micro-structural and nano-technology	1.19	1.44	0.95	0.20	0.26	0.12	5.12	9.21	8.99
23	Chemical engineering	31.93	26.43	22.29	1.43	2.00	1.65	4.50	5.60	6.84
24	Environmental technology	37.15	26.95	22.43	1.04	1.03	0.89	3.76	6.15	6.97
25	Handling	73.54	58.66	49.74	0.32	0.42	0.37	3.55	5.18	6.13
26	Machine tools	51.83	40.27	35.40	0.67	0.71	0.63	2.79	4.51	5.32
27	Engines, pumps, turbines	53.38	55.14	50.96	0.61	0.66	0.58	3.50	5.77	7.24
28	Textile and paper machines	66.24	56.20	45.01	0.54	0.58	0.59	1.68	2.78	2.95
29	Other special machines	64.79	50.83	43.77	1.06	1.29	1.13	4.42	7.32	6.95
30	Thermal processes and apparatus	37.35	31.76	29.57	0.36	0.42	0.41	2.31	3.72	4.29
31	Mechanical elements	57.77	56.49	45.93	0.38	0.49	0.37	2.50	4.17	5.82
32	Transport	71.51	81.40	72.13	0.59	0.79	0.71	1.61	3.00	4.05
33	Furniture, games	72.97	70.23	61.34	0.10	0.15	0.12	2.89	4.48	4.84
34	Other consumer goods	40.16	32.92	27.22	0.14	0.21	0.15	2.93	3.56	3.59
35	Civil engineering	79.77	56.77	46.01	0.57	0.66	0.59	1.15	2.78	3.21
	Average	56.62	53.71	46.65	1.22	1.53	1.37	3.89	5.91	6.76

Table 6a: Matrix of science field and technology class by academic inventors (2000-2003)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
	Science field (Elseviar ASJC)	Electrical machinery, apparatus, energy	Audio-visual technology	Telecommunications	Digital communication	Basic communication processes	Computer technology	IT methods for management	Semiconductors	Optics	Measurement	Analysis of biological materials	Control	Medical technology	Organic fine chemistry	Biotechnology	Pharmaceuticals	Macromolecular chemistry, polymers	Food chemistry	Basic materials chemistry	Materials, metallurgy	Surface technology, coating	Micro-structural and nano-technology	Chemical engineering	Environmental technology	Handling	Machine tools	Engines, pumps, turbines	Textile and paper machines	Other special machines	Thermal processes and apparatus	Mechanical elements	Transport	Furniture, games	Other consumer goods	Civil engineering
10	Multidisciplinary	0.01	0.00	0.00	0.01	0.01	0.01	-	0.02	0.01	0.02	0.03	0.00	0.01	0.03	0.10	0.05	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.02	0.00	0.02	0.00	0.04	0.00	0.00	0.00	0.01	0.01	0.00
11	Agricultural and Biological Sciences	0.02	0.00	0.00	0.00	0.00	0.02	-	0.01	0.01	0.07	0.11	0.01	0.06	0.22	0.70	0.40	0.07	1.71	0.35	0.02	0.01	0.02	0.06	0.15	0.02	0.04	0.04	0.04	0.37	0.05	0.01	0.01	0.07	0.02	0.03
12	Arts and Humanities	0.00	0.00	0.00	0.00	0.00	0.00	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	Biochemistry, Genetics and Molecular Biology	0.08	0.03	0.03	0.05	0.05	0.19	-	0.08	0.07	0.42	0.77	0.14	0.89	1.45	3.27	2.65	0.35	1.08	0.38	0.09	0.12	0.20	0.20	0.27	0.22	0.05	0.10	0.12	0.80	0.06	0.10	0.04	0.08	0.16	0.07
14	Business, Management and Accounting	0.00	0.00	0.00	0.00	0.00	0.00	-	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	Chemical Engineering	0.13	0.01	0.01	0.00	0.02	0.03	-	0.06	0.03	0.11	0.16	0.04	0.23	0.23	0.22	0.12	0.22	0.15	0.30	0.28	0.15	0.15	0.65	0.32	0.05	0.11	0.11	0.12	0.15	0.19	0.07	0.05	0.04	0.08	0.03
16	Chemistry	0.51	0.06	0.01	0.02	0.02	0.05	-	0.20	0.21	0.51	0.99	0.02	0.18	2.09	0.40	0.55	1.02	0.26	0.90	0.58	0.52	0.68	1.03	0.46	0.06	0.13	0.18	0.31	0.29	0.11	0.08	0.04	0.21	0.13	0.05
17	Computer Science	0.04	0.13	0.33	0.38	0.38	0.59	-	0.04	0.07	0.14	0.05	0.34	0.18	0.01	0.03	0.01	0.02	0.00	0.02	0.01	0.02	0.10	0.02	0.05	0.68	0.02	0.03	0.01	0.02	0.04	0.07	0.10	0.55	0.21	0.03
18	Decision Sciences	0.00	0.00	0.00	0.02	0.00	0.01	-	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00
19	Earth and Planetary Sciences	0.03	0.01	0.03	0.02	0.07	0.03	-	0.02	0.01	0.12	0.08	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.04	0.04	0.04	0.03	0.03	0.12	0.01	0.02	0.07	0.00	0.03	0.05	0.04	0.10	0.02	0.05	0.11
20	Economics, Econometrics and Finance	0.00	0.00	0.00	0.00	0.00	0.00	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	Energy	0.12	0.01	0.00	0.00	0.01	0.02	-	0.04	0.02	0.06	0.07	0.02	0.04	0.04	0.01	0.01	0.02	0.01	0.14	0.11	0.06	0.05	0.17	0.18	0.03	0.03	0.41	0.02	0.05	0.26	0.07	0.02	0.00	0.04	0.02
22	Engineering	0.66	0.51	0.86	0.58	1.10	0.61	-	0.57	0.69	0.85	0.58	0.78	0.77	0.08	0.11	0.07	0.11	0.02	0.28	0.38	0.53	0.62	0.41	0.39	1.75	0.68	0.88	0.12	0.22	0.63	1.05	0.77	1.13	0.92	0.38
23	Environmental Science	0.01	0.00	0.01	0.02	0.01	0.01	-	0.01	0.00	0.04	0.05	0.01	0.01	0.03	0.08	0.04	0.03	0.06	0.06	0.02	0.01	0.01	0.06	0.19	0.01	0.01	0.04	0.01	0.05	0.03	0.05	0.01	0.01	0.03	0.02
24	Immunology and Microbiology	0.01	0.00	0.00	0.02	0.01	0.03	-	0.02	0.01	0.08	0.14	0.02	0.08	0.22	0.85	0.58	0.04	0.24	0.07	0.01	0.01	0.01	0.03	0.12	0.01	0.01	0.01	0.02	0.36	0.02	0.01	0.01	0.00	0.00	0.03
25	Materials Science	0.77	0.25	0.07	0.05	0.25	0.11	-	1.03	0.50	0.52	0.75	0.10	0.61	0.44	0.17	0.20	1.43	0.06	0.94	2.21	1.69	1.42	0.80	0.44	0.15	0.99	0.36	0.52	0.69	0.23	0.39	0.07	0.18	0.26	0.12
26	Mathematics	0.01	0.01	0.01	0.05	0.04	0.07	-	0.01	0.01	0.02	0.01	0.06	0.03	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.03	0.01	0.02	0.00	0.00	0.01	0.01	0.01	0.06	0.08	0.01
27	Medicine	0.10	0.06	0.06	0.08	0.06	0.21	-	0.09	0.06	0.29	0.44	0.18	1.79	0.87	1.72	2.60	0.16	0.24	0.18	0.08	0.09	0.10	0.07	0.18	0.17	0.04	0.07	0.04	0.57	0.09	0.09	0.13	0.11	0.18	0.10
28	Neuroscience	0.01	0.02	0.01	0.01	0.01	0.04	-	0.01	0.01	0.06	0.09	0.03	0.21	0.17	0.32	0.47	0.03	0.05	0.01	0.01	0.02	0.02	0.02	0.04	0.04	0.01	0.01	0.00	0.18	0.01	0.01	0.01	0.07	0.05	0.01
29	Nursing	0.00	0.00	0.00	0.00	0.00	0.00	-	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
30	Pharmacology, Toxicology and Pharmaceutics	0.01	0.01	0.01	0.02	0.02	0.02	-	0.01	0.01	0.07	0.12	0.01	0.12	0.58	0.33	0.85	0.12	0.16	0.06	0.01	0.01	0.09	0.05	0.03	0.08	0.01	0.01	0.02	0.12	0.02	0.01	0.01	0.01	0.02	0.01
31	Physics and Astronomy	1.38	0.71	0.28	0.19	0.88	0.28	-	2.35	1.90	1.14	1.25	0.22	0.59	0.30	0.21	0.19	0.32	0.12	0.93	1.24	2.50	1.58	0.88	0.79	0.21	0.63	1.12	0.30	0.44	0.47	0.42	0.20	0.25	0.62	0.12
32	Psychology	0.00	0.00	0.00	0.00	0.00	0.00	-	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00
33	Social Sciences	0.00	0.00	0.00	0.00	0.00	0.01	-	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.05	0.00
34	Veterinary	0.00	0.00	0.00	0.00	0.00	0.00	-	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.06	0.05	0.00	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01	0.00	0.00
35	Dentistry	0.00	0.00	0.00	0.01	0.00	0.00	-	0.00	0.00	0.01	0.02	0.00	0.12	0.03	0.03	0.10	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
36	Health Professions	0.00	0.00	0.00	0.00	0.00	0.01	-	0.00	0.00	0.01	0.01	0.00	0.04	0.01	0.01	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.01	0.02	0.01	0.00	0.02	0.01	0.00

Table 6b: Matrix of science field and technology class by academic inventors (2008-2011)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
	Science field (Elseviar ASJC)	Electrical machinery,	Audio-visual technology	Telecommunications	Digital communication	Basic communication	Computer technology	IT methods for management	Semiconductors	Optics	Measurement	Analysis of biological materials	Control	Medical technology	Organic fine chemistry	Biotechnology	Pharmaceuticals	Macromolecular chemistry, polymers	Food chemistry	Basic materials chemistry	Materials, metallurgy	Surface technology, coating	Micro-structural and nano-technology	Chemical engineering	Environmental technology	Handling	Machine tools	Engines, pumps, turbines	Textile and paper machines	Other special machines	Thermal processes and apparatus	Mechanical elements	Transport	Furniture, games	Other consumer goods	Civil engineering
10	Multidisciplinary	0.01	0.01	0.01	0.00	0.02	0.02	0.02	0.03	0.02	0.04	0.07	0.01	0.03	0.07	0.17	0.11	0.02	0.10	0.03	0.03	0.03	0.06	0.02	0.05	0.01	0.02	0.03	0.01	0.07	0.01	0.02	0.01	0.00	0.00	0.01
11	Agricultural and Biological Sciences	0.03	0.02	0.01	0.01	0.02	0.06	0.05	0.01	0.02	0.11	0.21	0.02	0.07	0.25	0.91	0.43	0.07	2.74	0.44	0.03	0.03	0.03	0.09	0.22	0.04	0.09	0.05	0.07	0.96	0.07	0.03	0.04	0.03	0.05	0.16
12	Arts and Humanities	0.00	0.02	0.01	0.01	0.00	0.05	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
13	Biochemistry, Genetics and Molecular Biology	0.15	0.07	0.05	0.06	0.10	0.36	0.50	0.11	0.11	0.77	1.55	0.24	1.01	2.17	4.06	3.64	0.36	1.92	0.49	0.19	0.18	0.32	0.28	0.34	0.16	0.11	0.16	0.20	1.07	0.11	0.10	0.07	0.21	0.24	0.05
14	Business, Management and Accounting	0.00	0.00	0.00	0.01	0.00	0.02	0.05	0.00	0.00	0.00	0.00	0.03	0.01	0.01	0.01	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00
15	Chemical Engineering	0.20	0.04	0.01	0.01	0.03	0.04	0.02	0.10	0.06	0.19	0.29	0.14	0.23	0.47	0.33	0.19	0.39	0.19	0.38	0.42	0.21	0.43	0.84	0.40	0.07	0.09	0.18	0.22	0.23	0.35	0.17	0.05	0.08	0.03	0.04
16	Chemistry	0.65	0.10	0.03	0.04	0.07	0.09	0.07	0.37	0.28	0.55	1.02	0.09	0.19	2.08	0.64	0.66	1.36	0.41	0.95	1.00	0.70	1.35	1.29	0.48	0.06	0.17	0.37	0.37	0.46	0.22	0.13	0.04	0.20	0.19	0.05
17	Computer Science	0.22	0.76	1.27	1.65	1.35	2.51	1.63	0.19	0.34	0.57	0.22	2.10	0.69	0.06	0.14	0.07	0.07	0.05	0.12	0.08	0.09	0.27	0.10	0.14	2.39	0.17	0.26	0.06	0.11	0.19	0.60	0.45	1.23	0.59	0.24
18	Decision Sciences	0.00	0.00	0.00	0.01	0.01	0.03	0.05	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19	Earth and Planetary Sciences	0.03	0.01	0.02	0.02	0.08	0.04	0.06	0.03	0.05	0.20	0.15	0.02	0.04	0.02	0.05	0.02	0.01	0.02	0.08	0.08	0.02	0.03	0.08	0.27	0.02	0.07	0.19	0.00	0.05	0.09	0.11	0.21	0.05	0.00	0.34
20	Economics, Econometrics and Finance	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	Energy	0.37	0.01	0.02	0.01	0.08	0.03	0.10	0.09	0.02	0.12	0.12	0.11	0.04	0.05	0.05	0.02	0.04	0.02	0.21	0.25	0.11	0.08	0.27	0.27	0.04	0.07	0.99	0.04	0.06	0.40	0.13	0.17	0.03	0.01	0.09
22	Engineering	1.50	1.04	1.67	1.13	2.70	1.29	0.84	1.72	1.22	1.84	1.33	2.48	1.45	0.27	0.44	0.34	0.35	0.22	0.67	1.38	1.12	2.08	0.92	0.82	2.48	1.92	1.57	0.49	0.70	1.04	2.19	1.98	1.49	0.79	1.50
23	Environmental Science	0.04	0.01	0.01	0.00	0.01	0.02	0.09	0.01	0.01	0.09	0.12	0.02	0.04	0.08	0.14	0.06	0.07	0.14	0.15	0.08	0.03	0.04	0.17	0.49	0.01	0.04	0.16	0.03	0.17	0.30	0.06	0.03	0.02	0.03	0.15
24	Immunology and Microbiology	0.02	0.01	0.01	0.01	0.01	0.03	0.05	0.02	0.01	0.11	0.22	0.02	0.10	0.36	0.98	0.77	0.05	0.32	0.08	0.02	0.02	0.01	0.03	0.16	0.01	0.01	0.01	0.02	0.23	0.02	0.02	0.01	0.01	0.02	0.00
25	Materials Science	1.02	0.43	0.28	0.11	0.54	0.18	0.07	1.27	0.83	0.72	0.91	0.19	0.79	0.52	0.38	0.35	1.46	0.19	1.05	2.55	1.48	1.64	1.03	0.55	0.18	1.36	0.63	0.82	0.87	0.41	0.61	0.29	0.22	0.20	0.18
26	Mathematics	0.04	0.10	0.13	0.22	0.21	0.37	0.51	0.03	0.06	0.11	0.07	0.48	0.11	0.02	0.05	0.02	0.02	0.01	0.02	0.03	0.02	0.06	0.03	0.05	0.16	0.05	0.07	0.01	0.03	0.05	0.09	0.08	0.20	0.05	0.04
27	Medicine	0.15	0.14	0.10	0.07	0.10	0.41	2.08	0.16	0.13	1.11	2.19	0.45	3.49	2.28	3.55	5.11	0.18	0.60	0.26	0.15	0.16	0.15	0.16	0.45	0.15	0.09	0.22	0.15	0.93	0.12	0.17	0.14	0.60	0.43	0.09
28	Neuroscience	0.02	0.03	0.01	0.01	0.03	0.09	0.10	0.01	0.02	0.10	0.19	0.08	0.32	0.30	0.30	0.61	0.02	0.03	0.02	0.01	0.01	0.01	0.02	0.05	0.03	0.00	0.03	0.01	0.15	0.01	0.02	0.03	0.05	0.08	0.00
29	Nursing	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.01	0.02	0.03	0.03	0.02	0.07	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.04	0.01	0.00
30	Pharmacology, Toxicology and Pharmaceutics	0.02	0.01	0.01	0.00	0.02	0.04	0.10	0.02	0.02	0.12	0.24	0.03	0.14	0.83	0.45	1.18	0.09	0.16	0.08	0.03	0.02	0.26	0.08	0.04	0.02	0.01	0.02	0.02	0.10	0.02	0.01	0.01	0.02	0.05	0.01
31	Physics and Astronomy	1.51	1.18	0.76	0.37	1.28	0.50	0.18	3.05	2.69	1.58	1.75	0.67	0.97	0.35	0.45	0.31	0.46	0.23	1.03	2.03	2.07	2.15	1.43	2.10	0.23	1.03	2.27	0.40	0.70	0.86	1.31	0.39	0.17	0.72	0.21
32	Psychology	0.00	0.00	0.00	0.00	0.00	0.01	0.06	0.00	0.00	0.01	0.01	0.04	0.03	0.01	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.02	0.00
33	Social Sciences	0.01	0.03	0.13	0.08	0.05	0.10	0.11	0.01	0.04	0.03	0.02	0.11	0.03	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.05	0.00	0.01	0.01	0.01	0.03	0.02	0.03	0.01	0.03	0.03
34	Veterinary	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.00	0.03	0.05	0.10	0.13	0.00	0.09	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.04	0.00	0.00	0.00	0.00	0.01	0.00
35	Dentistry	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.01	0.01	0.24	0.05	0.07	0.14	0.01	0.00	0.00	0.02	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.03	0.01	0.01
36	Health Professions	0.00	0.01	0.00	0.00	0.00	0.03	0.04	0.00	0.00	0.01	0.01	0.02	0.11	0.02	0.02	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.04	0.01	0.00	0.01	0.01	0.02	0.00	0.01	0.01	0.10	0.02	0.00

Table 7: Results of the indicators of science and technology linkages

		Number of NPL citations			Via U-I joint patents (A*B)			Via academic PL citations (A*C)		
		2000-03	2004-07	2008-11	2000-03	2004-07	2008-11	2000-03	2004-07	2008-11
1	Electrical machinery, apparatus, energy	0.195	0.178	0.152	0.019	0.067	0.074	0.047	0.146	0.214
2	Audio-visual technology	0.161	0.207	0.247	0.005	0.012	0.017	0.012	0.050	0.091
3	Telecommunications	0.187	0.310	0.347	0.008	0.022	0.030	0.010	0.061	0.079
4	Digital communication	0.424	0.872	0.940	0.009	0.035	0.035	0.007	0.052	0.074
5	Basic communication processes	0.238	0.352	0.383	0.011	0.035	0.065	0.006	0.034	0.084
6	Computer technology	0.255	0.381	0.481	0.008	0.035	0.045	0.013	0.078	0.157
7	IT methods for management	0.603	0.421	0.580	0.000	0.041	0.055	0.000	0.089	0.173
8	Semiconductors	0.259	0.380	0.394	0.026	0.085	0.105	0.050	0.178	0.431
9	Optics	0.552	0.405	0.267	0.010	0.036	0.039	0.044	0.111	0.141
10	Measurement	0.207	0.256	0.268	0.068	0.181	0.223	0.093	0.322	0.538
11	Analysis of biological materials	2.352	2.673	2.156	0.196	0.737	1.194	0.251	1.245	1.930
12	Control	0.115	0.154	0.220	0.010	0.052	0.082	0.017	0.124	0.197
13	Medical technology	0.199	0.229	0.217	0.066	0.244	0.279	0.092	0.255	0.554
14	Organic fine chemistry	3.002	3.062	2.427	0.116	0.410	0.615	0.260	0.673	1.106
15	Biotechnology	9.866	7.756	6.322	0.759	1.220	1.833	1.202	2.997	6.045
16	Pharmaceuticals	4.079	3.908	3.370	0.243	0.899	1.148	0.527	1.807	2.676
17	Macromolecular chemistry, polymers	0.694	0.848	0.768	0.037	0.157	0.149	0.112	0.304	0.368
18	Food chemistry	0.472	0.655	0.672	0.088	0.220	0.240	0.325	0.781	0.890
19	Basic materials chemistry	0.668	0.704	0.618	0.048	0.164	0.203	0.211	0.361	0.491
20	Materials, metallurgy	0.299	0.421	0.482	0.097	0.287	0.375	0.220	0.597	0.867
21	Surface technology, coating	0.234	0.317	0.369	0.064	0.154	0.164	0.171	0.335	0.531
22	Micro-structural and nano-technology	1.010	0.739	0.399	0.158	0.542	0.642	0.416	0.870	1.139
23	Chemical engineering	0.200	0.252	0.300	0.082	0.212	0.337	0.258	0.503	0.837
24	Environmental technology	0.146	0.187	0.178	0.057	0.165	0.179	0.113	0.398	0.541
25	Handling	0.027	0.040	0.045	0.007	0.016	0.021	0.014	0.041	0.060
26	Machine tools	0.066	0.102	0.101	0.017	0.050	0.061	0.030	0.097	0.131
27	Engines, pumps, turbines	0.041	0.055	0.055	0.017	0.026	0.030	0.021	0.059	0.104
28	Textile and paper machines	0.257	0.243	0.211	0.005	0.013	0.018	0.020	0.038	0.051
29	Other special machines	0.211	0.281	0.243	0.045	0.129	0.120	0.086	0.294	0.355
30	Thermal processes and apparatus	0.034	0.040	0.054	0.010	0.022	0.028	0.013	0.043	0.066
31	Mechanical elements	0.035	0.048	0.037	0.009	0.023	0.033	0.010	0.032	0.064
32	Transport	0.033	0.038	0.040	0.007	0.014	0.016	0.008	0.027	0.051
33	Furniture, games	0.035	0.065	0.328	0.002	0.005	0.008	0.005	0.011	0.019
34	Other consumer goods	0.046	0.087	0.094	0.006	0.014	0.012	0.014	0.027	0.051
35	Civil engineering	0.045	0.080	0.082	0.010	0.038	0.043	0.015	0.058	0.085
	Average	0.778	0.764	0.681	0.066	0.182	0.243	0.134	0.374	0.605

Figure 4a: Science Intensity by Industry, 2000-2003 (papers per employees)

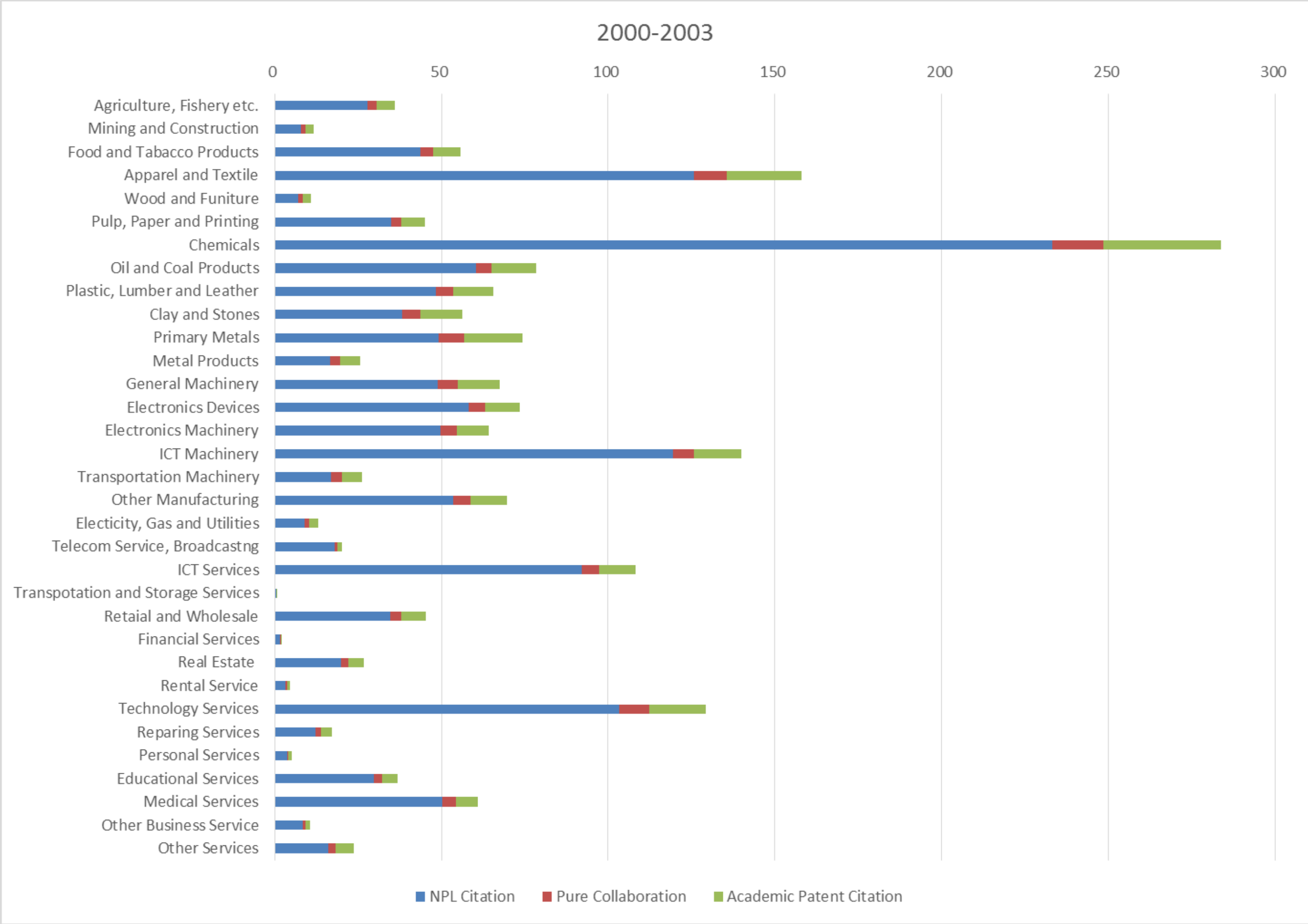
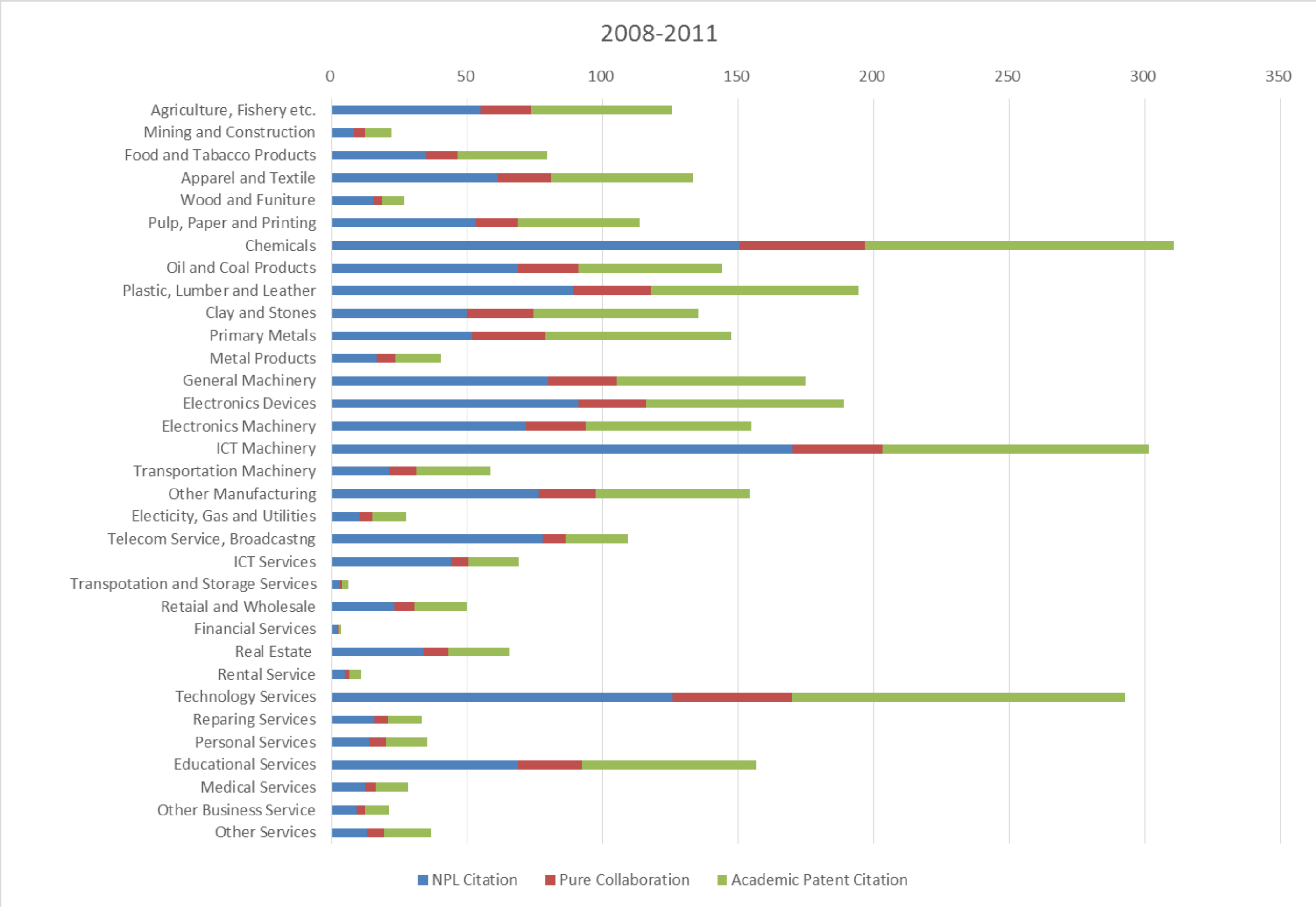


Figure 4b: Science Intensity by Industry, 2008-2011 (papers per employees)



Appendix 1. Disambiguation of Japanese patent inventors

In this appendix, we describe the method and the data to be used to identify (disambiguate) the inventors of patents filed in Japan Patent Office (JPO).

We utilize the patent data applied in 1995 or later from the IIP (Institute of Intellectual Property) Patent Database 2015 version (IIP-PD hereafter)[4]. IIP-PD is consist of a number of normalized tables and we use a tables for inventors, applications and applicants, they are named as inventor, ap and applicant, respectively. Since the names of the non-Japanese inventors are written in Katakana characters (one of Japanese syllabaries) in Japanese patent data and they contain a lot of spelling inconsistency, we use only the Japanese inventors for analysis. In order to extract Japanese inventors only, we take out the inventors whose name does not contain a Chinese character. The unit of the records of inventor table is patent-inventor and it contains 25,499,350 records in all but we extract 12,397,820 records from the table.

In advance to apply the disambiguation algorithm, we normalize the information of name and address of inventors and applicants. For inventor names, all spaces, including a space between sur name and given name of inventors, are removed and characters which have similar appearances are consolidated by each other. Addresses of inventors are divided into five reginal levels: prefecture (to, do, fu or ken), municipality (shi, ku, cho or son), city block (chome or aza), land number (banchi or ban) and land number extension (go). As the applicant information, we use the name of applicant and applicant identification number given by JPO. The identification number is replaced by a firm ID (NID) used in the "NISTEP Dictionary of Corporate Names Version 2015.1" developed by the National Institute of Science and Technology Policy (NISTEP) if the information can be successfully matched using a converter to IIP-PD also provided by NISTEP.[5]

Next, we apply a patent inventor disambiguation algorithm developed by Li et. al (2014) to the normalized data. The algorithm consists of the following process. First, patent-inventor level data set is prepared for analysis. Each record of "inventor" table in IIP-PD is the unit of analysis, which is identified by a combination of a patent application number (ida) and a sequential number of inventor for each patent application (seq). Second, the records are blocked by predetermined criteria that are likely to be

---

4 http://www.iip.or.jp/patentdb/

5 http://www.nistep.go.jp/research/scisip/data-and-information-infrastructure

satisfied by most matching records. We divided the records in which the inventor names are exactly same into a block. Third, for all pairs of the records within blocks, a vector of similarity of a record pair (which is called as similarity profile) is computed. Similarity profile for any two inventor-patent records i and j in a block is defined as the following multi-dimensional vector:

$$\mathbf{x}_{ij} = (x_{1,ij} \quad x_{2,ij} \quad \cdots \quad x_{k,ij} \quad \cdots \quad x_{K-1,ij} \quad x_{K,ij})$$

where $x_{k,ij}$ is the degree of similarity of record i and j on kth attribute. Table A x.1 represents the definition of the similarity profile in this study.

Table A1-1 Definition of similarity profile

Attributes	Values
Inventor name (x_1)	1 if names are completely same. 0 otherwise.
Co-inventors' names (x_2)	Number of common co-inventors, where more than 6 common co-inventors is set to a maximum value of 6.
Technology class (x_3)	4 if main IPCs are same at 4 digit level. 3 if main IPCs are same at 3 digit level. 2 if main IPCs are same at 1 digit level. 1 if main IPCs are not available. 0 if main IPCs are completely different.
Applicant (x_4)	3 if applicant identification numbers are equal. 2 if applicant names are same. 1 if either applicant identification number or applicant name are not available. 0 if both applicant identification numbers and names are different.
Address (x_5)	5 if matched at land number extension (go-level). 4 if matched at land number (banchi-level). 3 if matched at city block (chimei-level). 2 if matched at municipality-level. 1 if matched at prefecture-level. 0 otherwise.

Fourth, using predetermined training sets, we compute the likelihood that matching pairs and non-matching pairs could give rise to each similarity profile. Likelihood ratio (r-value) for a similarity profile $\mathbf{x}$ is defined as:

$$r(\mathbf{x}) = \frac{P(\mathbf{x}|M)}{P(\mathbf{x}|N)}$$

where $P(\mathbf{x}|M)$ and $P(\mathbf{x}|N)$ is the proportion of times that similarity profile $\mathbf{x}$ appeared in the match set and non-match set, respectively. In this study, we define match set as a group of record pairs of matched full inventor names defined as rare with respect to all inventor names and non-match set as a group of record pairs of non-matching full inventor names chosen from rare name list. We define rare names as names which does not appear in two or more times a year in the telephone directory published by NTT during 2000-2012.

Fifth, we estimate the posterior probability of a match for all record pairs using the likelihood ratio calculated from the training sets. Posterior probability are defined by Bayes' theorem as follows,

$$P\left(M_{ij}|\mathbf{x}_{ij}\right) = \frac{1}{1 + \frac{1 - P\left(M_{ij}\right)}{P\left(M_{ij}\right)} \frac{1}{r\left(\mathbf{x}_{ij}\right)}}$$

where $P\left(M_{ij}\right)$ is the prior probability of a match. The prior probability is calculated as the same as the original algorithm.

Finally, using the posterior match probability for all record pairs within the blocks and a set of thresholds, record pairs with relatively high probabilities are merged into a cluster iteratively. We used a set of seven thresholds (0.99, 0.95, 0.90, 0.8, 0.7, 0.6 and 0.5). Iterative clustering is stating from highest threshold (0.99) to the lowest threshold (0.5)[6].

The disambiguation algorithm used in Li et. al (2014) are publicly available at the GitHub website[7]. It is, however, developed for the patent data in the U.S., it is necessary to modify in order to apply to patent data in Japan. Table A.x.2 summarizes modified points. First of all, the original algorithm uses first name, middle name and last name as the inventor name attributes and allow misspelling or abbreviation in names by implementing several blocking rules. In contrast to the original program, we do not divide name attribute and do not allow any difference between inventor names. As the reasons of that, Japanese names usually do not contain middle name and abbreviation in

[6] Iterative clustering is a complex process and need to set some parameters. Following to the original program, "minimum threshold" is set to 0.3 and the "effective comparison count" is set to one fourth of the number of combinations of the members between two clusters. For details of iterative clustering, see Li et. al (2014).

[7] https://github.com/funginstitute/disambiguator

inventor names hardly occur in Japan.

For technology class, we use the international patent classification (IPC) while Li et. al (2014) use US technology class. Furthermore, although the original program allows multiple technology classes, because IIP-PD contains single main IPC code for each patent, we modify the definition of the similarity score for the technology class attribute.

Although Japanese patent has multiple applicants (assignees), the algorithm assume a single assignee. For that reason, we use only the information of the applicant appeared in the first place.

We significantly changed the training sets creation rules. Li et. al (2014) uses two types of training sets. One type of training sets are based on patent features and used for the learning of ambiguity in name features. Another type of training sets are created by name features and used for the learning of ambiguity in patent features. In this study, because we do not allow for difference in name attribute within a block, training sets for name features are not necessary. As the same as original algorithm, rare names are used to generate training sets for patent features. While the original algorithm determine rare names within patent inventors, we obtain the list of rare names from the telephone directory in order to improve the reliability of training sets.

Table A1-2. Modifications of disambiguation algorithm of Li et. al (2014)

	Li et. al (2014)	Our method
Attributes	Inventor name: First name, middle name and last name are distinguished Technology class: US class	Do not distinguish first and last name IPC
Blocking rule	7 steps	1 step: exact match of inventor name
Training sets	Two types: 1. Pairs of matched full inventor names defined as rare with respect to all inventor names. (Rare names are extract from patent inventors)	Pairs of matched full inventor names defined as rare with respect to all inventor names. (Rare names are extract from the telephone directory)

	2. Pairs sharing 2 or more common coauthors and technology classes.	

We run the modified program on the following system:

- CPU: 20Core Xeon E5–2660 v3 2.6GHz (10core x 2CPU)
- Memory: 64GB (8GBx8) ECC Registered DDR4–2133 Quad–Channel
- OS： Linux (Ubuntu) on Windows 10 using VMWare Workstation 12 Player
- CPLEX：IBM ILOG CPLEX Optimization Studio Version 12.6.2

Appendix2. NPL citations of JP patent applications

Most of the studies on science-linkage focus on NPL citations of US patents, because the US Patent Acts require applicants to disclose their knowledge of prior art documents and the databases of US patents are well organized.

The Japanese Patent Acts did not require information disclosure until 2002. Thus, prior art documents on front page references of Japanese patent gazette are listed up by patent examiners. And, citations by inventors/applicants are often embedded in text of detailed explanation of technical descriptions. In this paper, we used information of non-patent literature cited by inventors/applicants in Japanese patent applications, of which database we purchased from the Artificial Life Laboratory, Inc. They identified and extracted patent and non-patent literature documents cited in technical descriptions, using their text-mining algorithm based on Tamada et. al. 2006 and further developed. The database consists of Japanese patent application publications (including PCT applications; totally 8.2 million records), and Japanese granted patents (3.6 million records), of which gazette were published between 1993 and 2015.

In calculating average number of NPL citations in Table 3, we used the NPL citations of 3.4 million of patent application publications, of which applicant is identified as a Japanese firm and earliest priority year is between 2000 and 2011 (see Table A2).

Table A2. Sample size comparison

Eariest priority year	Number of appln. publications in Tamada DB			PATSTAT
		(1) Applicants identified as firms	(2) JP firms' appln. pub.with US patents	(3) Corresponding US patents
2000	415,323	336,976	28,239	32,355
2001	415,043	335,936	27,159	31,068
2002	400,676	320,521	27,557	31,569
2003	395,780	311,454	28,672	32,325
2004	402,884	313,361	31,100	34,143
2005	401,433	302,688	30,839	33,534
2006	372,849	281,975	32,055	34,223
2007	354,268	269,182	31,949	33,469
2008	346,554	264,113	30,529	32,263
2009	312,686	232,138	27,285	28,499
2010	300,912	222,778	24,891	25,536
2011	298,557	210,848	16,825	17,159
Total	4,416,965	3,401,970	337,100	366,143

Note: (1) used in Table 3. (2), (3) used in Figure A-1.

In order to compare information of NPL citations in Tamada DB with that of US patents, we used EPO PATSTAT database and extracted US patents in the DOCDB family of Japanese patent applications as the corresponding US patents. Figure A-1 shows average number of NPL citations of JP applications with that of the corresponding US patents by WIPO technology areas. We can observe very similar tendencies.

Figure A2. Comparison of NPL citations of JP and US patents