Technological Forecasting & Social Change

An International Journal

Knowledge combination modeling: the measurement of knowledge similarity between different technological domains

Hiroko Nakamura, Shinji Suzuki, Ichiro Sakata, Yuya Kajikawa

PII: S0040-1625(14)00279-0
DOI: doi: 10.1016/j.techfore.2014.09.009
Reference: TFS 18086

To appear in: *Technological Forecasting & Social Change*

Received date: 17 January 2014
Revised date: 7 April 2014
Accepted date: 14 September 2014

Please cite this article as: Hiroko Nakamura, Shinji Suzuki, Ichiro Sakata, Yuya Kajikawa, Knowledge combination modeling: the measurement of knowledge similarity between different technological domains, *Technological Forecasting & Social Change* (2014), doi: 10.1016/j.techfore.2014.09.009

KNOWLEDGE COMBINATION MODELING: THE MEASUREMENT OF KNOWLEDGE SIMILARITY BETWEEN DIFFERENT TECHNOLOGICAL DOMAINS

HIROKO NAKAMURA, SHINJI SUZUKI

Center for Aviation Innovation Research (CAIR), The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan, e-mail; techhn@mail.ecc.u-tokyo.ac.jp, Tel; +81-3-5841-0744

ICHIRO SAKATA

Innovation Policy Research Center, The University of Tokyo

YUYA KAJIKAWA

Graduate School of Innovation Management, Tokyo Institute of Technology

Abstract

This paper proposes the DB-Combination model that considers three different knowledge combinations in depth (D) and breadth (B) based on similarities of two technological knowledge domains. We also investigate three methodologies A1, A2 and A3 to highlight the three knowledge combinations. To identify technological knowledge domains, citation analysis on patent information was used for A1 and A2 and pre-existing patent classification analysis was used for A3. And to measure the similarity between identified technological knowledge domains, text similarity measurements, existing intra-industrial citation tracing and IPC share similarity comparison were used for A1, A2 and A3 respectively. The usability of the model and methodologies were demonstrated through a case study on technological knowledge of the automobile industry and the aircraft industry. While these methodologies still need to be improved, it was demonstrated that the three measurements can highlight candidates of the three knowledge combination proposed in DB-Combination model. This research contributes to accelerate breadth knowledge recombination in a complex technology industry.

Keywords

Patent analysis, citation analysis, bibliometrics, breadth search, knowledge recombination

1. INTRODUCTION

It is said that innovation comes from a recombination of knowledge (Dosi, 1982; Nelson and Winter, 1977; Schumpeter, 1934) and that combining one's own knowledge with that of different industry and different technology domains has the possibility of bringing new knowledge creation (Schoenmakers and Dysters, 2010; Gassmann and Zeschky, 2008; Dosi, 1982). On the other hand, the technological domain of industries with complex system has a wide range of technological sub-domains (Fleming and Sorenson, 2001, Eriksson, 2000). It is not easy for engineers of an industry to search a new candidate of knowledge combination in knowledge of another industry with complex system; firstly they should need to identify the technological sub-domains of the other, and then, to select sub-domains to look further. The information of sub-domains also must be updated frequently (Herrero et al., 2010). The cost of the collection and the integration of different knowledge (Nakamura et al., 2011a; Kajikawa et al. 2006; Tijssen, 1992) and the uncertainty of success (Schilling and Green, 2011; Moorthy and Polley, 2010) are problems that limit practitioners to explore new opportunities.

To support practitioners to have such exploration, namely breadth activities, and to bring innovation, the authors focus on one side of breadth activities; namely, searching the technological knowledge of other industries and integrating it to the own knowledge. We propose a knowledge combination model and discuss methodologies that effectively identify a technological sub-domain that can be combined between the two industries.

We focus on patents as the information source of technological knowledge of focused industries because firstly patent is considered to be the best available indicator for R&D invention related to technology and outcomes of innovation activities (OECD, 1994). And secondly we believed that there is a potential need in practitioners for scientometrics that can support breadth search with patents. Practitioners contacted in this study explained that, although a patent is essential to protect practitioners' intellectual products and investigate competitors' strategy, difficulty in searching information from patent data hinders frequent use of patents as a source of knowledge. And computer-based bibliometrics approaches is taken because it can process vast amount of data and it is expected to ease breadth search (Alavi and Leidner, 2001; Smallheiser and Swanson, 1998; Smalheiser, 2012; Herrero et al., 2010; Cantu and Ceballos, 2010; Fleming and Sorenson, 2001; Kostoff, 2008).

This paper is organized as follows: The next section review previous literature. The third section proposes a knowledge combination model and three measurements. The forth section conducts a case study on technological knowledge of automobile and aeronautic industry and shows identified technological domains and highlighted pairs of technological domains. The section also discusses the results with automobile and aeronautic experts. The Discussion section compares the three methodologies. The final section concludes this paper with the findings.

2. LITERATURE

Patent is often used in the innovation literatures. For example, patent is used as the indicators of technological knowledge of focused industries in the literatures mapping technological portfolios of a company or an industry. Leydesdorff et al. (2012) discussed methodologies to map the technological portfolios and the relation between the identified technology using International Patent Classification (IPC) and patent citation analysis approach. Schoen et al. (2012) discussed methodologies to map technological domains of major R&D companies and the dynamics of knowledge over firms and regions. Patent is also used as the indicators of technological knowledge of the focused company in the literatures investigating the impact of technological knowledge breadth and depth to company performance (ex. Schoenmakers and Duysters, 2010; Leiponen and Helfat, 2009; Moorthy and Polley, 2010; Fleming and Sorenson, 2001, Nesta and Saviotti, 2005). Patent is also used as the indicator of the efficiency of innovation policies of the focused countries, domains, or sectors in the literatures analyzing economic development, technological change, speed and change of industrial structures (ex. Criscuolo, 2006, Soete and Wyatt, 1983). Patent is used in these literatures because it is objective and their standard changes slowly (Griliches, 1990).

Our approach seems to be similar to the first group of literatures, such as Leydesdorff et al. (2012) and Schoen et al. (2012). However we use patent not as the indicators to obtain the overview of knowledge exist in a domain but as a source of technological knowledge to produce new knowledge combination. The aim of this paper is neither to describe technological trend and portfolio by utilizing patent as indicator nor to evaluate performance of firms and industries by utilizing patent as evidence. The aim of this paper is to explore opportunities for business development by utilizing patents as information resource.

Citation analysis that we use in this paper has been developed because of the need for scientific information retrieval and has established itself as one of the most effective approaches in identifying technological domains in academic documents and in creating the overview because a citation can reflect the self-organizing dynamics of scholars' communication (Leydesdorff, 2008; Kajikawa et al., 2008; Kajikawa et al., 2006; Nicolaisen, 2007; Cronin, 2001).

It is widely known that dissemination and exchange of communication is important for the development of science (Everett and Pecotich, 1993). Using citation information to evaluate scientific activities, nonetheless, brings various concerns such as risk of skewed analysis results due to the existence of citations for ciritics, self-citations, an English-bias, and availability of literatures (Martin, 2012; Everett and Pecotich, 1993; Garfield, 1979). Despite these concerns, citation analysis is utilized among researchers and decision-makers related to science and innovation management as quantifiable and objective approaches that can compensate and validate the experts' judgments (Nerur et al., 2008), and can be used in administration fields (see Garfield, 1979 for example), and its further development and use are expected by practitioners.

Because patent citations are made with the consideration of legal and economic matters, there is still criticism about applying citation analysis to patents because citation behavior is different between academic journals and patents and also citation analysis inclines too much toward documents with links (Leydesdorff, 2008; Michel and Bettels, 2001; Meyer, 2000; Kostoff, 1998). However, there are many attempts to apply such methods to patent analysis because the pre-existing classification system of patents is based on technological and functional characteristics and is often difficult to understand the overview of the system the patents are related from the classification (Griliches, 1990). Furthermore, when technology rapidly progresses and changes, it is not easy for existing classification schema to capture such a change. Narin (1994) discussed the high similarity between the analysis of scientific papers and analysis of patents because both are suited especially to the analysis of national productivity, inventor productivity, referencing cycles, and citation impact.

Considering the criticism and support regarding patent citation analysis, we conducted two different approaches to identify technological domains from patent data; namely, the citation analysis approach and another approach with an international classification standard of patents as we can see in the next section.

3. METHODOLOGY

3.1. Depth and Breadth Knowledge Combination Model

The methodology proposed in this paper aims to support the following knowledge recombination process of practitioners, that is, identifying the technological sub-domains of other industries, selecting sub-domains to combine and researching for bringing new knowledge. And to do that, firstly we propose a knowledge combination model between two technological domains, named the DB-Combination model (Fig. 1). We assume that, limited to the technological knowledge and the combination between different industries, knowledge recombination in depth and breadth discussed in the previous innovation literatures such as Alavi and Leidner, 2001, Gassmann and Zeschky, 2008; Schoenmakers and Dysters, 2010, Dosi, 1982, can be modeled depending on the "similarity" between the domains as Fig. 1. Figure 1 considers the similarity of any pairs of sub-domains of each technological domain; industry A and of industry B, in the horizon.

Figure 1. DB-Combination Model for knowledge integration between two technological domains in depth and breadth

Knowledge recombination in deep (D) search is categorized to DB-D in Fig. 1 that shows combining knowledge between a very similar pair of sub-domains. A deeper understanding and improvement of the technological domains will be brought by such combining. It is likely that such knowledge combining has been already challenged so that only incremental innovation is expected to occur (Katila and Ahuja, 2002). On the contrary, knowledge recombination in breadth are categorized to DB-T and DB-C because such recombination can be considered to be brought by either knowledge transfer (T) of unique technologies from one technology domain to another (pair DB-T) or knowledge combination (C) between different technology sub-domains with weak similarity (DB-C). Unlike knowledge combining in depth, knowledge combining in breadth is expected to broaden the scope of research and to provide a chance of new knowledge combination or transfer/ replacement (Katila and Ahuja, 2002).

Our proposal is similar to literature-based discovery (LBD) process studied in bibliometrics mining academic literature. According to Gordon et al. (2002), LBD can be either an open or closed discovery process. An open discovery process is characterized by the generation of a hypothesis, starting with a research question or scientific problem, and explores to extract plausible candidates relating with the starting concept (like DB-C in our model). On the other hand, a closed discovery process is the testing of a hypothesis, starting from two entities at both ends where common intermediate terms are extracted and evaluated to verify important connections between them (like DB-D in our model).

The similarity of sub-domains should have various factors of technological knowledge including problems that the technological knowledge aims to solve, the processes that the technological knowledge needs to solve a problem, the environment where the technological knowledge is used, and the components and materials of the technology itself. In the following sub-sections, we propose methodologies to measure the similarity of sub-domains of different industries, using patents. Finding sub-domains in different industries that are weakly similar to a sub-domain of an industry is more difficult for practitioners than finding very similar sub-domains or very unique sub-domains so that a similarity measurement of a DB-C combination, for example, is expected to be especially effective in helping practitioners bring forth an advancement in technology.

In order to investigate the effectiveness of the methodologies, we conduct a case study to analyze technological knowledge combination between the automobile industry and the aircraft industry, and interview experts on the results in order to evaluate the effectiveness of the model and methodologies. Automobile industry companies are facing severe global competition. The knowledge how to reduce the cost of production is critical and the relocation and offshoring of parts production and assembly are always investigated through the departments of business management (Herrero et al., 2010). Research, engineering, and product planning departments are also under pressure to increase productivity and creativity. One research center of a Japanese automobile company showed their interest in our research concept and asked for a system that can detect technological areas or trends that they themselves cannot recognize the importance of yet.

On the other hand, the aircraft industry is a high-tech industry and regarded as a source of technology renewal (Eriksson, 2000). There are many examples of the spillover of technology from the aircraft industry to the automobile industry, such as composite material production and head-up display. Recently, the aircraft industry has faced the need of radical changes to achieve environmental and business sustainability and to seek technological solutions for safe and economical carbon-neutral growth (Nakamura et al., 2012; Nakamura et al., 2011b). “More and All Electric” system concept in the aviation industry, that is, replacing more or all the heavy hydraulic systems of aircrafts by efficient electric systems, is one of the technological directions investigated in order to reduce fuel consumption (Nakamura et al., 2011b). The automobile industry can be considered as advanced in this area because the industry has already commercialized all-electric vehicles. Knowledge transfer or combination is now expected not only in the direction of the aircraft industry to the automobile industry, but also vice versa, from automobile to aircraft.

3.2. Data

Patent data were retrieved from Thomson Reuters’ Thomson Innovation with Derwent World

Patents Index (DWPI). Thomson Innovation is a comprehensive worldwide patent database, which covers patents recorded at more than 80 patent authorities and includes the United States Patent and Trademark Office (USPTO) (DWPI data is available from 1963), the World Intellectual Property Organization (WIPO) (DWPI from 1978), the European Patent Office (EPO) (DWPI from 1978) and the Japanese Patent Office (JPO) (DWPI from 1963). The DWPI bundles patents recorded at 47 worldwide patent authorities as a protection for the same invention as a sort of family so that duplicate data retrieval can be avoided during search of patents crossing over several patent offices' databases. DWPI also provides manually added English abstracts from patent documents issued in more than 30 foreign languages so that it can allow text analysis of inventions from non-English language sources.

We retrieved automobile and aviation patents which include one of following texts in applicant name, for the automobile industry: “honda motor", “honda giken", "toyota jidosha", "toyota motor", "mazda kk", "mazda motor", "nissan motor", "nissan jidosha", "mitsubishi jidosha" or "mitsubishi motor" and for aircraft industry: boeing, airbus, “rolls royce”, “pratt and whitney”, “rockwell collins”, “hamilton sundstrand”, “parker hannifin”, “messier and dowty” and mtu. 242,305 and 27,989 patents were retrieved in July 2012 for automobile and aircraft by using the queries above. About the queries for the automobile industry, we listed more than one name for one company. Standardization of applicant name is being carried in many authorities with the spread of online application but there are different representations of applicant names in the past data.

Compared to the aircraft industry, the automobile industry tends to patent more so that we limited the dataset for the automobile industry to patents of Japanese automobile companies. Queries for the aircraft industry cover major aircraft prime manufactures, major engine manufactures and major tier 1 suppliers of flight and data management, electric power systems, and mechanical components. To make the query process simple, we excluded General Electric or other major component suppliers that have other large business branches. We will discuss the limitation of our research as affected by queries in the discussion section, but experts of each industry we contacted in the survey validated that these queries are adequate for investigation of methodologies used to find knowledge integration candidates among the automobile and aircraft technology domains.

3.3. Methodologies

In order to identify a knowledge combination with the DB-Combination model from patent data, we take three computer-based bibliometrics approaches, A1, A2 and A3. Our DB-Combination model requires methodologies to identify technological sub-domains of industries from patent data and similarity measurements between identified sub-domains of two different industries. A1 and A2 take a citation analysis approach, and A3 takes IPC analysis approach for data structuring. For similarity measurement, A1 takes a text similarity measurement approach, A2 also takes an existing intra-industrial citation tracing approach, and A3 takes an IPC share similarity comparison approach (Table 1).

Table 1. Three approaches

3.3.1. Technological Sub-domain Identification

We conduct two different approaches to identify technological domains from the obtained patent data, using the citation analysis approach and an IPC analysis approach.

3.3.1.1. Citation Analysis Approach

The first approach is a citation network analysis. The analysis procedure is schematically illustrated in Fig. 2. The patent and citation data are converted into a non-weighted, non-directed network in which a patent is represented as a node and backward citations to patents as links. The maximum connected component (MC) of the network is extracted. To minimize noise data and the quantity of data, we regarded patents not citing or cited by other patents in the component as digressional from the mainstream of those technological domains and eliminated them. Finally, the network was divided into clusters depending on the density of links using a topological clustering method (Newman, 2004, Newman and Girvan, 2004). After clustering the network, we characterized each cluster by the expert-based approach. A1 and A2 (Table 1) take this approach.

Figure 2. Schematic diagram of the citation network analysis

3.3.1.2. IPC Analysis Approach

Classification is important in the patent system to facilitate a search for “prior art” (Leydesdorff, 2008). The Strasbourg Agreement concerning the International Patent Classification established the IPC in 1971. The IPC is a hierarchical system of language independent symbols which divides technology into eight sections with 70,000 subdivisions of twelve-digit codes. The appropriate symbols are allotted by the national or regional industrial property office of 57 states (WIPO homepage).

A3 takes the IPC subclass of four-digit codes. X_i is the share of subclass *i* in industry *X*. *num(X, IPC(i))* is the number of patents of subclass *i* of industry *X*, and *num(X)* is the total number of patents of industry *X*. A patent can have several IPC codes, where the total of the share for industry *X* exceeds 100%. This paper assumes the share as the rate of importance of the subclass in an industry and therefore, we filter minor subclasses with the share threshold.

$$X_i = \frac{num(X, IPC(i))}{num(X)} \qquad (1)$$

3.3.2. Similarity Measurement

3.3.2.1. Cosine text similarity measurement

A1 research takes the text similarity approach, assuming that the similarity in characteristics such as background problems, processes to solve the problem, operational conditions, or compounds, can be measured by similarity in the text.

First, the DWPI title and abstract of each text were analyzed and the frequency of word *i* in Cluster *s* ($FreW_{si}$) was evaluated by the following formula:

$$FreW_{si} = \frac{n_s}{n_s} \times \log\left(\frac{N}{N_i}\right) \qquad (2)$$

In (2), n_{si} represents the number of word *i* that appeared in the DWPI title and the abstract of the patents of Cluster *s* obtained in citation analysis. n_s represents the number of words that appeared in the title and the abstract of patents of Cluster *s*. *N* is the number of clusters in total. N_i is the number of clusters in which a patent contains the word *i* in the title and the abstract. The similarity of the text is evaluated by a cosine similarity that is often used in text mining and regards each text as a vector with the length of $FreW_{si}$. Cosine similarity *Cos(a,m)* between two clusters *a* and *m* is defined as (3). A large *Cos(a,m)* represents a relatively high similarity.

$$similarity = \cos(a,m) = \frac{\sum FreW_{ai} \times FreW_{mi}}{\sqrt{(FreW_{ai})^2}\sqrt{(FreW_{mi})^2}} \qquad (3)$$

3.3.2.2. Existing intra-industry citation tracing

In A2, we assume that the similarity of different technology sub-domains can be measured in citations between automobile patents and aircraft patents. We consider that, if there is a similarity, and a potential of knowledge combination, some practitioners must have already recognized the similarity and challenged it. In A2, unlike in A1 which identify technological sub-domains from automobile and aircraft patents separately in the citation analysis approach, these patents are combined and classified into clusters. Therefore, if there are patents with intra-industry citations, we must have clusters containing patents of both industries. We highlight such intra-industry clusters as similar technology sub-domains between the industries.

3.3.2.3. IPC share similarity comparison

It is obvious that patents with the same IPC are similar in some characteristics of technological knowledge. We highlight IPC domains that have a similar IPC share in two datasets as the formula (4). It is similar to Jaffe (1986) that characterized the technological position of a firm by the distribution of the firms' patents over patent classes. While Jaffe (1986) used a vector to see the overall characteristics, we focused on a factor or a patent class that have similar positions. If the share of IPC *i* over the firm's patents is similar between compared two industries, r_i is close to 0.5.

$$r_i = \frac{\sqrt{Avi_i Moto_i}}{Avi_i + Moto_i} \qquad (4)$$

3.3.3. Interviews

In order to research the effectiveness of three proposed approaches for practitioners to conduct breadth activities and to find possible knowledge combination domains, we discussed the results of the three approaches with experts. We interviewed four various levels of aeronautic researchers at Japan Aerospace Exploration Agency (JAXA) and two senior engineers at Toyota Central R&D Labs, INC in May 2012 in Japan. We visited their offices and had two-hour face-to-face interview for each center. We started the interview with questions about the importance and activities of breadth searches, and asked whether the highlighted areas in A1, A2, and A3 identified possible knowledge combination pairs and provided useful information that support practitioners in creating new knowledge.

4. RESULTS

4.1. Technological Sub-domain Identification

Among the three approaches (Table 1), the A1 and A2 approaches used citation analysis to identify technological domains from the data. The patent dataset of the automobile and aircraft industries was analyzed separately in the A1 research, and combined in the A2 research.

In the A1 research, the MC of automobile and aircraft consists of 60,458 patents and 8,281 patents, respectively, and were divided into 303 and 104 clusters. The number of patents in a cluster, that is, the size of clusters, varies from 6 to 5,187 for the automobile clusters and from 4 to 903 for the aircraft clusters. As the size of the top 4 clusters of the automobile was relatively large, we analyzed the sub-clusters of the four clusters, applying the same clustering approach. Each cluster was divided into 42, 29, 39, and 35 sub-clusters. Table 2 and 3 show the overview of the top 10 clusters in automobile and aircraft patents respectively. To identify the characterization of clusters, we analyzed the core patents that have frequent links (non-direction) with the other patents in the cluster and the frequent words that were calculated by the formula (3). The example of core patents is listed for the top 5 clusters in the Tables. The size and the average patented years can also provide characteristic trends of each cluster.

Table 2 Overview of automobile technology domains identified in approach A1

Table 3 Overview of aircraft technology domains identified in approach A1

In the A2 approach, 69,281 patents constructed the MC from the combined dataset of automobile and aircraft industry patents and were divided into 420 clusters. The size of clusters varied from 4 to 6,254. The overview of the top 10 clusters is listed in Table 4.

Table 4. Overview of the automobile and aircraft combined technology domains identified in approach A2

In the A3 approach, 8 sections, 112 classes and 676 subclasses and 8 sections, 119 classes and 554 subclasses were found in the automobile and aircraft dataset. To reduce the data volume, thresholds were set at two stages. First, IPC subclasses with more than 2% of share for either the automobile or aircraft were taken. Secondly, from the filtered IPC subclasses at the first stage, IPC subclasses with more than 0.5% of share for both the automobile and aircraft were taken. In total, 62 subclasses were taken as the major subclasses and cover 207,950 automobile and 22,231 aircraft patents. Table 5 lists the 10 biggest IPC subclasses found in automobile and aircraft patents. The titles appearing in this paper for each IPC code were adopted from the WIPO homepage.

Table 5. Major automobile and aircraft IPC sub-classes identified in approach A3

4.2. Similarity Measurement

4.2.1. A1 approach

In the A1 approach, the similarity measurement was conducted among the top 35 clusters and the top 15 sub-clusters of the top 4 clusters of the automobile and the top 25 clusters of aircraft. Table 6 shows part of the results. The average of cosine similarity except for the sub-clusters was 0.28 and the standard deviation was 0.11. The average of cosine similarity between the sub-clusters of the automobile and clusters of aircraft was 0.39, and the standard deviation was 0.1034. We highlighted cluster pairs with more than 0.5 of cosine similarity and automobile sub-clusters and aircraft cluster pairs with more than 0.6 of cosine similarity. As a result, 48 pairs of clusters were highlighted. We discussed the potential for the combination of knowledge.

Table 6. Cosine similarity results in approach A1

As a result, automobile clusters of driving control and hybrid car driving control such as in cluster 6 (Automatic Transmission Control) and the sub-clusters of cluster 1 (the Hybrid System) and 2 (Drive Control) were highlighted with the aircraft clusters of flight control such as in cluster 12 (Throttle Control), and 13 (Active Flow Control). The use of terms related to system control was similar among highlighted clusters.

However, according to the interviews with aeronautical researchers, basically, 'driving control for automobiles' represents actuating defined motion according to input while 'flight control for aircrafts' represents actuating motion according to operating conditions to achieve the inputted flight attitude. Based on the experts' judgments, the authors found that, in this case, the terms in patents are similar, but the engineering philosophy is very different so that there is less opportunity for knowledge recombination among highlighted automobile and aircraft control domains.

On the other hand, there was another type of pairs identified in A1, whose relations were not clear at a glance for authors and interviewers. For example, high similarity was detected between the sub-clusters of automobile cluster 3 (Exhaust Emission Control) and cluster 4 (IC Engine

Valve Control) and aircraft cluster 15 (Fiber-Reinforced Materials) and cluster 24 (Heat Blanket). We investigated the factors for this high similarity appeared in these very different pairs at a glance, and some similar use of terms were detected such as temperatures, heating, and pressure. The performance of an exhaust emission purifying system becomes lower at between 300 °C to 400 °C and over 500 °C (Patent JP2002-126453). Improvement of exhaust emission purifier efficiency in a wide- range of temperatures is expected in the automobile. On the other hand, application of fiber-reinforced materials such as carbon fiber-reinforced polymers and fiber-reinforced ceramics is important for aircraft development because light, strong, and heat-resistant characteristics of these materials are very suited to the aircraft operation condition. Investigating the relation between exhaust gas purification and composite materials further, some patents were found that focus on the heat resistance nature of fiber-reinforced ceramics and apply fiber-reinforced ceramics as exhaust purifying catalyst support and increase the performance of the purifier (ex. Patent JP2001-179110) under a wide range of temperatures.

At first, the relation between those clusters was unclear, but the similarity between problems to be solved in the exhaust emission purifier and in the properties of fiber-reinforced materials showed the possibility of knowledge combination and a breakthrough in the exhaust gas purifier.

When we showed the results of our investigation on such pairs, automobile engineers found that highlighting such similarity of problems is very useful for them to transfer their technology to other industries and vice versa.

4.2.2. A2 approach

A2 approach resulted 689 patents citing other industries and 612 patents cited by other industries in the automobile and aircraft dataset. In the MC of the combined dataset, 421 patents cite other industries and 345 patents are cited by other industries. We analyzed the top 35 clusters and highlighted clusters that have both automobile and aircraft patents. Table 7 shows part of the results and the percentage of each industry patent. 6 Clusters were highlighted in total.

Table 7. Percentages of Automobile and Aviation Patents in Identified Domains in Approach A2

At the results of A2, cluster 4 was the Aircraft System cluster, where many aircraft patents belong. According to sub-cluster analysis, the cluster consists of various sub-clusters such as Composite Structures, Composite Material Welding, Engine Mounting, Cabin Compartments, and Aircraft Wing System and Avionics. Among 4375 patents of the cluster 4, 110 patents are automobile patents. Many automobile patents are also found in Composite Structures and Composite Material Welding sub-clusters and the topics of the automobile patents found in these sub-clusters are the production of body structure with a fiber-reinforced composite, surface defect inspection, and stir welding. There are also automobile patents in Avionics and in that case the main topics of these automobile patents are navigation display processing. On the other hand, clusters 6 and 9 are the actual automobile clusters, such as the Fuel Cell System cluster and the Automatic Transmission Control cluster. Among 3,444 patents and 2,082 patents of clusters 6 and 9, 42 patens and 69 patents are aircraft patents.

Composite materials and the navigation system are well known as examples of technology transfer from the aircraft industry to the automobile industry. There is also technology transfer recognized in braking systems such as anti-lock systems and disk brakes from the aircraft industry to the automobile industry. In addition, the fuel cell and electric generation system is often considered to be more advanced in the automobile industry than in the aircraft industry so that aeronautic researchers are carefully watching technology development in the automobile industry. The A2 approach highlighted both past and emerging knowledge transfer from one industry to another and also highlighted knowledge exchanges in very similar processes such as in Assembly.

Figure 3 shows the sub-clusters of cluster 6 in A2 and visualizes the use of a large graph layout (LGL) (Adai et al., 2004)[1]. Researchers of JAXA showed interest in the results. When they needed to analyze patents related to Fuel Cell technology, they searched the data with keywords such as “fuel cell” and obtained many patents containing the keywords. However, the researchers had problems because it was difficult to categorize the obtained data in detail. The researchers said that results in Fig. 3 provided them with an interesting overview of patents related to automobile Fuel Cell technology that they would not have been able to analyze within their own limited resources.

Figure 3. Structure overview of fuel cell subclusters, Cluster 6 in A2

4.2.3. A3 approach

---

[1] LGL is based on a spring layout algorithm where links play the role of spring connecting nodes. As a result of this layout, the group of patents citing each other is located in closer positions and only the intra-cluster links for each cluster are shown with the same color, in the order in which the clusters are intuitively understood to clarify the position of each cluster.

In the A3 approach, IPC sub-classes with more than 0.48 of r_i were highlighted (Table 8). 19 sub-classes were highlighted.

Table 8. IPC-sub classes with high r_i

At the results of A3, it was easy to find common topics between the automobile and aviation industries in the highlighted IPC subclass because it is the very objective of IPC classification. For example, G01L is a Measuring Force class and how to exclude the impact of torsions of shafts in measuring the force and conciliation frequency of a hybrid engine shaft and turbine engine shaft is one of its common topics. H01L is semiconductor device class and a common topic in improving the efficiency of how to cool semiconductors. H04N became a Pictorial Communication and edge problem when several pictures were combined.

At the interview, automobile engineers commented that, even though this approach was simple, some of highlighted areas were interesting to look at. Furthermore, they commented that, to increase the value of information in this approach, for example, the analysis of the co-occurrence of IPC categories in patents of each industry, will be useful. As several IPC categories can be defined in one patent, highlighting differences in co-occurrence of IPC between two industries can give an industry new idea of technological application already challenged in other industries.

5. DISCUSSION

Table 9 compares A1, A2 and A3 results. For approaches of technological sub-domain identification, the citation analysis approach limited the identification to the MC so that the coverage of data were much inferior to the IPC approach. If a comprehensive overview of technological sub-domains in other industries is needed, the IPC approach can satisfy the needs as a list of technological knowledge. Moreover, the overview of the IPC approach can be obtained easily with spread sheet software. On the other hand, our citation analysis enables us to identify technology sub-domains as associated parts of systems so that the analysis can support engineers in obtaining breadth knowledge related to the technology sub-domains being transferred from or to and can accelerate a broadening scope of projects and adoption of technology.

However, wider data coverage of IPC approach does not promise effectiveness in knowledge combination. It can be interpreted from Table 4 that the Fuel Cell technology of cluster 6 is a very new domain compared to other listed clusters. In addition, analysis of sub-clusters of cluster 6 can dismantle different technological knowledge consisting of the Fuel Cell system as was appreciate by JAXA engineers (Fig.3). The Fuel Cell Structure shown in the citation analysis was descriptive and different from the IPC structure. Many fuel cell patents in cluster 6 in A2 were categorized to sub-classes: F02D (Controlling Combustion Engines), F02B (Internal-Combustion Piston Engines), F02M (Supplying Combustion Engines), or group F02D41/ F02D45 (Electrical control of supply of combustible mixture or its constituents), F02D13 (Controlling the engine output power by varying inlet or exhaust valve operating characteristics). IPC categories are periodically revised but not very flexible in adapting to the dynamism of technological knowledge. The principal aim of IPC is to facilitate a search of patents in function. Therefore, using IPC categories to see an overview of a system is difficult.

Table 9. A1, A2 and A3 data coverage

For measuring similarities between domains, compared to the other two approaches, the A1 approach had a higher likelihood of highlighting DB-C combinations. The A1 approach highlighted similarity either in functions (ex. control system) or properties (ex. heat resistance).

While it seemed difficult to combine knowledge of the same functions from different philosophies, bringing knowledge with different functions but the same property or the same type of problems together can have the possibility of success in knowledge recombination in breadth. To measure similarity in operating conditions or properties such as temperature, pressure or affinity, selecting terms to measure the text similarity is favorable and is expected to appear in future research.

On the other hand, as we have already discussed, the A2 approach highlighted past and emerging knowledge transfer from one industry to another (DB-T combination), such as in Avionics, Composite Materials, and Fuel Cell, and knowledge exchanges in very similar processes (DB-D combination) such as in Assembly. We can expect the A2 approach to provide useful technological information related to on-going technology transfer from one industry to another. We also expect that analysis of the development of links in past knowledge transfer will help to manage future knowledge transfer and such analysis is expected to be studied in the near future.

The A3 approach measured similarities, while it reduced the amount of data that needed to be looked at, and it is not surprising to find common topics between two industries' patents because they are categorized into the same IPC that have been established to facilitate finding DB-D pairs.

Figure 4. Three methodologies on the DB-Combination Model

While the methodologies still need to be improved, the three measurements can highlight DB-C, DB-T and DB-D pairs in the proposed DB-Combination model (Fig. 4). In our case, i.e., knowledge combination between automobile and aircraft industry, A1 approach is effective to explore DB-C, because it can capture implicit technological relationships by text-similarity measurement. On the other hand, A3 approach based on IPC similarity can extract explicit relationships and can be used to deepen existing common technologies, which we call DB-D. A2 approach also utilizes explicit technological relationships by citation-similarity measurement, but its effect is DB-D and DB-T. Therefore, we can conclude that practitioners should utilize each approach to supplement the others or select relevant approach to fit their purpose.

However, it must be noted the results of three approaches can largely depend on the dataset. In this paper, patents from a limited number of companies of the automobile industry and the aircraft industry were investigated. Expansion of the dataset is needed to explore patents in these industries more complete. And analysis on different sectors might give different results. In the above analysis, we used automobile patents only by Japanese automobile contractors. This is because according to our interview with experts, in automobile industry, patents tend to cite other patents issued by their competitor in the same country, while in aviation industry manufacturing is globalized and such a tendency is relatively not expected. And in the above analysis, we do not include major supplier of automobile industry like Aisin, Denso, and Bosch. This is because the difference of supply chain system between automobile and aircraft industry. Our aim is to capture the mainstream of technology on an industry. In the development of an aircraft, the aircraft prime manufactures often take risk and revenue sharing partnerships with suppliers and large part of technology development are done in suppliers. To avoid such a citation bias, we limited the scope of automobile patents and included patents by the suppliers in aviation industry.

However, there is a possibility that such a dataset construction strategy can affects the results. To test it, we added the following applicant names and reexamined the results; Ford or "General Motors" or GM or Chrysler or Daimlerchrysler or "Daimler Chrysler" or Daimler or "Bayerische Motoren Werke" or Peugeot or Citroen or Volkswagen or Renault or "Hyundai

Motor" or Aisin or Denso or Bosch. When we analyzed such an expanded dataset, the number of patents for the automobile increased from 243,305 to 573,788.

In the approach A1, the patents in the maximum component increased from 60,458 to 176,604. Even though the number of clusters also increased from 303 to 836, as we explained above, many clusters are the same technology domains but the different countries. For example, there are two large clusters of IC Engine Valve Control; one is mainly of Japanese applicants and the other is of US applicants. We also observed new technology domains; air conditioning system and multi function switch system. These technology domains are often applied to cars in upper segments and these patents are mostly of German companies. As a result, similarity measurements with the new dataset highlight not only the combination found in the original dataset but also new pairs such as air conditioning system. Therefore, automobile patent clusters are apparently affected by such a country citation bias.

In the approach A2, we obtained similar results with the original dataset. The patents in the maximum component increased from 69,281 to 188,032. In the original paper, 6 domains were highlighted and 5 of the 6 were found in the new dataset. The rest, Automatic Transmission domain, can be also highlighted by extracting a subcluster of a large automobile cluster. Therefore, the expansion of the dataset does not affect the result. But we must note that the number of the aircraft patents in the component also increased from 8,577 to 9,966, even though the same query is used for the aviation industry. Therefore, when we examine the details of such additional aviation patents, we expect to extract other technological domains where technologies in aviation industry have high relatedness with technologies in automobile industry.

In the approach A3 with the new dataset, 18 IPC sub-classes were highlighted in the same condition to the original paper and 13 sub-classes were common to the original results. The difference was occurred because the new dataset includes patents not from automobile companies. It reduces the share of automobile industry in technological domains such as painting and coating. However, if we set the threshold r_i of highlighting conditions from 0.48 to 0.42, the same 38 sub-clusters were highlighted in the both dataset.

As shown in the above, the addition of patents from automobile companies of other countries and from other companies increases opportunity of technology transfer but doesn’t affect the conclusion derived by the original dataset. The results with the expanded dataset has the same tendency with those with the original dataset in the comparison between A1, A2, and A3 approaches, while some differences appear because of country citation bias in A1 and similarity setting in A3. These must be tackled by future research to test reliability and demonstrate effectiveness of our methodology in different contexts and datasets.

We must also note that there is a criticism on using patent as the indicator of technological knowledge, even though patent is the best available indicator for R&D invention related to technology and outcomes of innovation activities (OECD, 1994). Knowledge of a company can be classified into tacit and explicit knowledge and a patent can represent only part of explicit knowledge (Jones and Miller, 2007; Nonaka, 1994; Nonaka, 1991). Not all the explicit knowledge appears in the form of patents because first, the benefit from patenting must exceed the cost of registration and maintenance of the patent and secondly, some industries tend to not publish but keep secret technological knowledge (Pavitt, 1985).

6. SUMMARY

To support practitioners to collect breadth technological knowledge in other industries and to combine it to their own knowledge, this paper proposed the DB-Combination model that considered the similarities of two technological knowledge domains and three different knowledge combinations in depth and breadth. We also investigated three methodologies A1, A2 and A3 to highlight the three combination pairs from patents by identifying technological knowledge sub-domains and measuring their similarities. The usability of the model and methodologies were demonstrated through a case study on the technological knowledge of the automobile industry and the aircraft industry. This paper used patent data as a source of technological knowledge and identified technological sub-domains from patent data in two approaches; citation analysis (A1, A2) and IPC analysis (A3), and then measured the similarity between different sub-domains in three approaches; text similarity measurement (A1), existing intra-industrial citation tracing (A2) and IPC share similarity comparison (A3).

A1 showed the potential of measuring DB-C combinations that are pairs of technology sub-domains with weak similarity that may bring a technological breakthrough if the similarities are of operating conditions or properties of technology. Using A2 it was possible to identify both past and current DB-T combinations that are a transfer of a unique technology sub-domain of one industry to another. A3 measured pairs of DB-D combinations.

The DB-Combination model with integration of A1, A2 and A3 methodologies can become an effective innovation designing methodology that allows engineers and product managers to find useful technological knowledge from different industries and explore opportunities of technological breakthrough in depth or breadth.

ACKNOWLEDGEMENT

A part of this research is financially supported by Research Institute of Science and Technology for Society (RISTEX) of Japan Science and Technology Agency (JST).

REFERENCES

Adai, AT, SV Date, S Wieland, and EM Marcotte (2004). LGL: creating a map of protein function with an algorithm for visualizing very large biological networks. Journal of molecular biology, 340(1), 179-190.

Alavi M., Leidner D. (2001) Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues, MIS Quarterly, 25, 1, p.107-136.

Cantu F.J., Ceballos H.G. (2010) A multiagent knowledge and information network approach for managing research assets, Expert Systems with Application, 37, p.5272-5284.

Criscuolo P. (2006) The 'home advantage' effect and patent families. A comparison of OECD triadic patents, the USPTO and the EPO, Scientometrics, 66, 1, p.23-41.

Cronin B. (2001) Bibliometrics and beyond: some thoughts on web-based citation analysis, Journal of Information Science, 27, 1, p.1-7.

Dosi G. (1982) Technological Paradigms and Technological Trajectories – a Suggested Interpretation of the Determinants and Directions of Technical Change. Research Policy, 11, p.147–162.

Eriksson S. (2000) Technology spill-over from the aircraft industry: the case of Volvo Aero, Technovation, 20, p.653-664.

Everett J.E., Pecotich A. (1993) Citation analysis mapping of journals in applied and clinical-psychology, Journal of Applied Social Psychology, 23, 9, p.750-766.

Fleming L., Sorenson O. (2001) Technology as a complex adaptive system: evidence from patent data, Research Policy, 30, p.1019–1039.

Garfield L. (1979) Is Citation analysis a legitimate evaluation tool? Scientometrics, 1, 4, p.359-375.

Gassmann O., Zeschky, M. (2008) Opening up the solution space: The role of analogical thinking for breakthrough product innovation, Creativity and Innovation Management, 17, 2, p.97-106.

Gordon, M.D., Lindsay, R.K., Fan, W. (2002) Literature based discovery on the world wide web. ACM Transactions on Internet Technology, 2, 4, p.261-275.

Griliches Z. (1990) Patent Statistics as Economic Indicators: A Survey, Journal of Economic Literature, 90, 28 4, p1661-1707.

Herrero A., Corchado E., Sáiz L, Abraham A. (2010) DIPKIP: A Connectionist knowledge management system to identify knowledge deficits in practical cases, Computational Intelligence, 26, 1, p.26–56.

Jaffe A.B., (1986). Technological Opportunity and Spillovers of R&D: evidence from Firm's Patents, Profits, and Market Value. American Economic Review, 76(5), pp. 984-1001.

Jones B., Miller B. (2007) Innovation diffusion in the new economy: the tacit component, New York:

Routledge.

Kajikawa Y., Abe K., Noda S. (2006) Filling the gap between researchers studying different materials and different methods: a proposal for structured keywords, Journal of Information Science, 32, p.511-524.

Kajikawa Y., Yoshikawa J., Takeda Y., Matsushima K. (2008) Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy, Technological Forecasting and Social Change, 75, 6, p.771-782.

Katila R., Ahuja G., (2002) Something old, something new: A longitudinal study of search behavior and new product introduction, Academy of Management Journal, 45, pp. 1183-1194.

Kostoff R.N. (1998) The use and misuse of citation analysis in research evaluation. Scientometrics, 43, 27-43.

Kostoff, R.N. (2008) Literature-related discovery (LRD): Introduction and background. Technological Forecasting and Social Change, 75, 2, p.165-185

Leiponen A., Helfat C.E. (2010) Innovation objectives, knowledge sources, and the benefits of breadth. Strategic Management Journal, 31, 2, p.224–236.

Leydesdorff L. (2008) Patent Classifications as Indicators of Intellectual Organization, Journal of the American Society for Information Science and Technology, 59, 10, 1582-1597.

Leydesdorff L., Kushnir D., Rafols, I., (2012) Interactive Overlay Maps for US Patent (USPTO) Data Based on International Patent Classifications (IPC). Scientometrics, DOI 10.1007/s11192-012-0923-2

Martin B.R. (2012) The evolution of science policy and innovation studies, Research Policy, 41, p.1219-1239.Michel J., Bettels B. (2001) Patent citation analysis. A closer look at the basic input data from patent search reports. Scientometrics, 51, 1, p.185-201.

Meyer, M. (2000) What is Special about Patent Citations? Differences between Scientific and Patent Citations, Scientometrics, 49, 93-123.

Michel J. & Bettels B. (2001) Patent citation analysis. A closer look at the basic input data from patent search reports. Scientometrics, 51, 1, p.185-201.

Moorthy S., Polley D.E. (2010) Technological knowledge breadth and depth: performance impacts, Journal of Knowledge Management, 14, 3, p.359-377.

Nakamura H., Suzuki S., Tomobe H., Kajikawa Y., Sakata I. (2011a) Citation lag analysis in supply chain research. Scientometrics, 87, 2, p.221-232.

Nakamura H., Kajikawa Y., Suzuki S. (2011b) Innovation for Sustainability in Aviation: World Challenges and Visions, In: F. Nobre, D. Walker, R. Harris (Eds) “Technological, Managerial and Organizational Core Competencies: Dynamic Innovation and Sustainable Advantage”, IGI Global.

Nakamura H., Kajikawa Y., Suzuki S. (2012) Multi-level perspectives with technology readiness measures for aviation innovation, Sustainability Science, Doi: 10.1007/s11625-012-0187-z.

Narin F. (1994) Patent bibliometrics, Scientometrics, 30, 1, p.147-155.

Nelson R.R., Winter S.G. (1977) In search of useful theory of innovation, Research Policy, 6, p.36-76.

Nerur S.P., Rasheed A.A., Natarajan V. (2008) The intellectual structure of the strategic management field: An author co-citation analysis, Strategic Management Journal, 29, p.319-336.

Nesta, L., Saviotti, P.P., (2005) Coherence of the Knowledge Base and the Firm's Innovative Performance: Evidence from the U.S. Pharmaceutical Industry, 8 (1), pp. 123-142.

Newman M.E.J (2004) Fast algorithm for detecting community structure in networks. Physical Review, 69, 066133.

Newman M.E.J., Girvan M. (2004) Finding and evaluating community structure in networks, Physical Review, 69, 026133.

Nicolaisen J. (2007) Citation Analysis, Annual Review of Information Science and Technology, 14, 3, p.609-641.

Nonaka I. (1991) The knowledge-creating company, Harvard Business Review, November-December, p. 96-104.

Nonaka, I. (1994) A dynamic theory of organizational knowledge creation. Organization Science, 5, 1, p.14-37.

OECD (1994) The Measurement of Scientific and Technological Activities: Using Patent Data as Science and Technology Indicators, Organisation for Economic Co-operation and Development, 108 pages.

Pavitt K. (1985) Patent statistics as indicators of innovative activities: Possibilities and problems. Scientometrics, 7, 1-2, p.77-99.

Schilling M.A., Green E. (2011) Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences, Research Policy, 40, 10, p.1321-1331.

Schoen A., Villard L., Laurens P., Cointet J-P., Heimeriks G., Alkemade F, (2012) The network structure of technological developments; Technological distance as a walk on the technology map, 17th International Conference on Science and Technology Indicators.

Schoenmakers W., Duijsters G.M. (2010) The technological origins of radical inventions. Research Policy, 39, 8, p.1051-1059.

Schumpeter J. (1934) The Theory of Economic Development, Cambridge, Harvard University Press.

Smalheiser N.R. (2012) Literature-Based Discovery: Beyond the ABCs, Journal of the American society for Information Scuebce and Technology, 63, 2, p.218–224.

Smallheiser, N.R., Swanson, D.R., (1998) Using ARROWSMITH: A computer-assisted approach to formulating and assessing scientific hypothesis. Computer Methods and Programs in Biomedicine, 57, 3, p.149-153.

Soete L.G., Wyatt S.M.E. (1983) The use of foreign patenting as an internationally comparable science and technology output indicator, Scientometrics, 5, 1, p.31-54.

Tijssen R.J.W. (1992) A quantitative assessment of interdisciplinary structures in science and technology: Co-classification analysis of energy research, Research Policy, 21, 1, p.27-44.

Tijssen, RJW (1992). A quantitative assessment of interdisciplinary structures in science and technology: Co-classification analysis of energy research, Res. Policy, 21 (1), 27-44.

***Biographical Note**

Dr. Hiroko Nakamura is a Project Researcher at the Centre for Aviation Innovation Research and also at the Global Leader Program for Social Design and Management at the University of Tokyo. She previously worked for Nissan Motor Co. Ltd., as a product planner. She received her bachelor's, and master's degree, and Ph.D from the same university and a Special Master's degree from the Ecole Centrale Paris. Her research interests include development of methodology for technology and innovation management.

Dr. Shinji Suzuki is a Professor at the Department of Aeronautics and Astronautics and the Director of the Center for Aviation Innovation Research at the University of Tokyo. He received his Ph.D. degree in Engineering from the University of Tokyo in 1986, after his research career at Toyota's Central R&D Labs, Inc., in Japan. He received his bachelor's, and master's degree from the same university. His main research interests are in the design and control aspects of air safety and unmanned aerial vehicles. He is currently an Executive Committee Member of ICAS.

Dr. Ichiro Sakata is a Professor at the Department of Technology Management for Innovation, the Director of Innovation Policy Research Center and the co-head of Presidential Endowed Chair for Electricity Network Innovation by Digital Grid at the University of Tokyo. He received his bachelor's degree and Ph.D from the same university, and master's degree from Brandeis University. He has a 20 years' working experience at the Japanese Ministry of Economy, Trade and Industry (METI). His research interests include technology management, technology roadmap and innovation network. He has published more than 100 papers in peer reviewed journals and conference proceedings.

Dr. Yuya Kajikawa is an Associate Professor at Graduate School of Innovation Management, Tokyo Institute of Technology. He is also a visiting researcher at Graduate School of Engineering, The University of Tokyo. He received his bachelor's, and master's degree, and Ph.D from the University of Tokyo. His research interests include development of methodology for technology and innovation management. He has a number of publications in peer-reviewed journals and conference proceedings, which cover a variety of disciplines including engineering, information science, environmental science, and technology and innovation management. He serves as an Associate Editor of Technological Forecasting and Social Change.

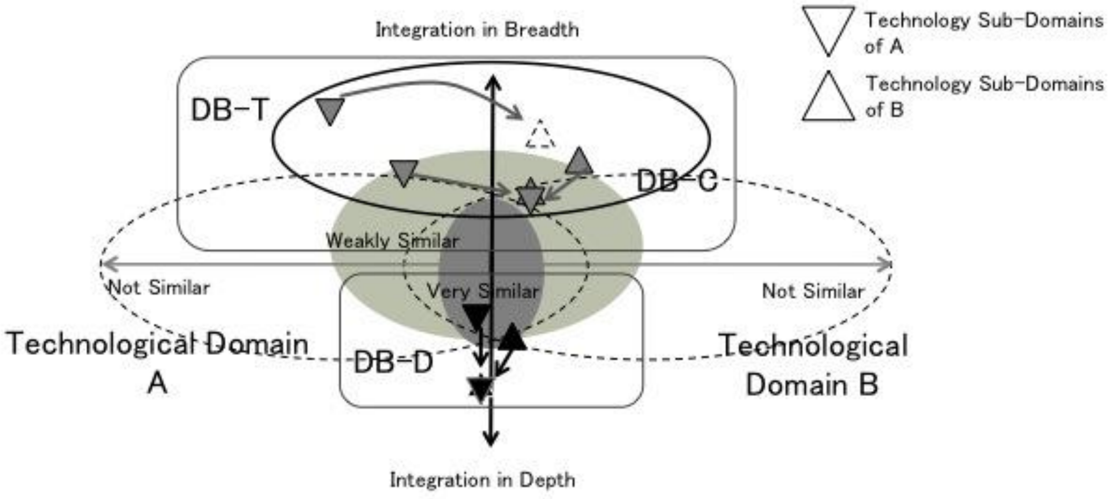
Integration in Breadth
Technology Sub-Domains of A
Technology Sub-Domains of B
DB-T
DB-C
Weakly Similar
Not Similar
Very Similar
Not Similar
Technological Domain A
DB-D
Technological Domain B
Integration in Depth

Fig. 1

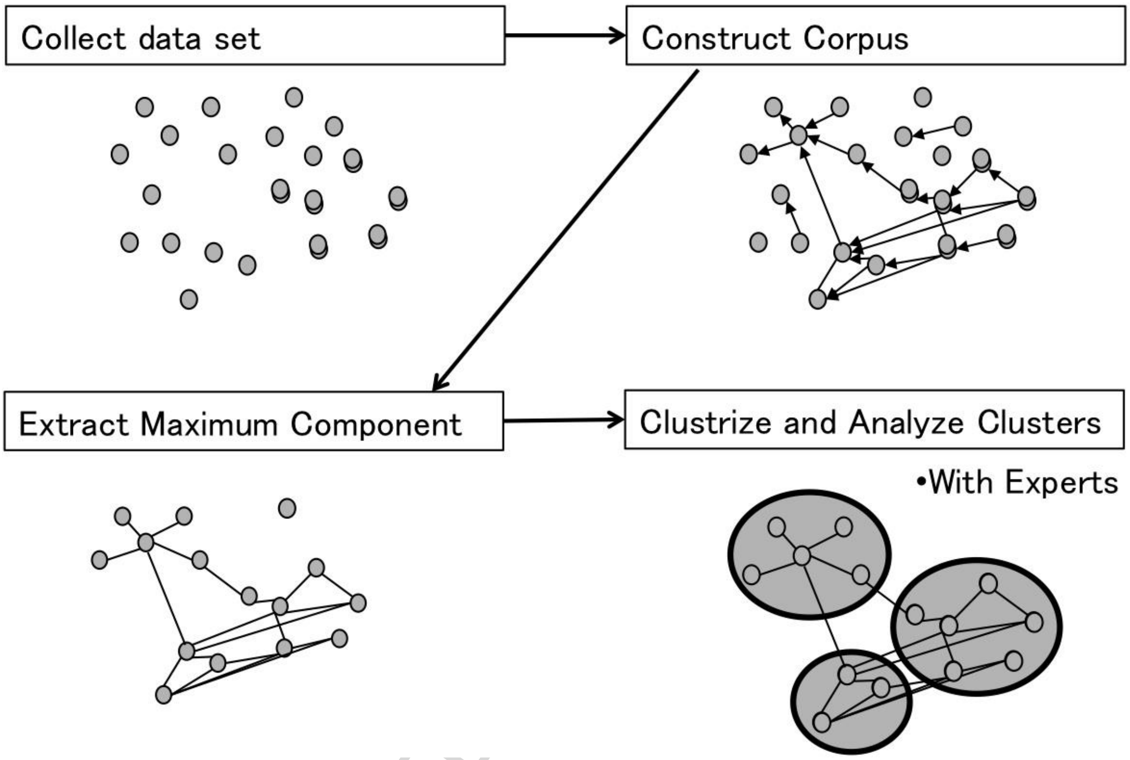
Collect data set
Construct Corpus
Extract Maximum Component
Clustrize and Analyze Clusters
•With Experts

Fig. 2


Separator/ Sealing
②
Gas Leakage
⑧
Drain water
⑦
⑥
Air shut valve
Hydrogen
filing
⑨
①
Scavenging
⑤
③
Idle Stop
Rapid warn-up
④
Fuel
Reforming
Temperature control
humidification

Fig. 3

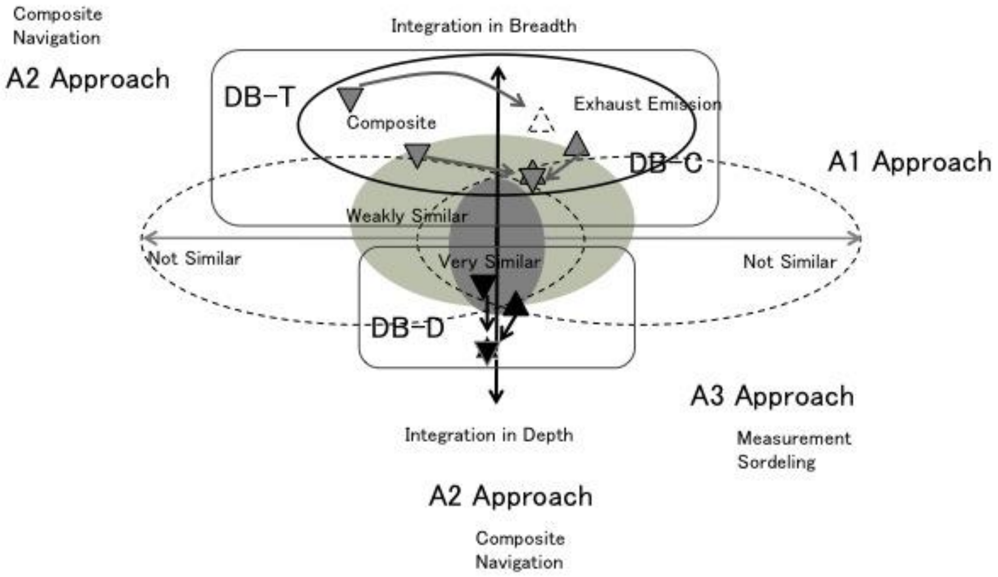
Composite
Navigation
A2 Approach
Integration in Breadth
DB−T
Composite
Exhaust Emission
DB−C
A1 Approach
Weakly Similar
Not Similar
Very Similar
Not Similar
DB−D
A3 Approach
Measurement
Sordeling
Integration in Depth
A2 Approach
Composite
Navigation

Fig. 4

Table 1. Three approaches

Method	Technological Sub-domains Identification	Similarity Measurement
A1	Citation Analysis	Text Similarity Measurement
A2	Citation Analysis	Intra-industrial Citation Tracing
A3	IPC Analysis	IPC Share Similarity Comparison

Table 2 Overview of automobile technology domains identified in approach A1

CL	**Size**	**Yr**	**Characteristic of Cluster**
	Freq. terms	**Core Patent**	**Core patent DWPI title example (For top 5 clusters)**
1	5187	2003.8	Hybrid System (Drive Control)
	hybrid, battery, hybrid vehicle, torque, engine, electric, clutch, braking, generator, drive	JP2002225578A	Hybrid vehicle powered both by internal combustion engine and electric motor, has power transmission unit that maintains continuity of input power level during power source switch over
		JP2000346187A	Power transmission for hybrid vehicle, includes drive shaft and output shaft whose revolution numbers are maintained more or less same, by controlling gear shift ratio
		JP2000002327A	Transmission speed controller of hybrid vehicles performs regenerative control of electric motor or generator, such that revolution number of transmission is decreased or increased, respectively when shift-up or shift-down is detected by speed change detector
2	4778	2003.3	Drive Control
	steering, lane, image, object, driver, road, preceding vehicle, assistance, information, preceding	JP10211886A	Power steering system for e.g. electrically driven vehicle, hydraulic vehicle has steering suppression unit that prevents steering operation after decision unit has determined degree of danger based on detected position, distance, and relative velocity of obstruction
		US20050125137A1	Vehicle deceleration control apparatus, has controller executing deceleration control by brake system that applies braking force to vehicle and shift operation that shifts automatic transmission of vehicle into low speed ratio
		JP2002267470A	Information presentation system calculates position of view point of driver and window of vehicle with respect to specific position, based on the information related to specific position
3	4350	2002.8	Exhaust Emission Control
	catalyst, exhaust, gas, exhaust gas, oxide, purification, nitrogen oxide, nitrogen, nox, fuel ratio	JP8338229A	Exhaust emission control device of diesel engine stores nitric oxide when exhaust gas temperature is low and converts into nitrogen dioxide when high using oxidation catalyst along with oil and heater
		JP7217474A	Exhaust gas purifying appts. for internal combustion engines comprises an air-fuel ratio controller to temporarily make rich exhaust gas flowing to nitrogen oxide absorber
		JP2004174490A	Manufacture of catalyst material, e.g. for exhaust gas purification, involves mixing acidic solution comprising ions of cerium, zirconium and catalyst metal with aqueous ammonia, for co-precipitating and baking obtained mixture
4	3072	2002.6	IC Engine Valve Control
	valve, engine, combustion, intake, internal combustion, internal combustion engine, valve timing, combustion, compression, internal	JP2003206771A	Internal combustion engine has camshafts comprising cams and movable bearings at axial portion, arranged on both sides of cylinder, to accommodate cam portions and bearings in holes formed in cylinder block and lower case
		JP7293216A	Valve gear of internal combustion engine for car has driving cam which is shut from suction valve supported by eccentric part in cylindrical cover
		JP2004218522A	Driving control apparatus of internal combustion engine, controls supply of air into cylinder corresponding machine compression ratio, by delaying/advancing valve closing time, based on driving load of engine
5	2881	2005.2	Fuel Cell System
	fuel cell, cell, fuel cell system, cell system, fuel, hydrogen, gas, cell stack, fuel cell stack, stack	JP7235324A	Fuel battery drive device increases dynamic pressure of oxygen gas supplied to electrode temporarily, when electrode gates get due to adhesion of solidified water drops
		JP2000243417A	Fuel cell apparatus comprises removal unit to remove impurities which reduce specific value of power generated

		JP2003331893A	Fuel cell system has control valve which supplies air to fuel cell stack, so that air purges through fuel electrode, during starting and electric power generation completion states
6	2557	1992.6	Automatic Transmission Control
	transmission, automatic transmission, clutch, automatic, shift, hydraulic, gear, valve, engine, pressure		
7	2416	2001.8	Body Structure
	frame, panel, bumper, front, vehicle body, collision, member, body, pillar, floor		
8	2397	2003.5	Combustion Control
	injection, fuel, combustion, fuel injection, ignition, engine, internal combustion, internal combustion engine, combustion, valve		
9	2007	1991.3	Air-Fuel Ratio Control
	air fuel, fuel ratio, air fuel ratio, fuel, engine, air, ratio, intake, fuel injection, injection		
10	2004	2004	Battery and Cooling System
	battery, cell, fuel cell, cooling, floor, battery pack, pack, seat, frame, fuel		

Table 3 Overview of aircraft technology domains identified in approach A1

CL	Size	Yr	Characteristic of Cluster
	Freq. terms	**Core Patent**	**Core patent DWPI title example (For top 5 clusters)**
1	903	2002.5	Cabin Equipment
	seat, cabin, door, passenger, compartment	US5083727A	Aircraft cabin system for selectively locating interior units with adjustable line connectable between aircraft cabin interior and adjacent service distribution coupling
	fuselage, gear, floor, luggage, overhead	US6464169B1	Overhead galley and crew rest facility for aircraft has cart transferring mechanism provided in overhead section and which supports galley cart between planar cart plane and cart lift
		US4055317A	Aft main deck split level galley is used in aircraft and has raised deck giving increased floor space in concave rear pressure bulkhead
2	665	2001.1	Structure Assembly/ Fabrication
	welding, stir welding, friction stir welding, stir, composite	US4622445A	Honeycomb panel inductive brazing placing alloy between face sheet and core and heating to alloy melting point
	friction stir, ceramic, friction, sol, workpiece	US5041321A	Mfg. fiber formed ceramic insulation useful in aerospace systems by multiple impregnation of soft felt mat formed from slurry of fiber with sol-gel glass binder, and adding fiber-reinforced glass layer
		US5645744A	Apparatus for induction processing of workpiece comprises forming dies non susceptible to induction heating combined with susceptor sheets susceptible to induction heating
3	577	2005.2	Flight Information System
	weather, display, radar, weather radar, radar system	US7675461B1	Terrain data circuit for use in aircraft, has display control circuit generating composite terrain image based on data from radar system and data from database, where display signal representative of terrain is provided based on image
	terrain, antenna, weather radar, signal, data	US8049644B1	Aircraft warning system for depicting terrain awareness and warning system alert information in terrain advisory display, has processing circuit providing indication of first or second warning signal, or caution signal on display
		US7965225B1	Method of adjusting position of antenna used in radar mounted in e.g. aircraft, involves reducing position error by adjusting antenna position using terrain angle within beam
4	568	2004.2	Composite Structure
	composite, tape, stringer, fuselage, skin	US20060060705A1	Shell structure for composite fuselage of aircraft, has fitting whose ends are attached to stiffeners and skins of primary and secondary panels, for joining primary panel to secondary panel
	mandrel, structure, material, composite, composite material	US20060108058A1	Composite shell structure manufacturing system for aircraft, has composite material applicator that moves with respect to support structure to apply composite material on interior mold surface of lay up mandrel
		US20090139641A1	Shell structure manufacturing method for fuselage of smaller aircraft e.g. fighter aircraft, involves positioning stiffeners on inner surface of skin lay-up, and concurring skin lay-up and stiffeners to bond stiffeners to skin lay-up
5	558	1998.7	Jet Engine Structure (Mounting, Nacell System, Nozzle)
	engine, nozzle, jet engine, jet, exhaust	US20100126139A1	Nozzle system for gas turbine engine having longitudinal axis for aircraft, has fan nozzle which can be pivoted about axis oriented transversely relative to longitudinal axis to vary fan duct nozzle throat area
	turbine, flow, gas, gas turbine, nacelle	US4044973A	Nacelle mounting for turbofan jet engine has rear nacelle structure formed in two D-shape ducts hinged to open as clam shells
		US4458863A	Strut supported inlet for turbofan has space frame thrust linkage transferring loads to wing strut
6	431	2003.4	Refueling System
	refueling, antenna, tanker aircraft, tanker, boom, phased array, phased, flight refueling, maintenance, array antenna		
7	428	1999.5	Wing System
	flap, wing, lift, edge flap, aircraft wing, high lift, slat, trailing edge, edge, trailing		

<table>
<tr><td rowspan="2">8</td><td>341</td><td>2004.4</td><td>Environment Control System, APU</td></tr>
<tr><td colspan="3">air, fuel cell, cooling, fuel, air conditioning, heat, conditioning, exchanger, heat exchanger, decompression</td></tr>
<tr><td rowspan="2">9</td><td>259</td><td>1998.6</td><td>Assembly tools</td></tr>
<tr><td colspan="3">drilling, workpiece, fastener, tool, hole, rivet, riveting, carriage, electromagnetic, machine</td></tr>
<tr><td rowspan="2">10</td><td>202</td><td>1999.8</td><td>Landing Gear System</td></tr>
<tr><td colspan="3">gear, brake, braking, wheel, landing gear, landing, control, braking system, command</td></tr>
</table>

Table 4. Overview of the automobile and aircraft combined technology domains identified in approach A2

CL	Size	Yr	Characteristic of Cluster
	Freq. terms	**Core Patent**	**Core patent DWPI title example (For top 5 clusters)**
1	6254	2003.9	Hybrid System (Drive Control)
	hybrid, hybrid vehicle, battery, engine, torque	JP2002225578A	Hybrid vehicle powered both by internal combustion engine and electric motor, has power transmission unit that maintains continuity of input power level during power source switch over
	electric, motor, drive, power, braking	JP2000346186A	Power transmission for hybrid vehicle, includes drive shaft and output shaft whose revolution numbers are maintained more or less same, by controlling gear shift ratio
		JP2000002327A	Transmission speed controller of hybrid vehicles performs regenerative control of electric motor or generator, such that revolution number of transmission is decreased or increased, respectively when shift-up or shift-down is detected by speed change detector
2	6138	1998.1	IC Engine System
	fuel, catalyst, exhaust, engine, injection	US4949695A	Fuel evaporative purge system malfunction detecting device has value in purge passage which opens when engine operates and sensor to detect negative vacuum and compare it with intake vacuum
	injection, air fuel, fuel ratio, combustion, echaust gas	US4571683A	Learning air-fuel ratio control system for electronic engine has second learning term corrected during idling period and first when predetermined load is exceeded
		US4561400A	Air-fuel ratio control method controlling fuel injection rate by learning correction coefficient for obstruction of air flow motor to compensate for ageing of meter
3	5282	2003.3	Driving Support System
	steering, lane, image, vehicle, driver	JP10211886A	Power steering system for e.g. electrically driven vehicle, hydraulic vehicle has steering suppression unit that prevents steering operation after decision unit has determined degree of danger based on detected position, distance, and relative velocity of obstruction
	driver, road, object, assistance, driving, preceding vehicle	JP2002267470A	Information presentation system calculates position of view point of driver and window of vehicle with respect to specific position, based on the information related to specific position
		JP2001310719A	Lane deviation prevention apparatus for vehicles, regulates yawing moment along specified direction, based on braking force difference between right and left wheels
4	4375	2002.1	Aircraft System
	aircraft, wing, composite, fuselage, skin	US7675461B1	Terrain data circuit for use in aircraft, has display control circuit generating composite terrain image based on data from radar system and data from database, where display signal representative of terrain is provided based on image
	panel, flap, flight, cabin, antenna	US8049644B1	Aircraft warning system for depicting terrain awareness and warning system alert information in terrain advisory display, has processing circuit providing indication of first or second warning signal, or caution signal on display
		US8077078B1	Measurement method for determining altitude of aircraft using altitude of runway and radar sweep by radar system, and based on vertical angle and range between runway and aircraft, and runway altitude
5	3467	2003.8	Combustion Control System
	fuel, combustion, injection, ignition, engine	JP10266878A	Controller for four-stroke engine has exhaust valve whose closing time is delayed when demand load of system becomes low

	fuel injection, internal comb., int. comb. Engine, comb. Engine, valve timing	JP2000179368A	Fuel supply procedure in gasoline internal combustion engine, involves increasing proportion of supply of fuel having high and low octane numbers during high and low load running respectively
		JP11210539A	Internal combustion engine has controller which regulates gas in combustion chamber to compression stroke end stage within target temperature range based on output of decision circuit that judges whether self-adhering fire of mixed air will increase
6	3444	2005.1	Fuel Cell System
	fuel cell, cell, fuel, fuel cell system, cell system, hydrogen, gas, cell stack, stack		
7	2473	2001.9	Body Structure
	front, frame, vehicle body, bumper, collision, member, body, rear, pillar, structure		
8	2297	2004.1	Battery and Cooling System
	battery, cell, fuel cell, fuel, cooling, floor, battery pack, stack, seat, cell stack		
9	2082	1993.5	Automatic Transmission Control
	transmission, automatic transmission, shift, gear, automatic clutch, hydraulic, variable transmission, speed, torque		
10	2008	2001.5	IC Engine Valve Control
	valve, compression ratio, engine, cam, combustion, internal comb., int. comb. Engine, comb. Engine, valve timing, timing		

Table 5. Major automobile and aircraft IPC sub-classes identified in approach A3

IPC	Class (ex. B23)	Sub-Class (ex. B23P)	S-Avi	S-M	
B23P	Machine Tools	Combined Operation	3.9%	1.0%	A
B29C	Working of Plastic	Shaping of Plastic	4.6%	2.1%	A
B32B	Layard Product	Layard Product	5.0%	0.8%	A
B60K	Vehicles in General	Arrangement of Propulsion Units	0.4%	10.2%	M
B60L		Propulsion of Electrically-Propelled Vehicles	0.3%	5.2%	M
B60R		Vehicles	0.6%	9.6%	M
B62D	Land Vehicles for Travelling	Motor Vehicles	0.2%	7.8%	M
B64C	Aircraft	Aeroplanes	15.4%	0.1%	A
B64D		Equipment	13.6%	0.0%	A
F01D	Machines or Engine	Non-positive-displacement	11.4%	0.2%	A
F01N		Exhaust Apparatus for Machines	0.5%	5.3%	M
F02B	Combustion Engines	Internal-Combustion Piston Engines	0.9%	5.0%	M
F02C		Gas-Turbine Plants	9.2%	0.2%	A
F02D		Controlling Combustion Engines	0.6%	13.6%	M
F02K		Jet-Propulsion Plants	3.9%	0.0%	A
F02M		Supplying Combustion Engines	1.1%	6.2%	M
F04D	Pumps for Liquids	Non-Positive-Displacement Pumps	3.5%	0.2%	A
F16H	Engineering Elements	Gearing	0.9%	6.9%	M
G06F	Computing	Electric Degital Data Processing	7.4%	2.9%	A
H01M	Basic Electric Elements	Batteries	1.1%	8.5%	M

Table 6. Cosine similarity results in approach A1

	Motor:1	Motor:2	Motor:3	Motor:4	Motor:5	Motor:6	Motor:7	Motor:8	Motor:9	Motor:10
Avi:1	0.22	0.19	0.19	0.23	0.10	0.24	0.41	0.17	0.23	0.50
Avi:2	0.19	0.13	0.19	0.19	0.14	0.19	0.28	0.17	0.21	0.31
Avi:3	0.20	0.36	0.15	0.17	0.11	0.20	0.12	0.20	0.21	0.20
Avi:4	0.18	0.21	0.14	0.19	0.07	0.18	0.36	0.12	0.16	0.31
Avi:5	0.35	0.18	0.41	0.40	0.11	0.34	0.32	0.37	0.47	0.36
Avi:6	0.27	0.25	0.17	0.17	0.10	0.30	0.14	0.19	0.25	0.17
Avi:7	0.30	0.23	0.19	0.30	0.20	0.27	0.21	0.25	0.24	0.27
Avi:8	0.33	0.20	0.32	0.38	0.41	0.36	0.18	0.37	0.53	0.46
Avi:9	0.28	0.23	0.17	0.19	0.08	0.25	0.22	0.26	0.28	0.27
Avi:10	0.46	0.44	0.21	0.22	0.07	0.45	0.18	0.22	0.27	0.15
Avi:11	0.23	0.31	0.19	0.20	0.08	0.23	0.39	0.17	0.16	0.39
Avi:12	0.43	0.33	0.23	0.44	0.16	0.55	0.17	0.28	0.45	0.22
Avi:13	0.32	0.34	0.28	0.37	0.23	0.37	0.24	0.33	0.41	0.38
Avi:14	0.25	0.19	0.31	0.30	0.08	0.22	0.36	0.28	0.24	0.42
Avi:15	0.25	0.13	0.32	0.29	0.26	0.31	0.21	0.29	0.27	0.26

* Automobile Technology Fields

Driving System	2, 33, 14, 34, 6 (AT), 18 (Motor), 19 (Suspension)
Engine	3 (Exhaust Purification), 4 (Valve Cntl), 8 (Combustion), 9, 11 (Fuel Injection), 31 (Cooling Sys.), 32 (Fuel Tank)
Structure	7, 21 (Collision Absorption), 15 (Airbag), 23 (Door Lock), 26 (Hood), 28 (Seat Storage), 35 (Assembry)
Hybrid/ Fuel System	1 (Drive Cntl.), 5, 16, 20 (Fuel Cell), 10, 25 (Battery)
Others	13 (Motor Cycle), 28 (Walking Robot), 30 (Nap Detector), 33 (Noise Control)

*Aircraft Technology Fields

Flight Control	3 (Flight Info. Sys.), 6 (Refueling Sys.), 7 (Wing Sys.), 10 (Landing Gear), 12 (Throttle Cntl.), 13 (Active Flow Cntl.)
Propulsion	5 (Nacell), 16 (Blade Containment), 19 (Gas Turbine Safely Devices), 20 (Gas Turbine Component), 23 (Combustion), 25 (Fuel Injection)
Structure	2 (Assembly Method), 4 (Composite Material), 9 (Assembly Tools), 11 (Composite Material, Fastening), 15 (Fiber Reinforced Material), 17 (Structure Inspection), 24 (Heat Blancket)
Others	1 (Cabine, Storage), 8 (APU Sys.), 14 (Harness), 18 (Starter Generator), 21 (Fiber-Opt Magnetic Field), 22 (Toilet Sys.)

Table 7. Percentages of Automobile and Aviation Patents in Identified Domains in Approach A2

	Motor	Aviation
1	99.90%	0.10%
2	99.90%	0.10%
3	100.00%	0.00%
4	2.50%	97.50%
5	99.70%	0.30%
6	98.80%	1.20%
7	99.80%	0.20%
8	99.60%	0.40%
9	96.70%	3.30%
10	99.90%	0.10%

Table 8. IPC-sub classes with high r_i

IPC	Class (ex. B23)	Sub-Class (ex. B23P)	S-Avi	S-M	R
B05D	Spraying	Processes for Applying Liquids	1.3%	0.9%	0.50
B21D	Mechanical Metal-Working	Processing of Sheet Metal	1.8%	1.1%	0.49
B22F	Casting	Working Metallic Powder	0.8%	0.6%	0.49
B23K	Machie Tools	Welding	2.9%	2.0%	0.49
C04B	Cements	Lime, Magnesia	0.6%	0.6%	0.50
C23C	Coating Metallic Material	Coating Metallic Material	1.8%	1.1%	0.48
F16C	Engineering Elements	Shafts	1.5%	1.1%	0.49
G01C	Measuring	Measuring Distance	2.5%	1.8%	0.49
G01L		Measuring Force	0.8%	0.7%	0.50
G01M		Testing Static	1.8%	1.1%	0.49
G01P		Measuring Linear	0.6%	0.6%	0.50
G01R		Measuring Electric Variables	1.4%	0.8%	0.48
G05B	Controling	Control System in General	1.6%	1.1%	0.49
G06T	Computing	Image Data Processing	0.6%	1.0%	0.48
H01L	Basic Electric Elements	Semiconductor Devices	1.5%	2.1%	0.49
H01R		Electrically Conductive Connections	1.0%	0.6%	0.48
H02K	Generation of Electric Power	Dynamo-Electric Machines	1.4%	2.0%	0.49
H02P		Control of Electric Motors	1.0%	1.5%	0.49
H04N	Electric Communication Technique	Pictorial Communication	0.7%	0.8%	0.50

Table 9. A1, A2 and A3 data coverage

	A1	A2	A3
Dataset	Automobile: 243,305, Aircraft: 27,989		
Patent Numbers that are used for technological sub-domain identification (% over the total)	Automobile MC: 60,458 Aircraft MC: 8,281 (25.4%)	Combined MC: 69,281 (25.6%)	270,294 (100%)
Identified Cluster	Automobile: 303 Aircraft: 104	420	676
Similarity Analyzed Cluster Numbers	Automobile: 35 Aircraft: 25 (20.6%)	35 (20.5%)	62 (85%)
Highlighted Similarity	48 pairs of clusters	6 clusters	19 sub-classes

Highlights

- We model knowledge combinations in depth and breadth between technological domains.
- We examine methodologies to measure modeled 3 combinations from patent information.
- We demonstrate the usability of the model and methodologies through a case study.
- We interview automobile and aircraft experts to discuss the results.
- The proposed measurements can highlight candidates of the modeled combination.