3A3-J-13-02

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Detecting Technology Portfolios in the Semiconductor Industry

Bohua Shao[*1] Kimitaka Asatani[*1] Ichiro Sakata[*1]

[*1] School of Engineering, the University of Tokyo

Semiconductor serves as the base for the Artificial Intelligence Society. We are interested in detecting technology trends and changes in the semiconductor industry because there are some large resource allocation and organization restructure in this industry recently. In this study, we narrowed our research boundary into the U.S. and Japan. We extracted patent information related with semiconductor and made a citation network. We used Louvain method to cluster the maximum connected component and considered several largest clusters. Results show that the technology portfolios among the two countries are different. We used "*tf-idf*" to detect keywords and features of these unbalanced clusters. In the future, we will link the applicant information in a patent database with mergers and acquisitions (M&A) information in a company database. We will compare and integrate findings from different sources, such as investment, M&A, technology features and industrial policies in order to have a comprehensive understanding.

1. Introduction

Semiconductor industry is considered important base for the Artificial Intelligence Society. Recently, by preparing the era of Industry 4.0, resource allocation and organization restructure were under way in the semiconductor industry. There are several influential Mergers and Acquisitions (M&A) in the semiconductor industry. For example, Avago Technologies purchased Broadcom Inc. with 37 billion U.S. dollars and Qualcomm, Inc. announced to purchase NXP Semiconductor with 44 billion U.S. dollars. These phenomena changed technology landscape in the semiconductor industry largely.

In this study, we compared semiconductor industries in the U.S. and Japan. Both two countries have been leading countries in the semiconductor industry for a long term. By comparing these two countries, it is easy to trace and detect the technology trends in the semiconductor industry.

Since technological intangible assets are difficult to measure, patent analysis is a dominant method for research in this field. [Lee 2009] Likewise, in this study, we used patent data for analyzing technology development trends in the semiconductor industry.

This paper is arranged as follows: Section 2 is Literature Review, Section 3 is Data, Section 4 is Experiment, Section 5 is Results and Section 6 is Summary.

2. Previous Literature

In 2009, Lee *et al.* proposed a keyword-based patent map. This map is generated through keywords of patents. These keywords in vectors are extracted by text mining. Then keyword vectors are reorganized by Principle Component Analysis. Finally, these vectors were projected onto 2-dimension surface. From the map, it is easy to detect where to invest. [Lee 2009]

In 2011, Wang *et al.* used patent co-citation information between Fortune 500 companies. According to the co-citation networks of different periods, companies are divided into different industry groups. Furthermore, companies' positions in the networks change among different periods. [Wang 2011]

Ma *et al.* proposed a comprehensive method for identifying technology-driven M&A targets. They used both qualitative and quantitative methods for analysis. They also invited policy makers and experts for evaluation. Finally, they used a company as an example to verify the effectiveness of their methods. This method provided a standard, sophisticated way for identifying M&A targets. [Ma 2017]

Shao *et al.* mainly focused on financial items of M&A in Japan. This paper deals with categorization of M&A in Japan. However, it did not take technology factors into consideration, even Japan is famous for its science, technology and trading. [Shao 2018]

3. Data

We used Derwent Innovation patent database and Derwent World Patent Index (DWPI) for extracting patent information. We made smart search topic being "semiconductor", Publication date (Basic) being from 1990.01.01 to 2018.01.01, Application Country/Region (Basic) being US or JP. 874005 items were retrieved. We chose this time period because large development in semiconductor industries in Korea and Taiwan happened in the 1990's. [Chiu 2014] We used "Basic" items because "Basic" records the first patent in the same DWPI patent family, which largely represents where and when the patent questioned came from. [Derwent Innovation 2019]

We extracted the following items from the database: Publication Number, Title, Title – DWPI, Publication Date, Cited Refs – Patent, Count of Cited Refs – Patent, Citing Patents, Count of Citing Patents, DWPI Family Members, DWPI Count of Family Members. We did not extract International Patent Classification (IPC) and Cooperative Patent Classification (CPC) data and we used the citation clustering method for classification. [Thurber 1918] In addition, technology trends and contents change rapidly nowadays. Hence, data merely from IPC or CPC did not assure accuracy.

Contact: Bohua Shao, School of Engineering, the University of Tokyo, shaobohua6@gmail.com

4. Experiment

4.1 Citation Network

Based on these patent data, we constructed a patent citation network. According to citation information of the patent database, we drew links between patents. Then we grouped them by DWPI Family Member information and built a new network. We assigned weights on edges between DWPI Family Members in the new network according to previous linkage information. The weight is the quantity of all links between any two patents which belong to the two DWPI Family pairs respectively.

4.2 Node Degree

We were interested in the network structure and we extracted the maximum connected component from the original citation network. We calculated the degree of each node in the maximum connected component and plotted the degree distribution in log-log scale.

4.3 Clustering

Louvain method [Blondel 2008] is an effective and high-speed method for dealing with large networks. We used Louvain method for clustering the maximum connected component because it is very large, with 612570 nodes.

4.4 Sum of "*tf-idf*"

In natural language processing, a very common method called "*tf-idf*" is widely used. The "*tf-idf*" aims at filtering out important terms of a document among a corpus. The "*tf-idf*" is calculated in the following way (1):

$$tf\text{-}idf_{t,d} = tf_{t,d} \cdot idf_t \tag{1}$$

where subscript t means a specific word (term) and d means a given document in the whole corpus. In order to calculate the importance of a word in the whole corpus, we sum up "$tf\text{-}idf_{t,d}$" of the same term across the whole corpus, as shown in (2). We have the "$tf\text{-}idf_t$" for the whole corpus.

$$tf\text{-}idf_t = \sum_d tf\text{-}idf_{t,d} \tag{2}$$

5. Results

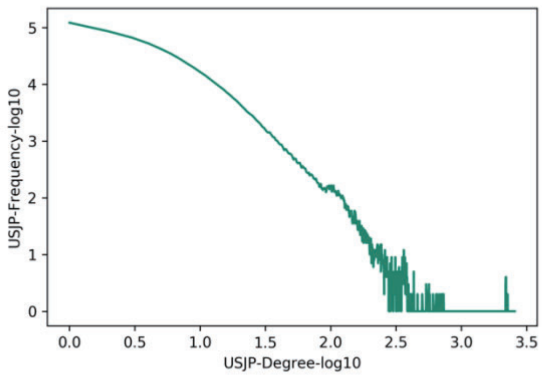

Figure 1. Patent degree distribution in log-log scale

Figure 1 shows the distribution of degrees in log-log scale. We took the logarithm to base 10. The horizontal axis is the degree in log scale whereas the vertical axis is the frequency in log scale. Different from the general understanding that companies only cite their own patents, patents related with semiconductor are cited by other companies as well. The line in Figure 1 is nearly straight in the middle part and we assumed that the maximum connected component has scale-free network features for nodes with degrees in the range of 20 to 80.

By Louvain method, we finally had 333 clusters. We present the largest 21 clusters and their contents in Table 1. [1]

Table 1 shows the quantities of patents grouped by application countries. The "USnum" column shows the quantities of patents from the U.S. by each cluster whereas the "JPnum" column shows the quantities of patents from Japan by each cluster. The column "US/JP-ratio" shows the ratios of the values in "USnum" to those correspondences in "JPnum". Similarly, column "JP/US-ratio" shows the ratios of the values in "JPnum" to those correspondences in "USnum".

We selected two leading countries in the semiconductor industry. Intuitively, ratios across different clusters did not fluctuate too much. However, as shown in this table, ratios are quite different. We mark and underline the ratios above 2 or below 0.5 in red. These unbalanced "Cluster_id" are 4, 14, 7, 20, 8, 16 and 23.

Table 1 Quantities of Patents in Each Cluster Grouped by Publication Countries

No.	Cluster_id	USnum	JPnum	US/JP-ratio	JP/US-ratio
1	6	33728	36138	0.933	1.071
2	1	24488	33412	0.733	1.364
3	5	18502	24441	0.757	1.321
4	12	19731	19964	0.988	1.012
5	3	18922	19082	0.992	1.008
6	4	29081	7798	3.729	0.268
7	14	22394	11061	2.025	0.494
8	15	16448	15092	1.090	0.918
9	7	22213	9053	2.454	0.408
10	2	16267	13532	1.202	0.832
11	0	15819	10578	1.495	0.669
12	9	13368	10941	1.222	0.818
13	19	15128	8130	1.861	0.537
14	20	3409	18297	0.186	5.367
15	17	12038	7522	1.600	0.625
16	10	7371	7381	0.999	1.001
17	13	5743	8856	0.648	1.542
18	8	2246	7895	0.284	3.515
19	16	7592	2030	3.740	0.267
20	22	5001	4112	1.216	0.822
21	23	6163	2012	3.063	0.326

Furthermore, patents granted from the U.S. are becoming more and more recently whereas patents from Japan are declining. From these results, investors can pay attention to the differences.

[1] The quantity of patents of the 22nd largest cluster declines to 3836. We also conducted Louvain method several times and results are slightly different. Here is a typical example.

Table 2 Sum of "*tf-idf*" in "Cluster_id" Being 4

word	tfidf-sum
film	950.1
gate	928.9
second	894.0
layer	835.5
region	758.0
silicon	741.6
dielectric	727.7
insulating	724.3
fin	662.7
material	657.5
oxide	657.0
metal	650.5
electrode	613.3
structure	593.9
trench	593.2

Table 3 Sum of "*tf-idf*" in "Cluster_id" Being 16

word	tfidf-sum
layer	262.4
memory	247.3
magnetic	227.7
second	210.1
material	185.3
resistance	181.7
electrode	181.6
change	177.0
cell	166.7
film	164.2
element	158.9
metal	150.3
conductive	149.9
line	146.5
phase	145.7

Table 4 Sum of "*tf-idf*" in "Cluster_id" Being 23

word	tfidf-sum
layer	208.9
gate	198.8
film	194.9
memory	189.4
second	188.8
pattern	180.0
mask	170.3
region	165.7
silicon	158.6
insulating	147.2
material	143.4
trench	140.6
line	136.7
conductive	136.2
forming	124.4

Let us consider several clusters marked red in Table 1. We calculated sum of "*tf-idf*" for clusters with "Cluster_id" being 4, 16 and 23 as examples because they all hold "US/JP-ratio" over 3. The corpus of each cluster is abstracts of patents belonging to the cluster questioned. Herein, we pasted words with top 15 highest sum of "*tf-idf*" values (keywords) of each cluster in Table 2, 3 and 4.

From Table 2, we only detected general terms used for semiconductor manufacturing. From Table 3 and 4, we argue that technology in the 2 clusters is related with memory storage. In all the 3 clusters, U.S. shows dominant power. We hold a hypothesis that U.S. is now allocating resources in manufacturing new generation memory in order to prepare for the big data era.

6. Conclusion

In this study, we extracted patent information in the semiconductor industry. We narrowed our research boundary inside the U.S. and Japan, the two leading countries around the world. We detected that the technology portfolios are different in the two countries. We investigated the technology differences and used "*tf-idf*" to filter out important words among these clusters.

We investigated 3 clusters with high "US/JP-ratio" values as examples. From these examples, we found that U.S. companies have been interested in memory manufacturing recently.

In this study, we only focused on citation information, which is far from enough. There are still other columns in the database, such as Publication Date, Assignee/Applicant and Abstract-DWPI. These columns provide the understanding of patents in a detailed way.

In the future, we want to deal with these columns to catch the current trend in the semiconductor industry. We will link Assignee/Applicant information of the Derwent Innovation patent database with company names in a company database. We will also incorporate social phenomena, such as M&A, industrial policies and taxes in the semiconductor industry so as to have both private and public perspectives.

References

[Lee 2009] Lee, Sungjoo, Byungun Yoon, and Yongtae Park, An approach to discovering new technology opportunities: Keyword-based patent map approach, Technovation 29 (6-7): 481-97, Elsevier.

[Wang 2011] Wang, Xianwen, Xi Zhang, and Shenmeng Xu, Patent co-citation networks of fortune 500 companies, Scientometrics 88 (3): 761-70, Springer Nature, 2011.

[Ma 2017] Ma, Tingting, Yi Zhang, Lu Huang, Lining Shang, Kangrui Wang, Huizhu Yu, and Donghua Zhu, Text mining to gain technical intelligence for acquired target selection: A case study for china's computer numerical control machine tools industry, Technological Forecasting & Social Change 116: 162-80, Elsevier, 2017

[Shao 2018] Shao, Bohua, Kimitaka Asatani, and Ichiro Sakata, Categorization of mergers and acquisitions in japan using corporate databases: A fundamental research for prediction. Paper presented at 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM 2018), Bangkok, Thailand, IEEE, 2018

[Chiu 2014] Chiu, Chien-Che, and Hsing-Ning Su, Analysis of patent portfolio and knowledge flow of the global semiconductor industry. 2014 Portland International Conference on Management of Engineering & Technology (Picmet): 3621-34, IEEE, 2014

[Derwent Innovation 2019] Derwent Innovation, Getting started in derwent innovation. Available from http://www.derwentinnovation.com/tip-innovation/support/help/introduction.htm., Accessed: 2019. 02.06.

[Thurber 1918] Thurber, William L., The need of a secondary patent classification based on industrial arts, Journal of the Patent Office Society 1 (3) (November 01): 106-11, Patent Office Society (U.S.), 1918

[Blondel 2008] Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre, Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008 (10): P10008, IOP Publishing, 2008